Универзитет Сингидунум

Наход Вуковић
Слађана Спасић

СТАТИСТИКА СА ПРАКТИКУМОМ

Београд, 2022.

УНИВЕРЗИТЕТ СИНГИДУНУМ

Проф. др Наход Вуковић
Проф. др Слађана Спасић

СТАТИСТИКА СА ПРАКТИКУМОМ

Прво издање

Београд, 2022.

СТАТИСТИКА СА ПРАКТИКУМОМ

Аутори:
Проф. др Наход Вуковић
Проф. др Слађана Спасић

Рецензенти:
Проф. др Милица Булајић
Проф. др Мирјана Шекарић

Издавач:
УНИВЕРЗИТЕТ СИНГИДУНУМ
Београд, Данијелова 32
www.singidunum.ac.rs

За издавача:
Проф. др Милован Станишић

Техничка обрада:
Слађана Спасић
Милош Вишњић

Дизајн корица:
Александар Михајловић

Година издања:
2022.

Тираж:
300

Штампа:
Бирограф, Београд

ISBN: 978-86-7912-776-1

ПРЕДГОВОР

Ова књига обрађује основе статистике, вероватноће и методе статистичке анализе обухваћене програмом предмета Статистика који се изучава на првој години Факултета за информатику и рачунарство Универзитета Сингидунум. Намењена пре свега студентима Универзитета Сингидунум, али и свима који желе да буду успешни у својим активностима, било да су информатичари, програмери, менаџери, предузетници, инжињери. Стечено и усвојено знање биће им од помоћи у проналажењу решења у свакодневним пословним одлукама које су базиране на статистичким подацима.

Последњи део уџбеника чини практикум са теоретским питањима и задацима. Збирка је прилагођена почетном курсу статистике и треба да послужи, уз коришћење уџбеника, материјала са предавања и вежби за самосталну проверу и утврђивање знања. Уз редован рад на предавањима и коришћење уџбеника студент може једноставно да провери тачност својих одговора па је зато збирка тест питања дата без обележених тачних одговора. Код сваког питања, дати су могући одговори. Нека питања су постављена у облику реченице коју треба допунити да би она постала тачна. У наредном потп-оглављу дати су мешовити задаци из читавог градива од којих су неки са решењима.

Захваљујемо се рецензентима проф. др Милици Булајић и проф. др Мирјани Шекарић.

Унапред захваљујемо свим студентима и другим читаоцима који би били љубазни да своја мишљења и запажања о збирци, сугестије и уочене недостатке пошаљу на sladjana.spasic@singidunum.ac.rs .

Београд, Аутори

Фебруар 2022. године

САДРЖАЈ

1. ОСНОВНИ ПОЈМОВИ

1.1. ПРЕДМЕТ СТАТИСТИЧКИХ ИСПИТИВАЊА

Циљ научних испитивања је да се опишу појаве које нас окружују и да се утврде законитости које владају међу њима. У целокупном нашем окружењу највећи број појава су појаве масовног карактера. То су појаве које могу бити описане једино преко великог броја информација нумеричког и описног својства које се односе на поједине елементе посматране појаве. На пример, ако се жели анализирати запосленост морају се посматрати сви запослени, ако се жели испитивати кретање становништва, мора се посматрати целокупно становништво, за испитивање производње у Србији морају се посматрати сви произвођачи и сл.

Статистичке методе се баве проучавањем појава изражених преко великог броја нумеричких и ненумеричких података који садрже информације о посматраним појавама.

Ако желимо да испитамо продуктивност у Србији морамо знати колика је продуктивност сваког привредног учесника у Србији. Другим речима, морамо посматрати скуп привредних учесника и код сваког елемента тог скупа „измерити" продуктивност. Исто тако кад хоћемо да посматрамо кретање личних доходака запослених, морамо посматрати цео скуп запослених и за сваки његов елемент утврдити висину личног дохотка. На тај начин добијамо статистичке податке који постају предмет статистичких испитивања.

Статистичко испитивање вршимо на једном скупу елемената кога ћемо звати **статистички скуп** или **популација** и на сваком елементу тог скупа „меримо" одређено својство које ћемо звати **статистичко обележје**.

Статистички скуп

Елементи који имају једну заједничку особину чине један скуп. Али заједничка особина тих елемената мора бити тако дефинисана да потпуно и недвосмислено одређује да ли посматрани елемент припада датом скупу или му не припада. На елементима потпуно одређеног скупа приступа се испитивању или мерењу једног или више обележја. При том посматрана обележја могу имати различите вредности код елемената датог скупа.

На пример, на скупу пословних субјеката у Србији у датом тренутку можемо испитивати број запослених, просечна лична примања, остварени доходак и сл. Исто тако на скупу студената Београдског Универзитета можемо посматрати физичку издржљивост, успех на појединим предметима, број прочитаних књига итд.

Статистички скуп или популација је онај скуп на чијим елементима меримо једно или више обележја чије вредности варирају од елемента до елемента посматраног скупа.

Статистички скуп мора бити дефинисан просторно и временски.

Временско дефинисање статистичког скупа утврђује тренутак или раздобље којим ће се обухватити они елементи који ће бити предмет статистичких испитивања. Тако на пример, није довољно рећи: посматрамо скуп путничких аутомобила у Београду. Морамо додати и време. На пример: дана 31.6.2006. године.

Просторно дефинисање статистичког скупа подразумева означавање простора на коме се налазе елементи датог скупа. Ако посматрамо скуп саобраћајних удеса морамо рећи да ли је то скуп удеса на посматраном путу, у Србији или на простору Балкана и сл.

У статистичкој анализи се веома често испитује део статистичког скупа и на основу особина елемената тог дела доносе закључци о целом статистичком скупу или популацији. Такав део статистичког скупа назива се **узорак**. Основно подручје **Аналитичке статистике** представљају узроци и статистичка анализа базирана на њима.

Статистичко обележје

Под статистичким обележјем подразумева се оно својство елемената статистичког скупа које варира од елемента до елемента датог скупа. За сваки елемент статистичког скупа утврђује се „вредност” посматраног обележја.

Статистичка обележја се деле на ***квалитативна*** и ***квантитативна***.

Квалитативна обележја изражавају квалитативна својства елемената статистичког скупа. На пример: боја косе, школска спрема, пол, место рођења и сл. Свако квалитативно обележје има одређене категорије и елементи посматраног скупа сврставају се у одговарајуће категорије.

Квантитативна обележја изражавају квантитативна својства елемената статистичког скупа. Вредности таквих обележја су реални бројеви, а утврђивање тих вредности врши се „мерењем”. На пример, код скупа домаћинстава можемо „измерити” број чланова домаћинстава, или код скупа становника Београда може се мерити висина или тежина сваког Београђанина.

Квантитативна обележја се деле на ***прекидна*** и ***непрекидна*** обележја.

Ако статистичко обележје може узети неку вредност из коначног скупа могућих вредности онда кажемо да је то **прекидно обележје.** На пример: број повреда на послу, реализована производња, број чланова домаћинства и сл.

Кад посматрано статистичко обележје може узети било коју вредност из коначног или бесконачног интервала, кажемо да је то **непрекидно обележје**. На пример: век трајања електронске цеви, висина студента, температура и сл.

Матрица података

Предмет изучавања статистике су статистички скупови и обележја елемената статистичког скупа. У пракси су то велики скупови са великим бројем обележја. На пример, скуп становника Београда, Србије, и сл. са обележјима: пол, старост, место рођења, школска спрема, економски статус, итд. Реч је о најмање 2 милиона становника са најмање 20 обележја, а то је огромна маса података. Исто тако, скуп предузећа у Србији са свим подацима о њиховим карактеристикама и активностима, биће описан са огромном масом података.

Означимо са N број елемената статистичког скупа, а са $X_1, X_2, ..., X_k$, k обележја. Измерене вредности свих обележја не сваком елементу статистичког скупа можемо приказату у виду матрица реда N x k, у којој ће врсте представљати елементе скупа, а колона одговарајуће вредности посматраних обележја. Тако ћемо добити **матрицу података**

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nk} \end{bmatrix}$$

са онолико врста колико има статистички скуп елемената и колона колико обележја меримо на елементима статистичког скупа. Ова матрица представља **статистичку грађу** и предмет је статистичких проучавања.

Основни проблем је непрегледност статистичких података, па је циљ да се прикупљени подаци **среде и учине прегледнијим**, а затим да се пређе на њихову **анализу**.

1.2. ПОДЕЛА СТАТИСТИКЕ

Статистика или статистичке методе се могу поделити у две класе:

- **Дескриптивна статистика**
- **Аналитичка статистика.**

Под **дескриптивном статистиком** се подразумевају оне статистичке методе које се баве прикупљањем, срећивањем и анализом статистичких података. Њен циљ је да се опише измерена појава или процес **у домену мерења.**

Аналитичка статистика има за циљ да на основу мерења утврди општије законитости које важе и **у домену оног што није било мерено**. Наиме, за потпуно описивање појаве или процеса потребно је бесконачно много мерења. Пошто увек располажемо са резултатима коначног броја мерења морамо користити научне методе којима ћемо на егзактан начин описати комплетну појаву или процес.

На пример, ако се жели почети производња неког новог производа мора се утврдити време потребно за његову израду. Извршиће се одређен број мерења у предпроизводњи и методама дескриптивне статистике анализирати добијени резултати. Да би тачно одредили потребно време израде тог производа и у будућој производњи морамо користити методе аналитичке статистике којима ћемо на основу мерења у предпроизводњи ,,предвидети" време које ће бити потребно за тај производ у наредном периоду.

Исто тако, ако посматрамо неку економску категорију можемо методама дескриптивне статистике утврдити особине те категорије у домену мерења (на датом простору и у датом времену). Али за потпуно испитивање те категорије и њених законитости и изван домена мерења морамо се користити методама аналитичке статистике.

Основа аналитичке статистике је **теорија вероватноће** која математичким моделима описује случајне експерименте.

Аналитичка статистика се састоји из следећих делова:

- *Теорије статистичког оцењивања и*
- *Теорије статистичког одлучивања.*

Дескриптивна статистика обухвата следеће гране:

- *Статистику становништва и друштвених делатности,*
- *Привредну статистику,*
- *Статистику привредних биланса и*
- *Финансијску статистику.*

Статистика становништва и друштвених делатности

Ова статистичка грана обухвата неколико различитих подручја. Један од њених важнијих задатака је ***проучавање и праћење управљања***. Статистичка испитивања из области управљања прате:

- избор и структурни састав свих органа управљања на свим нивоима друштвено-политичких заједница и месних заједница,
- избор и структурни састав свих управних органа предузећа и интересних заједница,
- активности месних заједница, скупштина општина, и заједница пословних субјеката,
- изборе делегата за друштвено-политичке заједнице,
- међуопштинску сарадњу и
- изборе пословодних органа.

Друга важна функција ове статистичке гране обавља се кроз ***статистику становништва*** или ***демографску статистику*** која се бави прикупљањем, срећивањем и анализом података који садрже информације о следећем:

- укупан број и територијална расподела становништва,
- укупан број и расподела домаћинстава и породица,
- природни прираштај,
- структуре становништва према разним обележјима,
- састав и друштвено-економски положај домаћинстава и породица.

Статистика становништва располаже веома великим серијама података добијеним пописом становништва и свеобухватним праћењем рађања и умирања, склапања и развода бракова и сл. Она има изузетну важност у планирању укупног друштвеног развоја.

Статистика рада обухвата испитивања из области запослености, запошљавања, заштите на раду и личних доходака запослених. Базу статистике рада чине законом прописане евиденције у области рада и матичне евиденције осигураника пензијског и здравственог осигурања. Статистичка испитивања из области рада и личних доходака базирају се на подацима које дају статистички органи и Служба за запошљавање.

Статистика образовања се бави испитивањима која омогућавају да се сагледа организациона структура и територијални распоред школа према врстама и смеровима, кретање и успех ученика и студената према разредима односно према годинама студија, токови наставних процеса и кретање наставног особља. Ови се подаци прикупљају редовно у годишњој периодици од свих школа.

Статистика научноистраживачке делатности пружа информације о организационим облицима, областима и гранама истраживања, о кадровима, о приходима научноистраживачких и развојних организација, о броју завршених научноистраживачких радова, броју проналазака и патената и сл.

Статистика културе и уметности обухвата испитивања о квантитативним аспектима културних и уметничких збивања: број активности радничких и народних универзитета и домова културе, културно-уметничких и културно-просветних друштава, позоришта, ансамбала народних игара, филхармонија и професионалних оркестара, архива, библиотека, музеја, галерија, уметничких збирки и биоскопа. Посебно се прати и врши анализа издавања књига, брошура и часописа као и производња, увоз и извоз филмова.

Статистика физичке и техничке културе постаје све значајнија. Обухвата проучавања спортских организација и спортску активност као и спортско рекреационе објекте. Организације техничке културе постају бројне и значајне по активности и по броју чланова које окупљају па је неопходно праћење и анализа кретања у овој делатности.

Статистика здравственог, пензијског и социјалног осигурања представља изузетно важну статистичку делатност. Бави се прикупљањем, сређивањем и анализом података о здравственој служби и здравственој заштити, о броју, положају и кретању пензионера, о социјалној заштити и о заштити деце и омладине.

Статистика правосуђа обухвата веома широку област статистичких истраживања о откривањима и кажњавањима противправних радњи и дела, њиховим учиниоцима, о казнама, затим о грађанским и управним споровима, управним споровима и радним споровима, привредним преступима као и о другим појавама које се решавају правосудним поступком.

Привредна статистика

Ова статистичка грана пружа информације потребне за вођење целокупне економске политике, за планирање на свим нивоима друштвене и економске делатности и за обавештавање стручне, научне и најшире јавности о свим привредним кретањима. Привредна статистика посматра укупну привредну активност по привредним делатностима и у целини. Обухвата праћење и анализу података који исказују:

- основна средства,
- запосленост,

- доходак,
- инвестиције,
- личну и општу потрошњу,
- извоз и увоз,
- цене и
- интеграциона кретања.

Привредна статистика се дели на следеће области:

- Статистика индустрије и рударства,
- Статистика пољопривреде,
- Статистика шумарства,
- Статистика водопривреде,
- Статистика грађевинарства,
- Статистика стамбене и комуналне делатности, насеља и животне средине,
- Статистика саобраћаја и веза,
- Статистика унутрашње трговине,
- Статистика угоститељства и туризма,
- Статистика економских односа са иностранством и
- Статистика цена.

Статистика привредних биланса

Основни задатак статистике привредних биланса је анализа и приказивање међусобне повезаности система биланса, функционисања друштвеног процеса репродукције у целини и посебно у његовим фазама. Главне делове ове статистике представљају следећи биланси:

- биланс формирања, расподеле и употребе друштвеног производа,
- табеле међусобних односа привредних делатности (тзв. input-output табеле);
- биланси финансијских рачуна и
- биланс народног богатства.

Највећи део података за статистику привредних биланса пружају банке и Народна Банка Србије.

Финансијска статистика

Основна подручја финансијске статистике су:

- финансијско пословање предузећа и корисника друштвених средстава,
- извори средстава и њихова употреба за заједничку и општедруштвену потрошњу,
- утрошена средства за инвестиције и стање основних и обртних фондова,
- новац и кредити
- финансијско пословање са иностранством.

Финансијска статистика користи податке о евиденцији корисника друштвених средстава (платни налози, промет на текућим и другим рачунима, наплате прихода буџета и сл.) и на књиговодственим евиденцијама поменутих корисника.

Испитивања којима се бави финансијска статистика заснивају се на Закону о статистичким истраживањима. Основне податке пружају банке, Народна банка Србије и друге институције финансијског тржишта.

2. ДЕСКРИПТИВНА СТАТИСТИКА

2.1. АНАЛИЗА РЕЛАТИВНИМ БРОЈЕВИМА

Прикупљени статистички подаци представљају основу статистичких испитивања. Први задатак у тим испитивањима је припрема података за анализу и интерпретација добијених резултата. Другим речима, масу података морамо свести на мањи број нумеричких показатеља који ће бити погодни за анализу и интерпретацију, помоћу којих ћемо моћи да објаснимо посматране појаве или процесе везане за дати статистички скуп. У многим истраживањима ова област је веома значајна и широко распрострањена, а позната је и под називом ***експлораторна анализа.***

Статистички подаци представљају апсолутне бројне износе или атрибутивне карактеристике појединих елемената статистичког скупа или његових појединих делова. Међусобна поређења апсолутних бројева или карактеристика не дају нам праву слику о посматраној појави. У поређењима нас интересују релативни односи одговарајућих апсолутних износа који омогућавају поређења више различитих показатеља посматране појаве у статистичком скупу. Тако је на пример, далеко боље утврдити тзв. стопу незапослености него број незапослених. Исто тако пожељно је знати остварени доходак по запосленом, а не само укупан остварен доходак у предузећу.

У статистичкој пракси се користе три категорије релативних бројева: **пропорције, релативни показатељи** и **индекси**.

Пропорција

У анализи статистичких података који се односе на квалитативна својства статистичког скупа служимо се најчешће пропорцијама. Претпоставимо да својство X има k категорија, $(k \geq 2)$ и да N_1 елемената статистичког скупа припада првој категорији, N_2 елемената припада другој категорији, ..., k-toj категорији припада N_k елемената посматраног скупа. Означићемо са N укупан број елемената статистичког скупа. Тада је

$$\sum_{i=1}^{k} N_i = N \,.$$

Количници

$$P_i = \frac{N_i}{N}, \quad i = 1,2, \ldots, k, \tag{2.1}$$

називају се **пропорције** *одговарајућих категорија својства X. Често се пропорције изражавају у процентима. Процентуално изражене пропорције се добијају ако се* P_i *помножи са 100.*

Податке о пропорцијама дајемо у виду табела. Тако је у Табели 2.1. дата структура запослених у Србији у 2006. години према сменама у којима раде запослени. Структура је дата преко одговарајућих пропорција.

Смена	Број	Пропорција	Пропорција у %
Прва смена	943.255	0,686	68,6
Друга смена	256.371	0,186	18,6
Трећа смена	176.267	0,128	12,8
У к у п н о	1.375.893	1,00	100,0

Табела 2.1. Запослено особље према сменама

Пропорције су саставни део статистичких табела публикованих података у скоро свим статистичким публикацијама. Пропорције се могу користити и у анализи квантитативних обележја. Веома су погодне за поређења вредности обележја X на статистичким скуповима са великим бројем елемената.

Релативни показатељи

Многе се појаве могу изразити преко релативних односа два различита обележја. Тако се на пример, посматра густина насељености као количник броја становника и величине територије, доходак *per capita* (по главни становника) као показатељ оствареног дохотка земље, чист доходак по запосленом итд.

Релативни показатељ појаве је количник два обележја на посматраном статистичком скупу.

Изабрана обележја у релативним показатељима треба да су најважнија за објашњавање посматране појаве. Вредности тих обележја могу бити директно утврђена „мерењем" на елементима статистичког скупа. Али је често потребно те вредности кориговати и на тај начин добити боље показатеље дате појаве. На пример, у посматрању проблема заштите на раду, основни показатељ је број повреда. Бољи показатељ је број изгубљених радних сати због повреда. Ово обележје треба поредити са бројем запослених у организацији. Наравно, боље је поређење извршити према броју запослених у производњи. Ради упоредивости ова два обележја треба израчунати укупан број радних сати свих запослених у производњи и са њим поредити број изгубљених сати због повреда. Тако ће се добити најбољи показатељи организованости заштите на раду који ће бити упоредив са таквим показатељима у различитим предузећима и у различитим временима.

За релативне показатеље је од изузетне важности база показатеља, а то је обележје са којим се дели друго обележје. Тако се избором базе може доћи до сасвим супротних закључака о посматраној појави на основу истих вредности једног обележја X. У Табели 2.2. дати су подаци о саобраћајним несрећама у САД за 1950. годину и 1964. годину. Из четврте колоне се види да је у 1964. години дошло до повећања броја погинулих за 37% у односу на 1950. годину. Такође је број погинулих према броју становника порастао за 8%. Међутим, код броја погинулих према броју аутомобила дошло је до смањења за 23%. Слично је и са бројем погинулих према укупном броју пређених километара.

Број погинулих	1950.	1964.	Стопе промена
Укупан број	34.763	47.700	+37
Према броју становника	23,0	24,9	+ 8
Према броју аутомобила	7,1	5,5	-23
Према броју пређених километара	7,6	5,7	-25

Извор: National Safety Council Accident Facts, 1965.

Табела 2.2. Број погинулих у аутомобилским несрећама

Индекси

Један од важнијих задатака статистичких испитивања је праћење и анализа динамичких кретања друштвених, економских и природних појава. За овакве анализе подаци се прикупљају тако што се у датим временским интервалима врше мерења посматране појаве преко статистичких обележја и формирају се тзв. **временске серије** посматраних обележја.

За најједноставнију анализу таквих серија служимо се индексима који представљају показатеље промена посматраних обележја у датим временским интервалима (дневне промене, месечне, годишње и сл.).

Индекс *је број који представља количник вредности посматраног обележја X у тренутку t,* (X_t)*, и вредности тог обележја у неком другом тренутку t',* $(X_{t'})$,

$$I(t,t') = \frac{X_t}{X_{t'}} \tag{2.2}$$

Постоје две врсте индекса: ***ланчани*** и ***базни*** индекси.

Ланчани индекси *су количници вредности обележја X у тренутку t,* (X_t)*, и његове вредности у претходном тренутку мерења* $t-1, (X_{t-1})$.

Ако временска серија садржи податке о вредностима обележја X у тренуцима мерења t, при чему је

$$t \in [1,2,...,T],$$

тада су ланчани индекси дати низом бројева

$$\frac{X_2}{X_1}, \frac{X_3}{X_2}, ..., \frac{X_T}{X_{T-1}} \tag{2.3}$$

Најчешће се ови бројеви множе са 100 па се ланчани индекси изражавају процентуално. Низ ланчаних индекса (2.3) изражава узастопне промене обележја X. Означаваћемо их са

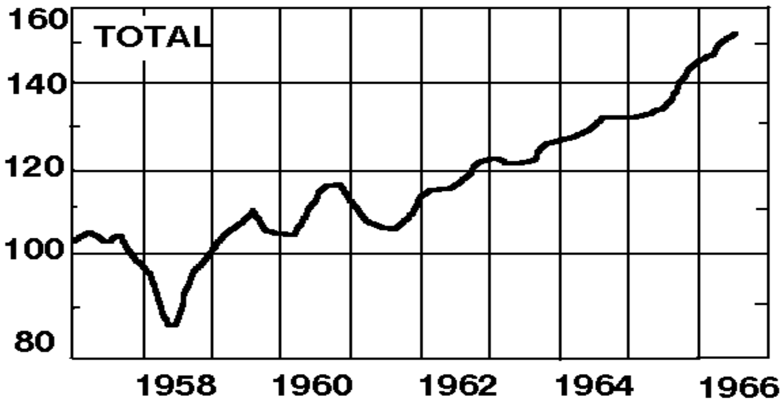
$$I_2, I_3, ...I_T.$$

За временску серију од T података добија се (T-1) ланчаних индекса јер се први податак у серији не може поредити са претходним.

Ако од процентуално израженог индекса у тренутку t одузмемо 100 добићемо процентуално изражену промену обележја X од момента (t-1) до момента t коју називамо ***стопа промене***. На пример, ако је I_t= 141,5 то значи да је обележје X порасло за 41,5%, а ако је I_t=92,3 онда се обележје X смањило за 7,7%.

Низ ланчаних индекса се може представити графички. На хоризонталној оси се наноси време (дани, месеци, године и сл.), а на вертикалној оси одговарајуће вредности ланчаних индекса. Добијене тачке се споје полигоналном линијом. Тако се добије графикон који на очигледан начин показује кретање посматраног обележја временске серије. На Слици 2.1. приказано је кретање индустријске производње преко месечних индекса у периоду 1958-1966. године.

Ланчани индекси се користе за анализу временских серија различитих обележја. На пример, ако се посматра кретање броја запослених и вредност производње онда се било каква анализа ова два обележја базира на поређењима одговарајућих ланчаних индекса датих временских серија.



Слика 2.1. Графикон кретања индустријске производње

Базни индекси *служе за исказивање промене посматраног обележја у временској серији у односу на један тренутак мерења са којим желимо да поредимо настале промене.*

Базни индекси су количници вредности обележја X у тренуцима мерења t, при чему је

$$t \in [1,2,...,T],$$

и вредности обележја X у тренутку t_0. Базни индекси изражени у процентима добијају се ако се количници помноже са 100.

Ако су вредности обележја X у посматраним тренуцима $X_1, X_2, ..., X_T$, а X_0 његова вредност у тренутку t_0, тада низ количника

$$\frac{X_1}{X_0}, \frac{X_2}{X_0}, \;\ldots\ldots \frac{X_T}{X_0}$$

представља базне индексе посматране временске серије.

2.2. РАСПОДЕЛА ФРЕКВЕНЦИЈА

Посматраћемо статистички скуп од N елемената на којима ћемо испитивати квантитативно обележје X. Претпоставимо да обележје X на једном елементу статистичког скупа може узети саму једну вредност из скупа бројева

$$\{x_1, x_2, ..., x_k\},$$

где је k број могућих различитих вредности посматраног обележја, коначан је, али не сувише велики број.

На пример, ако се посматра број чланова домаћинства на скупу домаћинстава у једном региону онда је скуп могућих вредности овог обележја скуп бројева

$$\{1,2,...,10\}.$$

Исто тако број смена у предузећима може узети једну од вредности

$$\{1,2,3\}.$$

Код таквог обележја све његове вредности на посматраном скупу можемо изразити ***расподелом апсолутних фреквенција*** која се добија тако што се за сваку могућу вредност обележја x_i утврди колико елемената статистичког скупа узима ту вредност.

Означимо са f_i број елемената статистичког скупа код којих је $X = x_i$.

Тада се скуп парова вредности

$$(x_1, f_1), (x_2, f_2), ..., (x_k, f_k)$$

назива **расподела апсолутних фреквенција** *посматраног обележја X на датом статистичком скупу.*

Најчешће се, ради прегледности, расподела даје у виду табеле (Табела 2.2.).

Вредност обележја *X* (X)	Апсолутне фреквенције (f)
x_1	f_1
x_2	f_2
.	.
.	.
.	.
x_k	f_k
Укупно	$\boldsymbol{N}$

Табела 2.2. Расподела обележја X

По правилу је статистички скуп велики, а број могућих вредности обележја *X* мали у односу на величину статистичког скупа. Зато нам расподела фреквенција чини масу података прегледнијом, а исто тако и погоднијом за даљу анализу. Сад се уместо *N* вредности обележје *X* посматра *k* парова вредности (x_i, f_i) и при томе се не губи ништа у укупној количини информација која је садржана у тим подацима.

Претпоставићемо да је обележје *X* непрекидног типа и да може узети било који реалан број из интервала (*a,b*). Интервал (*a*,*b*) може бити коначан или бесконачан.

Поделићемо интервал (a, b) на k подинтервала

$$\boldsymbol{a_0\text{-}a_1,\ a_1\text{-}a_2\ ,...,\ a_{k-1}\text{-}a_k,}$$

при чему је $\boldsymbol{a_0 = a,\ a_k = b}$, *које ћемо звати* **групни интервали**.

На основу измерених вредности обележја *X* на елементима статистичког скупа одређује се расподела фреквенција тако што се за сваки групни интервал a_{i-1}-a_i утврди број елемената статистичког скупа код којих вредност X припада том интервалу, тј. код којих је

$$\boldsymbol{a_{i-1} < X \leq a_i}$$

за свако $i \in \{1,2, ...,k\}$. Означимо број тих елемената са f_i. Добијена расподела фреквенција приказана је табеларно (Табела 2.3.).

Групни интервали	Апсолутне фреквенције
$a_0 - a_1$	f_1
$a_1 - a_2$	f_2
.	.
.	.
.	
$a_{k-1} - a_k$	f_k
Укупно	**N**

Табела 2.3. Расподела обележја X

Групни интервали у Табели 2.3. не морају бити једнаке дужине. Најчешће су први и последњи интервал највећи и са најмањим фреквенцијама. Дужине осталих (k-2) интервала зависе од њиховог броја и од појаве коју изражава обележје X. Одређивање броја групних интервала се на различите начине. Треба водити рачуна да превелики број интервала отежава анализу и чини податке непрегледним. Са друге стране, превише мали број интервала не садржи довољну количину информација о статистичком скупу.

При одређивању групних интервала треба се придржавати следећег принципа: ***одредити број и дужине групних интервала тако да они садрже максималну количину информација уз њихов минималан број.***

Релативне фреквенције

Расподела апсолутних фреквенција је подесна за анализу и посматрање појединих обележја одвојено. Међутим, ако треба поредити обележје X на два или више статистичких скупова, онда апсолутне фреквенције не одражавају упоредиве реалне износе. Исто тако, ако треба поредити обележје X са неким другим обележјем Y тада нам апсолутне фреквенције не омогућавају таква поређења. Зато је потребно апсолутне бројеве претворити у релативне и анализу вршити на основу тзв. релативних фреквенција**.**

Претпоставимо да обележје X има неку од расподела апсолутних фреквенција као у Табелама 2.2. и 2.3. За сваки групни интервал и за сваку вредност обележја X одредимо релативну фреквенцију као количник

$$p_i = \frac{f_i}{N}$$

при чему је $i \in \{1,2,...k\}$, а са p_i смо означили вредности релативне фреквенције. Релативне фреквенције су бројеви који задовољавају следећа два услова:

1) $p_i \geq 0$, за свако $i \in \{1, 2, ..., k\}$ и

2) $p_1 + p_2 + ... + p_k, = 1$, tj. $\sum_{i=1}^{k} p_i = 1$.

Релативне и апсолутне фреквенције се најчешће дају у једној истој табели као што је то у Табели 2.4 приказано. Процентуално изражавање релативних фреквенција чини их погоднијим за поређења и за графичко представљање.

Вредност обележја, X	Апсолутне фреквенције, f	Релативне фреквенције, p	Релативне фреквенције у процентима , %
$a_0 - a_1$	f_1	$p_1 = \frac{f_1}{N}$	$p_1 \cdot 100$
$a_1 - a_2$	f_2	$p_2 = \frac{f_2}{N}$	$p_2 \cdot 100$
...	...	...	...
$a_{k-1} - a_k$	f_k	$p_k = \frac{f_k}{N}$	$p_k \cdot 100$
Укупно	$\boldsymbol{N}$	**1,00**	**100,00**

Табела 2.4. Расподела релативних фреквенција[1)]

Кумулативне фреквенције

У анализама неког обележја X на статистичком скупу често нас интересује одговор на следеће питање: Колико има елемената у статистичком скупу код којих обележје X има вредност мању од неког унапред датог броја x? На пример, интересује нас колико има радника са личним дохотком мањим од 40000 дин, или колико има општина са националним дохотком по глави становника мањим од просечног националног дохотка у Републици Србији и сл.

Да би лакше дошли до одговора на оваква питања служимо се тзв. кумулативним фреквенцијама које се одређује на основу апсолутних или релативних фреквенција.

[1)] Ако је обележје X прекидног типа онда у првој колони уместо групних интервала стоје одговарајуће вредности тог обележја.

Из расподеле апсолутних фреквенција вредност кумулативне функције за i-ти групни интервал добија се кад се све фреквенције за претходне интервале саберу. Тако ће се добити да је

$$c_i = f_1 + f_2 + \ldots + f_i,$$

за $i \in \{1, 2, \ldots, k\}$, *а* c_i *означава* **кумулативну фреквенцију** *i-тог групног интервала.*

Кад се кумулативне фреквенције c_i поделе са бројем елемената статистичког скупа добиће се релативне кумулативне фреквенције које означавамо са F_i и које су дате са

$$F_i = \frac{c_i}{N}$$

за свако $i \in \{1, 2, \ldots, k\}$.

Расподела кумулативних фреквенција приказана је у Табели 2.5.

Групни интервали, X	Апсолутне фреквенције, f	Кумулативне фреквенције, c	Релативне кумулативне фреквенције, F
$a_0 - a_1$	f_1	$c_1 = f_1$	$F_1 = \frac{c_1}{N}$
$a_1 - a_2$	f_2	$c_2 = f_1 + f_2$	$F_2 = \frac{c_2}{N}$
.	.	.	.
.	.	.	.
.	.	.	.
$a_{k-1} - a_k$	f_k	$c_k = f_1 + f_2 + \ldots . f_k$	$F_k = \frac{c_k}{N} = 1$
Тотал	**N**		

Табела 2.5. Расподела кумулативних фреквенција

Релативне кумулативне фреквенције се могу одредити и из релативних фреквенција. Наиме, из израза за C_j и F_i види се да је

$$F_i = \frac{f_1}{N} + \frac{f_2}{N} + \ldots + \frac{f_i}{N},$$

односно

$$F_i = \sum_{j=1}^{i} p_j,$$

за свако $i \in \{1, 2, \ldots, k\}$.

Релативне кумулативне фреквенције

$$F_1, F_2, \ldots, F_k$$

чине један неопадајући низ реалних бројева из интервала [0,1].

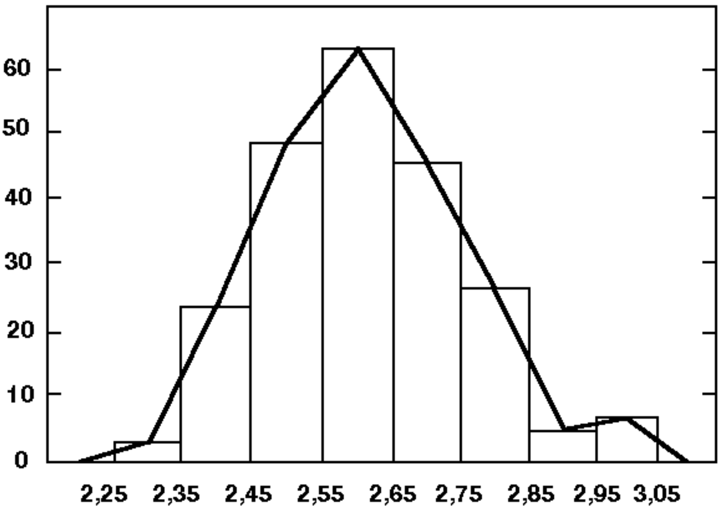
2.3. ГРАФИЧКО ПРИКАЗИВАЊЕ РАСПОДЕЛА ФРЕКВЕНЦИЈА

Да би лакше уочили и истакли главне карактеристике расподеле фреквенција неког обележја X на посматраном статистичком скупу служимо се графиконима.

Графиконе расподела фреквенција конструишемо тако што на хоризонталној оси наносимо вредности обележја X (ако је X прекидног типа) или границе групних интервала (ако је X непрекидног типа). Уместо граница на оси x се могу наносити средине групних интервала. Код графикона апсолутних фреквенција на осу y се наносе вредности одговарајућих фреквенција за групни интервал или за вредности X и конструишу правоугаоници изнад групних интервала са висинама једнаким фреквенцијама. Ако се тачке које одговарају срединама групних интервала и фреквенција споје дужима добиће се тзв. ***полигон апсолутних фреквенција.*** Пример: Табела 2.6. и Слика 2.2.

X	f	P	C
2,25 – 2,35	2	0,01	2
2,35 – 2,45	23	0,11	25
2,45 – 2,55	49	0,23	74
2,55 – 2,65	63	0,29	137
2,65 – 2,75	45	0,21	182
2,75 – 2,85	25	0,12	207
2,85 – 2,95	3	0,01	210
2,95 – 3,05	4	0,02	214
Укупно	**214**	**1,00**	

Табела 2.6. Расподела обележја X



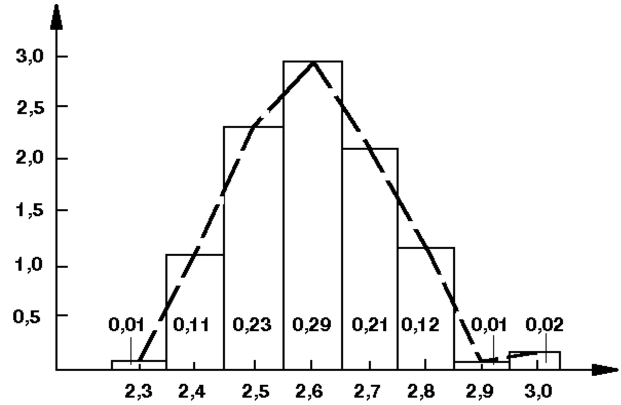
Слика 2.2. Полигон и расподела фреквенција обележја X

За приказивање расподела релативних фреквенција користе се ***хистограми***. Конструишу се тако што се на осу x наносе средине групних интервала, а затим се око тих тачака конструишу правоугаонци чије ће површине бити једнаке вредностима релативних фреквенција.

Ако је дужина i-тог интервала $d_i = a_i - a_{i-1}$, онда висина правоугаоника изнад тачке $x_i = \frac{a_i + a_{i-1}}{2}$, треба да је $h_i = \frac{p_i}{d_i}$, при чему је $i = 1,2, ...,k$, а p је одговарајућа релативна фреквенција.

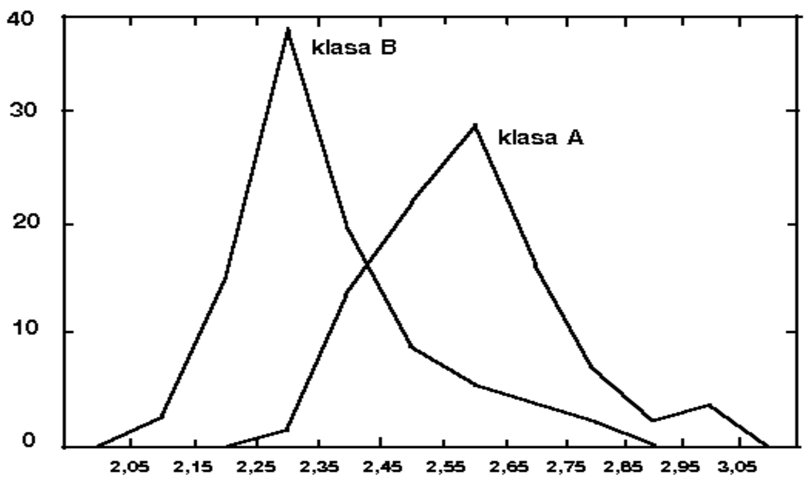
Тако се за податке из Табеле 2.6. добија хистограм приказан на Слици 2.3.

$$h_i = \frac{p_i}{0{,}1}$$



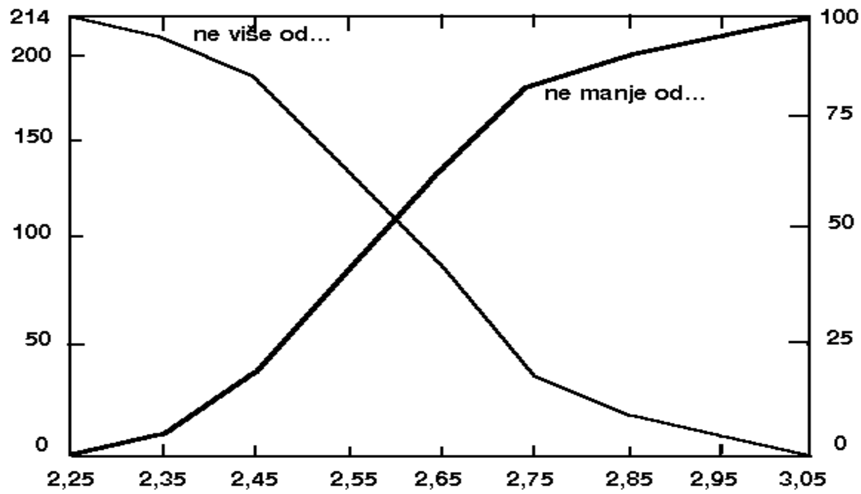
Слика 2.3. Хистограм релативних фреквенција

Хистограм је веома погодан за поређење расподела фреквенција једног обележја на више статистичких скупова. Тако се на пример са Слике 2.4, на којој су дати хистограми обележја X – две класе А и В, види да постоје разлике у расподелама. У класи А веће вредности обележја X су код великог броја елемената посматраног скупа, а у класи В је обрнуто.



Слика 2.4. Расподела обележја X на два скуng

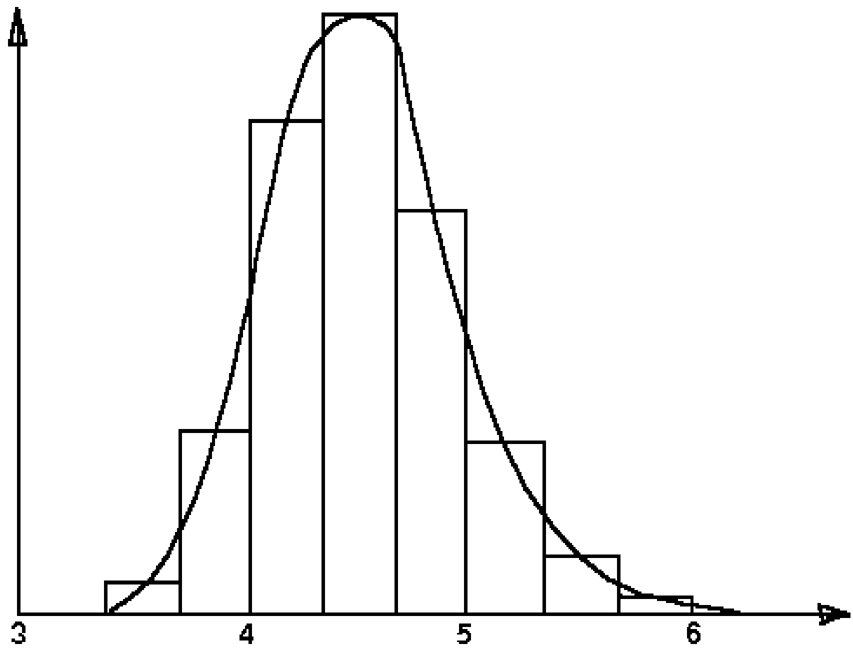
Кумулативне фреквенције се такође могу приказати и графички као што је дато на Слици 2.5. на којој су приказани подаци из Табеле 2.6.



Слика 2.5. Графикон кумулативних фреквенција

На основу хистограма који одговара расподели релативних фреквенција непрекидног обележја X, може се изабрати крива линија која ће апроксимирати посматрану расподелу фреквенција. Та крива линија се зове **закон вероватноће** *или* **функција густине** *и представља гранични облик хистограма ако се број елемената у статистичком скупу неограничено повећава, а дужине групних интервала на основу којих правимо хистограм неограничено смањују.*

Такав један закон вероватноћа дат је на Слици 2.6. Одређивање закона вероватноћа на основу података о вредностима у коначном статистичком скупу представља један од основних проблема аналитичке статистике. Кад се утврди тачно закон вероватноћа који одговара датим мерењима може се сматрати да су успешно решени скоро сви проблеми везани за одлучивања на основу посматраног обележја.



Слика 2.6. Закон вероватноћа

2.4. СРЕДЊЕ ВРЕДНОСТИ

Користећи релативне бројеве и расподелу фреквенција може се стећи известан глобални утисак о посматраној појави и посматраном статистичком скупу. Ипак за даљу и сврсисходнију анализу потребне су нам прецизније методе којима ћемо масу статистичких података обрадити тако да постане употребљива у процесу доношења одлука.

Анализу статистичких података можемо вршити тако што ћемо дефинисати извесне показатеље или параметре, чије ће нам вредности изражавати одређене сумарне карактеристике датих података. Вредности утврђених параметара омогућиће доношење закључака о одређеној појави или процесу који су изражени посматраним подацима.

Прва група таквих параметара су тзв. ***средње вредности*** или ***просеци.*** Веома често се користе и у свакодневном животу. Наиме, сваком је познато шта значи просечан лични доходак, просечна стопа раста, или просечна продуктивност и сл. Ови параметри показују неку централну вредност посматраног обележја X на елементима статистичког скупа. Пожељно је да средње вредности имају следеће особине:

- Ако су све вредности посматраног обележја X на статистичком скупу међусобно једнаке, онда и њихова средња вредност треба да је једнака тој вредности.
- У датом статистичком скупу постоји најмања и највећа вредност посматраног обележја X. Средња вредност треба да је већа од најмање, а мања од највеће вредности обележја X.
- Средња вредност зависи од свих вредности обележја X на целом статистичком скупу.

У ову групу параметара спадају: **аритметичка**, **хармонијска** и **геометријска средина**, затим **модус** и **медијана**.

Аритметичка средина

У свакодневном животу као средња вредност најчешће се користи аритметичка средина. Зато се често под појмом „просек” подразумева аритметичка средина.

Аритметичка средина *низа бројева је број који се добије када се њихов збир подели са укупним бројем чланова тог низа.*

Основни метод израчунавања аритметичке средине је исти за негруписане и груписане податке, а формуле се мало разликују.

Посматрајмо статистички скуп од N елемената на којима испитујемо квантитативно обележје X. Нека су утврђене следеће вредности обележја X код елемената статистичког скупа:

$$X_1, X_2, ..., X_N.$$

Тада је аритметичка средина обележја X једнака

$$\bar{x} = \frac{1}{N}\left(X_1 + X_2 + \ldots + X_N\right),$$

или краће написано

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{2.4}$$

при чему је $\bar{x}$ („икс бар”) ознака за аритметичку средину обележја X.

Ако је обележје X прекидног типа са k могућих различитих вредности

$$x_1, x_2, ..., x_k$$

тада су вредности обележја X на статистичком скупу дате расподелом фреквенција,

X	x_1 x_2 ... x_k
F	f_1 f_2 ... f_k

при чему је

$$\sum_{i=1}^{k} f_i = N \, .$$

Аритметичка средина овог обележја је

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot f_i \, , \qquad (2.5)$$

јер у изразу (2.4) међу бројевима X_1, X_2, ..., X_N који одговарају вредностима обележја X на елементима статистичког скупа, f_1 ових бројева имаће вредност x_1, f_2 ће имати вредност x_2, итд., f_k ће имати вредност x_k, па у њиховом укупном збиру треба свако x_i помножити са f_i. Тако се добија формула дата изразом (2.5).

И коначно претпоставићемо да је обележје X непрекидног типа и да су његове вредности на статистичком скупу дате расподелом фреквенција, при чему је

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2	...	f_k

$$\sum_{i=1}^{k} f_i = N \qquad (2.6)$$

Да би тачно одредили аритметичку средину за ово обележје X сматраћемо да сви елементи статистичког скупа код којих обележје X упада у један групни интервал, имају исту вредност. Најприхватљивије је да за ту вредност узмемо средину одговарајућег интервала.

Зато ћемо у случају кад је дужина првог и последњег групног интервала приближно једнака дужини осталих интервала, i-том интервалу придружити вредност обележја x_i која је једнака

$$x_i = \frac{a_{i-1} + a_i}{2} \qquad (2.7)$$

за $i = 1, 2, ..., k$.

Ако се први и последњи групни интервал знатно разликују по дужини од осталих интервала, или ако је неки од њих бесконачан, тада ћемо за x_1 и x_k узетих вредности

$$x_1 = a_1 - \frac{\overline{d}}{2}, \quad x_k = a_{k-1} + \frac{\overline{d}}{2}, \tag{2.8}$$

где је $\overline{d}$ просечна дужина преосталих (k-2) интервала. Тако ће се добити следећи низ тачака на реалној оси.

$$x_1 = a_1 - \frac{\overline{d}}{2} \qquad x_2 \qquad x_3 \qquad x_k = a_{k-1} + \frac{\overline{d}}{2}$$

$$a_0 = a \quad a_1 \quad a_2 \quad a_3 \qquad a_{k-1} \quad a_k = b$$

Аритметичка средина обележја X датог расподелом (2.6) једнака је

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{k} x_i \cdot f_i \tag{2.9}$$

при чему су вредности x_i одређене преко (2.7) и (2.8).

Аритметичку средину можемо одредити и из расподеле релативних фреквенција. С обзиром да је

$$p_i = \frac{f_i}{N},$$

за свако $i \in (1,2, ..., k)$, то се израз (2.5), односно (2.6) за аритметичку средину може написати и у облику

$$\overline{x} = \sum_{i=1}^{k} x_i \cdot p_i \tag{2.10}$$

Аритметичка средина има све три особине које су пожељне за једну средњу вредност. Заиста

1. Ако су све вредности обележја X на посматраном скупу једнаке, тј. ако је за свако i=1,2,...,N , $X_i = c$, тада је

$$\bar{x} = \frac{1}{N} N \cdot c = c ,$$

што значи да $\bar{x}$ има особину 1.

2. Нека је x најмања вредност обележја X у посматраном статистичком скупу. Лако је проверити да је $\bar{x} > x$. Ако је x' највећа вредност обележја X у посматраном скупу, тада је $\bar{x} < x'$, а то значи да $\bar{x}$ има особину 2, тј. већа је од најмање, а мања од највеће вредности посматраног обележја у датом скупу.
3. Из израза (2.4) којим је дефинисана аритметичка средина види се да постоји директна функционална зависност од свих вредности обележја *X*.

Поред ових особина аритметичка средина има и неке друге корисне особине.

Особина 1. *Збир одступања вредности обележја X на посматраном скупу од њихове аритметичке средине је једнако нули, тј.*

$$\sum_{i=1}^{N} (X_i - \bar{x}) = 0 , \qquad (2.11)$$

односно

$$\sum_{i=1}^{k} (x_i - \bar{x}) f_i = 0$$

кад је X дато расподелом фреквенција.

Доказ. Растављањем на два збира леве стране израза (2.11) добиће се

$$\sum_{i=1}^{N} (X_i - \bar{x}) = \sum_{i=1}^{N} X_i - N \cdot \bar{x} .$$

Са друге стране из израза којим је дефинисана аритметичка средина добија се

$$\sum_{i=1}^{N} X_i = N \cdot \bar{x}$$

тако да је

$$\sum_{i=1}^{N} (X_i - \bar{x}) = N \cdot \bar{x} - N \cdot \bar{x} = 0, \qquad (2.12)$$

а то је и требало проверити.

Особина 2. *Збир квадрата одступања вредности обележја X на елементима статистичког скупа од било ког броја* a *је најмањи ако је* $a = \bar{x}$ *, тј.*

$$\sum_{i=1}^{N} (X_i - a)^2 = \min, \qquad (2.13)$$

за $a = \bar{x}$.

Доказ. Означимо са $F(a)$ функцију која представља леву страну израза (2.13). Да би одредили минимум функције $F(a)$ можемо користити изводе. Али ћемо овде то урадити једноставније. Функција $F(a)$ је

$$F(a) = \sum_{i=1}^{N} (X_i - a)^2 \qquad (2.14)$$

Додавањем и одузимањем $\bar{x}$ у малој загради у функцији F(a), квадрирањем и растављањем на три збира добиће се

$$F(a) = \sum_{i=1}^{N} (X_i - \bar{x})^2 - 2\sum_{i=1}^{N} (X_i - \bar{x})(\bar{x} - a) + \sum_{i=1}^{N} (\bar{x} - a)^2,$$

односно

$$F(a) = \sum_{i=1}^{N} (X_i - \bar{x})^2 - 2(\bar{x} - a)\sum_{i=1}^{N} (X_i - \bar{x}) + N(\bar{x} - a)^2, \qquad (2.15)$$

Из особине 1. следи да је други члан у (2.15) једнак нули тако да је

$$F(a)=\sum_{i=1}^{N}(X_i-\overline{x})^2+N(\overline{x}-a)^2, \tag{2.16}$$

Први члан на левој страни израза (2.16) не зависи од a и има ненегативну вредност. Други члан је такође ненегативан и зависи од a. Минималну вредност функција $F(a)$ ће имати за оно a за које други члан има најмању вредност. То ће бити у случају кад је $\overline{x}-a=0$, односно $\overline{x}=a$.

Особина 3. *Ако између обележја X и Y која се испитују ђg једном статистичком скупу, постоји линеарна зависност, тада постоји иста таква зависност и између њихових аритметичких средина, тј. ако је*

$$Y=aX+b, \tag{2.17}$$

тада је

$$\overline{y}=a\overline{x}+b, \tag{2.18}$$

при чему су a *и* b *реални бројеви (*$a\neq 0$*),* $\overline{x}$ *и* $\overline{y}$ *аритметичка средина обележја X, односно Y.*

Доказ. Из линеарне зависности (2.17) следи да је за сваку вредност обележја X_i, одговарајућа вредност обележја Y_i једнака

$$Y_i=aX_i+b \tag{2.19}$$

Аритметичка средина обележја Y је

$$\overline{y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$$

одакле се после замене (2.19) добија

$$\overline{y}=\frac{1}{N}\left\{\sum_{i=1}^{N}aX_i+N\cdot b\right\},$$

односно

$$\overline{y} = a \cdot \frac{1}{N}\sum_{i=1}^{N} X_i + b \qquad (2.20)$$

Пошто је

$$\frac{1}{N}\sum_{i=1}^{N} X_i = \overline{x}$$

то значи да је из израза (2.20)

$$\overline{y} = a\overline{x} + b,$$

а то је и требало проверити.

Особина 4. *Претпоставимо да на једном статистичком скупу од N_1 елемената обележје X има аритметичку средину $\overline{x}_1$, а на другом скупу од N_2 елемената има аритметичку средину $\overline{x}_2$. Тада је аритметичка средина за X на оба статистичка скупа заједно једнака тзв.* ***пондерисаној аритметичкој средини***

$$\overline{x} = \frac{N_1\overline{x}_1 + N_2\overline{x}_2}{N_1 + N_2} \qquad (2.21)$$

Доказ. Ако са $X_1, X_2, ..., X_{N_1}$, односно $X_1^`, X_2^`, ..., X_{N_2}^`$ означимо вредности обележја X на елементима првог, односно другог скупа, аритметичка средина оба скупа биће једнака

$$X = \frac{1}{N_1 + N_2}\left\{\sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} X_i^`\right\} \qquad (2.22)$$

Пошто је

$$\sum_{i=1}^{N_1} X_i = N_1 \cdot \overline{x}_1 \text{ i } \sum_{i=1}^{N_2} X_i^` = N_2 \cdot \overline{x}_2 \qquad (2.23)$$

заменом у израз (2.22) добиће се (2.21), односно особина 4.

Особина 5. *Посматрајмо s статистичких скупова са N_1, N_2,..., N_S, елемената респективно и са аритметичким срединама $\overline{x}_1$, $\overline{x}_2$, ...$\overline{x}_s$.*

Тада је аритметичка средина обележје X на свим скуповима заједно једнака пондерисаној аритметичкој средини

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{s} N_i \cdot \bar{x}_i \,, \qquad (2.24)$$

при чему је N укупан број елемената свих статистичких скупова.

Доказ. Тачност израза (2.24) може се проверити на основу особине 4. користећи принцип математичке индукције.

Наведене особине аритметичке средине омогућавају у извесним случајевима знатно лакша и тачнија рачунања. Ако је обележје X непрекидног типа и дато расподелом фреквенција

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2		f_k

тада се погодним избором два броја A и d може увести смена

$$Y = \frac{X - A}{d} \qquad (2.25)$$

тако да се уместо бројева $x_1, x_2, ..., x_k$ добију мањи бројеви $y_1, y_2, ..., y_k$ који ће бити погодни за рачунање.

Кад су групни интервали једнаки, онда за d треба узети дужину групних интервала, а за A узети средину оног интервала који има највећу фреквенцију. У том случају вредности за Y ће бити цели бројеви 0, ±1, ±2,... па је лако за Y срачунати његову аритметичку средину. Из релације (2.25) и Особине 3. следи да је аритметичка средина X једнака

$$\bar{x} = d\bar{y} + A \,.$$

У пракси су чести случајеви кад нам је позната средина једног обележја, а желимо да одредимо средину неког другог обележја.

На пример, ако је просечни нето лични доходак у једном предузећу $\bar{x}$ = 50000 динара, какав би био просечан лични доходак после повећања? Нека је по завршном рачуну прихваћен предлог за 5%

повећања и плус 1500 динара за сваког радника за топли оброк. Колики ће бити просечан лични доходак? Ако је X садашњи лични доходак, а Y нови лични доходак, тада је

$$Y = 1{,}05X + 1500\,,$$

те се из *Особине 3*. добија да је

$$\bar{y} = 1{,}05\bar{x} + 1500\,,$$

односно

$$\bar{y} = 54000 \text{ динара.}$$

То је просечни лични доходак после повећања. Као што се види, за његово одређивање није нам било потребно рачунање личних доходака свих радника, па на основу тих рачунања утврђивање просека.

Често се статистички скупови састоје из тзв. ***стратума.*** *Ако је цео скуп подељен на подскупове који се међусобно иҗкључују, онда кажемо да је то стратификовани статистички скуп, а одговарајући подскупови су стратуми.*

За стратификовани скуп статистички подаци се прикупљају одвојено по стратумима и врши се њихова обрада такође по стратумима. Некада је анализа целог статистичког скупа немогућа, те се морају знати методе за обраду података које ће омогућити да се на основу резултата анализе по стратумима дође до резултата за цео скуп.

Геометријска средина

У анализама временских серија најпогоднија средња вредност је геометријска средина.

Геометријска средина *низа бројева је N-ти корен из производа његових чланова. Нека су X_1, X_2, ..., X_N вредности посматраног обележја X на елементима статистичког скупа. Геометријска средина обележја X је једнака*

$$G = \sqrt[N]{X_1 \cdot X_2 \cdot \ldots \cdot X_N} \qquad (2.26)$$

Кад је обележје X дато расподелом фреквенција,

X	x_1	x_2	...	x_k
f	f_1	f_2	...	f_k

геометријска средина је

$$G = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_k^{f_k}}\,, \tag{2.27}$$

јер се свака вредност обележја x_i појављује f_i пута.

Да би одредили геометријску средину за свако N вредности обележја X морају бити позитивне. Зато је и употреба геометријске средине ограничена само на она обележја која су позитивна.

Логаритмовањем израза (2.26) добиће се

$$\log G = \frac{1}{N} \sum_{i=1}^{N} \log X_i\,, \tag{2.28}$$

што значи да је логаритам геометријске средине једнак аритметичкој средини логаритама вредности обележја X на елементима статистичког скупа.

На сличан начин се и за обележје X које је дато расподелом фреквенција из (2.27) добија

$$\log G = \frac{1}{N} \sum_{i=1}^{k} f_i \cdot \log x_i \tag{2.29}$$

Из расподеле релативних фреквенција добиће се да је логаритам геометријске средине једнак

$$\log G = \sum_{i=1}^{k} p_i \cdot \log x_i\,, \tag{2.30}$$

при чему је за $i \in \{1,2, ...,k\}$

$$p_i = \frac{f_i}{N}\,.$$

Све особине аритметичке средине поседује и логаритам геометријске средине јер је то такође једна аритметичка средина.

За временску серију обележја X

$$X_1, X_2, ..., X_T$$

ланчани индекси чине низ бројева

$$\frac{X_2}{X_1}, \frac{X_3}{X_2}, ..., \frac{X_T}{X_{T-1}},$$

који показују узастопне промене обележја X. Да би добили средњу вредност тих промена за цео период T, одредићемо геометријску средину овог низа

$$G = \sqrt[T-1]{\frac{X_2}{X_1} \cdot \frac{X_3}{X_2} \cdot ... \cdot \frac{X_T}{X_{T-1}}} \tag{2.31}$$

После скраћивања под кореном у изразу (2.31) добија се

$$G = \sqrt[T-1]{\frac{X_T}{X_1}} \tag{2.32}$$

одакле се види да геометријска средина зависи од тзв. ***стопе раста*** у посматраном интервалу. Њено рачунање је доста једноставно. Довољно је познавати вредност обележја X на почетку и на крају временске серије.

Геометријска средина је погодна за одређивање просечних промена са непотпуним подацима, тј. за серије у којима не знамо вредности обележја X за све године. Тако се на пример на основу пописа становништва добијају вредности у годинама пописа, али се на основу тих вредности морају извршити процене вредности и за године између пописа. Прва анализа таквих података укључује у себе одређивање просечне стопе раста.

У Табели 2.7. дати су резултати пописа становништва бивше СФРЈ и одговарајуће геометријске средине.

Година пописа	Број становника X (у мил.)	log X	Логаритми ланчаних индекса	Геометријска средина
1948	15842	4,199814	-	-
1953	16991	4,230216	0,030402	1,014
1961	11549	4,268318	0,038102	1,011
1971	20523	4,312243	0,043925	1,010

Табела 2.7. Рачунање просечних стопа раста становништва бивше СФРЈ

Из последње колоне Табеле 2.7 види се да је просечна стопа раста у периоду 1948-1953. године била 1,4%, у периоду 1953-1961. године 1,1% а у периоду 1961-1971. године 1,0%.

Хармонијска средина

За анализу оних појава чији је интензитет обрнуто пропорционалан вредностима посматраног обележја, најпогоднија средња вредност је хармонијска средина. На пример, при испитивању продуктивности мери се време потребно за израду неког производа, али је продуктивност обрнуто пропорционална том времену, па зато у овом случају треба користити хармонијску средину као средњу вредност.

Хармонијска средина *низа бројева је реципрочна вредност аритметичке средине реципрочних вредности чланова тог низа. Ако су*

$$X_1, X_2, \ldots X_N$$

вредности обележја X на N елемената статистичког скупа тада је хармонијска средина једнака

$$H = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \ldots + \frac{1}{X_N}}, \tag{2.33}$$

односно

$$H = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}}.$$

Из (2.33) добија се реципрочна вредност хармонијске средине

$$\frac{1}{H} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{X_i}, \tag{2.34}$$

што значи да је реципрочна вредност хармонијске средине једнака аритметичкој средини реципрочних вредности посматраног обележја *X*.

За обележје *X* дато расподелом фреквенција добиће се следећи израз за хармонијску средину

$$\frac{1}{H} = \frac{1}{N} \sum_{i=1}^{k} \frac{f_i}{x_i} \tag{2.35}$$

а преко релативних фреквенција хармонијска средина ће се добити из формуле

$$\frac{1}{H} = \frac{1}{N} \sum_{i=1}^{k} \frac{1}{x_i} \cdot p_i \tag{2.36}$$

На пример, да би се одредило потребно време за вожњу аутобуса од места А до места В извршено је 40 пробних вожњи и добијени су следећи подаци:

Утрошено време	4	5	6	10
Број вожњи	5	5	20	10

За даљу анализу потребно је одредити средњу вредност за ова мерења. Како је време обрнуто пропорционално жељеном ефекту вожње, то је хармонијска средина најпогоднија средња вредност. Тако ћемо добити

$$\frac{1}{H}=\frac{1}{40}\left(\frac{1}{4}\cdot 5+\frac{1}{5}\cdot 5+\frac{1}{6}\cdot 20+\frac{1}{10}\cdot 10\right),$$

односно H = 6,08 па у даљој анализи као показатељ „просечно” потребног времена треба узети ову вредност. Интересантно је поредити аритметичку, геометријску и хармонијску средину.

То поређење омогућава тзв.

Кошијева (Cauchy) теорема: *Хармонијска средина је мања од геометријске, а геометријска је мања од аритметичке средине, тј.*

$$H < G < m \tag{2.37}$$

Модус и медијана

У многим испитивањима интересују нас вредности обележја X које се најчешће појављују у посматраном скупу. На пример, ако се посматра једна телефонска централа онда нас интересује који је најчешћи број позива у посматраном временском интервалу. Или нас може интересовати величина личног дохотка коју има највећи број запослених и сл. Параметар обележја X који изражава ову особину зове се *модус*.

Модус *је она вредност обележја X која има највећу фреквенцију у посматраном статистичком скупу или она вредност у чијој се околини најчешће појављују „измерене” вредности X на статистичком скупу***.**

Модус је добар показатељ средње вредности за оне статистичке скупове који су хомогени. Највише се користи у процесу доношења разних пословних одлука. Кад је обележје X груписано и дато расподелом фреквенција, за модус се узима средина оног групног интервала који има највећу фреквенцију.

При испитивањима једног статистичког скупа можемо се интересовати за ону средњу вредност обележја X која дели цео скуп на два једнака дела.

Медијана *је она вредност обележја X која дели уређени статистички скуп на два једнака дела.*

Медијану ћемо означавати са M_e и то је она вредност обележја X за коју пола елемената статистичког скупа има обележје X не веће од M_e, а друга половина има вредности које нису мање од M_e.

Нека су

$$X_1, X_2, \ldots, X_N \tag{2.38}$$

измерене вредности обележја X на N елемената статистичког скупа. Да бисмо одредили медијану прво ћемо вредности (2.38) поређати у неопадајући низ

$$X_1^` \le X_2^` \le \ldots \le X_N^` .$$

Медијана је једнака

$$M_e = \begin{cases} X'_{\frac{N+1}{2}} & \text{ако је } N \text{ непаран број.} \\ \frac{1}{2}\left(X'_{\frac{N}{2}} + X'_{\frac{N}{2}+1} \right) & \text{ако је } N \text{ паран број.} \end{cases} \tag{2.39}$$

За обележје X које је груписано и дато расподелом фреквенција медијану одређујемо тако што прво одредимо групни интервал у коме се налази медијана M_e, а затим ћемо по договору изабрати једну вредност из тог интервала за медијану.

Нека је обележје X дато расподелом.

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2		f_k

За одређивање интервала медијане, одредићемо онај индекс s за који ће бити задовољене следеће две једначине

$$\sum_{i=1}^{s} f_i \le \frac{N}{2}$$

$$\sum_{i=1}^{s+1} f_i > \frac{N}{2}$$

За медијану ћемо узети вредност обележја X у интервалу $a_s - a_{s+1}$ тако што ћемо на леву границу тог интервала додати једну вредност која ће зависити од дужине тог интервала, фреквенције и разлике у збиру фреквенција до тог интервала и половине елемената статистичког скупа.

Тако ћемо добити

$$M_e = a_s + \frac{a_{s+1} - a_s}{f_{s+1}} \left(\frac{N}{2} - \sum_{i=1}^{s} f_i \right) \qquad (2.40)$$

На основу података за 429 саобраћајних предузећа у једној години, посматраћемо број запослених и одредити медијану. Подаци и резултати су дати у Табели 2.8.

Број запослених X	Број предузећа f	Кумулативне фреквенције c
до 15	22	22
15 - 30	10	32
30 - 60	27	59
60 - 125	68	127
125 - 250	77	204
250 - 500	100	304
500 - 1000	65	369
1000 - 2000	38	407
преко 2000	22	429
Укупно	**429**	

Табела 2.8. Број запослених у саобраћајним предузећима

Из колоне за кумулативне фреквенције види се да је

$$\sum_{i=1}^{5} f_i = 204 < 214{,}5 \quad \sum_{i=1}^{6} f_i = 304 > 214{,}5 \, ,$$

тако да је s = 5, а медијана је

$$M_e = 250 + \frac{500 - 250}{100}\left(\frac{429}{2} - 204\right)$$

односно $M_e = 276{,}25$.

2.5. МЕРЕ ВАРИЈАБИЛИТЕТА

Другу групу параметара којима се описују резултати посматраног обележја *X* на датом статистичком скупу чине параметри који изражавају варијабилитет посматраног обележја. То су они показатељи који зависе од укупних разлика у вредностима обележја X на елементима статистичког скупа. Та група параметара су тзв. **мере варијабилитета** или **мере дисперзије**.

У мере варијабилитета спадају следећи параметри:

- Размак варијације,
- Квартилна девијација,
- Средња девијација,
- Варијанса и стандардна девијација и
- Коефицијент варијације.

Вредности параметара који спадају у мере варијабилитета треба да укажу на две чињенице. Прво, њихов износ говори колико су средње вредности обележја *X* добри представници свих његових вредности на посматраном скупу, друго, ови параметри треба да покажу колико се сви елементи на посматраном скупу међусобно разликују у односу на дато обележје *X*.

Ако је варијабилитет посматраног обележја мали, онда то значи да средња вредност може бити добар представник посматраног обележја, а ако је његов варијабилитет велики то значи да ће средња вредност бити лош представник тог обележја.

На пример, ако се утврди да је варијабилитет у личним примањима запослених велики, онда то значи да ће просечна лична примања бити лош показатељ примања појединих радника. Исто тако, ако се утврди за домаћинства у Београду да је варијабилитет њихових издатака за основне животне потребе (стан, храна и сл.) јако мали, то ће значити да ће просечни издаци домаћинства бити доста добар показатељ издатака свих домаћинстава у Београду.

Са друге стране, ако се зна колики је варијабилитет неке појаве може се утицати на ту појаву у циљу њене стабилизације. На пример, промене, односно варијабилитет у температури, пулсу, крвном притиску и сл. служе као основни показатељи у постављању дијагнозе и одређивању терапије код разних обољења.

Размак варијације

Размак варијације је најједноставнија мера варијабилитета самим тим и најлакше га је одредити. То је размак између највеће и најмање вредности обележја X на посматраном статистичком скупу.

Означимо са X_{MAX} највећу вредност обележја X, а са X_{MIN} његову најмању вредност на статистичком скупу. Тада је ***размак варијације*** *једнак*

$$R = X_{MAX} - X_{MIN} \tag{2.41}$$

где смо са R означили размак варијације.

Нека је обележје X груписано и дато расподелом фреквенција,

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2		f_k

(2.42)

при чему је

$$\sum_{i=1}^{k} f_i = N.$$

Да бисмо тачно одредили размак варијације морамо поред расподеле (2.42) знати најмању вредност обележја X у првом интервалу a_0 - a_1 и највећу вредност обележја X у последњем интервалу $a_{k-1} - a_k$. Ако нам ове вредности нису познате не можемо ни одредити размак варијације. Размак варијације је погодна мера варијабилитета за она обележја код којих се највећа и најмања вредност не разликују много од осталих вредности обележја X. Исто тако, то је важан показатељ за оне појаве које су описане великим и малим вредностима обележја.

На пример у временским извештајима важна је разлика између највише и најниже температуре, или између највишег и најнижег водостаја. Највећу примену размак варијације има у контроли квалитета у индустријској производњи.

Овај показатељ је лоша мера варијабилитета за оне статистичке скупове код којих постоји мали број елемената са малим или са великим вредностима обележја X, док се вредности X код осталих елемената не разликују много. Тако би на пример у посматрањима цена свих приватних кућа у граду, размак варијација био лош показатељ варијабилитета јер су поједине луксузне куће веома скупе у односу на најјефтиније.

Квартилна девијација

Да би елиминисали утицај екстремних вредности у посматраном статистичком скупу можемо податке редуковати, тако што ћемо из посматрања искључити оне елементе статистичког скупа код којих обележје има те вредности. Од укупног броја елемената статистичког скупа треба искључити један одређен проценат. Најчешће то чинимо преко квартила.

Означимо са $X_{0,25}$ ону вредност обележја X за коју 25% елемената посматраног скупа има X мање од $X_{0,25}$. Са $X_{0,75}$ означимо ону вредност обележја X за коју 75% елемената статистичког скупа има вредност X мању од $X_{0.75}$, или ону вредност за коју 25% елемената има вредност X већу од $X_{0.75}$.

Вредност $X_{0.25}$ назива се **доњи квартил** *обележја X, а $X_{0.75}$* **горњи квартил** *обележја X.*

Доњи и горњи квартил деле цео статистички скуп на три дела. Једна четвртина елемената су они елементи код којих је $X < X_{0.25}$, друга четвртина су елементи код којих је $X > X_{0.75}$ а преосталих 50% елемената има вредност обележја X која је $X_{0.25} < X < X_{0.75}$

Квартилна девијација Q је број одређен изразом

$$Q = \frac{X_{0,75} - X_{0,25}}{2}. \qquad (2.43)$$

Доњи и горњи квартили се рачунају слично као и медијана. Ако су вредности обележја X код N елемената статистичког скупа $X_1, X_2, ..., X_N$ поређане по величини, треба одредити оне вредности у низу $X_1, X_2, ..., X_N$ за које је четвртина низа мања од $X_{0.25}$, а четвртина већа од $X_{0.75}$.

Аналогно одређивању медијане за груписане податке одредићемо и квартиле. Нека је обележје X дато расподелом,

X	$a_0 - a_1$	$a_1 - a_2$	...	$a_{k-1} - a_k$
f	f_1	f_2		f_k

при чему је

$$\sum_{i=1}^{k} f_i = N.$$

За одређивање доњег квартила одредићемо прво групни интервал у коме ће се налазити $X_{0.25}$. То ће бити интервал са редним бројем s за који су испуњене следеће неједначине

$$\sum_{i=1}^{s} f_i < \frac{N}{4} \qquad \sum_{i=1}^{s+1} f_i \geq \frac{N}{4}. \qquad (2.44)$$

Тада је доњи квартил

$$X_{0,25} = a_s + \frac{a_{s+1} - a_s}{f_{s+1}} \left(\frac{N}{4} - \sum_{i=1}^{s} f_i\right). \qquad (2.45)$$

На сличан начин одређујемо и горњи квартил. Прво одредимо интервал са редним бројем p за који су испуњене неједначине

$$\sum_{i=1}^{p} f_i < \frac{3N}{4} \qquad \sum_{i=1}^{p+1} f_i \geq \frac{3N}{4}.$$

Горњи квартил је

$$X_{0,75} = a_p + \frac{a_{p+1} - a_p}{f_{p+1}} \left(\frac{3N}{4} - \sum_{i=1}^{p} f_i\right) \tag{2.46}$$

Квартиле ћемо најлакше израчунати ако се користимо табелом података и кумулативним фреквенцијама.

Посматрајмо пример расподеле броја запослених у угоститељским предузећима. Расподела је дата у Табели 2.9.

Број запослених X	Број предузећа, f	Кумулативне фреквенције, c
- 6	73	73
6- 15	51	124
15- 30	45	169
30- 60	63	232
60- 125	72	304
125- 250	54	358
250- 500	30	388
500-1000	14	402
1000-2000	2	404
Укупно	404	

Табела 2.9. Расподела броја запослених у угоститељству

Овде је N= 404, па је N/4= 101 тако да се из треће колоне Табеле 2.9. види да је доњи квартил у интервалу 6-15. На основу обрасца (2.45) добиће се

$$X_{0,25} = 6 + \frac{15-6}{51}(101-73) = 11.$$

Горњи квартил се налази у интервалу 60-125, и на основу формуле (2.46) има вредност

$$X_{0,75} = 60 + \frac{125-60}{72}\left(303-232\right).$$

Квартилна девијација за број запослених на основу (2.43) има вредност Q=25,5.

Варијанса и стандардна девијација

Видели смо да варијабилитет обележја X можемо изразити преко одступања појединих његових вредности од средње вредности. Та одступања могу бити позитивна или негативна, а за укупан варијабилитет није нам важан знак одступања. Зато можемо посматрати квадрате тих одступања и на основу њих дефинисати параметар варијабилитета. Тако се дефинишу *варијанса и стандардна девијација*.

Варијанса *је аритметичка средина квадрата одступања вредности обележја X од њихове аритметичке средине. Позитивна вредност корена варијансе назива се* **стандардна девијација**.

Нека су вредности обележја X на N елемената статистичког скупа

$$X_1, X_2, ..., X_N.$$

Тада је низ квадрата одступања вредности обележја X од његове аритметичке средине

$$\left(X_1-\overline{x}\right)^2, \left(X_2-\overline{x}\right)^2, ..., \left(X_N-\overline{x}\right)^2. \tag{2.47}$$

Аритметичка средина низа (2.47) јЛ

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{x})^2, \tag{2.48}$$

и назива се **варијанса** *обележја X, а*

$$S = +\sqrt{S^2} \tag{2.49}$$

је његова **стандардна девијација**.

Нека је обележје X груписано и дато расподелом фреквенција,

X	$a_0 - a_1$ x_1	$a_1 - a_2$ x_2	… …	$a_{k-1} - a_k$ x_k
f	f_1	f_2		f_k

тада је варијанса обележја *X* дата изразом

$$S^2 = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot f_i \, , \tag{2.50}$$

Варијанса обележја X се може изразити и преко релативних фреквенција

$$S^2 = \sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot p_i \; . \tag{2.51}$$

Ради тачности рачунања и прегледности варијансу и стандардну девијацију треба рачунати преко табеле (Табела 2.10).

X	f	$X \cdot f$	$(X - \overline{x})$	$(X - \overline{x})^2 \cdot f$
x_1	f_1	$x_1 \cdot f_1$	$x_1 - \overline{x}$	$(x_1 - \overline{x})^2 \cdot f_1$
x_2	f_2	$x_2 \cdot f_2$	$x_2 - \overline{x}$	$(x_2 - \overline{x})^2 \cdot f_2$
…	…	…	…	…
x_k	f_k	$x_k \cdot f_k$	$x_k - \overline{x}$	$(x_k - \overline{x})^2 \cdot f_k$
Укупно		$\sum_{i=1}^{k} x_i f_i$		$\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot f_i$

Табела 2.10. Рачунање варијансе

Из података се прво попуне прва и друга колона табеле 2.10. Трећа колона се добија множењем прве две колоне и из њеног збира се одређује $\overline{x}$. Затим се од прве колоне одузима $\overline{x}$ и добија се четврта колона. Квадрирањем четврте колоне и множењем са другом добија се пета колона чији збир треба поделити са N и на тај начин је одређена варијанса S^2.

Видели смо да је аритметичка средина имала следећу особину: збир квадрата одступања вредности обележја X од било ког броја најмањи је за $a = \overline{x}$. Зато се аритметичка средина $\overline{x}$ користи за одређивање варијансе и стандардне девијације обележја X.

Навешћемо неке особине варијансе, које ће омогућити њену употребу у даљој статистичкој анализи и које ће олакшати рачунања.

Особина 1. *Варијанса обележја X је једнака нули, ако и само ако је обележје X константно на свим елементима статистичког скупа, тј.*

$$\left\{S^2 = 0\right\} \Leftrightarrow \left\{X_i = c, \forall i \in [1, 2, N]\right\}.$$

Доказ. Пошто је варијанса збир позитивних бројева, онда ће она имати вредност нулу ако и само ако је сваки тај број једнак нули. То ће бити испуњено једино кад је $X = c$.

Особина 2. *Варијанса обележја X се може изразити формулом*

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{x}^2, \tag{2.52}$$

а ако је обележје X груписано, варијанса је

$$S^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 f_i - \overline{x}^2. \tag{2.53}$$

Доказ. Пошто је варијанса дата изразом

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{x}\right)^2,$$

после квадрирања и растављања десне стране на три сабирка добија se

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \frac{2}{N}\sum_{i=1}^{N} X_i \cdot \overline{x} + \frac{1}{N}\sum_{i=1}^{N}\overline{x}^2 \tag{2.54}$$

У другом члану израза (2.54) $\overline{x}$ не зависи од индекса и па се може извући пред знак сигма, а у трећем члану збир је једнак $N \cdot \overline{x}$. Пошто је

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} X_i \ ,$$

то се после замене (2.54) своди на

$$S^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - 2 \cdot \overline{x} \cdot \overline{x} + \frac{1}{N} N\overline{x}^2$$

одакле се после скраћивања и сабирања добија образац (2.52).

Особина 3. *Нека обележје X има варијансу* S_x^2*, а обележје Y има варијансу* S_y^2 *и нека између X и Y постоји линеама веза облика*

$$Y = aX + b. \tag{2.55}$$

Тада је варијанса обележја Y једнака

$$S_y^2 = a^2 S_x^2. \tag{2.56}$$

Доказ. За сваку вредност обележја $X = X_i$, из релације (2.55) добиће се одговарајућа вредност обележја Y

$$Y_i = aX_i + b\,. \tag{2.57}$$

Аритметичка средина обележја *Y* је

$$\overline{y} = a\overline{x} + b\,. \tag{2.58}$$

Замењујући (2.57) и (2.58) у израз за варијансу обележја Y добиће се

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{y}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(aX_i - a\overline{x}\right)^2$$

тј.

$$S_y^2 = a^2 \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \overline{x}\right)^2$$

односно израз (2.56), што доказује Особина 3.

Особина 4. *Претпоставимо да један статистички подскуп има N_1 елемената са аритметичком средином $\overline{x}_1$ и варијансом обележја X једнаком S_1^2. Нека други статистички подскуп има N_2 елемената, аритметичку средину $\overline{x}_2$ и варијансу S_2^2. Тада је варијанса статистичког скупа једнака*

$$S^2 = \frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2} + \frac{N_1 N_2}{(N_1 + N_2)^2}(\overline{x}_1 - \overline{x}_2)^2 \tag{2.59}$$

Доказ. Означимо са

$$X_1, X_2, ..., X_{N_1}$$

вредност обележја X на првом подскупу, а са

$$X_1^{`}, X_{2}^{`}{}_2, ..., X_{N_2}^{`}$$

његове вредности на другом подскупу.

Аритметичка средина целог скупа једнака је пондерисаној аритметичкој средини

$$\overline{x} = \frac{N_1\overline{x}_1 + N_2\overline{x}_2}{N_1 + N_2}, \tag{2.60}$$

а његова варијанса је

$$S^2 = \frac{1}{N_1 + N_2}\left(\sum_{i=1}^{N_1}\left(X_2 - \overline{x}_1\right)^2 + \sum_{i=1}^{N_2}\left(X_i^{`} - \overline{x}\right)^2\right) \tag{2.61}$$

Додавањем и одузимањем $\overline{x}_1$ у загради код првог члана на десној страни израза (2.61), а у другом члану додавањем и одузимањем $\overline{x}_2$ добиће се

$$S^2 = \frac{1}{N_1 + N_2}\left(\sum_{i=1}^{N_1}\left(X_2 - \overline{x}_1\right)^2 + \sum_{i=1}^{N_2}\left(X_i^{`} - \overline{x}\right)^2 + N_1\left(x_1 - \overline{x}\right)^2 + N_2\left(\overline{x}_2 - \overline{x}\right)^2\right) \tag{2.62}$$

Из (2.60) се добија

$$\overline{x}_1 - \overline{x} = \frac{N_2\left(\overline{x}_1 - \overline{x}_2\right)}{N_1 + N_2} \qquad \overline{x}_2 - \overline{x} = \frac{N_1\left(\overline{x}_2 - \overline{x}_1\right)}{N_1 + N_2}$$

па се после замене у (2.62) добија

$$S^2 = \frac{1}{N_1 + N_2}\left(N_1 S_1^2 + N_2 S_2^2 + \frac{N_1 N_2}{N_1 + N_2}\left(\overline{x}_1 - \overline{x}_2\right)^2\right)$$

јер је

$$\sum_{i=1}^{N_1}\left(X_i - \overline{x}_1\right)^2 = N_1 S_1^2 \qquad \sum_{i=1}^{N_2}\left(X_i^{`} - \overline{x}_2\right)^2 = N_2 S_2^2$$

Тиме је и доказана особина 4.

Ову особину можемо уопштити и на случај кад се статистички скуп састоји од више подскупова.

Особина 5. *Нека се статистички скуп састоји од k подскупова величина N_1, N_2, ... , N_k чије аритметичке средине су $\overline{x}_1, \overline{x}_2, ..., \overline{x}_K$, а варијансе*

$$S_1^2, S_2^2, ..., S_K^2.$$

Тада је варијанса статистичког скупа једнака

$$S^2 = \frac{1}{N}\sum_{i=1}^{K} N_i S_i^2 + \frac{1}{N}\sum_{i=1}^{K} N_i\left(\overline{x}_i - \overline{x}\right)^2,$$

при чему је $N = \sum_{i=1}^{k} N_i$ величина статистичког скупа, а $\overline{x}$ је његова аритметичка средина.

Доказ. Применом принципа математичке индукције лако се из Особине 4 добија Особина 5.

Коефицијент варијације

Често смо у ситуацији да треба да поредимо варијабилитет два различита обележја. На пример варијабилитет у висини и тежини ученика. Пошто су сви до сада наведени параметри варијабилитета изражени у јединицама у којима је мерено посматрано обележје, онда немогуће је вредности тих параметара поредити за обележја која су мерена у различитим јединицама.

За оваква поређења погоднији су релативни показатељи, какав је *коефицијент варијације.*

Коефицијент варијације *је процентуално изражен однос стандардне девијације и аритметичке средине посматраног обележја X.*

Нека је $\bar{x}$ *аритметичка средина обележја X, а S његова стандардна девијација. Тада је* **коефицијент варијације V** *једнак*

$$V = 100 \cdot \frac{S}{\bar{x}} \tag{2.63}$$

На пример, нека је на једном скупу студената мерена њихова висина и тежина. Нека X представља висину, а Y тежину и нека је аритметичка средина висине $\bar{x}$, а S_x, њена стандардна девијација. Даље, нека је аритметичка средина тежине $\bar{y}$, а њена стандардна девијација S_y. Претпоставимо да су се из резултата мерења добиле следеће вредности

$\bar{x}$ = 174 cm; S_x = 11 cm, $\bar{y}$ = 69 kg; S_y = 7 kg.

Интересује нас да ли је код студената већи варијабилитет у тежини од варијабилитета у висини или обрнуто.

Пошто су S_x и S_y изражени у различитим јединицама, немогуће је њихово поређење. Зато треба одредити коефицијенте варијације.

За X се добија

$$V_1 = 100 \cdot \frac{11}{174} = 6{,}32\,\%,$$

а за Y

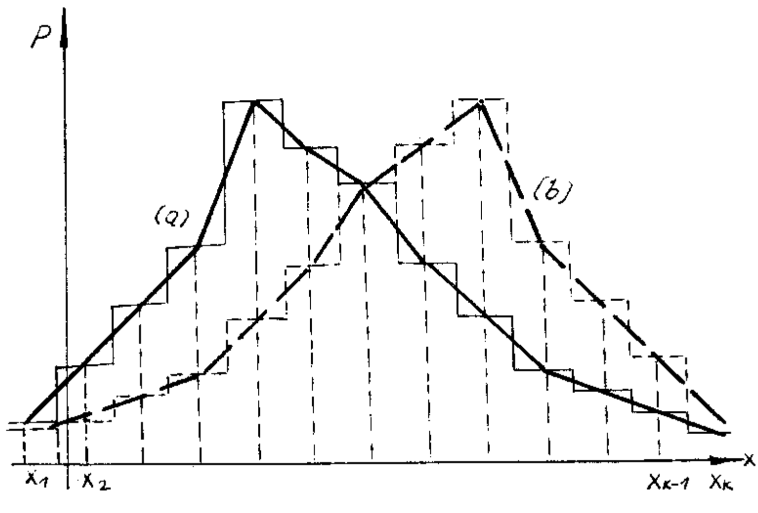
$$V_2 = 100 \cdot \frac{7}{169} = 10{,}14\,\%,$$

па се може закључити да је за посматрани скуп студената већи варијабилитет у тежини од варијабилитета у висини.

2.6. КОЕФИЦИЈЕНТИ АСИМЕТРИЈЕ И СПЉОШТЕНОСТИ

Битна карактеристика расподеле обележја X на статистичком скупу је симетричност. Може се десити да су поједине вредности обележја X симетрично распоређене у односу на неку одређену вредност. Мање вредности обележја X, међутим, могу имати мање фреквенције у односу на веће вредности или обрнуто. На Слици 2.7. су приказане две расподеле. Код расподеле (а) мање вредности обележја X имају веће фреквенције, а код расподеле (b) је обрнуто.

За обе расподеле кажемо да су асиметричне.



Слика 2.7. Асиметричне расподеле обележја X

За мерење асиметричности расподеле обележја X служи једноставан параметар који је базиран на разлици између аритметичке средине $\bar{x}$ и медијане M_e. Наиме, ако је расподела обележја X симетрична, онда су аритметичка средина и медијана једнаке. А кад постоји асиметричност онда ће постојати и разлика између $\bar{x}$ и M_e. Та разлика је већа што је асиметричност већа јер медијана зависи само од фреквенција и помераће се ка вредностима за X које имају веће фреквенције. Аритметичка средина зависи и од фреквенција и од вредности обележја X па ће њено померање ка вредностима обележја X које имају веће фреквенције, бити спорије у односу на померање медијане.

Коефицијент асиметрије *је количник*

$$K_A = \frac{3(\bar{x} - M_e)}{S}, \tag{2.64}$$

при чему је $\bar{x}$ *аритметичка средина,* M_e *медијана, а S стандардна девијација обележја X.*

Коефицијент асиметрије је једнак нули за симетричне расподеле. Ако је расподела *асиметрична у леву страну* (расподела (a) на Слици 2.7) коефицијент асиметрије *K_A је позитиван*, а ако је расподела *асиметрична у десну страну* (расподела (b) на Слици 2.7) коефицијент асиметрије *K_A је негативан.* Напоменимо да је коефицијент асиметрије K_A релативни број без димензије који нам омогућава поређења расподела обележја X и Y мерених у различитим јединицама мере.

Мада је коефицијент асиметричности веома једноставан за рачунање ипак се у пракси чешће користи тзв. Пирсонов (Pearson) први коефицијент као мера асиметрије. Дефинисан је на одступањима појединих вредности обележја X од њихове аритметичке средине.

Одредимо прво тзв. трећи централни моменат. То је аритметичка средина низа

$$(X_1 - \bar{x})^3, (X_2 - \bar{x})^3, \ldots, (X_N - \bar{x})^3$$

који представља трећи степен одступања појединих вредности обележја X од њихове аритметичке средине.

Трећи централни моменат *је једнак*

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{x})^3 \tag{2.65}$$

Ако је обележје X груписано и дато расподелом фреквенција

X	x_1	x_2	...	X_k
f	f_1	f_2	...	f_k

тада је трећи централни моменат једнак

$$\mu_3 = \frac{1}{N}\sum_{i=1}^{k}\left(x_i - \bar{x}\right)^3 \cdot f_i\ , \tag{2.66}$$

или преко релативних фреквенција

$$\mu_3 = \sum_{i=1}^{k}\left(x_i - \bar{x}\right)^3 \cdot p_i\ . \tag{2.67}$$

Први Пирсонов коефицијент *је дефинисан преко трећег централног момента и стандардне девијације обележја X као количник*

$$\beta_1 = \frac{\mu_3}{S^3}. \tag{2.68}$$

Ако је расподела обележја X симетрична, коефицијент β_1 је једнак нули. Усвојићемо и обрнуто: ако је $\beta_1 = 0$ рећи ћемо да је расподела обележја X симетрична. Ако је расподела асиметрична у леву страну коефицијент β_1 је позитиван, а ако је расподела асиметрична у десну страну коефицијент β_1 је негативан. што је асиметричност већа то је апсолутна вредност коефицијента β_1 већа.

Пример: Нека је обележје X дато расподелом

X	-6	-5	-4	-3	-2	-1	0	1,5	3,5	5,5	7,5	9,5
f	7	15	34	19	19	19	6	9	10	6	1	3

Израчунати коефицијент асиметрије!

Табела 2.10. Рачунање коефицијента асиметрије

X	f	C	Xf	$(X-\bar{x})$	$(X-\bar{x})^3$	$((X-\bar{x})^3 \cdot f)$	$(X^2 f)$
-6	7	7	-4.2	-4.12	-68.92	-482.45	252
-5	25	32	-125	-3.1	-29.79	-744.78	625
-4	34	66	-136	-2.1	-9.26	-314.87	544
-3	19	85	-57	-1.1	-1.33	-25.29	171
-2	19	104	-38	-0.1	-0.00	-0.01	76
-1	19	123	-19	0.9	0.73	13.85	19
0	6	129	0	1.9	6.86	41.16	0
1.5	9	138	13.5	3.4	39.30	353.74	20.25
3.5	10	148	35.0	5.4	157.46	1574.64	122.50
5.5	6	154	33.0	7.4	405.22	2431.34	181.50
7.5	1	155	7.5	9.4	830.58	830.58	56.25
9.5	3	158	28.5	11.4	1481.54	4444.63	270.75
Укупно	158		-299.5			8122.54	2338.25

Из треће колоне Табеле 2.10. се види да медијана обележја X припада групном интервалу (-3,5; -2,5) тако да је медијана једнака

$$M_e = -3{,}5 + \frac{1}{19}(79 - 66) = -2{,}816 .$$

Из четврте колоне се добија да је аритметичка средина обележја X једнака

$$\bar{x} = \frac{1}{158}(-229{,}5) = -1{,}896 .$$

Варијансу и стандардну девијацију добијаћемо из последње колоне

$$S^2 = \frac{1}{158} 2338,25 - (-1,896)^2 = 11,204\,,$$

тако да је стандардна девијација $S = 3,347$.

Коефицијент симетрије K_A има вредност

$$K_A = \frac{3(-1.896 + 2.816)}{3.347} = 0.824\,.$$

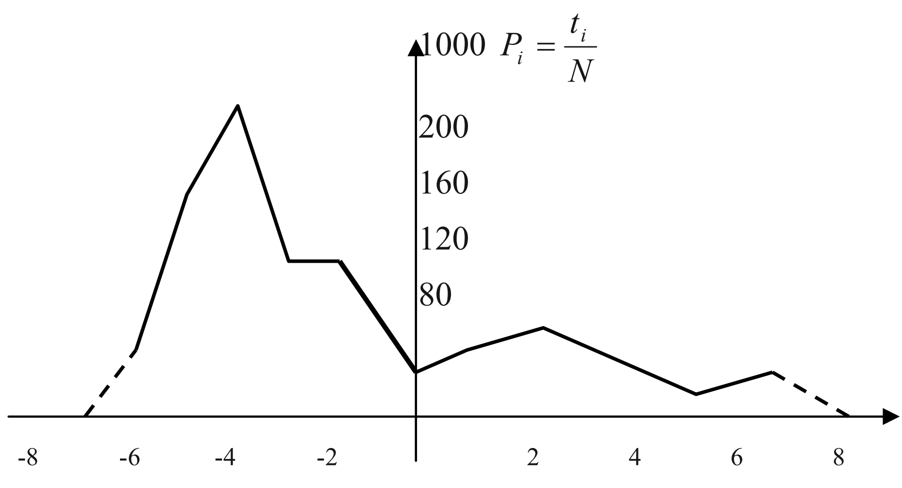
Из Табеле 2.10. добија се трећи централни моменат

$$\mu_3 = \frac{1}{158} 8122.54 = 51.41\,,$$

па је први Пирсонов коефицијент једнак

$$\beta_1 = \frac{51.411}{3.347^3} = 1.37\,.$$

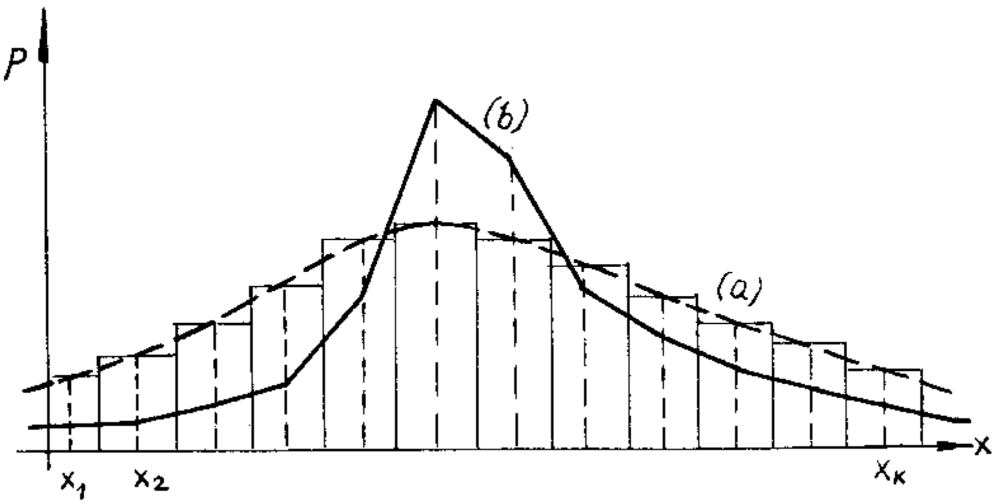
Као што се види оба коефицијента су позитивна, што значи да је расподела асиметрична у лево. То се види и на слици 2.8. на којој је приказана расподела посматраног обележја X.



Слика 2.8. Расподела релативних фреквенција за X

Мада два обележја X и Y могу имати исту асиметричност или чак бити потпуно симетрична, њихове расподеле могу бити сасвим различите. На Слици 2.9 приказане су две такве расподеле. Оне се разликују у спљоштености.

Коефицијент спљоштености *служи као мера спљоштености једне расподеле. Дефинише се на основу тзв. четвртог централног момента.*



Слика 2.9. Расподеле са различитим спљоштеностима

Посматрајмо одступања појединих вредности обележја X од њихове аритметичке средине и низ

$$\left(X_1-\bar{x}\right)^4,\left(X_2-\bar{x}\right)^4,\ldots,\left(X_N-\bar{x}\right)^4. \tag{2.69}$$

Четврти централни моменат је аритметичка средина низа (2.69), тј.

$$\mu_4=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^4 . \tag{2.70}$$

За груписано обележје дато расподелом фреквенција четврти централни моменат се своди на

$$\mu_4=\frac{1}{N}\sum_{i=1}^{k}\left(x_i-\bar{x}\right)^4\cdot f_i , \tag{2.71}$$

односно

$$\mu_4 = \frac{1}{N}\sum_{i=1}^{k}\left(X_i - \bar{x}\right)^4 \cdot p_i \ , \qquad (2.72)$$

за расподелу релативних фреквенција.

Други Пирсонов коефицијент *или* **коефицијент спљоштености** *је количник*

$$\beta_2 = \frac{\mu_4}{S^4} \qquad (2.73)$$

при чему је S стандардна девијација обележја X.

Ако је коефицијент $\beta_2 = 3$, кажемо да *обележје X има нормалну спљоштеност*. Кад је $\beta_2 > 3$ кажемо да *X има већу спљоштеност од нормалне*, а кад је $\beta_2 < 3$ кажемо да је *спљоштеност обележја X мања од нормалне*.

Уместо коефицијента спљоштености β_2 може се користити коефицијент

$$\gamma_2 = \beta_2 - 3$$

који има вредност нулу кад обележје X има нормалну спљоштеност. У осталим случајевима коефицијент γ_2 је или позитиван или негативан.

3. СЛУЧАЈНИ ДОГАЂАЈИ И ВЕРОВАТНОЋЕ

Циљ дескриптивне статистике је био да се прикупе подаци о тзв. масовним појавама и да се изврши анализа посматраног статистичког скупа на основу прикупљених података. Таква анализа је у великој мери непотпуна јер не објашњава посматрану појаву или процес са становишта узрочно-последичних односа. Са друге стране, методе дескриптивне статистике не омогућавају утврђивање законитости које важе у масовним појавама, а то би требало да буде основни циљ сваког испитивања.

Поред тога многе податке је немогуће у потпуности обухватити због њихове масовности. Зато је потребно изучавати појаве на великим статистичким скуповима и то на основу мањег броја података, а затим извршити уопштавање добијених резултата и на цео статистички скуп.

Теорија вероватноће представља основну научну дисциплину која омогућава решавање ова два проблема статистичке анализе. Зато су следећа поглавља посвећена основама теорије вероватноће и то у оној мери која је потребна за даља статистичка истраживања.

Развој статистичке методологије почиње онда када су статистичари почели користити дефиниције, постулате и теореме математичке ***<u>теорије вероватноће</u>***. Ова математичка дисциплина се бави изучавањем законитости случајних појава и представља основу статистике. Почела је да се развија средином XVII века и то на основу извесних проблема везаних за хазардне игре. Захваљујући узајамном стимулативном деловању теорије и примене дошло је до развитка опште математичке теорије која има велику примену у веома широком спектру различитих области природних и друштвених појава.

3.1. СИГУРНИ, НЕМОГУЋИ И СЛУЧАЈНИ ДОГАЂАЈИ

Изучавање појава или процеса, било у природи или у друштву, врши се тако што се посматрају догађаји везани за ту појаву. Утврђивањем одређених релација међу разним догађајима описују се појаве и одређују законитости везане за ту појаву или процес.

Све догађаје везане за једну појаву или процес делимо у три класе:

- **Немогући догађаји,**
- **Сигурни догађаји и**
- **Случајни догађаји**.

У природи постоје појаве које се могу потпуно описати помоћу сигурних и немогућих догађаја. То су појаве детерминистичког карактера. На пример, познавајући законе кретања небеских тела у сунчевом систему могуће је сасвим тачно утврдити положај сваке планете у сваком тренутку, предвидети помрачење сунца, утврдити удаљеност и брзину кретања планета и сл. Исто тако, знајући Њутнов закон гравитације могуће је утврдити брзину падања тела, на основу времена и брзине утврдити положај тела и сл.

Проучавање оваквих појава се заснива на утврђивању законитости из којих следе сигурни или немогући догађаји. Међутим, у природи, а нарочито у друштву далеко је већи број појава које се не могу потпуно описати помоћу сигрних и немогућих догађаја.

Узмимо најједноставнији пример бацања коцке. Јасно је да се може утврдити да ће коцка сигурно пасти на сто, да је немогуће да се задржи у ваздуху и сл. Таква тврђења нам ни издалека не описују тај експеримент. Знамо да ће при бацању пасти један од бројева {1,2,3,4,5,6}, али не знамо који. Ако при бацању коцке посматрамо, на пример, догађај: пао је број 4, јасно је да се тај догађај може, али и не мора остварити.

Овакву врсту догађаја који се у датој појави може остварити, али и не мора називамо **случајним догађајем.** *Појаве или процесе које ћемо проучавати преко случајних догађаја зваћемо* **експерименти**.

Скуп елементарних догађаја

У изучавањима случајних догађаја прво утврђујемо експеримент или појаву коју ћемо испитивати, а затим скуп могућих резултата посматраног експеримента. Посматрајмо следећих неколико примера.

Пример 1. При бацању правилног новчића утврђујемо да се експеримент састоји у бацању новчића и да је скуп могућих резултата експеримента:

{писмо, грб}.

Пример 2. Расподела три службеника у три канцеларије врши се тако што се за првог службеника бира случајно једна од канцеларија, затим се за другог бира једна од преосталих и за трећег трећа канцеларија. Скуп могућих резултата овог експеримента је скуп следећих расподела службеника **a**, **b** и **c** по канцеларијама:

1.|**a**|**b**|**c**|; 4. |**b**|**c**|**a**|;

2.|**a**|**c**|**b**|; 5. |**c**|**b**|**a**|;

3.|**b**|**a**|**c**|; 6. |**c**|**a**|**b**|.

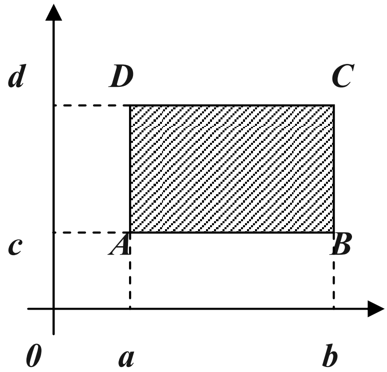
Пример 3. На једном месту се бележи број саобраћајних несрећа. Ако се посматрају дани у недељи и расподела n несрећа по данима, онда је скуп могућих резултата експеримента скуп расподела n несрећа у 7 дана. Различитих могућих резултата има укупно n^7.

Пример 4. У мету се гађа 10 пута. Ако се бележи број погодака, скуп могућих резултата експеримента биће скуп бројева

{0,1,2,3,4,5,6,7,8,9,10}.

Пример 5. Баца се коцка до првог падања шестице. Кад се бележи број бацања, скуп могућих резултата је скуп бројева {1,2,3,... }.

Пример 6. Из скупа брачних парова бира се на случајан начин један брачни пар и бележи се старост оба супружника. Скуп могућих резултата је било која тачка (x, y) из равни $x0y$ у правоугаонику ABCD датом на слици 3.1.



Слика 3.1.

Као што се види из ових примера, за сваки експеримент везујемо један скуп могућих резултата експеримента. У једном извођењу експеримента оствариће се тачно један од могућих резултата. Унапред се не може знати који ће то резултат бити. Кажемо да се ови могући резултати међусобно искључују. Поред тога сваки од њих представља случајан догађај јер се у извођењу експеримента може остварити, а и не мора.

Скуп могућих исхода експеримента зовемо **скуп елементарних догађаја**. *Било који резултат експеримента потпуно се описује једним и само једним елементом скупа могућих исхода.*

Из наведених примера се види да скуп елементарних догађаја могу бити слова, расподеле елемената, реални бројеви и сл. Поред тога скуп елементарних догађаја може бити: ***коначан, бесконачан, али пребројив*** и ***бесконачно непребројив.***

Скуп елементарних догађаја може бити скуп тачака у једнодимензионалном простору, скуп тачака у дводимензионалном, тродимензионалном или вишедимензионалном простору.

На пример, при мерењу века трајања једне сијалице скуп елементарних догађаја је скуп тачака на реалној оси таквих да је $\{t \mid t \geq 0\}$.

Ако се на пример на случајан начин из скупа привредних предузетника бира једно предузеће и испитује се:

- број запослених,
- укупан доходак,
- трошкови пословања,
- просечан лични доходак,
- број повреда на раду;

онда ће скуп елементарних догађаја бити скуп тачака у петодимензионалном простору који одговара ограничењима везаним за свако посматрано обележје.

Посматрајмо сада произвољни случајни догађај A. За сваки елемент скупа елементарних догађаја знамо да ли његова реализација значи остварење догађаја A, или његово „не остварење”. Другим речима, скуп елементарних догађаја можемо поделити на два подскупа:

1. подскуп елементарних догађаја чија реализација значи реализацију догађаја A, и
2. подскуп елементарних догађаја чија реализација значи да се догађај A није реализовао.

Први подскуп елементарних догађаја потпуно описује догађај A. Са друге стране, било какав подскуп скупа елементарних догађаја који садржи једну или више тачака може се посматрати као случајан догађај A који се остварује ако се оствари једна од тачака посматраног подскупа.

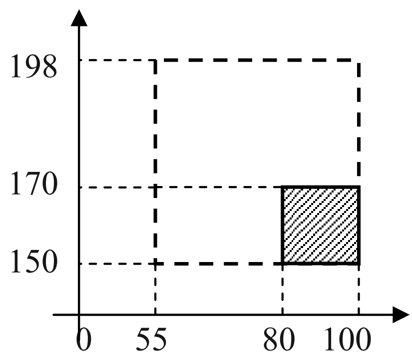
Према томе, случајни догађај се може дефинисати на следећи начин:

Дефиниција 3.1. **Случајни догађај** *је подскуп скупа елементарних догађаја.*

Случајни догађај може бити исказан речима. Његово проучавање могуће је једино ако одредимо одговарајући експеримент, скуп елементарних догађаја и подскуп који одговара датом случајном догађају.

На пример, нека случајни догађај А значи: при бацању коцке пао је број дељив са три. Овде је експеримент бацања коцке, а скуп елементарних догађаја је скуп бројева $\{1,2,3,4,5,6\}$. Случајном догађају А одговара подскуп чији су елементи бројеви $\{3,6\}$, јер остварење било ког од њих значи и остварење догађаја А. Са друге стране догађај А се остварује једино онда кад се оствари један од елемената посматраног подскупа.

Посматрајмо други пример. Нека се из једне велике групе студената бира, на случајан начин, један студент и нека се мере његова тежина *X* и висина *Y*. Ако се тежина свих посматраних студената креће у интервалу [55, 100], а висина у интервалу [150, 198], шта ће представљати догађај А: „изабрани студент је тежи од 80 kg, а није виши од 180 cm".



Слика 3.2. Пример у 2 димензије

На слици 3.2. догађај А је скуп тачака у шрафираној површини.

Ако се у једном експерименту догађај не може остварити, тј. ако је то немогућ догађај, онда њему одговара подскуп скупа елементарних догађаја који нема ни један елемент, а то је празан подскуп. Према томе, немогућ догађај треба дефинисати на следећи начин:

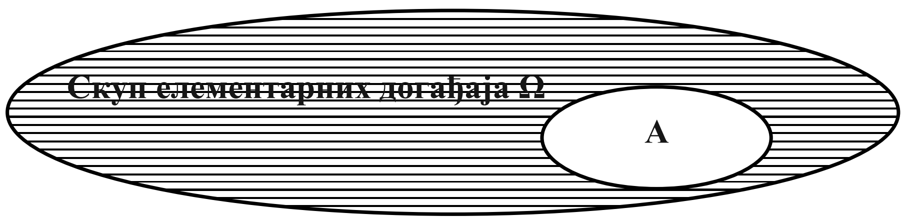
Дефиниција 3.2. **Немогућ догађај** *је подскуп скупа елементарних догађаја који нема ниједан елемент и означава се са Ø.*

Операције са случајним догађајима

Случајне догађаје дефинишемо као подскупове и међу њима можемо дефинисати извесне операције аналогне операцијама на скуповима. Наравно, о тим операцијама говорићемо као о операцијама међу случајним догађајима.

Дефиниција 3.3. *Нека је дат случајан догађај A.* **Супротан догађај** *је догађај* $\overline{A}$ *који се оствари онда и само онда кад се A не оствари.*

Супротном догађају $\overline{A}$ одговара подскуп елементарних догађаја који не припадају одговарајућем подскупу догађаја A. На слици 3.3. то су тачке у шрафираној површини.



Слика 3.3. Супротан догађај

На пример, бацају се истовремено четири новчића. Бележимо колоко пута је пало писмо. Нека случајни догађај A значи: „ни једанпут није пало писмо". Тада ће супротан $\overline{A}$ догађај значити „бар једанпут је пало писмо". Случајни догађај A је подскуп са једним елементом $A=\{0\}$, а супротан догађај је подскуп $\overline{A}=\{1,2,3,4\}$.

Дефиниција 3.4. *Супротан догађај немогућег догађаја је* **сигуран догађај**. *Означаваћемо га са* $\mathbf{\Omega}$.

Немогућем догађају одговара празан подскуп. Сви елементи који нису у празном подскупу су елементарни догађаји. Зато је сигуран догађај у ствари скуп елементарних догађаја.

Дефиниција 3.5. *Дата су два случајна догађаја A и B. Ако се сваки пут кад се оствари догађај A остварује и догађај B кажемо да* ***A*** **имплицира** ***B*** *и означавамо то са*

$$A \subseteq B . \tag{3.1}$$

Релација (3.1) значи да сви елементи подскупа који одговара догађају A припадају и подскупу који одговара догађају B, тј. кад је A садржано у B.

Дефиниција 3.6. *За два случајна догађаја* ***A*** **и** ***B*** *кажемо да* **су једнаки** *ако истовремено A имплицира B и B имплицира A, тј. ако је*

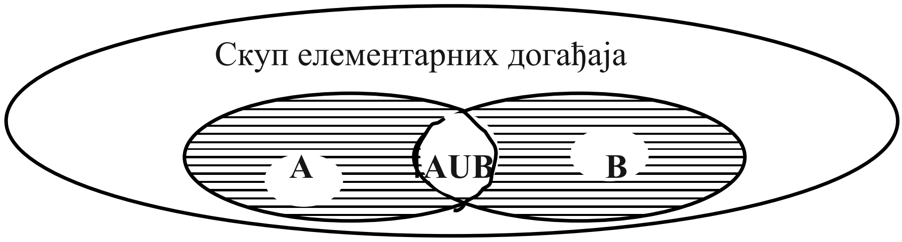
$$A \subseteq B \, и \, B \subseteq A \tag{3.2}$$

Као што се на реалним бројевима дефинишу операције сабирања и множења тако се и на случајним догађајима дефинишу две операције. То су унија и пресек случајних догађаја.

Операција уније: *Нека су дати случајни догађаји A и B*, **унија догађаја** ***A*** **и** ***B*** *је догађај који се оствари онда и само онда кад се оствари бар један од догађаја A и B. Унију два догађаја означаваћемо са*

$$A \cup B \tag{3.3}$$

Случајним догађајима A и B одговарају два подскупа елементарних догађаја. Њиховој унији $A \cup B$ ће одговарати подскуп елементарних догађаја који припадају бар једном од подскупова везаних за догађаје A и B. На слици 3.4. су то тачке шрафиране површине.



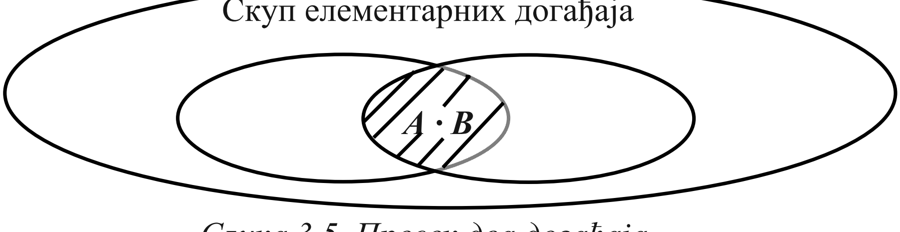
Слика 3.4. Унија два догађаја

На пример, зна се да један производ може имати дефект прве врсте и дефект друге врсте. Нека случајни догађај A значи: „производ има дефект прве врсте”, а догађај B значи: „производ има дефект друге врсте”. Њихова унија $A \cup B$ ће бити случајни догађај који ће значити: „производ има дефект ”, ili „производ је неисправан ”.

<u>Операција пресека:</u> *Дата су два случајна догађаја A и B,* **пресек догађаја *A* и *B*** *је случајан догађај који се остварује онда и само онда кад се остваре оба догађаја, тј. кад се истовремено оствари и догађај A и догађај B. Пресек два догађаја означаваћемо са*

$$A \cdot B \text{ или } A \cap B. \tag{3.4}$$

Догађајима A и B одговарају два подскупа елементарних догађаја. Њиховом пресеку $A \cdot B$ одговараће подскуп елементарних догађаја који истовремено припадају одговарајућим подскуповима догађаја A и B. То су уствари заједнички елементи подскупова A и B, као што се види на слици 3.5.



Слика 3.5. Пресек два догађаја

Нека се статистички скуп састоји од N студената међу којима је један број пушача. На случајан начин се бира један студент. Нека догађај A значи: „изабрана је студенткиња", а догађај B нека значи: „изабран је пушач". Њихов пресек $A \cap B$ биће случајан догађај који ће значити: „изабрана је студенткиња пушач". Подскуп који одговара догађају A је подскуп студенткиња, а подскуп који одговара догађају B су пушачи (студенти и студенткиње). Подскуп који одговара пресеку $A \cap B$ су студенткиње пушачи, тј. они елементи који припадају A и B.

Дефиниција 3.7. *За два догађаја* **A и B** *кажемо да* **се међусобно искључују** *ако је њихов пресек немогућ догађај, тј. ако је*

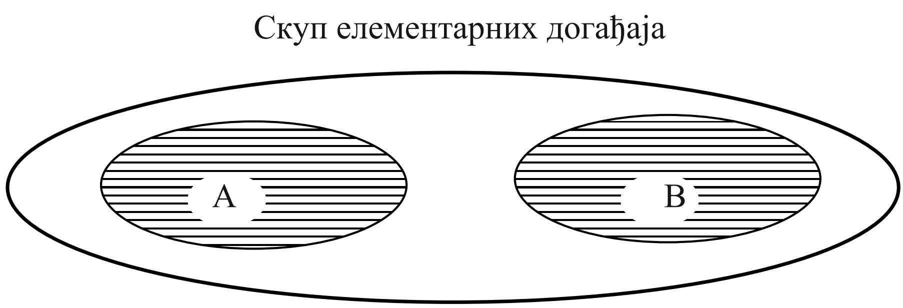
$$A \cap B = \varnothing . \tag{3.5}$$

Догађајима који се међусобно искључују одговарају подскупови елементарних догађаја који немају заједничких тачака.

Ако се два догађаја A и B међусобно искључују тада је њихова унија збир тих догађаја, што означавамо са

$$A \cup B = A + B . \tag{3.6}$$

Тада њиховој унији одговара подскуп елементарних догађаја који припадају или подскупу A или подскупу B. То су тачке у шрафираној површини на слици 3.6.



Слика 3.6. Збир два догађаја A+B

За операције уније и пресека који су дефинисани на случајним догађајима важе следећа три закона:

1. ***Комутативност**:* $A \cdot B = B \cdot A$ и $A \cup B = B \cup A$;

2. ***Асоцијативност***: $\left(A\cup B\right)\cup C = A\cup\left(B\cup C\right)$ и $\left(A\cdot B\right)\cdot C = A\cdot\left(B\cdot C\right)$;
3. ***Дистрибутивност***: $A\cdot(B\cup C)=(B\cdot A)\cup(A\cdot C)$.

За наведене дефиниције и операције на случајним догађајима важе следеће особине :

Особина 1. *А и* $\overline{A}$ *се међусобно искључују јер је*

$$A\cdot\overline{A}=\varnothing. \tag{3.7}$$

Њихова унија је сигуран догађај

$$A+\overline{A}=\Omega. \tag{3.8}$$

Особина 2. *Сваки случајан догађај А имплицира сигуран догађај* Ω

$$A\subseteq\Omega.$$

Особина 3. *Ако А имплицира В, тада је*

$$A\cup B=B \quad \text{и} \quad A\cdot B=A.$$

Особина 4. *Унија случајног догађаја А и немогућег догађаја* $\varnothing$ *је*

$$A\cup\varnothing=A,$$

g њихов пресек је

$$A\cdot\varnothing=\varnothing.$$

Особина 5. *Унија и пресек случајног догађаја А са самим собом је сам догађај А, тј.*

$$A\cup A=A \quad \text{и} \quad A\cdot A=A.$$

Особина 6. *Посматрајмо случајни догађај*

$$A \cup B$$

који се остварује кад се оствари бар један од њих. Супротан догађај је догађај који се оствари кад се не оствари ниједан од њих, тј. кад се оствари пресек $\overline{A}$ *и* $\overline{B}$ *. Зато је*

$$\overline{A \cup B} = \overline{A} \cdot \overline{B} \, .$$

На сличан начин се може закључити да је

$$\overline{A \cdot B} = \overline{A} \cup \overline{B} \, .$$

У решавањима разних проблема везаних за случајне догађаје користе се наведене особине и операције и помоћу једноставнијих случајних догађаја описују се неки други случајни догађаји.

На пример, гађа се мета 3 пута и посматрају се случајни догађаји A_1, A_2 и A_3 који значе погађање мете у првом, другом и трећем гађању, респективно. Тада ће случајни догађај A који значи: „**мета је погођена са три поготка**” бити једнак

$$A = A_1 \cdot A_2 \cdot A_3 \, ,$$

а догађај B: „**мета погођена бар једанпут**” једнак је

$$B = A_1 \cup A_2 \cup A_3 \, ,$$

док случајни догађај C значи: „**мета није уопште погођена**” и једнак је

$$C = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3} \, ,$$

односно

$$C = \left(\overline{A_1 \cup A_2 \cup A_3}\right).$$

3.2. ВЕРОВАТНОЋА

Као што је већ речено, за описивање тзв. масовних појава и појава са случајним исходима служимо се случајним догађајима. Сад смо у стању да тачно опишемо случајне догађаје и да успоставимо извесне релације међу њима. Оно што је битно за посматрану појаву су законитости које треба да утврдимо. То ћемо моћи да урадимо онда кад будемо у стању да на известан начин **„меримо”** случајност код случајних догађаја. На основу **„измерених”** случајности могуће је квантитативно поређење случајних догађаја, а то ће нам омогућити и потпуније описивање посматране појаве. Случајност код случајних догађаја се испитује и **„мери”** вероватноћом која је у различитим тренуцима свог развоја била дефинисана различито. Ради бољег схватања суштине, упознаћемо се са три дефиниције вероватноће и то редоследом историјског развоја ове теорије.

Класична дефиниција вероватноће

Класична дефиниција вероватноће одређује тзв. **вероватноћу *à priori*** а појам вероватноће се дефинише преко тзв. **једнако вероватних догађаја**. Класична дефиниција вероватноће омогућава да се одреди вероватноћа случајних догађаја једне релативно уске класе експеримената, а то је класа експеримената са **коначним бројем једнако вероватних могућих резултата.**

Посматрајмо експеримент коме одговара коначан број могућих резултата, рецимо ***n***. Нека је сваки од ***n*** могућих резултата подједнако вероватан у следећем смислу: на основу експеримента немамо никаквог разлога да верујемо да ће **неки резултат наступити пре него неки други**.

На пример, ако бацамо **правилну** коцку, скуп могућих резултата су бројеви {1,2,3,4,5,6}.Немамо никаквог разлога да верујемо да ће рецимо 2 пасти пре него 5 (под условом да је коцка правилног облика). Ако коцка није правилног облика не можемо унапред прихватити резултате као подједнако вероватне.

Дакле, претпоставићемо да скуп елементарних догађаја има ***n*** подједнако вероватних елемената. Посматрајмо случајни догађај A. Нека је ***m*** број елемената подскупа који одговара случајном догађају A. Кажемо да је **m** број „повољних” резултата експеримента за случајни догађај A.

<u>**Дефиниција 3.8.**</u> **Вероватноћа случајног догађаја** ***A*** *је однос броја „повољних” резултата за догађај A и броја могућих резултата експеримента. Вероватноћу случајног догађаја означаваћемо са*

$$P(\cdot),$$

при чему ћемо у загради уписати догађај и читати „Вероватноћа догађаја...”.

Из дефиниције 3.8. следи да је

$$P(A)=\frac{m}{n}, \tag{3.9}$$

*при чему је **m** број „повољних”, а **n** број могућих резултата експеримента. Вероватноћа (3.9) се зове* **вероватноћа à priori .**

Пошто су ***m*** и ***n*** цели ненегативни бројеви, није тешко проверити да *вероватноћа à priori* има следеће особине:

Особина 1. *Вероватноћа било ког догађаја је ненегативан број, тј.*

$$P(A)\geq 0. \tag{3.10}$$

Особина 2. *Вероватноћа сигурног догађаја је једнака јединици. Ако је Ω сигурни догађај онда је за Ω број „повољних” резултата $m=n$, тако да је из (3.9)*

$$P(\Omega)=1. \tag{3.11}$$

Особина 3. Ако се случајни догађаји A_1 и A_2 међусобно искључују, онда је број „повољних” резултата за њихову унију A_1+A_2 једнак

$$m = m_1 + m_2,$$

при чему је m_1 број „повољних” резултата за A_1, а m_2 број „повољних” за A_2.

На основу (3.9) добиће се да је

$$P(A_1 + A_2) = \frac{m_1}{n} + \frac{m_2}{n}, \text{ односно}$$

$$P(A_1 + A_2) = P(A_1) + P(A_2). \tag{3.12}$$

Особина 4. *Вероватноћа немогућег догађаја једнака је нули, тј:*

$$P(\varnothing) = 0. \tag{3.13}$$

Особина 5. *Вероватноћа супротног догађаја једнака је*

$$P(\overline{A}) = 1 - P(A). \tag{3.14}$$

При одређивању à priori вероватноће треба бити доста опрезан. Прво што треба утврдити је број могућих резултата експеримента n. Друго, треба проверити да ли су ти резултати подједнако вероватни.

На пример, два пута се баца правилна коцка и бележи се збир тачкица у ова два бацања. Скуп могућих резултата су бројеви

$$\{2,3,4,5,6,7,8,9,10,11,12\}.$$

Ови резултати нису подједнако вероватни, јер на пример резултат ће бити 2 ако је пала јединица у првом и другом бацању. Резултат бацања ће бити седам ако су пали следећи парови бројева у бацањима:

(1,6) (2,5) (3,4) (4,3) (5,2) (6,1).

Зато је вероватније да ће резултат бити седам него да ће бити два. У овом примеру број могућих резултата је $n=6^2$, и то су парови бројева у бацањима

(1.1) (2,1) ... (6,1)

(1.2) (2,2) ... (6,2)

$$\vdots \quad \vdots \quad \vdots \qquad (3.15)$$

$$(1,6)\ (2,6)\ \ldots\ (6,6).$$

Сваки од резултата (3.15) је подједнако вероватан. Зато је вероватноћа да ће збир бити 2 једнака

$$P(2) = \frac{1}{36},$$

а вероватноћа да ће збир бити 7 једнака је

$$P(7) = \frac{6}{36},$$

Често се непотпуном анализом лако прихвати закључак о једнако вероватним исходима, што може довести до погрешних закључака.

Такав је био случај са тзв. проблемом **CHEAVALIER DE MÉRÉ**: Посматрају се два експеримента:

а) баца се правилна коцка **4 пута** и посматра се догађај да **бар једанпут** ***падне шестица;***

б) бацају се истовремено две коцке **24 пута** и посматра се догађај да **бар једанпут падне пар** ***шестица***.

Проблем је да се одреди који од два догађаја је вероватнији.

Извесно време се сматрало да су оба догађаја једнако вероватна јер се закључивало на следећи начин: број различитих резултата са две коцке 6 пута је већи од броја резултата са једном коцком. Пошто у а) бацамо коцку 4 пута а у б) две коцке бацамо 6 пута више (тј. 24 пута) онда су исте шансе да падне једна шестица у 4 бацања једне коцке као и падања пара шестица у 24 бацања пара коцки.

Међутим, кад је била организована игра на бази ова два експеримента утврђено је било да у великом броју играња чешћи је био добитак базиран на игри под а) од добитка базираног на игри под б).

Заиста, у бацању једне коцке број могућих резултата је

$$n = 6^4,$$

а број резултата да ниједанпут не падне 6 је

$$m = 5^4 .$$

Зато је тражена вероватноћа

$$P = 1 - \left(\frac{5}{6}\right)^4 = 0.516 \tag{3.16}$$

као вероватноћа супротног догађаја.У случају бацања пара коцки, број могућих резултата је

$$n = 36^{24},$$

а међу њима број резултата да не падне ниједанпут пар шестица је

$$m = 35^{24} .$$

Тражена вероватноћа је једнака :

$$P = 1 - \left(\frac{35}{36}\right)^{24} = 0.491 . \tag{3.17}$$

Из (3.16) и (3.17) се види да је вероватнији догађај под а) од догађаја под б), што је сасвим у складу са експериментално утврђеном чињеницом на бази игара.

Класичну дефиницију вероватноће могуће је у извесним случајевима проширити и на експерименте са бесконачно могућих резултата. Тако се добија тзв. ***геометријска вероватноћа***.

На пример, нека је у равни *x0y* дата ограничена област G и унутар ње друга област g. Замислимо следећи експеримент: произвољно мала куглица се на случајан начин баци у област G и посматра се тачка у коју ће пасти куглица. Свака тачка области G је могући резултат експеримента и све тачке су „подједнако вероватне”. Посматрајмо случајни догађај A који значи: „бачена куглица је пала у тачку области g”. Тада је вероватноћа догађаја А једнака

$$P(A) = \frac{Pov(g)}{Pov(G)}, \tag{3.18}$$

тј. вероватноћа је једнака односу површина области g и G.

Дефиниција геометријске вероватноће се може проширити и на једнодимензионалан, тродимензионалан или вишедимензионалан простор. У једнодимензионалном простору посматра се коначан интервал дужине D и интервал дужине d. Ако је свака тачка интервала D „подједнако вероватна” као могућ резултат експеримента онда је вероватноћа да ће се случајно изабрана тачка наћи у интервалу чија је дужина d једнака

$$P(A)=\frac{d}{D}, \tag{3.19}$$

при чему је интервал дужине d садржан у интервалу дужине D.

У тродимензионалном простору, геометријска вероватноћа је однос одговарајућих запремина

$$P(A)=\frac{V(g)}{V(G)}, \tag{3.20}$$

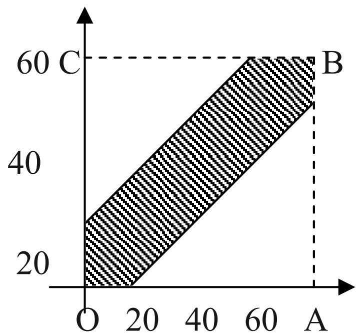
а у вишедимензионалном простору то је однос одговарајућих хиперзапремина.

На пример, у рулету кружница је подељена на 37 поља која су наизменично обојена црвеном и црном бојом. Једно поље је остало беле боје. Свако поље носи број 1, 2, ..., или 36, а бело поље има број 0. Учесници у игри стављају одређен износ на поље које желе. Куглица кружи ивицом круга који се такође врти, али у обрнутом смеру од куглице. Свака тачка кружнице је подједнако вероватна. Зато свако поље има исту вероватноћу која је једнака

$$P(A)=\frac{1}{37},$$

а то је однос дужине кружнице која одговара исечку и обима кружнице.

ПРОБЛЕМ СУСРЕТА. Посматрајмо следећи пример. Два лица се договоре да се нађу на договореном месту у временском интервалу од једног сата. Први који дође на уговорено место чека 20 минута, после чега одлази. Ако сваки од њих може доћи у било којем тренутку посматраног сата, а њихови доласци су независни међусобно, колика је вероватноћа сусрета? На слици 3.7. приказан је посматрани експеримент.



Slika 3.7 Проблем сусрета

На оси x нанешен је тренутак доласка првог лица, а на оси y тренутак доласка другог лица. Могући резултати су све тачке у правоугаонику OABC.

До сусрета ће доћи ако је $|x - y| < 20$,

а то су тачке у шрафираној површини. Зато је тражена вероватноћа

$$P = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

Емпиријска вероватноћа

Посматрајмо случај кад изаберемо једно лице и испитујемо његову старост. Ако је T време које доживи посматрано лице, онда је скуп могућих резултата „експеримента” скуп ненегативних бројева t, при чему је $0 \le t \le T_0$, где је T_0 највећа старост коју може доживети неко лице. Из искуства знамо да је вероватније да ће посматрано лице доживети 60 година, него да ће доживети, рецимо, 100 година. Према томе, скуп елементарних догађаја не можемо одредити тако да сваки елемент буде подједнако вероватан, па самим тим не можемо одредити ни вероватноће случајних догађаја везаних за овај „експеримент”.

Зато је и дефинисана ***емпиријска вероватноћа*** или ***вероватноћа à posteriori*** која ће нам омогућити одређивање вероватноћа и за ону класу експеримената који доводе до неједнако вероватних резултата.

Претпоставимо да експеримент можемо поновити неограничен број пута. Посматраћемо случајни догађај A везан за тај експеримент.

Ако експеримент поновимо ***n*** пута и при том се посматрани догађај A реализовао тачно ***n(A)*** пута, тада се количник

$$f = \frac{n(A)}{n} \tag{3.21}$$

назива *релативна фреквенција догађаја А.*

__**Дефиниција 3.9**__. *Ако број понављања експеримента неограничено расте и ако релативна фреквенција догађаја А тежи неком коначном броју, онда ту граничну вредност називамо вероватноћом догађаја А.*

*Вероватноћа одређена овом дефиницијом назива се **емпиријска вероватноћа** или **вероватноћа à posteriori**. Често се та вероватноћа зове и **статистичка вероватноћа**.*

У ствари, **емпиријска вероватноћа** случајног догађаја A је једнака

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}. \tag{3.22}$$

Основни проблем у дефинисању емпиријске вероватноће је проблем егзистенције и једнозачности граничне вредности (3.22). Ако та **гранична вредност не постоји и није једнозначна** за неке случајне догађаје, онда нема смисла ни вероватноћа дефинисана том граничном вредношћу. Гранична вредност (3.22), односно *à posteriori* вероватноћа може се одредити експерименталним путем.

Тако је на пример, Буфон извео експеримент који се састојао у 4040 бацања правилног новчића. Добио је, да је релативна фреквенција појављивања „Писма” била 0,5080.

Пирсон је овај експеримент поновио 12000 пута и добио је релативну фреквенцију 0,5016*,* а на основу 24000 понављања овог експеримента добио је вредност 0,5005. На основу ових података се може закључити да је вероватноћа падања „Писма” при бацању правилног новчића једнака 0,5.

Или на пример, таблица случајних бројева се прави тако што се на случајан начин бирају бројеви {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} и свакој цифри се приписује иста вероватноћа која је једнака 0,10.

Ако се у таблици случајних бројева посматра релативна фреквенција појављивања рецимо седмице добиће се следећи резултати:

Број цифара у таблици случајних бр.	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
Фреквенција појављивања броја 7	0,095	0,088	0,095	0,120	0,095	0,099	0,082	0,089	0,111	0,102

Види се да је гранична вредност релативних фреквенција појављивања седмице једнака 0,10, односно вероватноћа појављивања седмице је једнака 0,10.

Међутим, то ипак није потпуно решење проблема егзистенције емпиријске вероватноће. Ови примери потврђују претпоставку да постоји гранична вредност (3.22) за посматране експерименте и посматране случајне догађаје. Потпуно решење овог проблема могуће је имати тек онда кад се докаже да гранична вредност (3.22) постоји и за све друге експерименте и за све друге случајне догађаје. То ће бити предмет изучавања 4. поглавља (Закон великих бројева).

За *à posteriori* вероватноћу може се лако проверити да поседује све особине као і *à priori* вероватноћа:

1. $0 \le P(A) \le 1$,
2. $P(\Omega) = 1$,
3. $P(\varnothing) = 0$,

4. $P(A_1 + A_2) = P(A_1) + P(A_2)$,

5. $P(\overline{A}) = 1 - P(A)$,

6. $P(A \cup B) = P(A) + P(B) - P(AB)$.

Аксиоматска дефиниција вероватноће

Да би одредили емпиријску вероватноћу случајних догађаја потребно је извести велик број пута посматрани експеримент и тек онда утврдити граничну вредност (3.22), односно вероватноћу. Међутим, у пракси најчешће нисмо у стању да многе експерименте поновимо више пута.

На пример, ако се одређује вероватноћа судара космичког брода са васионским честицама немогуће је то урадити експерименталним путем. Исто тако, ако треба одредити вероватноћу да одређени производ има жељену тражњу што такође не можемо урадити експериментално. Према томе, још увек остаје проблем како одредити вероватноће случајних догађаја везаних за неке експерименте.

Савремена дефиниција вероватноће омогућава развој теорије која даје решења за одређивање вероватноће било ког случајног догађаја. Посматра се скуп случајних догађаја и пресликавање тог скупа на скуп реалних бројева, тако што ће се сваком случајном догађају A придружити реалан број $P(A)$.

Дефиниција 3.10. *Вероватноћа случајних догађаја, је свака функција P дефинисана на случајним догађајима која пресликава случајне догађаје у реалне бројеве и која има следеће* ***три особине****:*

1. Ненегативност. *Сваком случајном догађају A функција P придружује ненегативан број P(A) тј.*

$$P(A) \geq 0 . \tag{3.23}$$

2. Нормираност. *Сигурном догађају функција P придружује број један, тј.*

$$P(\Omega) = 1 . \tag{3.24}$$

3. Адитивност. *Ако се случајни догађаји* A_1, A_2, ... *међусобно искључују, онда функција P придружује њиховој унији број који је једнак збиру бројева које функција P придружује догађајима* A_1, A_2, ..., *тј.*

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots \tag{3.25}$$

Ове три особине представљају **аксиоме** који одређују вероватноћу. Све остале особине и примена целе теорије вероватноће базирана је на овим аксиомима.

Одмах се намеће ***питање:*** Да ли *à priori* вероватноћа и емпиријска вероватноћа имају особине 1. 2. и 3. којима је дефинисана вероватноћа? Другим речима, да ли се може сматрати да су те две вероватноће које смо раније дефинисали за одређене класе експеримената, уједно и вероватноће које одговарају савременој дефиницији вероватноће?

Одговор је позитиван. Из дефиниција 3.7. и 3.8. лако се проверава да ове две вероватноће имају особину ненегативности, нормираности и адитивности.

Из аксиома ненегативности, нормираности и адитивности следе особине вероватноће.

Особина 1. *Вероватноћа супротног догађаја је једнака*

$$P(\overline{A}) = 1 - P(A). \tag{3.26}$$

Заиста, посматрајмо догађаје A и $\overline{A}$. То су догађаји који се међусобно искључују. На основу аксиома адитивности имали би

$$P(A + \overline{A}) = P(A) + P(\overline{A}) \tag{3.27}$$

Са друге стране њихова унија је сигуран догађај тј.

$$\overline{A} + A = \Omega,$$

па је на основу аксиома нормираности

$$P(A + \overline{A}) = 1. \tag{3.28}$$

Кад се (3.28) замени у (3.27) добиће се (3.26).

Особина 2. *Вероватноћа немогућег догађаја је једнака нули, тј.*

$$P(\varnothing) = 0 . \qquad (3.29)$$

Пошто се догађаји Ω и $\varnothing$ међусобно искључују и такви су да је

$$\Omega + \varnothing = \Omega$$

на основу аксиоме адитивности следи

$$P(\Omega) + P(\varnothing) = P(\Omega),$$

односно (3.29).

Особина 3. *Вероватноћа било ког случајног догађаја је број који се налази између 0 и 1, тј.*

$$0 \le P(A) \le 1 . \qquad (3.30)$$

Особина 4. *Ако догађај A имплицира догађај B, тада су њихове вероватноће такве да је*

$$P(A) \le P(B). \qquad (3.31)$$

Заиста из $A \subseteq B$ следи да је

$$B = A + \overline{A} \cdot B ,$$

па на основу адитивности следи

$$P(B) = P(A) + P(\overline{A} \cdot B) \qquad (3.32)$$

Како је на основу особине ненегативности

$$P(\overline{A} \cdot B) \ge 0 ,$$

из (3.32) следи (3.31).

Особина 5. Вероватноћа уније два догађаја *A* и *B* је једнака

$$P(A \cup B) = P(A) + P(B) - P(A \cdot B). \qquad (3.33)$$

Са слике 4.9. се види да је

$$A \cup B = A + \overline{A} \cdot B$$

$$B = A \cdot B + \overline{A} \cdot B \tag{3.34}$$

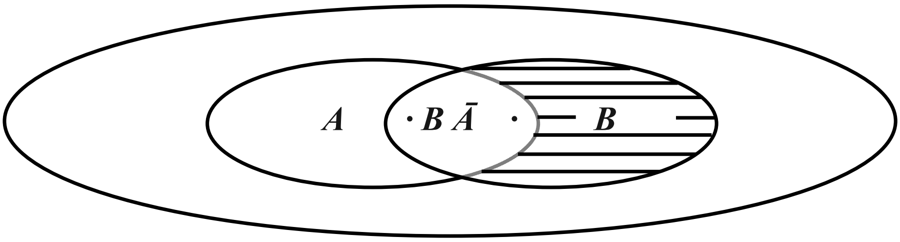
при чему се A и $\overline{A} \cdot B$ међусобно искључују. Исто тако се і $A \cdot B$ и $\overline{A} \cdot B$ такође међусобно искључују.

Због адитивности из (3.34) се добија

$$P(A \cup B) = P(A) + P(\overline{A} \cdot B)$$

$$P(B) = P(A \cdot B) + P(\overline{A} \cdot B) \tag{3.35}$$

Одузимањем од прве једначине другу у (3.35) добиће се (3.33).



Слика 3.8. Унија догађаја

3.3. УСЛОВНЕ ВЕРОВАТНОЋЕ И НЕЗАВИСНОСТ

У испитивањима одређених појава и случајних догађаја везаних за појаве често се поставља проблем међусобне зависности случајних догађаја, тј. поставља се проблем како реализација једног случајног догађаја утиче на реализацију другог.

Посматрајмо два случајна догађаја A и B. Постоје две крајности у њиховом међусобном односу. Прво, догађаји A и B су такви да из реализације догађаја B следи сигурна реализација догађаја A. Друго, реализација догађаја B искључује могућност реализације догађаја A. Остали међусобни односи догађаја A и B налазе се између ове две крајности.

Међусобна веза догађаја А и В у теорији вероватноће проучава се преко тзв. ***условне вероватноће***.

Означимо са

$$P(A/B)$$

условну вероватноћу реализације догађаја A под условом да се реализовао догађај B.

Посматрајмо пример једног експеримента са коначним бројем могућих исхода који су подједнако вероватни. Нека је $\boldsymbol{n}$ укупан број могућих исхода, а $\boldsymbol{n(B)}$ број исхода у којима се реализује догађај B. Нека је $\boldsymbol{n(AB)}$ број могућих исхода који доводе до реализације оба догађаја.

Тада је вероватноћа догађаја В једнака

$$P(B) = \frac{n(B)}{n}, \tag{3.36}$$

а вероватноћа реализације и догађаја А и догађаја В је

$$P(AB)=\frac{n(AB)}{n}. \tag{3.37}$$

Сад посматрајмо догађај A/B, тј. догађај да се А реализује под условом да се B реализовао. Ако је услов испуњен онда постоји *n(B)* могућих исхода а међу њима је *n(AB)* повољних за догађај А.

Зато је условна вероватноћа догађаја A/B једнака

$$P(A/B)=\frac{n(A\cdot B)}{n(B)}. \tag{3.38}$$

Ако у изразу (3.38) и бројилац и именилац поделимо са n и заменимо вредности из (3.36) и (3.37) добићемо релацију

$$P(A/B)=\frac{P(A\cdot B)}{P(B)}. \tag{3.39}$$

Одавде се види да условну вероватноћу треба дефинисати преко количника (3.39).

Дефиниција 3.11. *Условна вероватноћа догађаја А, под условом да се реализује догађај B једнака је количнику*

$$P(A/B)=\frac{P(A\cdot B)}{P(B)},$$

тј. једнака је количнику вероватноће истовременог остварења догађаја А и B и вероватноће догађаја B, $(P(B)>0)$.

Условна вероватноћа задовољава сва три аксиома којима је дефинисана вероватноћа. Наиме, из количника (3.39) се види:

1. $P(A/B)\geq 0$
2. $P(A/B)=1$ и то у случају кад догађај B имплицира А, тј. кад је $B\subseteq A$.
3. Ако се А може изразити као унија догађаја $A_1, A_2, \ldots$ који се међусобно искључују, онда се и условни догађај

A/B може изразити као унија условних догађаја $A_1/B,\ A_2/B,\ \ldots$, па је

$$P(A/B)=\sum_{k=1} P(A_k/B).$$

За одређивање вероватноће пресека два догађаја користи се условна вероватноћа. Из (3.39) се добија да је

$$P(AB)=P(B)\cdot P(A/B). \tag{3.40}$$

Ако A и B замене места добиће се релација

$$P(AB)=P(A)\cdot P(B/A). \tag{3.41}$$

У практичним проблемима често је лакше одредити условне вероватноће, па се релације (3.40) и (3.41) могу користити за одређивање вероватноћа истовременог остварења догађаја A и B.

Формула потпуне вероватноће. У многим примерима вероватноћу случајног догађаја је тешко одредити без посматрања других случајних догађаја. На пример, из једне кутије у којој има 6 белих и 4 црне куглице извлаче се две куглице једна за другом без враћања и посматра се догађај да у другом извлачењу буде извучена бела куглица. Наравно, реализација тог догађаја зависи од тога која је куглица била извучена у првом извлачењу. Зато овај догађај морамо изразити преко догађаја везаних за прво извлачење, а затим на основу тих догађаја одредити и вероватноћу посматраног догађаја. За решавање оваквих проблема служи тзв. ***формула потпуне (тоталне) вероватноће:***

Нека су

$$B_1,\ B_2,\ \ldots,\ B_n$$

случајни догађаји који се међусобно искључују и чија унија је сигуран догађај, тј.

$$B_i\cdot B_j=\varnothing$$

за свако $i\neq j,\ i,j\in\{1,\ 2,\ \ldots,\ n\}$,

$$\sum_{k=1}^{n} B_k = \Omega .$$

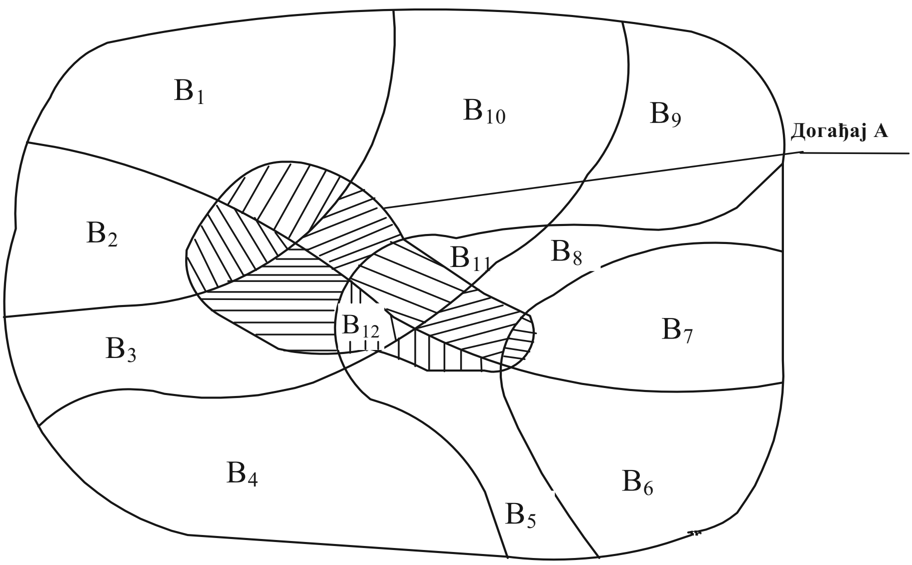
Тада је вероватноћа догађаја А једнака

$$P(A) = \sum_{k=1}^{n} P(B_k) \cdot P(A/B_k), \tag{3.42}$$

при чему су $P(B_k)$ вероватноће догађаја B_k, а $P(A/B_k)$ условне вероватноће догађаја А под условом да се реализовао B_k.

Доказ. Догађаји B_k чине разбијање скупа могућих исхода Ω. Подскуп који одговара догађају А се састоји од подскупова који су заједнички са А и са B_k, (Слика 3.9.) тако да је

$$A = AB_1 + AB_2 + \cdots + AB_k . \tag{3.43}$$



Слика 3.9.

Вероватноћа сваког од догађаја на десној страни једначине (3.43) може се изразити преко условне вероватноће

$$P(AB_k) = P(B_k)P(A/B_k),$$

па на основу аксиома адитивности и релације (3.43) следи (3.42).

Узмимо следећи пример. Један производ се припрема за завршну операцију која се обавља успешно ако је припрема производа била добра. Припрема се обавља на једној од четири машине. Нека је поузданост машина 0,90; 0,91; 0,90 и 0,92 респективно за прву, другу, трећу и четврту машину.

Расподела производа по машинама за припрему се врши случајно и то тако да се другој и четвртој машини додељује по два посто више производа него машинама један и три. Колика је вероватноћа да ће производ бити успешно обрађен у завршној операцији?

Да би одредили тражену вероватноћу означићемо са:

- B_1 − догађај да ће производ у припреми доћи на 1. машину;
- B_2 − догађај да ће производ у припреми доћи на 2. машину;
- B_3 − догађај да ће производ у припреми доћи на 3. машину;
- B_4 − догађај да ће производ у припреми доћи на 4. машину.

Са А ћемо означити догађај да ће производ бити успешно обрађен. Тада је

$$P(B_1) = 0{,}244$$

$$P(B_2) = 0{,}256$$

$$P(B_3) = 0{,}244$$

$$P(B_4) = 0{,}256,$$

а условне вероватноће су

$$P(A/B_1) = 0{,}90$$

$$P(A/B_2) = 0{,}91$$

$$P(A/B_3) = 0{,}90$$

$$P(A/B_4) = 0{,}91.$$

Догађаји B_1, B_2, B_3 и B_4 се међусобно искључују и њихова унија је сигуран догађај. Зато се вероватноћа догађаја A може одредити преко формуле потпуне вероватноће, тако да је

$$P(A) = 0{,}244 \cdot 0{,}90 + 0{,}256 \cdot 0{,}91 + 0{,}244 \cdot 0{,}90 + 0{,}256 \cdot 0{,}91,$$

тј.

$$P(A) = 0{,}90512.$$

Посматраћемо још један **пример игре**. Баца се правилан новчић. Играч пре бацања „погађа” која ће страна новчића пасти. Ако погоди страну добија динар, а ако не погоди плаћа динар. Играч поседује X динара и решио је да игра све док не повећа свој износ на A динара. Колика је вероватноћа да ће играч изгубити целу суму X пре него што оствари жељу?

Означићемо са

$$p(X)$$

вероватноћу да ће играч изгубити суму X. После првог бацања његова сума или ће бити (X+1) динара или (X-1).

Вероватноћа да у наредним бацањима изгуби суму X+1 биће

$$p(X+1)$$

а вероватноћа да изгуби суму X-1 биће

$$p(X-1)$$

Нека је B_1 догађај да играч у првом бацању погоди, а B_2 догађај да не погоди. Тада се B_1 и B_2 искључују и њихова унија је сигуран догађај. Ако је A догађај да играч изгуби суму X онда је на основу формуле потпуна вероватноћа

$$p(X) = 0{,}5p(X+1)+0{,}5p(X-1)$$

Решење ове једначине је линеарна функција

$$p(X) = C_1+C_2X \qquad (3.44)$$

при чему C_1 и C_2 треба одредити из почетних услова. Пошто је вероватноћа да играч изгуби суму од нула динара једнака

$$p(0) = 1 \qquad (3.45)$$

а вероватноћа да изгуби суму од А динара

$$p(A) = 0 \qquad (3.46)$$

(јер кад оствари суму од А динара престаје да игра), то значи да ће се из (3.44), (3.45) и (3.46) добити

$$C_1 = 1;\ C_2 = -\frac{X}{A}$$

Тражена вероватноћа губитка почетне суме од X динара је једнака

$$p(X) = 1 - \frac{X}{A}\cdot$$

Бајесова (Bayes) формула. Посматрајмо пример продавнице која продаје један производ. Нека тај производ продавница набавља од два произвођача и то у односу 2:1. Зна се да је 1% производа првог произвођача неисправно, а 2% код другог произвођача. За један производ случајно изабран у продавници утврди се да је неисправан. Колика је вероватноћа да је то производ другог произвођача?

Означимо са B_1, B_2 и A следеће догађаје

- B_1: изабрани производ је производ првог произвођача
- B_2: изабрани производ је производ другог произвођача
- А: изабрани производ је неисправан.

Тада је вероватноћа догађаја B_1 једнака

$$P(B_1) = \frac{2}{3},$$

а вероватноћа догађаја B_2 је

$$P(B_2) = \frac{1}{3}.$$

Условне вероватноће неисправности производа су

$$P(A/B_1) = 0{,}01$$
$$P(A/B_2) = 0{,}02\ ,$$

па је вероватноћа догађаја A

$$P(A) = \frac{2}{3}0{,}01 + \frac{1}{3}0{,}02\ .$$

Међутим, зна се да се догађај A реализовао и тражи се вероватноћа да је изабрани производ другог произвођача тј. тражи се условна вероватноћа догађаја B_2 под условом кад се реализовао A. Значи тражи се

$$P(B_2 / A).$$

На основу дефиниције условне вероватноће, тражена вероватноћа је једнака

$$P(B_2 / A) = \frac{P(A \cdot B_2)}{P(A)},$$

односно

$$P(B_2 / A) = \frac{P(B_2) \cdot P(A/B_2)}{P(A)} \tag{3.47}$$

јер је $P(A \cdot B_2) = P(B_2) \cdot P(A/B_2)$.

На тај начин ћемо добити

$$P(B_2 / A) = \frac{\frac{1}{3}0{,}02}{\frac{2}{3}0{,}01 + \frac{1}{3}0{,}02} = 0{,}5\,.$$

Израз (3.47) познат је под називом **Бајесова формула**. *Служи за одређивање условне вероватноће извесних догађаја који су претходили реализацији једног случајног догађаја.*

Посматраћемо општији случај. Нека је дат низ случајних догађаја

$$B_1,\ B_2,\ ...,\ B_n,$$

који се међусобно искључују и чија унија је сигуран догађај Ω, тј. таквих да је

$$B_i \cdot B_j = \varnothing\,,$$

за свако $i \neq j,\ i, j \in \{1,\ 2,\ ...,\ n\},$

$$B_1 + B_2 + \cdots + B_n = \Omega\,.$$

Нека је A неки случајни догађај везан за посматрани експеримент. Тада је за сваки догађај B_k, условна вероватноћа једнака

$$P(B_k / A) = \frac{P(A \cdot B_k)}{P(A)}\,. \tag{3.48}$$

Бројилац у изразу (3.48) може се изразити са

$$P(A \cdot B_k) = P(B_k) \cdot P(A / B_k),$$

а за именилац се може користити формула потпуне вероватноће. Тако се добија:

<u>Бајесова формула.</u> *За сваки догађај B_k условна вероватноћа његове реализације кад се реализовао догађај А је једнака*

$$P(B_k / A) = \frac{P(B_k) \cdot P(A / B_k)}{\sum_{i=1}^{n} P(B_i) \cdot P(A / B_i)} . \qquad (3.49)$$

Нека се на пример у једној лабораторији налазе три различите полице са епруветама у којима је крв из три различите крвне групе. У првој полици су 2 епрувете са првом крвном групом, 1 са другом и 3 са трећом. У другој полици су 3 са првом, 2 са другом и 4 са трећом, а на трећој полици су 4 са првом, 3 са другом и 2 са трећом. Позитивна реакција се постиже једино ако се помешају трећа и друга крвна група. Лаборант је на случајан начин изабрао једну полицу, а затим са те полице изабрао две епрувете. Експеримент је показао позитивну реакцију помешане крви из те две епрувете. Колика је вероватноћа да су то биле епрувете са прве полице?

Означимо са B_1, B_2 и B_3 случајни догађај да ће изабране епрувете бити са прве, друге, односно треће полице. Са А ћемо означити случајан догађај да ће се постићи позитивна реакција. Тада је

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3} .$$

Условне вероватноће су

$$P(A / B_1) = \frac{3}{6} \cdot \frac{1}{5} + \frac{1}{6} \cdot \frac{3}{5} = \frac{1}{5}$$

$$P(A / B_2) = \frac{2}{9}$$

$$P(A / B_3) = \frac{1}{6} \cdot$$

Примењујући Бајесову формулу добиће се

$$P(B_1 / A) = \frac{18}{53} .$$

Независни догађаји. Посматраћемо два експеримента који се изводе независно један од другог.

Нека је A_1 случајни догађај везан за први експеримент, а A_2 случајни догађај везан за други експеримент. Тада реализација догађаја A_1 не утиче на реализацију догађаја A_2 и кажемо да су догађаји A_1 и A_2 ***међусобно независни.***

Наравно, основни проблем везан за оваква два догађаја је проблем одређивања вероватноће њиховог истовременог остварења, тј. проблем одређивања вероватноће пресека независних догађаја A_1 и A_2.

Други проблем представља утврђивање независности за оне догађаје који су везани за један исти експеримент. Одредићемо вероватноћу пресека независних догађаја преко емпиријске вероватноће, а затим ћемо тај резултат уопштити и за остале случајеве.

Претпоставимо да два посматрана експеримента можемо поновити неограничен број пута. Нека смо истовремено поновили оба експеримента *n* пута.

Означимо са $n(A_2)$ број понављања експеримената у којима се реализовао догађај A_2, а са $n(A_1A_2)$ број понављања експеримената у којима су се реализовали истовремено и A_1 и A_2. Тада је емпиријска вероватноћа приближно једнака

$$\begin{aligned} P(A_2) &\approx \frac{n(A_2)}{n} \\ P(A_1A_2) &\approx \frac{n(A_1A_2)}{n}. \end{aligned} \tag{3.50}$$

А сад посматрајмо само она понављања у којима се реализовао догађај A_2. Њих има укупно $n(A_2)$. Међу њима биће и неких у којима се реализовао и догађај A_1. Таквих има укупно $n(A_1A_2)$. Ако је $n(A_2)$

довољно велико, тада је емпиријска вероватноћа догађаја A_1 приближно једнака

$$P(A_1) = \frac{n(A_1 A_2)}{n}. \tag{3.51}$$

Вероватноћа пресека $A_1 A_2$ у (3.50) се може написати у облику

$$P(A_1 A_2) = \frac{n(A_1 A_2)}{n} \cdot \frac{n(A_2)}{n}$$

одакле се добија да је

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) \tag{3.52}$$

Израз (3.52) служи за дефиницију независности догађаја.

Дефиниција 3.12. *Нека су A_1 и A_2 било каква два случајна догађаја. Рећи ћемо да су A_1 и A_2,* **међусобно независни,** *ако и само ако је*

$$P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) \tag{3.53}$$

Независност случајних догађаја може се посматрати и преко условних вероватноћа. Наиме, ако су догађаји A_1 и A_2 међусобно независни онда је условна вероватноћа, једнака

$$\begin{aligned} P(A_1 / A_2) &= P(A_1) \\ P(A_2 / A_1) &= P(A_2) \end{aligned}. \tag{3.54}$$

Обрнуто, ако је условна вероватноћа једнака

$$P(A_1 / A_2) = P(A_1)$$

тада је

$$P(A_1 \cdot A_2) = P(A_2) \cdot P(A_1 / A_2) = P(A_1) \cdot P(A_2)$$

тј. тада су A_1 и A_2 независни. Зато се може рећи:

Догађаји A_1 и A_2 су **међусобно независни**, *ако и само ако је условна вероватноћа једног од њих, под условом да се реализовао други, једнака вероватноћи тог догађаја.*

У неким примерима није лако одмах на основу експеримента утврдити независност случајних догађаја. Тада се мора проверавати тачност израза (3.53) или (3.54) и тек онда утврдити независност.

На пример, из шпила од 36 карата на случајан начин се бира карта. Нека догађај A_1 значи: „изабрана карта је треф"; а A_2: „изабрана карта је дама". На основу експеримента не можемо одмах рећи да ли су догађаји A_1 и A_2 независни. Проверићемо њихову независност на основу дефиниције (3.53).

Вероватноће догађаја A_1 и A_2 су

$$P\left(A_1\right) = \frac{9}{36},\ P\left(A_2\right) = \frac{4}{36}\cdot \tag{3.55}$$

Догађај $\left(A_1 \cdot A_2\right)$ значи да је изабрана карта дама треф и његова вероватноћа је

$$P\left(A_1 \cdot A_2\right) = \frac{1}{36}. \tag{3.56}$$

Из (3.55) и (3.56) се види да је задовољена релација (3.53), што значи да су то независни догађаји.

Није тешко проверити и независност следећих догађаја. Ако су A_1 и A_2 међусобно независни тада су независни и следећи догађаји

$$\overline{A_1} \text{ i } \overline{A_2};\ A_1 \text{ i } \overline{A_2};\ \overline{A_1} \text{ i } A_2.$$

Дефиниција независности два догађаја се може проширити и на више догађаја.

Нека је $A_1,\ A_2,\ \dots$ низ случајних догађаја. Ако је за сваки изабрани низ

$$A_{i1},\ A_{i2},\ ...,\ A_{in}$$

вероватноћа пресека једнака производу вероватноће, тј. ако је

$$P(A_{i1} \cdot A_{i2} \cdot ... \cdot A_{in}) = P(A_{i1}) \cdot P(A_{i2}) \cdot ... \cdot P(A_{in}) \qquad (3.57)$$

тада кажемо да је низ $A_1,\ A_2,\ ...$ ***низ независних догађаја***.

На пример, истовремено се бацају две коцке. Нека су дати следећи догађаји:

- A_1 : на првој коцки пао је непаран број,
- A_2 : на другој коцки пао је непаран број,
- A_3 : збир тачкица на обе коцке је непаран број.

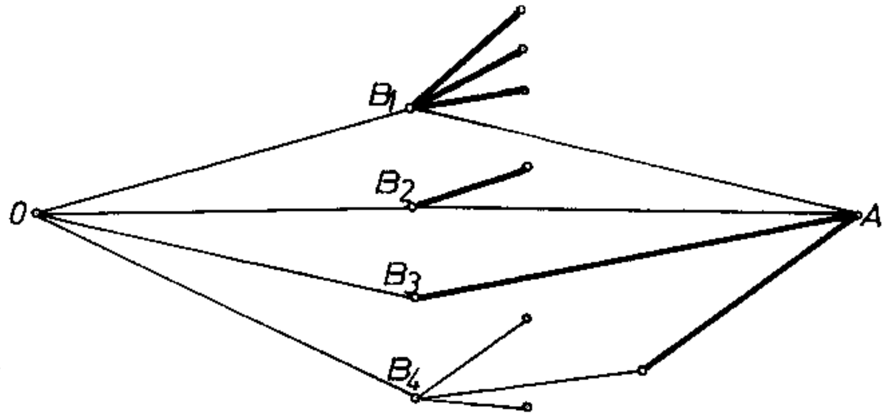
Лако је видети да је

$$P(A_3 / A_1) = 0{,}5 \qquad P(A_3 / A_2) = 0{,}5 \qquad P(A_3) = 0{,}5$$

што значи да су догађаји A_1 и A_3 међусобно независни, а исто тако су независни и догађаји A_2 и A_3.

Међутим, ако су се истовремено остварили и A_1 и A_2, тада се догађај A_3 не може уопште остварити. То значи да се не може сматрати да је A_3 независан од A_1 и A_2 па се не може рећи да су догађаји A_1, A_2 и A_3 независни догађаји.

Пример 1. Путник креће из места О у место А. У сваком месту на случајан начин бира пут за даље кретање. Колика је вероватноћа да ће путник стићи у место А без враћања назад ако су путеви дати на шеми?



Да би путник дошао у A мора проћи кроз једно од места B_1, B_2, B_3 или B_4. Зато је

$$P(B_i) = \frac{1}{4},\ i = 1, 2, 3, 4.$$

Условне вероватноће су

$$P(A/B_1) = \frac{1}{4},$$

$$P(A/B_2) = \frac{1}{2},$$

$$P(A/B_3) = 1,$$

$$P(A/B_4) = \frac{1}{3}.$$

Тражена вероватноћа је

$$P(A) = \frac{1}{4}\left(\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{3}\right)\frac{25}{48}.$$

Пример 2. У кутији се налази n_1 белих и n_2 црних куглица. Случајно се бира једна куглица, а затим се у кутију додаје m куглица оне боје које је била извучена куглица и k куглица друге боје (m и k могу бити и негативни бројеви што значи да се из кутије одузимају куглице). Одредити вероватноћу да се у другом извлачењу извуче бела куглица. Ако је у другом извлачењу извучена црна куглица, колика је вероватноћа да је у првом извлачењу била извучена бела куглица?

Пример 3. Један стрелац гађа мету n пута. Вероватноћа поготка мете у сваком гађању је једнака p, $(0 < p < 1)$. Колика је вероватноћа да ће мета бити погођена бар једанпут? Колико треба извршити гађања да би са вероватноћом 0,999 тврдили да ће мета бити погођена бар једанпут?

Вероватноћа промашаја мете у сваком гађању је $(1-p)$, а вероватноћа промашаја у n гађања је

$$(1-p)^n$$

због независности. Тражена вероватноћа је

$$P(A) = 1-(1-p)^n \tag{3.58}$$

Да би одредили n, вероватноћу (3.58) изједначимо са 0,999. Тако добијамо да n треба да задовољава једнакост

$$(1-p)^n = 0{,}001,$$

односно

$$n = \frac{\log 0{,}001}{\log(1-p)}.$$

Пример 4. Наћи вероватноћу да се строј који ради у тренутку t, неће покварити до тренутка $t_0 + t$ ако:

а) вероватноћа квара зависи само од дужине временског интервала,

б) вероватноћа рада без квара за интервал Δt је пропорционална Δt са тачношћу до бесконачно мале вредности вишег реда,

в) исправност строја у интервалима који се не секу независне су међусобно.

Означимо са $p(t)$ вероватноћу квара строја од t_0 до $t_0 + t$. Тада је вероватноћа рада за време Δt једнака

$$1 - p(\Delta t) = a\Delta(t) + o(\Delta t),$$

при чему је a константа.

Рад од t_0 до $t_0 + t + \Delta t$ и вероватноћа квара строја у том интервалу је једнака

$$p(t + \Delta t)$$

Интервал $(t + \Delta t)$ се састоји од два интервала дужина t и Δt који се међусобно не секу. Зато је

$$p(t + \Delta t) = p(t)p(\Delta t),$$

тако да је

$$\frac{p(t + \Delta t) - p(t)}{\Delta t} = -ap(t) - o(1)$$

Гранична вредност овог израза, кад $\Delta t \to 0$, је једначина

$$\frac{dp(t)}{dt} = -ap(t). \qquad (3.59)$$

Тражена вероватноћа ће бити решење једначине (3.59), а то је функција

$$p(t) = e^{-at}$$

Пример 5. Вероватноће за тзв. *таблице смртности*.

Претпоставке:

1. Вероватноћа да ће неко лице умрети у времену од t до $t + \Delta t$ једнака је

$$p(t, t + \Delta t) = a(t)\Delta t + o(\Delta t). \qquad (3.60)$$

при чему је $a(t)$ ненегативна и непрекидна функција;

2. Сматра се да умирање неког лица у интервалу (t_1, t_2) не зависи од тога шта је било до момента t_1;
3. Вероватноћа смрти у моменту рађања једнака је нули.

Треба одредити вероватноћу да ће једно лице умрети кад доживи старост t. Означимо са $P(t)$ вероватноћу да ће лице доживети старост t. Са $P(t+\Delta t_1,\, t)$ означимо вероватноћу да ће лице које доживи старост *t* доживети и старост $t+\Delta t$. Тада је

$$P(t+\Delta t)=P(t)P(t+\Delta t,\, t),$$

Из (3.60) добија се даје

$$P(t+\Delta t)=P(t)\{1-p(t,\, t+\Delta t)\},$$

односно

$$P(t+\Delta t)=P(t)\{1-a(t)\Delta t-o(\Delta t)\} \tag{3.61}$$

Из израза (3.61) добиће се да је

$$\frac{P(t+\Delta t)-P(t)}{\Delta t}=-a(t)P(t)-0(1)$$

чија је гранична вредност кад $\Delta t \to 0$ једнака

$$\frac{dP(t)}{dt}=-a(t)P(t) \tag{3.62}$$

Решење диференцијалне једначине (3.62) је функција

$$P(t)=e^{-\int_0^t a(z)dz}$$

на основу које се добијају тражене вредности. Специјално, ако је $a(t)=\alpha+\beta e^{\gamma t}$ тада је тражена вероватноћа једнака

$$P(t)=e^{-\alpha t}-\frac{\beta}{\gamma}\, e^{\gamma t}-1).$$

4. ЗАКОН ВЕЛИКИХ БРОЈЕВА

Аксиоматска дефиниција вероватноће случајних догађаја не одређује начин на који ћемо у неком реалном експерименту одредити вероватноће случајних догађаја. Зато нас и интересује одговор на следеће питање: Да ли постоје извесне законитости у практичним извођењима неког експеримента и одговарајућих вероватноћа случајних догађаја? Другим речима да ли се на основу резултата извођења експеримента може одредити одговарајућа вероватноћа?

Зато је наредно поглавље посвећено разматрању проблема постојања тих законитости да би се дошло до позитивних одговора на постављено питање. Пошто ћемо се у наредним извођењима користити Стирлинговом (Stirling) формулом и Гаусовом (Gauss) кривом, укратко ћемо дати њихов преглед и основне особине.

4.1. СТИРЛИНГОВА ФОРМУЛА И ГАУСОВА КРИВА

Посматрајмо $\boldsymbol{n}$ елемената једног скупа. **Пермутација** $\boldsymbol{n}$ елемената је редослед ових елемената тако да се зна који елеменат је на првом месту, који на другом, итд.

Број пермутација $\boldsymbol{n}$ елемената је једнак факторијелу броја $\boldsymbol{n}$**.** Тај број се означава са

$$n!$$

и при томе је

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

За велике вредности броја $\boldsymbol{n}$ одређивање $n!$ је доста отежано. Стирлингова формула служи за приближно одређивање тог броја, а такође има велику важност у решавањима различитих теоријских питања везаних за граничне вредности низова који зависе од $n!$.

Стирлингова формула *се може представити следећом изразом:*

$$n! = \sqrt{2\pi n} \cdot n^n \cdot e^{-n}\left[1 + O(n)\right], \tag{4.1}$$

при чему $O(n)$ *означава остатак који тежи нули кад* $n \to \infty$.

Стирлингова формула се може написати и у облику

$$n! = \sqrt{2\pi n} \cdot n^n \cdot e^{-n} \cdot e^{\theta_n} \quad ,$$

при чему θ_n задовољава неједначину

$$\left|\theta_n\right| \le \frac{1}{12n}.$$

Зато узимамо да је приближна вредност за $n!$ *једнака*

$$n! \approx \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$$

Стирлингова формула даје доста блиске вредности за $n!$ чак и за мале вредности броја n. Тако се за $n = 1$ добија да је

$$1! \approx 1{,}0023,$$

за $n = 2$ добија се

$$2! \approx 2{,}0007,$$

за $n = 5$

$$5! \approx 120{,}01.$$

За $n = 10$ тачна вредност ће бити

$$10! = 3628800,$$

а на основу Стирлингове формуле се добија

$$10! \approx 3598600$$

што у односу на праву вредност чини грешку од свега 0,8%.

Доказ Стирлингове формуле може се наћи код W. Feller-a у [2], или R.von Mises [6].

У примени теорије вероватноће велику улогу има функција позната под именом ***Гаусова*** или ***Гаус-Лапласова (Gauss-Laplace) крива*** која се често зове и ***закон вероватноћа нормалне расподеле***.

Гаусова крива је крива дата функцијом

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{4.2}$$

која је дефинисана за свако реално x из интервала $(-\infty, +\infty)$.

Функција $\phi(x)$ је симетрична функција која има максимум у тачки $x = 0$. Максимална вредност ове функције је

$$\phi(0) = \frac{1}{\sqrt{2\pi}} \approx 0{,}3989\,.$$

Превојне тачке функције $\Phi(x)$ су тачке $x = \pm 1$ у којима функција има вредност

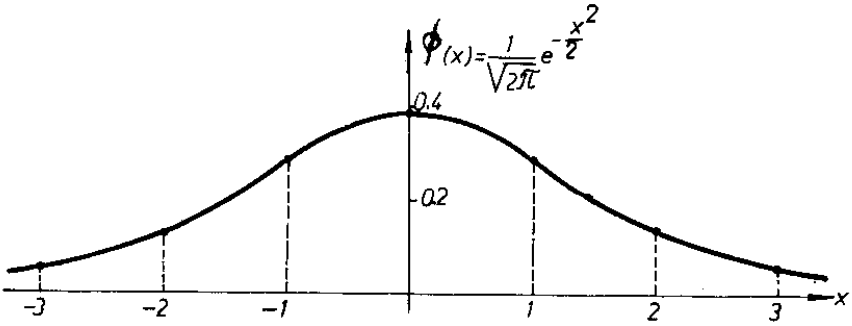
$$\phi(\pm 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}} \approx 0{,}2420\,.$$

Поред тога је

$$\lim_{x \to +\infty} \phi(x) = 0\,,$$

$$\lim_{x \to -\infty} \phi(x) = 0\,.$$

Гаусова крива је приказана на слици 4.1.



Слика 4.1. Гаусова крива $\phi(x)$

Површина између криве $q\left(x\right)$ и осе x је уствари једнака интегралу

$$\int_{-\infty}^{+\infty}\phi(x)dx$$

и има вредност једнаку јединици, тј.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx=1 \tag{4.3}$$

Да је једначина (4.3) задовољена није тешко проверити. Треба посматрати двоструки интеграл

$$\frac{1}{2\pi}\iint e^{-\frac{x^2}{2}-\frac{y^2}{2}}dxdy \tag{4.4}$$

дуж целе $x0y$ равни, који је уствари једнак квадрату интеграла

$$\int_{-\infty}^{+\infty}\phi(x)dx$$

Интеграл (4.4) се лако решава преласком на поларне координате.

Вредности функције $\phi(x)$ дају се табеларно. У табели А-1 (види Прилог, Табела А-1) дате су њене вредности за свако x из интервала [0, 3,99]. За x = 3,99 функција има вредност 0,0001, а за вредности x веће од 3,99 функција $\phi(x)$ има мале вредности које су практично једнаке 0.

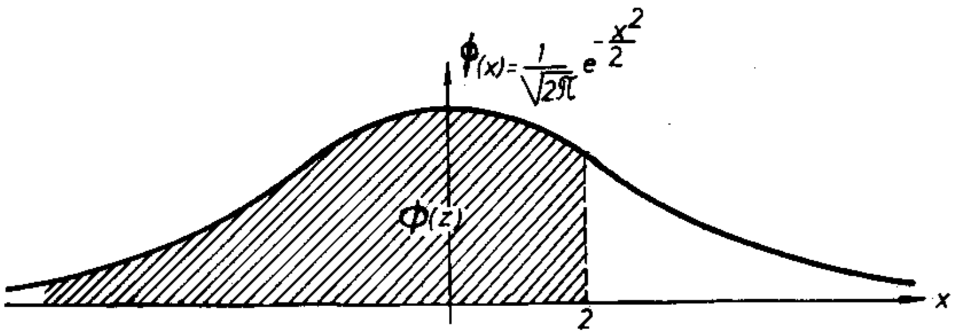
На основу Гаусове криве дефинисана је тзв. **функција нормалне расподеле** $q\,(z)$ чије су вредности дате табеларно и која служи за одређивање површине између криве $\phi(x)$ и осе x у неком произвољном интервалу [a, b].

<u>**Дефиниција 4.1.**</u> **Функција нормалне расподеле** $q\,(z)$ *дефинисана је за свако* $z \in (-\infty, +\infty)$ *једначином*

$$\Phi(z) = \int_{-\infty}^{z} \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$$

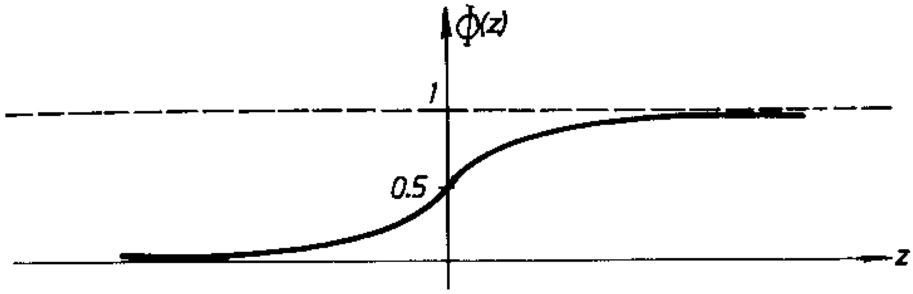
и представља површину између ṛ̈ иве $\phi(x)$ *и осе x у интервалу* $(-\infty, z]$.

Графички приказ преко површине функције $\phi(x)$ дат је на слици 4.2.



Слика 4.2. Функција ђормалне расподеле

Функција нормалне расподеле $\Phi(z)$ је монотоно растућа функција која има асимптоте осу x и праву y = 1. Графички је приказана на слици 4.3.



Слика 4.3. Функција Φ(z)

Вредности функције нормалне расподеле $\Phi(z)$ дате су табеларно за вредности z из интервала $[0, +\infty)$. За негативне вредности функцију $q(z)$ одређујемо из релације

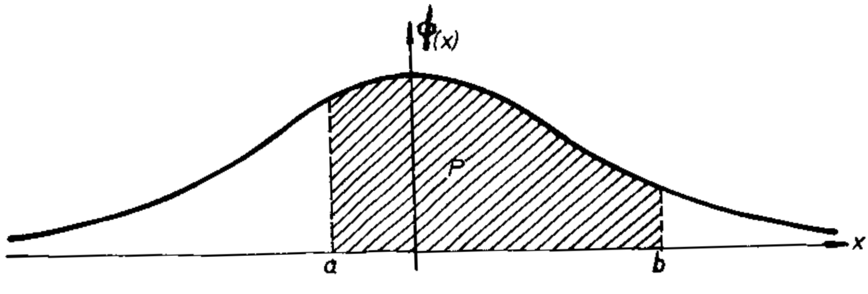
$$q(-z) = 1 - q(z)$$

коју није тешко проверити, на основу дефиниције (4.1).

У табели А-2 (види Прилог, Табела А-1) дате су вредности функције за z из интервала [0, 3,49]. За вредности веће од 3,49 функција Φ практично има вредност 1.

Да би одредили ***површину између Гаусове и осе x*** *у произвољном интервалу* $(a, b]$ *користимо се функцијом* $q(z)$. *Са слике 4.4. види се да је та површина дата са*

$$P = \frac{1}{\sqrt{2u}} \int_a^b e^{-\frac{x^2}{2}} dx = q(b) - q(a)$$



Slika 4.4. Приказ површине и функције $q\,(z)$

Најчешће се посматра симетричан интервал $[-z,\, z]$. Површина између функције $\phi(x)$ и осе x, у том интервалу једнака је

$$P = \frac{1}{\sqrt{2u}} \int_{-z}^{z} e^{-\frac{x^2}{2}}\, dx = 2q\,(z) - 1\,.$$

Тако се добија да је површина у интервалу $(-1,\, 1]$ једнака

$$P = 0{,}6826 = 68{,}3\%\,,$$

а у интервалу $(-2,\, 2]$ површина је једнака

$$P = 0{,}9544 = 95{,}4\%\,,$$

док је за интервал $(-3,\, 3]$

$$P = 0{,}9974 = 99{,}7\%\,,$$

што значи да се 99,7% од укупне површине између Гаусове криве $\phi(x)$ и осе x, налази у интервалу (-3, 3] док је изван тог интервала свега 0,03% те површине.

Ове проценте треба запамтити јер ће касније бити интерпретирани преко вероватноће.

4.2. БЕРНУЛИЈЕВА ШЕМА НЕЗАВИСНИХ ПОНАВЉАЊА ЕКСПЕРИМЕНАТА

Изузетно важну улогу у теорији вероватноће, њеној примени и развоју статистике, одиграла је једна класа експеримената познатих под именом ***Бернулијеви (Bernoulli) експерименти*** или класа независних понављања експеримента. Одређивање вероватноћа случајних догађаја везаних за Бернулијеве експерименте представља тзв. **Бернулијев проблем**.

Претпоставимо да један експеримент можемо поновити неограничено пута. При томе се мисли да је могуће остварити неограничено пута један исти комплекс услова. Сваки пут кад се оствари тај комплекс услова неки догађај A се може остварити или не тј. догађај A је случајан догађај. Кад се експеримент понавља више пута претпостављамо да резултат у једном понављању не утиче на резултате експеримента у другим понављањима. Кажемо да су понављања експеримента **међусобно независна**. Поред тога се предпоставља да је у сваком понављању експеримента **вероватноћа остварења догађаја A иста**.

Ево неколико примера таквих експеримената.

Пример 1. У статистичком скупу од N елемената $p\%$ њих има одређено својство, док преосталих $(1-p)\%$ нема то својство. На случајан начин се из посматраног скупа бира један елемент и бележи се да ли он има посматрано својство или не. При томе се претпоставља да се извлачење врши са враћањем, тј. сваки пут се извучени елемент враћа у посматрани скуп па се поново из целог скупа бира други елемент, или се претпоставља да је број елемената N јако велики.

Интересује нас проблем да се одреди вероватноћа да ће међу n извучених елемената бити тачно k са посматраним својством. Исто тако нас интересује вероватноћа да ће се количник k/n разликовати од p за не више од једног унапред датог броја. Специјално нас интересује чему ће тежити количник k/n кад број извлачења n тежи бесконачности.

Пример 2. Баца се неограничено пута правилан новчић и бележи се број падања „писма”. Интересује нас чему ће тежити фреквенција појављивања „писма” кад неограничено пута бацамо новчић.

Пример 3. Једна машина аутоматски обрађује неки производ са вероватноћом погрешне обраде једнаком p. Ако се у великој серији од 10000 комада или 100000 комада појави велики број неправилно обрађених комада, да ли то значи да се поузданост машине променила?

Пример 4. Зна се да је међу свим порођајима $p\%$ близанаца. Ако се у једној клиници на коју долазе породиље изабране на случајан начин нађе 1000 породиља, колики се може очекивати број близанаца?

У свим овим примерима посматра се један експеримент и случајан догађај А који се остварује са вероватноћом p, при чему је $q = 1 - p$ вероватноћа да се догађај А неће остварити. При томе је $p + q = 1$.

За низ независних понављања овог експеримента рећи ћемо да представља **Бернулијев експеримент** *ако се у сваком од њих догађај А остварује са вероватноћом* p *и при том вероватноћа* p *остаје непромењена у сваком понављању.*

Означимо са n број понављања посматраног експеримента. Најједноставнији и за нас најинтересантнији проблем је ***одређивање вероватноће да ће се у n понављања посматрани догађај А реализовати тачно одређен број пута, рецимо k пута***. То значи да се у преосталих $(n - k)$ понављања догађај А неће реализовати. При томе је k један одређен број из скупа бројева $\{0, 1, 2, ..., n\}$. Означићемо ту вероватноћу са $P_n(k)$.

Пре него пређемо на одређивање вероватноће $P_n(k)$ у општем случају, посматрајмо један пример.

Нека се у једној кутији налази 5 сијалица за које се зна да је 0,95 вероватноћа за сваку од њих да буде исправна. Колика ће бити вероватноћа да ће се у овој кутији наћи три исправне и две неисправне сијалице?

Догађај A који посматрамо у овом „експерименту" је догађај: „сијалица је исправна", а супротан догађај је $\overline{A}$: „сијалица је неисправна". При томе је $p = 0.95$, $q = 0.05$, а број понављања је једнак броју сијалица, тј. $n = 5$. Тражена вероватноћа је $P_5(3)$. Ако свакој сијалици придружимо догађај A или $\overline{A}$, онда ће се догађај да међу пет сијалица буду три исправне реализовати ако се реализовала једна од могућности писања низа од пет словних места у којима је на 3 места слово А а на два места слово $\overline{A}$. Нека је једна од тих могућности на пример низ

$$A\,A\,\overline{A}\,A\,\overline{A}\,.$$

Вероватноћа да се оствари овај низ је једнака

$$0{,}95 \cdot 0{,}95 \cdot 0{,}05 \cdot 0{,}95 \cdot 0{,}05 = 0{,}95^3 \cdot 0{,}05^2,$$

јер су догађаји у овом низу независни. Оваквих могућих низова има укупно 10. То су

1. $AAA\overline{A}\,\overline{A}$	6. $A\overline{A}A\overline{A}A$
2. $AA\overline{A}A\overline{A}$	7. $\overline{A}AA\overline{A}A$
3. $A\overline{A}AA\overline{A}$	8. $A\overline{A}\,\overline{A}AA$
4. $\overline{A}AAA\overline{A}$	9. $\overline{A}A\overline{A}AA$
5. $AA\overline{A}\,\overline{A}A$	10. $\overline{A}\,\overline{A}AAA$.

Сваки од ових низова има исту вероватноћу која је једнака

$$0{,}95^3 \cdot 0{,}05^2\,.$$

Сви ови низови се међусобно искључују јер се у једном пакету од 5 сијалица може реализовати тачно један од њих. Посматрани догађај се реализује кад се реализовао један од ових низова. Зато је његова вероватноћа једнака збиру вероватноћа сваког од њих, тј.

$$\mathrm{P}_5(3) = 10 \cdot 0{,}95^3 \cdot 0{,}05^3 .$$

На сличан начин се могу одредити и вероватноће

$$\mathrm{P}_5(0), \mathrm{P}_5(1), \mathrm{P}_5(2), \mathrm{P}_5(4), \mathrm{P}_5(5).$$

А сад ћемо посматрати општи случај. Ако се n пута понавља експеримент, колика је вероватноћа да ће се у k понављања догађај А реализовати, а у преосталих $(n-k)$ неће се реализовати?

Сваком извођењу експеримента приписаћемо слово A или $\overline{A}$, зависно од тога да ли се A реализовао или није. Тако ћемо, низу од n понављања експеримента придружити низ од n словних места, који је састављен од слова A и $\overline{A}$. Посматрани догађај ће се реализовати ако се реализује низ у коме ће на k места бити слово A, а на преосталих $(n-k)$ места, биће слово $\overline{A}$.

Вероватноћа сваког таквог низа је једнака:

$$p^k q^{n-k},$$

због независности догађаја.

Са друге стране, сви ти низови се међусобно искључују, јер једном Бернулијевом експерименту од n понављања одговара тачно један од тих могућих низова у којима се посматрани догађај A реализовао k пута. Таквих различитих низова има онолико колико има могућности да се од n словних места изабере k места на која се ставља слово A, (на преостала места се ставља $\overline{A}$), а тај број је једнак:

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Зато је тражена вероватноћа једнака:

$$P_n(k)=\binom{n}{k}p^k q^{n-k}=\frac{n!}{k!(n-k)!}p^k q^{n-k}.$$

Низ вероватноћа

$$P_n(0),\ P_n(1),\dots,\ P_n(n)$$

је **низ биномних вероватноћа** *и представља вероватноће да се у Бернулијевом експерименту догађај А реализовао тачно 0 пута, 1 пута,..., n пута.*

Биномне вероватноће $P_n(k)$ имају мале вредности за велики број *n*, чак и за $n=50$, те вероватноће су доста мале. Види се то из Табеле 4.1. Највећа вероватноћа, за $n=50$ і $p=\frac{1}{3}$, једнака је *0,1178*. Ако је $n=50$ і $p=0{,}5$, највећа вероватноћа је мања од *0,08*. Из табеле 4.1. види се да биномне вероватноће расту до извесне вредности, а затим опадају.

	$P_n(k)$	k	$P_n(k)$
5	0,0001	17	0,1178
8	0,0033	18	0,1080
10	0,0157	19	0,0910
12	0,0470	20	0,0704
14	0,0879	22	0,0332
15	0,1077	25	0,0059
16	0,1178	30	0,0001

Табела 4.1. Биномне вероватноће за $n=50$ *и* $p=\frac{1}{3}$.

Значи да ће се, за једну вредност *k*, добити највећа вредност биномних вероватноћа. Интересује нас која је то вредност.

Биномне вероватноће за k и за $(k+1)$ су једнаке

$$P_n(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

$$P_n(k+1) = \frac{n!}{(k+1)!(n-k-1)!} p^{k+1} q^{n-k-1}$$

Њихов количник је број

$$\frac{P(k+1)}{P(k)} = \frac{n-k}{k+1} \frac{p}{q} \tag{4.5}$$

Ако је овај количник већи од један значиће да са порастом k бинома вероватноћа расте. То ће бити случај кад је

$$\frac{n-k}{k+1} \frac{p}{q} \geq 1,$$

односно кад је

$$k \leq np - q.$$

За

$$k > np - q$$

количник (4.5) је мањи од један, што значи да ће за $k > np - q$ биномне вероватноће опадати.

Ако је $(np - q)$ цео број, количник биномних вероватноћа за $(np - q + 1)$ i $(np - q)$ биће једнак

$$\frac{P_n(np-q+1)}{P_n(np-q)} = \frac{n-(np-q)}{np-q+1} \frac{p}{q}$$

Односно

$$\frac{P_n(np-q+1)}{P_n(np-q)} = 1.$$

На основу изложеног се може рећи:

Ако $np-q$ *није цео број, онда је највероватнији број понављања у којима се реализује случајан догађај A, једнак:*

$$k=\left[np-q\right],$$

при чему је $[\ ]$ *ознака за највећи цео број који није већи од* $np-q$.

Ако је $np-q$ *цео број, онда постоје две вредности k које су највероватније, а то су вредности:*

$$k=np-q \qquad\qquad k+1=np-q+1$$

За те вредности, биномне вероватноће су међусобно једнаке, и веће су од свих осталих вероватноћа.

Пошто ћемо се у каснијим извођењима веома често служити збировима појединих биномних вероватноћа, доказаћемо једну теорему која даје горње границе тих збирова.

Теорема 4.1. *Ако је s цео број који није већи од* np, *tj. ако је*

$$s\le np$$

тада збир биномних вероватноћа од 0 до s је ограничен са горње стране са

$$\sum_{i=0}^{s}P_n(i)<P_n(s)\frac{(n-s+1)p}{(n+1)p-s}, \tag{4.6}$$

а ако је r цео број који испуњава услов

$$r\ge np$$

тада је збир бинимних вероватноћа од r до n ограничен са горње стране са

$$\sum_{i=r}^{n}P_n(i)<P_n(r)\frac{(r+1)q}{(r+1)-(n-1)p}. \tag{4.7}$$

Израз (4.6) представља горњу границу вероватноће да појављивања експеримената у којима се реализовао догађај A неће бити већи од s. Израз (4.7) представља горњу границу вероватноће да број појављивања експеримената у којима се реализује догађај A неће бити мањи од r.

Ако волиш проверу свог логичког размишљања прођи доказ теореме. Можда ћеш открити неку грешку. Ако не, ти веруј да је тачна и запамти је!

Доказ. Количник биномних вероватноћа за $(k+1)$ и k дат изразом (4.5) може се написати и на следећи начин

$$\frac{P_n(k+1)}{P_n(k)} = \left(\frac{n+1}{k+1} - 1\right)\frac{p}{q}$$

одакле се види да он опада кад k расте. То значи да ће реципрочна вредност овог количника расти. Зато низ количника

$$\frac{P_n(0)}{P_n(1)}, \frac{P_n(1)}{P_n(2)}, \ldots\ldots, \frac{P_n(s-1)}{P_n(s)} \tag{4.8}$$

монотоно расте. Сваки од њих има вредност која је мања од последње вредности у том низу. Последња вредност низа (4.8) је

$$\frac{P_n(s-1)}{P_n(s)} = \frac{s}{(n-s+1)}\frac{q}{p}$$

тако да је једначина

$$\frac{P_n(k)}{P_n(k+1)} < \frac{s}{(n-s+1)}\frac{q}{p} \tag{4.9}$$

испуњена за свако

$$k = s-1, s-2, \ldots\ldots, s-i \tag{4.10}$$

Када у (4.9) стављамо редом вредности индекса k из низа (4.10) добићемо и неједначине

$$\frac{P_n(s-1)}{P_n(s)} < \frac{s}{(n-s+1)}\frac{q}{p}$$

$$\frac{P_n(s-2)}{P_n(s-1)} < \frac{s}{(n-s+1)}\frac{q}{p} \qquad (4.11)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{P_n(s-i)}{P_n(s-i+1)} < \frac{s}{(n-s+1)}\frac{q}{p}$$

Множећи неједначине у систему (4.11) добиће се да је

$$\frac{P_n(s-i)}{P_n(s)} < \left\{\frac{s}{(n-s+1)}\frac{p}{q}\right\} \qquad (4.12)$$

Неједначина (4.12) је задовољена за свако $i = 0,1,2,......,s,$ тако да се добије систем од $(s+1)$ неједначина.

$$\frac{P_n(s)}{P_n(s)} < 1$$

$$\frac{P_n(s-1)}{P_n(s)} < \frac{s}{(n-s+1)}\frac{q}{p} \qquad (4.13)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{P_n(0)}{P_n(s)} < \left\{\frac{s}{(n-s+1)}\frac{q}{p}\right\}^s$$

Сабирањем неједначина у систему (4.13) добићемо на десној страни геометријску прогресију са количнком

$$\frac{s}{(n-s+1)}\frac{q}{p} \qquad (4.14)$$

Ако је $s \le np$, количник (4.14) није већи од 1. Зато је збир на десној страни мањи од

$$\frac{1}{1-\dfrac{1}{(n-s+1)}\dfrac{p}{q}}$$

односно, мањи од

$$\frac{(n-s+1)p}{(n+1)p-s} \tag{4.15}$$

На основу (4.15) и збира система (4.13) добија се израз (4.6).

Сличним закључивањем се може проверити и неједначина (4.7), чиме је теорема 4.1. доказана.

4.3. БЕРНУЛИЈЕВ И БОРЕЛ-КАНТЕЛИЈЕВ ЗАКОН ВЕЛИКИХ БРОЈЕВА

Посматраћемо Бернулијеву шему независних понављања експеримента са случајним догађајем A и вероватноћом p реализације догађаја A у сваком понављању. Проблем који нас интересује је следећи: ***Ако се n пута понови посматрани експеримент, шта се дешава са фреквенцијом остварења догађаја и шта се може закључити о међусобној вези фреквенције и вероватноће p?***

Одговор на постављено питање дао је J. Бернули (J. Bernoulli, 1654-1705. године) у познатом делу „*Ars Conjectandi*" објављеном 1713. године. Бернулијеве резултате уопштио је Абрахам Д Муавр (Abraham De Moivre) 1718. године, а затим је Пијер Лаплас (Pierre S. Laplace) 1812. године дао најопштију формулацију одговора на постављено питање. Ради бољег разумевања правог значења Бернулијевих резултата прво ћемо размотрити Бернулијеву формулацију.

Означимо са X број понављања експеримента у којима се реализовао догађај A. Ако се експеримент понови n пута X ће узети једну вредност из низа могућих вредности

$$\{0,1,2,...,n\}.$$

Догађај $D=\{X=k\}$ је случајан догађај који се у Бернулијевом експерименту реализује са вероватноћом

$$P_n(k)=P\{X=k\}=\frac{n!}{k!(n-k)!}p^k q^{n-k} \tag{4.16}$$

Количник

$$\frac{X}{n} \tag{4.17}$$

се назива **фреквенција остварења догађаја** ***A*** *у* ***n*** *пута поновљеном експерименту.*

Фреквенција остварења догађаја А може узети једну вредност из низа могућих вредности

$$\{0, \frac{1}{n}, \frac{2}{n}, ..., 1\}.$$

Догађај $D` = \left\{ \frac{X}{n} = \frac{k}{n} \right\}$ је случајан догађај који се у Бернулијевом експерименту реализује са вероватноћом датом изразом (4.16).

Интересује нас каква је разлика између фреквенције (4.17) и вероватноће p. Зато ћемо посматрати апсолутну вредност те разлике.

$$\left| \frac{X}{n} - p \right|. \tag{4.18}$$

Ако n пута поновимо експеримент, разлика (4.18) ће узети једну вредност из интервала [0, 1] . Кад број понављања повећамо шта ће бити са овом разликом?

Изаберимо произвољно мали позитиван број ε и посматрајмо догађај да ће разлика (4.18) бити већа од тог броја, тј. посматрајмо догађај

$$D = \left\{ \left| \frac{X}{n} - p \right| > \varepsilon \right\}, \tag{4.19}$$

и одредимо вероватноћу остварења овог догађаја у n понављања експеримента.

Догађај D ће се остварити кад је

$$X > n(p + \varepsilon) \tag{4.20}$$

или кад је

$$X < n(p - \varepsilon) \tag{4.21}$$

(Види слику 4.1.).

Slika 4.1.

Вероватноћа догађаја датог изразом (4.20) је једнака збиру биномних вероватноћа

$$P\{X > n(p+\varepsilon)\} = \sum_{i=r}^{n} P_n(i) \tag{4.22}$$

при чему је r најмањи цео број који је већи од производа $n(p+\varepsilon)$.

Из теореме 4.1. (поглавље 4.2) следи да је та вероватноћа ограничена са горње стране тако да је

$$P\{X > n(p+\varepsilon)\} < P_n(r)\frac{n(p+\varepsilon)+q}{n\varepsilon+q}\,. \tag{4.23}$$

Кад број понављања експеримента n неограничено расте, израз на десној страни (4.23) остаје ограничен. Из дефиниције биномних вероватноћа се види да је њихова гранична вредност једнака

$$\lim_{n\to\infty} P_n(r) = 0\,,$$

Зато се из (4.23) може закључити да је гранична вредност вероватноће остварење догађаја (4.20) једнака нули, тј. може се закључити да је

$$\lim_{n\to\infty} P\{X > n(p+\varepsilon)\} = 0\,, \tag{4.24}$$

На сличан начин се може утврдити да је гранична вредност вероватноће догађаја датог изразом (4.21) такође једнака нули, тј. може се утврдити да је

$$\lim_{n\to\infty} P\{X < n(p-\varepsilon)\} = 0. \tag{4.25}$$

Граничне вредности дате са (4.24) и (4.25) представљају граничну вредност догађаја D датог изразом (4.19), тако да је

$$\lim_{n\to\infty} P\left\{\left|\frac{X}{n} - p\right| > \varepsilon\right\} = 0,$$

одакле следи

Теорема 4.2. ***(Бернулијева теорема).*** *Вероватноћа да ће фреквенција остварења догађаја A одступити од његове вероватноће p за вредност мању од произвољно малог броја ε тежи јединици кад број понављања експеримента n неограничено расте тј.*

$$\lim_{n\to\infty} P\{|\frac{X}{n} - p| \le \varepsilon\} = 1 \tag{4.26}$$

Теорема 4.2. представља **Бернулијев закон великих бројева**.

Ако је број понављања n довољно велики, теорема тврди да је скоро сигурно да ће разлика релативне фреквенције остварења догађаја A и његове вероватноће бити произвољно мале. То значи да се може на основу релативне фреквенције скоро са сигурношћу утврдити вероватноћа неког случајног догађаја. Кад би смо експеримент поновљали бесконачно много пута, тада би са сигурношћу утврдили вредност вероватноће p.

Тек кад је доказан овај закон, било је могуће прихватити дефиницију емпиријске вероватноће и на тај начин експерименталним путем одређивати вероватноће случајних догађаја. Овај закон одиграо је пресудну улогу у развоју и примени теорије вероватноће, односно у развоју статистике, јер је тек сад постало јасно да постоје законитости код појава и процеса које нису детерминистичког карактера, тј. код појава и процеса стохастичког карактера. Са развојем теорије вероватноће утврђене су и општије формулације закона великих бројева. Са становишта одређивања самог појма вероватноће и њене суштине интересантно је уопштење које су дали **Е. Борел** (E. Borel) и

Ф. П. Кантели (F. P. Cantelli) које је познато под називом „***јаки закон великих бројева***" и који гласи:

Ако је X број понављања експеримента у којима се случајни догађај A остварио, вероватноћа да ће гранична вредност фреквенција остварења догађаја A бити једнака вероватноћи p је једнака јединици, тј.

$$P\{\lim_{n\to\infty}\frac{X}{n}=p\}=1 \tag{4.27}$$

Закон дат изразима (4.26) и (4.27) потпуно описује фундаментално својство случајности које је могуће проучавати преко вероватноће.

5. ЈЕДНОДИМЕНЗИОНАЛНЕ СЛУЧАЈНЕ ПРОМЕНЉИВЕ

Проучавање појава у природи или друштву, базирано је на успоставаљању и откривању релација између догађаја везаних за дате појаве. Потпуније изучавање доводи до три категорије догађаја:

- немогући;
- сигурни и
- случајни догађаји.

За описивање појава или експеримената који генеришу случајне догађаје користе се ***пробабилистички модели*** засновани на ***Теорији вероватноће.*** Најширу, и за праксу најважнију класу пробабилистичких модела, чини класа модела базирана на *случајним промењљивим*. ***Случајне промењљиве проучавају оне експерименте*** чији резултати се могу изразити нумеричким вредностима: У сваком „извођењу" експеримента „нешто" се пребројава или мери. Резултат експеримента је број из коначног, бесконачног и пребројивог или бесконачно-непребројивог скупа реалних бројева. Другим речима, резултат експеримента је тачка на реалној оси. Такву класу случајних промењљивих представљају *једнодимензионалне случајне промењљиве*.

У овој књизи ћемо се ограничити на проучавање једнодимензионалних случајних промењљивих и упознати се са основним елементима теорије вероватноће једнодимензионалних случајних промењљивих. То су они елементи ове теорије, који су *неопходни за примену у* ***статистичким закључивањима*** - а то је главни циљ статистичких проучавања природних и друштвених промена.

5.1. РАСПОДЕЛА СЛУЧАЈНИХ ПРОМЕЊИВИХ

Кад се баца коцка на равну површину, онда је скуп могућих резултата скуп бројева

$$\{1,2,3,4,5,6\} \tag{5.1}$$

Ако са X означимо резултат експеримента, онда кажемо да промењива величина X може узети једну вредност из скупа реалних бројева (5.1), и то на ***случајан начин.***

Нека је дат статистички скуп од N елемената и нека обележје X има расподелу дату табелом

X	x_1	x_2	...	x_n
f	f_1	f_2	...	f_n

при чему су бројеви

$$\{x_1, x_2, \ldots, x_n\}$$

могуће вредности обележја X на елементима статистичког скупа, а бројеви

$$\{f_1, f_2, \ldots, f_n\}$$

су фреквенције појављивања појединих вредности X у посматраном скупу. Ако на случајан начин бирамо један елемент из посматраног скупа и утврђујемо вредност обележја X код тог елемента, тада је промењива величина X случајна промењива која узима једну вредност из низа могућих *вредности статистичког скупа и то на случајан начин.*

На телефонској централи се посматра број позива у временском интервалу. Промењива величина, која представља број позива, може узети једну од вредности из низа реалних бројева.

$$\{0,\ 1,\ 2,\ \ldots\} \tag{5.2}$$

и то на случајан начин.

При обради једног производа, на тачност неке димензије утиче низ фактора. Зна се да ће димензија X бити у неким утврђеним границама, рецимо у интервалу

$$(x_1, x_2) \tag{5.3}$$

Код сваког обрађеног производа, димензија X ће *на случајан начин* узети једну вредност из интервала (5.3).

Посматра се век трајања једног уређаја. Резултат може бити било који реалан број из интервала

$$(0, \infty) \tag{5.4}$$

Сваки случајни догађај, везан за овакве појаве, може се изразити преко променљиве X. Тако се, на пример, у бацању коцке, случајни догађај A: „Пао је број не мањи од 3”, може изразити са

$$A = \{X \geq 3\} = \{3,4,5,6\}$$

Вероватноћа догађаја A може се одредити лако ако се зна вероватноћа са којом X узима поједине вредности из скупа могућих вредности $\{1,2,3,4,5,6\}$.

Ако је телефонска централа ограниченог капацитета, рецимо највише 100 позива у датом интервалу, онда се случајни догађај A: „Телефонска централа неће бити заузета”, може изразити преко променљиве X са

$$A = \{0 \leq X < 100\}$$

при чему је X број позива у посматраном временском интервалу. Ако знамо вероватноћу појединих вредности променљиве X, лако ћемо одредити вероватноћу случајног догађаја A.

У сваком експерименту, резултатима експеримента придружујемо променљиву величину X која у једном извођењу експеримента може узети само једну вредност из датог скупа реалних бројева. Унапред не знамо коју ће вредност узети X. Једино што можемо утврдити је скуп могућих вредности за ту променљиву.

У даљим разматрањима оваквих променљивих величина, први задатак је утврдити *како често*, односно *са којом вероватноћом* ће променљива X узимати поједине вредности из датог скупа могућих вредности.

У пракси се посматрају две категорије случајних променљивих и то *прекидног* и *непрекидног* типа. Посматраћемо прво случајне величине прекидног типа и на основу њихове дефиниције и основних особина утврдити потребне елементе за дефиницију и изучавање непрекидних случајних променљивих.

Закон вероватноћа прекидне случајне променљиве

Дефиниција 5.1: *Променљива величина* **X је случајна променљива прекидног типа**, *ако вредности, које она може узети, образују коначан или пребројив низ реалних бројева*

$$\{x_1, x_2, x_3, ...\}, \tag{5.5}$$

а узимање сваке од ових вредности је случајан догађај са одређеном вероватноћом.

Посматрајмо низ случајних догађаја

$$D_1=(X=x_1),\ D_2=(X=x_2),\ D_3=(X=x_3), \tag{5.6}$$

и означимо са

$$p_1,\ p_2,\ p_3,\ ... \tag{5.7}$$

одговарајуће вероватноће догађаја $D_1, D_2, ...,$ при чему је p_i вероватноћа да ће X узети вредност x_i, за i=1,2, ...тј.

$$p_i=P(X=x_i).$$

Случајни догађаји (5.6) се међусобно искључују. Њихова унија је сигуран догађај, јер се у једном извођењу експеримента реализује тачно један од њих. Зато низ одговарајућих вероватноћа (5.7) представља низ ненегативних бројева чији је збир једнак јединици, тј. низ је такав да је

$$\sum_{i=1}^{\infty} p_i = 1 . \tag{5.8}$$

У случају када је низ могућих вредности случајне променљиве (5.5) коначан низ од *n* бројева, збир у (5.8) биће збир од 1 до *n*, тако да је

$$\sum_{i=1}^{n} p_i = 1 . \tag{5.9}$$

Дефиниција 5.2: *Скуп могућих вредности случајне променљиве X и одговарајућих вероватноћа, називамо закон вероватноћа случајне променљиве X и најчешће га приказујемо у виду табеле*

X	x_1	x_2	x_3	...
$P(x)$	p_1	p_2	p_3	...

при чему је

$$\forall i \in \{1,2,3,\cdots\} \Rightarrow p_i = P(X = x_i) .$$

Кажемо да је ***законом вероватноћа*** *одређена* ***расподела случајне променљиве.***

Расподела случајне променљиве *X* садржи максималну количину информација о посматраном експерименту, односно, о посматраној појави или процесу. Поред тога, ова расподела омогућава комплетну анализу посматране појаве.

На пример, бацају се истовремено два новчића и региструје број појављивања „писма”. Ако означимо са *X* резултат експеримента, онда ће *X* бити случајна променљива са могућим вредностима $\{0,1, 2\}$

и одговарајућим вероватноћама

$$\left\{\frac{1}{4},\frac{2}{4},\frac{1}{4}\right\}.$$

Закон вероватноћа случајне промењиве X је дат табелом:

X	0	1	2
$p(x)$	0.25	0.50	0.25

Ако се, на пример, гађа неки циљ све до момента његовог погађања, може се посматрати број извршених гађања X. Нека је, у сваком гађању, вероватноћа поготка циља једнака p, при чему је $0<p<1$. Тада је, због независности, вероватноћа да број гађања X буде једнак, рецимо k, једнака

$$P(X=k)=\left(1-p\right)^{k-1}p\,,$$

јер ће се догађај

$$D=\left(X=k\right)$$

реализовати једино ако је у (k-1) гађања циљ промашен, а погођен у k-том гађању. Закон вероватноћа случајне промењиве X је дат следећом табелом:

X	1	2	3	...
$p(x)$	p	$(1-p)p$	$(1-p)^2p$	...

Посматрајмо статистички скуп од N елемената и расподелу обележја X на том скупу, дату табелом

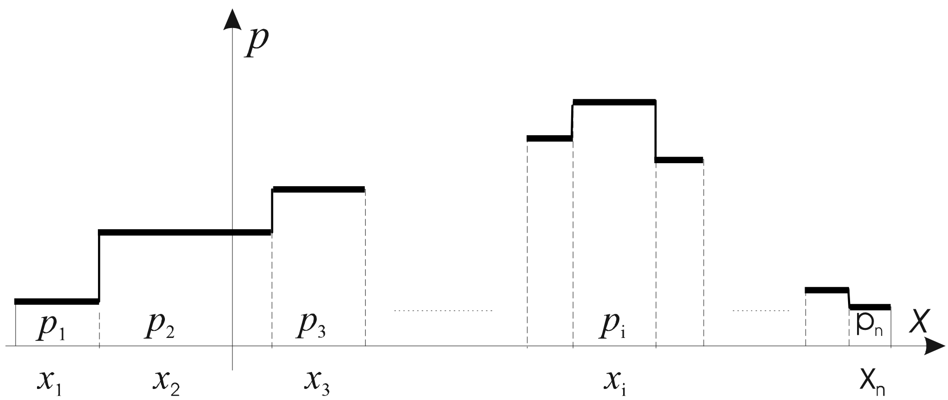
X	x_1	x_2	...	x_n
F	f_1	f_2	...	f_n

Ако на случајан начин из посматраног скупа бирамо један елеменат, тад је вредност обележја X на изабраном елементу случајна промењива са расподелом

X	x_1	x_2	...	x_n
$p(x)$	f_1/N	f_2/N	...	f_n/N

Ова расподела није ништа друго него расподела релативних фреквенција обележја X на посматраном статистичком скупу. Зато се сваки ĩgкључак који изведемо за случајну променљиву X односи и на расподелу обележја X на статистичком скупу.

Расподела случајне променљиве *X* представља се графички у виду **хистограма** приказаног на сл 5.1. На оси *X* се наносе могуће вредности случајне променљиве *X,* а затим се изнад сваке тачке конструише правоугаоник чија је површина једнака одговарајућој вероватноћи.



Слика 5.1. Закон вероватноћа случајне променљиве X

Пошто је

$$\sum_{i=1}^{n} p_i = 1,$$

на хистограму расподеле случајне променљиве, цела површина између осе *X* и криве којом је ограничен хистограм, треба да је једнака јединици.

На основу расподеле случајне променљиве *X* прекидног типа, могуће је одредити расподелу било које случајне променљиве *Y* која је дефинисана преко *X*. Ако између случајних променљивих *X* и *Y* постоји веза

$$Y = g(X),$$

тада је закон вероватноћа Y дат у следећој табели

Y	$y_1=g(x_1)$	$y_2=g(x_2)$	$y_3=g(x_3)$	...
$p(y)$	p_1	p_2	p_3	...

Најчешће се посматра линеарна функција

$$Y = aX + b,$$

при чему је $a \neq 0$. Тада је закон вероватноћа случајне промењиве Y дат табелом.

Y	ax_1+b	ax_2+b	ax_3+b	...
$p(y)$	p_1	p_2	p_3	...

Ако се зна закон вероватноћа случајне промењиве X, могуће је лако одредити вероватноће било ког случајног догађаја везаног за тај експеримент. Наиме, случајни догађај је подскуп скупа могућих вредности, тј. подскуп скупа реалних бројева

$$\{x_1, x_2, x_3, \ldots\}$$

Када одредимо подскуп који одговара догађају, на основу вероватноћа појединих тачака тог подскупа, лако се одређује вероватноћа посматраног догађаја.

Пример 1. Ако се сматра да је рађање девојчице случајан догађај са вероватноћом 0.5, а дечака такође, са вероватноћом 0,5, колика је вероватноћа да ће у породици са четворо деце бити бар две девојчице? Означимо са X број девојчица у породици са четворо деце. Тада X може узети једну вредност из скупа

$$\{0, 1, 2, 3, 4\},$$

са вероватноћама {0,0625; 0,2500; 0,3750; 0,2500; 0,0625}.

Тако је закон вероватноћа за X дат табелом

X	0	1	2	3	4
$p(x)$	0.0625	0.2500	0.3750	0.2500	0.0625

Посматрани догађаји ће се реализовати ако се реализује једна од вредности из подскупа $\{2, 3, 4\}$.

Тражена вероватноћа је једнака

$$P(2 \leq X \leq 4) = P(X = 2) + P(X = 3) + P(X = 4)$$

односно,

$$P(2 \leq X \leq 4) = 0{,}3750 + 0{,}2500 + 0{,}0625 = 0{,}6875 .$$

Ако, у посматраном примеру, знамо да у породици има један дечак и једна девојчица колика би била вероватноћа да ће у тој породици бити бар два дечака? Тражена вероватноћа је условна вероватноћа

$$P(A/B),$$

при чему је

$$A = (0 \leq X \leq 2)$$

$$B = (1 \leq X \leq 3)$$

Пошто је

$$A \cdot B = (1 \leq X \leq 2)$$

из дефиниције условне вероватноће следи да је

$$P(A/B) = \frac{P(A \cdot B)}{P(B)} = \frac{0{,}2500 + 0.3750}{0{,}2500 + 0{,}3750 + 0{,}2500} = 0{,}7143.$$

Пример 2. Ради утврђивања квалитета у великој серији, узима се један производ и врши комплетна анализа. Ако изабрани производ, у потпуности, одговара жељеном квалитету, даља анализа се не врши, а ако не одговара, бира се следећи и врши његова анализа, итд. Нека је вероватноћа за сваки производ, да одговара жељеном квалитету, једнака 0.8, а трошкови анализе су 2000 динара. Интересује нас вероватноћа да ће трошкови контроле серије бити мањи од 5000 динара.

Означимо са X број производа који ће у контроли серије, бити анализирани. Тада X има закон вероватноће

X	1	2	3	...
$p(x)$	0.8	$0.2 \cdot 0.8$	$0.04 \cdot 0.8$	...

Означимо са Y трошкове контроле серије. Тада је

$$Y = 2000 \cdot X$$

тако да је закон вероватноћа за трошкове Y дат табелом:

Y	2000	4000	6000	...
$p(y)$	0.8	0.16	0.032	...

Трошкови контроле ће бити мањи од 5000, ако је анализиран један или два производа, па је вероватноћа да трошкови контроле буду мањи од 5000, једнака

$$P(Y < 5000) = P(X=1) + P(X=2)$$

односно,

$$P(Y < 5000) = 0.96.$$

Закон вероватноћа непрекидне случајне промењиве

При гађању мете мери се удаљеност поготка од центра. Промењива величина X биће један од реалних бројева из интервала (0, a), а при једном „извођењу” експеримента (у једном гађању), X ће узети само једну вредност из тог интервала.

Век трајања сијалице је промењива величина која ће узети вредност из интервала $(0, +\infty)$. За једну сијалицу (једно „извођење” експеримента), век трајања ће бити само један број из овог интервала.

У оваквим експериментима, посматрана променљива величина *X* може узети једну вредност из неког интервала. Сматраћемо да је то бесконачни интервал

$$(-\infty,\ +\infty). \tag{5.10}$$

Ако је у експерименту интервал коначан тада ћемо сматрати да су тачке изван тог интервала немогући догађаји са вероватноћама једанаким нули.

Случајна променљива непрекидног типа треба да је тако одређена да је могуће тачно утврдити њену расподелу, као и код променљивих прекидног типа. По аналогији са случајним променљивим прекидног типа, и овде се мора дефинисати закон вероватноћа тако да је могуће да за сваки случајни догађај везан за тај експеримент, буде одређена његова вероватноћа.

У овом случају, интервал $(-\infty;\ +\infty)$, поделимо на подинтервале произвољно мале дужине Δx. Уместо тачака $x \in (-\infty;\ +\infty)$ посматраћемо интервале $\left(x-\frac{\Delta x}{2};x+\frac{\Delta x}{2}\right)$. За тако одређене интервале могуће је одредити једну функцију *f*(*x*), преко које ћемо дефинисати вероватноће.

<u>Дефиниција 5.3:</u> *За променљиву величину X кажемо да је случајна променљива непрекидног типа ако узимање било које вредности из интервала* $(-\infty;\ +\infty)$ *представља случајан догађај и ако постоји функција f(x) таква да је вероватноћа, да ће се X наћи у призвољно малој околини тачке x, тј. у интервалу*

$$\left(x-\frac{\Delta x}{2};x+\frac{\Delta x}{2}\right) \tag{5.11}$$

једнака производу $f(x)\cdot\Delta x$ *; тј. ако је*

$$P\left(x-\frac{\Delta x}{2}<X\le x+\frac{\Delta x}{2}\right)=f(x)\cdot\Delta x \tag{5.12}$$

и за сваку произвољно малу вредност Δx.

Функција f(x) назива се ***закон вероватноћа случајне променљиве X.***

Према томе, случајна променљива непрекидног типа биће одређена функцијом $f(x)$, односно, *законом вероватноћа*.

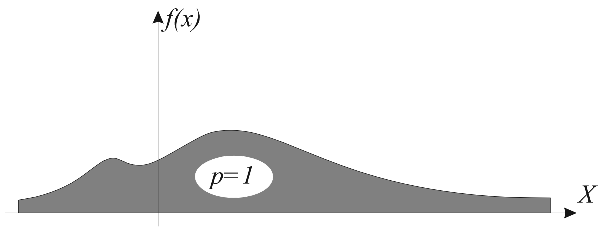
Закон вероватноћа $f(x)$ мора задовољавати следећа два услова:

1. $f(x) \geq 0$, за свако $x \in (-\infty, +\infty)$;

2. $\int_{-\infty}^{+\infty} f(x)dx = 1$. (5.13)

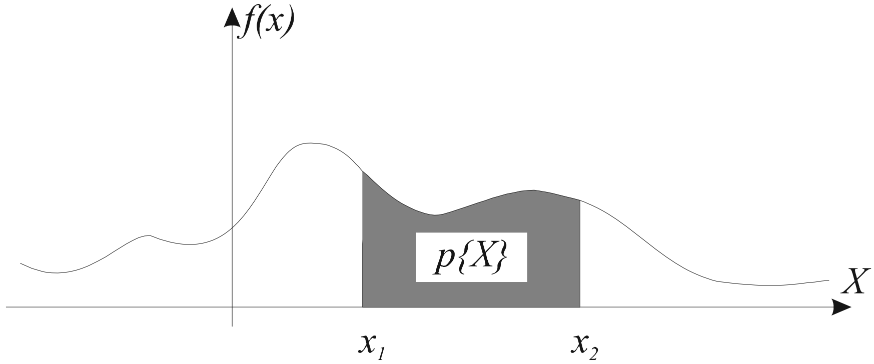
Први услов следи из (5.12) и особине ненегативности вероватноћа.

Пошто се у посматраном експерименту реализује тачно једна вредност x из $(-\infty, +\infty)$, онда је „збир” вероватноћа за све тачке у том интервалу једнак јединици (види сл. 5.2), а тај „збир” није ништа друго до интеграл (5.13). Геометријски, то је површина између X осе и закона вероватноћа $f(x)$.



Слика 5.2. Закон вероватноћа случајне променљиве X.

Најчешће нас интересује одређивање вероватноћа да ће се случајна променљива наћи у неком интервалу, рецимо интервалу (x_1, x_2). Тражену вероватноћу ћемо наћи тако што ћемо, за све тачке из интервала (x_1, x_2), „сабрати” одговарајуће вероватноће дате са (5.12). Тај збир није ништа друго до одређени интеграл функције $f(x)$ у границама (x_1, x_2). На слици 5.3. то је шрафирана површина.



Слика 5.3. Вероватноћа да се X нађе у интервалу (x_1, x_2).

Тражена вероватноћа једнака је

$$P\left(x_1 < X \leq x_2\right) = \int_{x_1}^{x_2} f\left(x\right)\, dx\,. \tag{5.14}$$

Сваки случајни догађај, везан за посматрани експеримент, може се изразити преко интервала, а вероватноће интервала одредићемо преко (5.14). Зато ћемо, на основу закона вероватноћа $f(x)$, моћи одредити и вероватноће било ког случајног догађаја.

На пример, произвођач аутомобилских гума тврди да су његове гуме веома поуздане. Зна се да је за сваку гуму њен "век трајања" случајна променљива са законом вероватноћа

$$f(x) = \begin{cases} 25 \cdot 10^{-6}\, e^{25 \cdot 10^{-6} x} & за\ x > 0 \\ 0 & за\ x \leq 0 \end{cases} \tag{5.15}$$

Функција (5.15) је, заиста закон вероватноћа. Задовољава услове:

1. $f(x) \geq 0$,

2. $\int_{-\infty}^{+\infty} f\left(x\right) dx = \int_{-\infty}^{+\infty} 25 \cdot 10^{-6}\, e^{-25 \cdot 10^{-6} x} \cdot dx = 1\,.$

Тражи се вероватноћа да ће купљена гума издржати 40.000 km, односно, $4 \cdot 10^4$ km. То ће се реализовати када је $X \geq 4 \cdot 10^4$, тако да је тражена вероватноћа једнака

$$P(X > 40000) = 25 \cdot 10^{-6} \int_{4 \cdot 10^4}^{\infty} e^{-25 \cdot 10^{-6} x} dx = 0.3678789,$$

а вероватноћа да ће гума издржати 20.000 km, је

$$P(X > 20.000) = 0.6065$$

Узмимо још један пример. Нека X има закон вероватноћа

$$f(x) = \frac{2(b - x)}{b^2}$$

за $x \in [0, b]$. Потребно је одредити z тако да, са вероватноћом 0.5, тврдимо да ће случајна промењљива X узети вредност већу од z.

Тражена вредност z је број за који је

$$P(X > z) = 0,5$$

односно,

$$\int_{z}^{b} f(x)\, dx = 0,5$$

Решење интеграла даје једначину

$$\left(\frac{z\text{-}b}{b}\right)^2 = 0,5$$

одакле се добија да је

$$z = b\left(1\text{-}\sqrt{0,5}\right).$$

Функција расподеле случајне промењиве

У практичним одређивањима вероватноћа случајних догађаја везаних за случајне промењиве, често смо заинтересовани за једну класу догађаја који се формулишу на следећи начин:

Случајна величина X узеће вредност не већу од z, при чему је *z* унапред одређен број из интервала $(-\infty,\ +\infty)$.

На пример, интересује нас догађај да ће тражња за једним производом бити мања од једне одређене вредности, рецимо 1.000 јединица у датом временском интервалу.

Или нас интересује догађај да ће, у аутоматској централи број позива бити мањи од неког броја, који представља капацитет централе.

За лакше одређвање вероватноћа оваквих догађаја, служи функција расподеле случајне промењиве *X*, која се дефинише на следећи начин:

<u>Дефиниција 5.4.:</u> *Функција расподеле случајне промењиве X је функција F(z) која, за свако*

$$z \in (-\infty, +\infty) \tag{5.16}$$

представља вероватноћу да случајна промењива X неће узети вредност већу од z, тј.

$$F(z) = P(X \leq z). \tag{5.17}$$

Пошто су вредности функције расподеле *F(z)* вероватноће, то значи да ће та функција имати вредности између нула и један, тј. за свако $z \in (-\infty, +\infty)$

$$0 \leq F(z) \leq 1 \tag{5.18}$$

Функција расподеле случајне промењиве *X* је функција са следећим особинама:

***Особина* 1.** $F(-\infty)=0$.

$F(-\infty)$ је вероватноћа да ће X бити мање од $-\infty$, а то је немогућ догађај чија је вероватноћа једнака 0.

***Особина* 2.** $F(+\infty) = 1$.

$F(+\infty)$ је вероватноћа да ће X бити мање од $+\infty$, а то је сигуран догађај са вероватноћом једнаком један.

***Особина* 3.** *Функција F(z) је монотоно неопадајућа функција.*

Заиста, ако су z' и z'' две вредности, такве да је $z' \leq z''$, онда догађај

$$D' = (-\infty < X \leq z')$$

имплицира догађај

$$D'' = (-\infty < X \leq z''),$$

што значи да је

$$P(D') \leq P(D'') . \tag{5.19}$$

Пошто је

$$P(D') = F(z')$$

и

$$P(D'') = F(z''),$$

онда, из (5.19) следи да је

$$F(z') \leq F(z'') \tag{5.20}$$

Неједнакост (5.20) одређује монотоност функције.

***Особина* 4.** *Код случајне промењљиве X прекидног типа, функција расподеле F(z) је степенаста функција која у тачкама*

$$x_1, x_2, x_3, \ldots$$

има скокове једнаке одговарајућим вероватноћама.

Наиме, за свако

$$
\begin{aligned}
&z \in (-\infty,\ x_1) \Rightarrow F(z) = 0,\\
&z \in (x_1, x_2) \Rightarrow F(z) = P(X{=}x_1){=}p_1,\\
&z \in (x_2, x_3) \Rightarrow F(z) = P(X{=}x_1){+}P(X{=}x_2){=}p_1 + p_2\\
&\dots\\
&z \in (x_k, x_{k+1}) \Rightarrow F(z) = \sum_{i=1}^{k} P(X{=}x_i){=}\sum_{i=1}^{k} p_i\\
&\dots
\end{aligned}
\tag{5.21}
$$

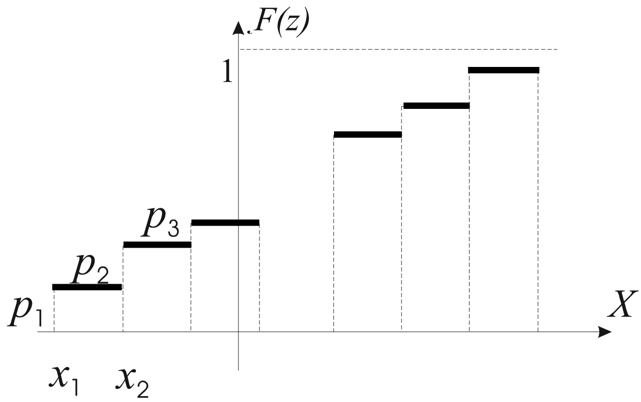
У овом случају, функција расподела *F(z)* је функција која има облик приказан на слици 5.4.

На основу израза (5.21) може се лако одредити фунцкија расподеле *F(z)*, када се зна закон вероватноћа случајне променљиве *X*.

Обрнуто, ако је позната функција расподела *F(z)*, онда се из (5.21) добија да је

$$
\begin{aligned}
&p_1 = F(x_1)\\
&p_2 = F(x_2) - F(x_1)\\
&p_3 = F(x_3) - F(x_2)\\
&\dots\dots\dots\dots\dots\dots\dots\\
&p_k = F(x_k) - F(x_{k-1})\\
&\dots\dots\dots\dots\dots\dots\dots,
\end{aligned}
\tag{5.22}
$$

са чиме је потпуно одређен закон вероватноћа случајне промељиве *X*.



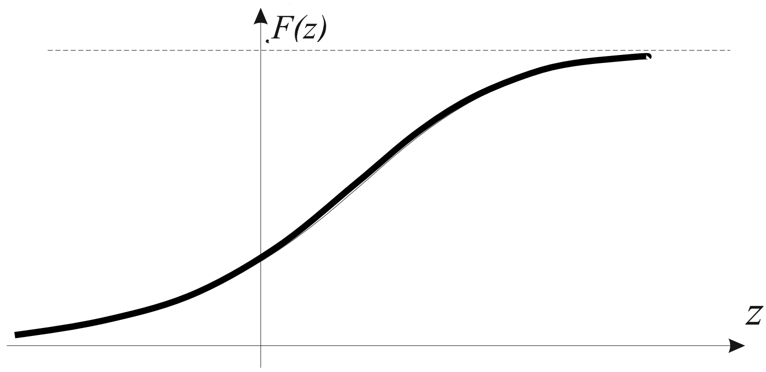
Слика 5.4. Функција расподеле случајне променљиве X прекидног типа.

Особина 5. *Ако је X ЂЛрекидна случајна промељива са законом вероватноћа f(x), тада је, за свако* $z \in (-\infty, +\infty)$

$$F(z) = \int_{-\infty}^{z} f(x)\, dx \,. \tag{5.23}$$

Функција расподеле *F(z)* је непрекидна и монотоно неопадајућа функција, или је прекидна за коначно много тачака прекида.

Функција расподеле непрекидне случајне променљиве *X* има облик криве, приказане на слици 5.5.



Слика 5.5. Функција расподеле непрекидне случајне променљиве X.

Када нам је познат закон вероватноће *f(x)* једне случајне промењиве, решавањем интеграла у (5.23), одредићемо њену функцију расподеле.

Обрнуто, ако нам је позната функција расподеле *F(z),* закон вероватноћа *f(x)* одредићемо преко извода функција *F(z)*. Тада је за свако $z \in (-\infty, +\infty)$

$$f(x) = \frac{dF(z)}{dz}\Big|_{z=x} \tag{5.24}$$

***Закључак**: Под расподелом случајне променљиве X треба подразумевати, или њен **Закон вероватноћа**, или **Функцију расподеле**. Када знамо закон вероватноћа, лако одређујемо функцију расподеле, и обрнуто, када знамо функцију расподеле случајне промељиве X, лако одређујемо њен закон вероватноћа. Расподелом случајне променљиве одређена је максимална количина информација о експерименту са случајним исходима.*

Посматрајмо следећи пример. Познато је да је свака четврта особа алергична на антибиотике. Ради провере новог лека, на случајан начин је изабрана група од седам лица. Одредићемо закон вероватноћа случајне променљиве *X*, који представља број изабраних лица који су алергични на антибиотике.

Случајна променљива *X* је прекидног типа и може узети једну од вредности:

0, 1, 2, 3, 4, 5, 6, 7.

За сваку изабрану особу, вероватноћа да буде алергична на антибиотике има вредност

$$p = 0.25,$$

а избор особа вршимо независно. Значи, седам пута „понављамо” експеримент, и у сваком „понављању” посматрамо случајни догађај *A*: „Изабрано лице је алергично на антибиотике”. Догађај *A* се, у сваком „понављању” реализује са вероватноћом 0,25. Случајна промељива *X* представља број „понављања експеримента” у којима се догађај A реализовао. Зато је вероватноћа да X узме неку вредност k једнака

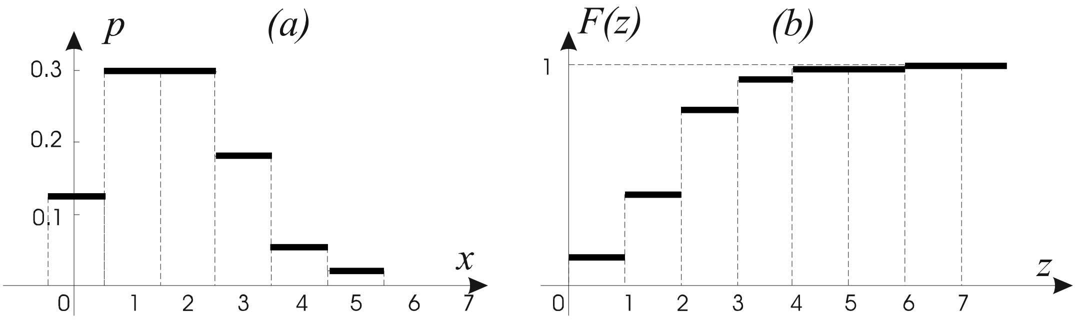
$$P(X=k)=P_7(k)=\binom{7}{k}\ 0.25^k\ \ 0.75^{7-k}$$

тј. вероватноће за поједине вредности случајне промељиве X су тзв. *биномне вероватноће $P_7(k)$.*

Закон вероватноћа случајне променљиве X дат је табелом:

X	0	1	2	3	4	5	6	7
$p(x)$	0.133	0.311	0.311	0.173	0.057	0.011	0.001	0.0001

Његов хистограм је приказан на слици 1.6.(а).



Слика 5.6. Закон вероватноћа и функција расподеле.

Функција расподеле ове случајне променљиве је дата са:

$$z \in (-\infty,\ 0) \Rightarrow F(z) = 0$$
$$z \in [0,\ 1) \quad \Rightarrow F(z) = 0{,}1335$$
$$z \in [1,\ 2) \quad \Rightarrow F(z) = 0{,}4450$$
$$z \in [2,\ 3) \quad \Rightarrow F(z) = 0{,}7565$$

$$z \in [3,\ 4) \quad \Rightarrow F(z) = 0{,}9295$$
$$z \in [4,\ 5) \quad \Rightarrow F(z) = 0{,}9872$$
$$z \in [5,\ 6) \quad \Rightarrow F(z) = 0{,}9987$$
$$z \in [6,\ 7) \quad \Rightarrow F(z) = 0{,}9999$$
$$z \in [7,\ -\infty) \Rightarrow F(z) = 1{,}000$$

Ова функција приказана је на Слици 5.6 (b).

На сличан начин, посматраћемо пример непрекидне случајне промељиве X са законом вероватноћа

$$f(x) = \begin{cases} \dfrac{1}{4}(3-x) \ \text{за}\ 0<x<2 \\ 0 \qquad\qquad \text{за}\ x \le 0,\ x \ge 2 \end{cases} \tag{5.25}$$

Функција расподеле $F(z)$ је одређена на следећи начин:

$$\text{ĩg}\ z \in (-\infty, 0] \Rightarrow F(z) = 0\,;$$

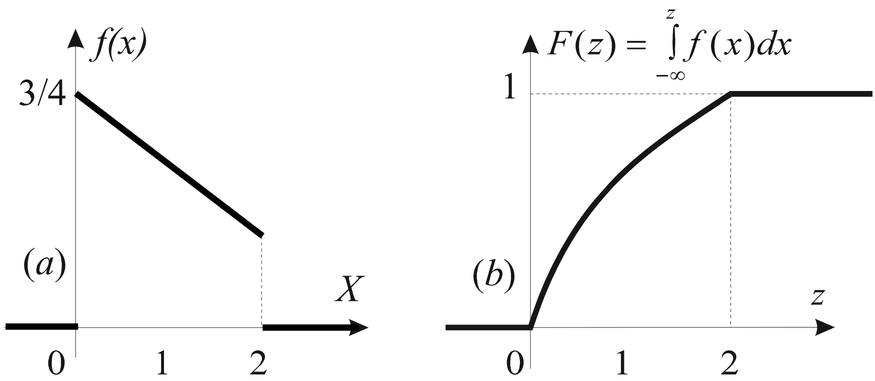
$$\text{ĩg}\ z \in (0,2) \Rightarrow F(z) = \frac{1}{4}\int_0^z (3 - x)dx = \frac{z}{4}(3 - \frac{z}{2})\,;$$

$$\text{ĩg}\ z \in [2,+\infty) \Rightarrow F(z) = \frac{1}{4}\int_0^2 (3 - x)dx = 1\,.$$

Према томе, функција расподеле $F(z)$, за случајну промељиву X, са законом вероватноћа (5.25), дата је функцијом

$$F(z) = \begin{cases} 0 & \text{за}\ z \le 0 \\ \dfrac{z}{4}\left(3 - \dfrac{z}{2}\right) & \text{за}\ 0<z<2 \\ 1 & \text{за}\ z \ge 2. \end{cases} \tag{5.26}$$

Функције (5.25) и (5.26) приказане су на слици 5.7. (а) и (b).



Слика 5.7. Закон вероватноћа и функција расподеле случајне промењљиве.

5.2. АНАЛИЗА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Карактериситке једне случајне променљиве X садржане су у њеној расподели датој законом вероватноћа или функцијом расподеле. Зато се, у употреби случајних променљивих, мора радити са расподелом, а то отежава примењивост теорије вероватноће базираној на случајним променљивим.

Са друге стране, у пракси, најчешће не знамо расподелу случајне променљиве. У таквим случајевима се показало да је познавање одређених нумеричких карактеристика довољно за решавање практичних проблема примене теорије вероватноћа. Зато је потребно одредити нумеричке карактеристике по унапред одређеним правилима, па уместо целог закона вероватноћа, односно фунцкије расподеле случајне променљиве, посматрати одређени број параметара, који ће, у довољној мери, описивати посматрану променљиву.

Као што се, пре употребе неког грађевинског материјала, прво морају утврдити његове карактеристике, и то тако што ће се извршити мерење вредности посматраних карактеристика, тако се и у употреби случајних променљивих прво морају "измерити" одређене карактеристике. Претходно се мора обезбедити процедура или начин на који се врши „мерење” жељених карактеристика.

Одређивање појединих карактеристика, односно параметара случајне променљиве X, врши се преко ***очекиване вредности*** или ***математичког очекивања***. Зато ћемо се, укратко, упознати са овим појмом, а затим ћемо прећи на одређивање појединих параметара случајних променљивих.

Очекивана вредност

У анализи статистичких података на појавама масовног карактера, служили смо се расподелом фреквенција и могућим вредностима посматраног обележја. При томе смо посматрали поједине вредности, или њихове функције посматраног обележја, и њих смо множили апсолутним фреквенцијама, затим сабирали и делили са укупним бројем елемената статистичког скупа. Тако смо, на пример одредили аритметичку средину

$$\overline{x}=\sum_{i=1}^{n} x_i \cdot p_i, \tag{5.28}$$

при чему су p_i релативне фреквенције, $p_i = \dfrac{f_i}{N}$.

Пошто се расподела статистичког скупа може сматрати расподелом случајне промењљиве *X*, онда аритметичка средина може представљати „просечну вредност” случајне промењљиве, па ће нам (5.28) послужити и за дефиницију очекиване вредности случајне промењљиве.

Дефиниција 5.5.: *Нека је дата расподела случајне промењљиве X*

X	x_1	x_2	x_3	...
$p(x)$	p_1	p_2	p_3	...

Ако ред

$$\sum_{i=1}^{\infty} x_i \cdot p_i$$

конвергира, његову суму називамо очекиваном вредношћу, или краће очекивањем случајне поменљиве X *и означавамо га са* $E(X)$, *па је*

$$E(X)=\sum_{i=1}^{\infty} x_i \cdot p_i \tag{5.29}$$

<u>Дефиниција 5.6.:</u> *Нека је f(x) закон вероватноћа случајне променљиве X. Ако постоји интеграл*

$$\int_{-\infty}^{\infty} x f(x)\, dx,$$

његова вредност се назива очекиваном вредношћу, или краће, очекивањем случајне променљиве X и означава се са $E(X)$, тако да је

$$E(X)= \int_{-\infty}^{\infty} x f(x)\, dx. \tag{5.30}$$

Ако се, преко случајне променљиве *X*, дефинише нова случајна променљива *Y* функцијом

$$Y = g(X)$$

тада је *очекивана вредност променљиве Y* одређена преко расподеле променљиве *X*, као сума реда

$$E\left[g(X)\right]=\sum_{i=1}^{\infty} g(x_i)p_i \tag{5.31}$$

односно, као интеграл

$$E\left[g(X)\right]=\int_{-\infty}^{\infty} g(x)\, f(x)\, dx \tag{5.32}$$

Очекивање случајне променљиве *X* и очекивање функција случајне променљиве *X*, је основни апарат у анализи случајних променљивих.

Пример. Типичан проблем **залиха** је проблем одређивања броја јединица неког артикла. Тражња за тим артиклом се може сматрати случајном променљивом. Потражње су нам алтернативе за набавку, али нам нису познати трошкови, односно, зараде за поједине алтернативе. Очекивани трошкови, односно, очекивана зарада ће нам послужити као критеријум за избор најбоље алтернативе.

Предпоставимо да продавница треба да купи 50, 60 или 70 сезонских хаљина, а да је њена зарада 100 динара по комаду. На крају сезоне непородате хаљине се могу продати по сниженој цени која ће довести до губитака од 75 динара по непродатој хаљини. Нека се зна да је вероватноћа да се прода 50 хаљина, 0,30, а да се прода 60, је 0,50, док је вероватноћа продаје 70 хаљина 0,20. У Табели 5.1. дате су очекиване зараде за посматране одлуке.

Из последње врсте Табеле 5.1, види се да је најбоља алтернатива набавити 60 хаљина. Код ње је очекивана зарада 5475 динара.

		Алтернативе					
Тражња	Вероватноћа	Набавити 50 хаљина		Набавити 60 хаљина		Набавити 70 хаљина	
		Условна зарада	Очекивана зарада	Условна зарада	Очекивана зарада	Условна зарада	Очекивана зарада
50	0,30	5000	1500	4250	1275	3500	1050
60	0,50	5000	2500	6000	3000	5250	2625
70	0,20	5000	1000	6000	1200	7000	1400
Укупна очекивана зарада			5000		5475		5075

Табела 5.1. Рачунање очекиване зараде

Ако случајна величина X има расподелу

X	10	20	30	40
$p(x)$	0.1	0.3	0.5	0.1

тада је њено очекивање једнако

$$E(X) = 10 \cdot 0.1 + 20 \cdot 0.3 + 30 \cdot 0.5 + 40 \cdot 0.1 = 26$$

а очекивана вредност функције

$$g(X) = 2X^2 - 900$$

једнака је

$$E\left[g(X)\right]=(2\cdot 10^2-900)\cdot 0.1+(2\cdot 20^2-900)\cdot 0.3+$$
$$+(2\cdot 30^2-900)\cdot 0.5+(2\cdot 40^2-900)\cdot 0.1=560.$$

За случајну променљиву *X*, код које је закон вероватноћа

$$f(x)=\begin{cases}\frac{1}{50}x & za\ 0<1\leq 10\\ 0 & za\ x\leq 0,\ x>10.\end{cases}$$

Очекивана вредност случајне променлјиве *X* је једнака

$$E(x)=\frac{1}{50}\int_0^{10}x^2dx=20/3.$$

Очекивана вредност функције

$$g(x)=2X^2-4$$

је једнака интегралу

$$E\left[2X^2\text{-}4\right]=\frac{1}{50}\int_0^{10}(2x^2\text{-}4)x\,dx=28/3.$$

Ради лакшег одређивања очекиване вредности, навешћемо неке њене особине које ће, уједно, помоћи и бољем разумевању самог појма.

Особина 1: *Очекивана вредност константе **c** једнака је тој константи **c**.*

Наиме, ако случајна величина *X* може узети само једну вредност, онда је њена расподела

X	*c*
p(x)	*1*

тако да је њена очекивана вредност једнака

$$E(x) = c \cdot 1 = c. \quad (5.33)$$

Особина 2: *Очекивање производа константе с и функције* $g(X)$ *једнако је производу те константе с и очекивања функције* $g(X)$*, тј.*

$$E\left[c \cdot g(X)\right] = c \cdot E\left[g(X)\right] \quad (5.34)$$

Заиста, ако X има закон вероватноћа $f(x)$, тада је

$$E\left[c \cdot g(X)\right] = \int_{-\infty}^{\infty} cg(X)\, f(x)\, dx = c \int_{-\infty}^{\infty} g(X)\, f(x)\, dx = c \cdot E\left[g(X)\right].$$

На сличан начин се може проверити особина 2 и за случајну величину X прекидног типа.

Особина 3: *Очекивање збира две функције случајне промењиве, једнако је збиру њихових очекивања.*

Наиме, ако су $g(X)$ и $h(X)$ функције, тада је

$$E\left[g(X) + h(X)\right] = E\left[g(X)\right] + E\left[h(X)\right] \quad (5.35)$$

Из ових особина следи следећа

Особина 4: *Очекивања линеарне функције случајне промењиве једнако је линеарној функцији очекивања те промењиве.*

Нека је Y промењива дата са

$$Y = aX + b\,,$$

тада је њена очекивана вредност једнака

$$E(Y) = aE(X) + b\,. \quad (5.36)$$

Особина 5: *Очекивање линеарне комбинације две случајне промењиве, једнако је линеарној комбинацији њихових очекивања.*

Наиме, ако је

$$Z = a_1 X + a_2 Y\,,$$

очекивање Z биће једнако

$$E(Z) = a_1 E(X) + a_2 E(Y). \tag{5.37}$$

Специјално, очекивање збира, или разлике две случајне променљиве једнако је збиру, односно, разлици њихових очекивања, тј.

$$E(X \pm Y) = E(X) \pm E(Y). \tag{5.38}$$

На пример, бележећи потражњу за својим производима, пекар је утврдио да она варира на следећи начин

Потражња *(комада)*	10000	11000	12000	13000	14000
Број дана	18	90	120	60	12

Користећи ***закон великих бројева*** можемо посматрати потражњу X као случајну променљиву са распоредом

X	10000	11000	12000	13000	14000
$P(x)$	18/300	90/300	120/300	60/300	12/300

Очекивана потражња једнака је

$$E(X) = 11.860$$

Нека су трошкови производње 0,15 дин/ком а продајна цена 0,20 динара. Непродати производи се други дан продају по цени 0,09 динара. Одредићемо очекивану зараду. Ако је Q произведена количина, X потражња а Y зарада, добиће се

$$Y = 0{,}05 \qquad \text{за } X \geq Q$$

$$Y = 0{,}05 - 0{,}06(Q - X) \qquad \text{за } X < Q$$

У првом случају је очекивана зарада

$$E(Y) = 0{,}05 \cdot Q \tag{5.39}$$

а у другом случају

$$E(Y) = 0{,}11E(X) - 0{,}06 \cdot Q = 1304{,}6 - 0{,}06 \cdot Q. \tag{5.40}$$

Користећи (5.39) и (5.40), може се за сваку произведену количину, одредити очекивана зарада, а на основу тога изабрати најбоља одлука о количини коју треба произоводити.

Посматрајмо случајну промењиву *X* са законом вероватноћа

$$\mathrm{f}_1(\mathrm{x}) = \begin{cases} 2\mathrm{x} & \text{за } 0 < \mathrm{x} < 1 \\ 0 & \text{за } \mathrm{x} \leq 0 \text{ и } \mathrm{x} \geq 1, \end{cases}$$

и случајну промењиву *Y* са законом вероватноћа

$$f_2(y) = \begin{cases} e^{-y} & \text{за } y > 0 \\ 0 & \text{за } y \leq 0. \end{cases}$$

Очекивана вредност функције

$$Z = 3X + 2Y \ ,$$

једнака је

$$E(Z) = 3E(X) + 2E(Y) = 3\int_0^1 2x^2 dx + 2\int_0^\infty xe^{-x} dx = 2 + 2 = 4.$$

Моменти случајне промењиве

Веома важну класу очекивања функција случајних промењивих, чине тзв. *моменти*. Практично, скоро цела анализа случајних промењивих базирана је на моментима и функцијама момената.

Разликују се две врсте момената: *обични и центални моменти*.

<u>Дефиниција 5.7.:</u> *Обични моменат k-тог реда случајне промењиве X је математичко очекивање k-тог степена те промењиве.*

Обичне моменте реда *k* означићемо са m_k, тако да је

$$m_k = E(X^k) \quad \text{за} \quad k = 0,\ 1,\ 2,\ \ldots \tag{5.41}$$

Ако је X прекидна случајна променљива, обични моменат реда k, једнак је збиру

$$m_k = \sum_{i=1}^{\infty} x_i^k \, p_i \, , \tag{5.42}$$

а ако је X непрекидног типа, обични моменат је једнак интегралу

$$m_k = \int_{-\infty}^{\infty} x^k f(x)\, dx \, . \tag{5.43}$$

Обичан моменат нултог реда једнак је јединици, јер је

$$m_0 = E(X^0) = E(1) = 1.$$

Обични моменат првог реда је, уствари, *очекивана вредност* или краће, *очекивање* случајне величине X. Означаваћемо га са

$$m = E(X) \, . \tag{5.44}$$

Дефиниција 5.8.: *Централни моменат k-tog реда случајне променљиве X, је очекивана вредност k-тог степена разлике X и очекиване вредности m.*

Централни моменат реда k означаваћемо са μ_k, тако да је

$$\mu_k = E(X - m)^k \qquad za\ k=0,\ 1,\ 2,\ ... \tag{5.45}$$

За случајну променљиву прекидног типа, централни моменти су једнаки збировима

$$\mu_k = \sum_{i=1}^{\infty} (x_i - m)^k \, p_i \, , \tag{5.46}$$

а за непрекидну случајну променљиву, централни моменти су једнаки интегралима

$$\mu_k = \int_{-\infty}^{\infty} (x-m)^k f(x)\, dx \ . \tag{5.47}$$

И централни моменат нултог реда је једнак један, јер је

$$\mu_0 = E(X-m)^0 = E(1) = 1,$$

а ***централни моменат првог реда једнак је нули***. Наиме, из особина очекивања, следи да је

$$\mu_1 = E(X-m) = E(X) - m = m\text{-}m = 0.$$

Централни моменти се лакше одређују преко обичних момената. Ако k-ti степен разлике $(X-m)$, изразимо преко биномног образца добићемо

$$(X-m)^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} X^{\mathrm{i}},$$

па је, због особина очекиване вредности

$$E(X-m)^k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} E\left(X^i\right),$$

односно

$$\mu_k = \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} m^{k-i} m_i \qquad \text{за } k=0,\ 1,\ 2,\ \ldots \tag{5.48}$$

Специјално

$$\begin{aligned} &\mu_2 = m_2 - m^2 \\ &\mu_3 = m_3 - 3m_2 m + 2m^3 \\ &\mu_4 = m_4 - 4m_3 m + 6m_2 m^2 - 3m^4. \end{aligned} \tag{5.49}$$

Обрнуто, користећи везе (5.48), можемо из централних момената одредити обичне моменте. Зато се, практично, из расподеле случајне променљиве X, одреди једна класа момената, а затим се из израза (5.48), одређује друга класа.

На пример, нека случајна променљива X има тзв. *хипергеометријску расподелу*

X	0	1	2	...	k
$p(x)$	p_0	p_1	p_2	...	p_k

при чему је за свако i=0,1,2,3....,k.

$$p_i = \frac{\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}} \tag{5.50}$$

Одредићемо њена прва два обична и централна момента.

Због особине закона вероватноћа, из (5.50) добија се

$$\sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}=\binom{m+n}{k}.$$

Први обичан моменат је

$$E(X)=\frac{\sum_{i=0}^{k} i\binom{m}{i}\binom{n}{k-i}}{\binom{m+n}{k}}=\frac{1}{\binom{m+n}{k}}\sum_{i=1}^{k} m\binom{m-1}{i-1}\binom{n}{k-i}=$$

$$=\frac{1}{\binom{m+n}{k}}\sum_{j=0}^{k-1} m\binom{m-1}{j}\binom{n}{k-1-j}=\frac{m}{\binom{m+n}{k}}\binom{m-1+n}{k-1}=\frac{mk}{m+n}.$$

Сличним операцијама може се одредити и други моменат, с тим што у збиру

$$\sum_{i=0}^{k} i^2 p_i \ ,$$

замењујемо

$$i^2 = i(i-1) + i,$$

а збир растављамо на два збира. Тако ће се добити

$$m_2 = E(X^2) = \frac{m(m-1)\ k(k-1)}{(m+n)(m+n-1)} + \frac{mk}{m+n}$$

Први централни моменат је једнак нули, а други централни моменат је

$$\mu_2 = m_2 - [E(X)]^2 = \frac{mnk(m+n-k)}{(m+n)^2(m+n-1)}.$$

На сличан начин би одредили трећи и четврти обични и централни моменат.

Посматрајмо пример непрекидне случајне променљиве са законом вероватноћа

$$f(x) = \begin{cases} 2(1-x) & \text{за } 0 < x < 1 \\ 0 & \text{за } 0 \leq x \leq 1. \end{cases}$$

Обични моменат k-тог реда, једнак је

$$m_k = E(X^k) = 2\int_0^1 x^k (1-x)dx = \frac{2}{(k+1)(k+2)}$$

Специјално

$$m_1 = \frac{2}{2\cdot 3} = \frac{1}{3}; \qquad m_2 = \frac{2}{3\cdot 4} = \frac{1}{6}$$

$$m_3 = \frac{2}{4\cdot 5} = \frac{1}{10}; \qquad m_4 = \frac{2}{5\cdot 6} = \frac{1}{15},$$

тако да је

$$\mu_1=0$$

$$\mu_2=\frac{1}{6}-\frac{1}{9}$$

$$\mu_3=\frac{1}{10}-3\frac{1}{6}\cdot\frac{1}{3}+2\cdot\left(\frac{1}{3}\right)^3$$

$$\mu_4=\frac{1}{10}-4\frac{1}{10}\cdot\frac{1}{3}+6\cdot\frac{1}{6}\left(\frac{1}{3}\right)^2-3\left(\frac{1}{3}\right)^4.$$

Параметри случајних промењљивих

Како смо испитивали статистичка обележја преко параметара које смо рачунали по одређеним правилима, тако ћемо и случајне промењљиве испитивати преко параметара добијених из њихових расподела. Биће то групе параметара које ће изражавати одређене особине расподела, а то су: мере централне тенденције, мере варијабилитета, мере асиметрије и мере спљоштености.

Мере централне тенденције

Прву групу параметара случајних промењљивих чине параметри који показују *централну тенденцију расподеле*. То су: средине (очекивана, хармонијска и геометријска средина), модус и медијана. Сви ови параметри су дефинисани за статистичка обележја. Овде ћемо само уопштити дате дефиниције.

Очекивана вредност, или краће ***средина***, је очекивана вредност случајне величине X, тј.

$$m = E(X). \tag{5.51}$$

Хармонијска средина је дата преко очекивања реципрочне вредности случајне промењљиве, изразом

$$\frac{1}{H} = E\left(\frac{1}{X}\right), \tag{5.52}$$

Геометријска средина је дефинисана преко очекиване вредности логаритма случајне пременљиве *X*, тако да је

$$\log G = E(\log X). \tag{5.53}$$

Модус случајне променљиве *X* је она вредност M_0 за коју расподела, односно закон вероватноћа има највећу вредност.

Медијана случајне променљиве *X* је она вредност (означимо је са M_e) за коју је

$$P\{X \leq M_e\} = P\{X > M_e\}.$$

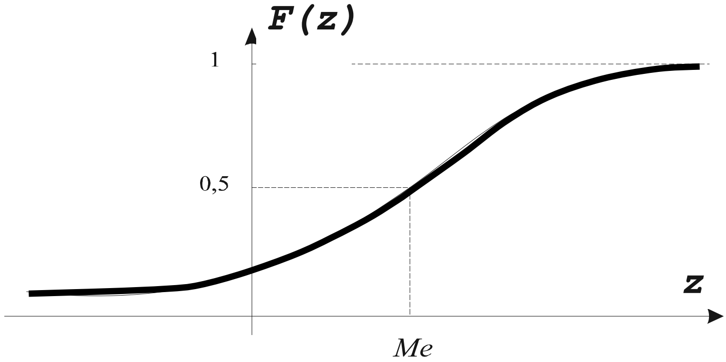
Вероватноћа да ће *X* бити мање од медијане, једнака је 0.5, а толика је и вероватноћа да ће *X* бити веће од медијане.

За случајну променљиву непрекидног типа, медијана је решење једначине дате интегралом

$$\int_{-\infty}^{M_e} f(x)\,dx = 0{,}5\,.$$

Ако је позната функција расподеле $F(z)$, онда је медијана она вредност за коју је $F(M_e)=0{,}5$

То је, у ствари, апсциса пресечне тачке праве $F(z) = 0.5$ и функције расподеле $F(z)$. (види слику 5.8).



Слика 5.8. Медијана и функција расподеле.

Мере варијабилитета

Друга група параметара случајних променљивих су параметри који показују варијабилитет. То су следећи параметри: средња девијација, варијанса, стандардна девијација и коефицијент варијације.

Средња девијација је очекивана вредност апсолутне вредности разлике X и његовог очекивања, тј.

$$e_m = E(|X-m|). \tag{5.54}$$

Ако је X прекидног типа, средња девијација је једнака збиру

$$e_m = \sum_{i=1}^{\infty} |x_i - m| p_i ,$$

а ако је непрекидног типа, средња девијација је једнака интегралу

$$e_m = \int_{-\infty}^{\infty} |x-m| \, f(x) \, dx.$$

Варијанса случајне променљиве *X је други централни моменат променљиве X.*

Означавамо је са σ^2. У ствари, варијанса је дефинисана изразом

$$\sigma^2 = \mu_2 = E(X\text{-}m)^2 = m_2\text{-}m^2 \ . \qquad (5.55)$$

Лако је проверити да варијанса има следеће особине:

Особина 1**:** Варијанса случајне променљиве X је једнака нули онда и само онда када је $X = const$.

Особина 2**:** Ако је случајна променљива Y линеарна функција

$$Y = aX + b \ ,$$

тада је варијанса Y једнака производу a^2 и варијансе X, тј.

$$\sigma^2(Y) = a^2\sigma^2(X) \ .$$

Особина 3: *Очекивана вредност квадрата разлике X и константе c, имаће минимум ако је та константа* $c = m$ *, а та минимална вредност је једнака варијанси случајне променљиве X, тј.*

$$\min_c \ E(X\text{-}c)^2 = E(X\text{-}m)^2 = \sigma^2 \ .$$

Стандардна девијација случајне променљиве X је позитивна вредност корена варијансе, тј.

$$\sigma = +\sqrt{\sigma^2}$$

Коефицијент варијације *случајне променљиве X је процентуално изражени количник стандардне девијације и средине случаја променљиве X, тј.*

$$V = \frac{\sigma}{m} \cdot 100\%$$

Коефицијенти асиметрије и спљоштености

Као ***мера асиметрије*** *расподеле, најчешће се користи тзв. први Пирсонов коефицијент дат количником*

$$L_1 = \frac{\mu_3}{\sigma^3}$$

где је μ_3*, трећи централни моменат случајне променљиве* X.

Ако је расподела симетрична, β_1 има вредност нула. Усвојићемо је обрнуто, ако је $\beta_1 = 0$ рећи ћемо да X има симетричну расподелу.

Други Пирсонов коефицијент је количник

$$L_2 = \frac{\mu_4}{\sigma^4}$$

где је μ_4, четврти централни моменат. Представља ***меру спљоштености***.

Код расподеле са „Нормалном спљоштеношћу", β_2 је једнако 3. У противном, β_2 је, или веће или мање од 3.

Чебишевљева теорема

Често се у пракси, одређује приближна вредност вероватноће да ће се случајна величина X наћи у некој околини свог очекивања. Чебишевљева теорема омогућава да се одреди та вероватноћа.

Теорема 5.2. Код сваке случајне променљиве X, која има коначну варијансу σ^2, за свако произвољно ε испуњена је следећа неједначина.

$$P\{|X - m| \geq \varepsilon\} \leq \frac{\sigma}{\varepsilon^2} \qquad (5.56)$$

Доказ. Нека је X непрекидног типа са законом вероватноће $f(x)$. Тада је, због ограничења функције

$$|X - m| \geq \varepsilon,$$

односно,

$$(X-m)^2 \geq \varepsilon^2,$$

$$\int_{-\infty}^{\infty} (x\text{-}m)^2 f(x)\, dx \qquad \int_{(x-m)^2 \geq \varepsilon^2} (x\text{-}m)^2 f(x)\, dx, \tag{5.57}$$

па је

$$\int_{-\infty}^{\infty} (x\text{-}m)^2 f(x)\, dx \geq \varepsilon^2 \int_{(x-m)^2 \geq \varepsilon^2} f(x)\, dx$$

Интеграл на левој страни (5.57) је варијанса случајне промењиве X, а на десној страни, интеграл представља вероватноћу дату у (5.56). Зато из (5.57) следи (5.56).

Из Чебишевљеве теореме добија се следећа неједнакост за вероватноћу да ће се X наћи у интервалу

$$(m\text{-}\alpha\sigma;\ m+\alpha\sigma],$$

при чему је α произвољно одабрани број,

$$P\{m-\alpha\sigma < X \leq m+\alpha+\} \geq 1\text{-}\frac{1}{\alpha^2} \tag{5.58}$$

Специјално

$$P\{m-2\sigma < X \leq m+2\alpha\} \geq 0.75$$
$$P\{m-3\sigma < X \leq m+3\sigma\} \geq 0.8889.$$

шо представља тзв. „**правило две (односно 3) сигме**”.

5.3. ЗАДАЦИ

1. Случајна променљива *X* има закон вероватноћа

$$f(x) = \begin{cases} \dfrac{1}{(x+1)^2} & \text{за } x > 0 \\ 0 & \text{за } x \le 0. \end{cases}$$

Нека су дати следећи догађаји:

$$A_1 = (-\infty < X < 0)$$
$$A_2 = (0 \le X \le \infty)$$
$$A_3 = (0 \le X \le 1)$$
$$A_4 = (-6 \le X \le 0)$$

Наћи вероватноће следећих случајних догађаја:

a) $\overline{A_1}$;	*d)* $A_3 \cup A_1$;	g) $(A_3 \cup A_4)(A_1 \cup A_2)$;
b) $\overline{A_1} \cup A_2$;	*e)* $A_1 \cup A_2$;	h) $\overline{A}_3$;
c) $A_2 \cup A_3$;	*f)* $A_1 \cup \overline{A_4}$;	i) $\overline{A_1 \cup A_3}$.

2. За случајну променљиву *X* чији закон вероватноћа је

$$f(x) = \begin{cases} e^{-x} & \text{за } x > 0 \\ 0 & \text{за } x \le 0, \end{cases}$$

одредити функцију расподеле. Нацртај *f(x)* и функцију расподеле, а затим одреди вероватноћу случајних догађаја датих у задатку 1.

3. Аутоматски струг исеца кружну лимену плочицу. Дијаметар исечене плочице је случајна променљива са законом вероватноћа

$$f(x) = \begin{cases} c(x - 0.24)^2 (x - 0.26)^2 & \text{за } 0.24 \le x \le 0.26 \\ 0 & \text{за } x < 0.24 \quad x > 0.26. \end{cases}$$

а) Одредити константу *c*.

б) Плочице се морају дорађивати ако је њихов дијаметар одступио од прописане вредности 0.25 за више од 0.008. Колики се може очекивати проценат плочица за дораду?

4. Баца се новчић све док не падне писмо. Одреди расподелу случајне променљиве која представља број бацања новчића.

5. Играч *A* има 2 динара, а играч *B* 1 динар. Један од играча баца новчић, а други погађа страну новчића која ће пасти у бацању. Уколико погоди, добија динар од другог играча. У противном, плаћа динар. Игра се завршава када један од играча добије сва 3 динара. Одреди расподелу случајне променљиве која представља број бацања новчића до завршетка игре.

6. Машина производи ексере са 1% дефектних у укупној производњи. Како изгледа расподела броја неисправних ексера у узорку од 60 комада.

7. Правилна коцка се баца све док се не појави 6. Одреди вероватноћу да ће се коцка бацати више од пет пута.

8. Закон вероватноћа случајне променљиве *X* је функција

$$f(x) = \begin{cases} cx^3 & \text{за } 0 < x < 1 \\ 0 & \text{за } x \le 0; x \ge 1 \end{cases}$$

а) Одредити константу *c*.

б) Одреди ону вредност *X* за коју ће вероватноћа да случајна променљива *X* буде већа од ње, бити једнака вероватноћи да ће *X* бити мање од те вредности.

в) Одреди број *A* тако да са вероватноћом 0.05 тврдимо да ће случајна променљива бити већа од *A*.

9. Случајна променљива X има тзв. ***Gama расподелу*** дату законом вероватноћа

$$f(x) = \begin{cases} \dfrac{\lambda^{\beta}}{\Gamma(\beta)} x^{\beta-1} e^{-\lambda x} & \text{за } x \geq 0 \\ 0 & \text{за } x < 0. \end{cases}$$

Испитај функцију $f(x)$ и провери да је то заиста функција која представља закон вероватноћа.

10. На основу модела из задатка бр. 9.

   а) за $\beta = 640$, $\lambda = 20$, провери да је

$$P(X = 30) = 0.057\,.$$

   б) Ако је $\lambda = 20$, а вероватноћа

$$P(X \leq 30) = 0.01\,,$$

   колико треба да је β *?*

11. Одреди параметре расподеле случајне величине X са законом вероватноћа

$$f(x) = ce^{-\frac{|x-a|}{b}}\,,\ b > 0.$$

12. Случајна променљива X може узети само две вредности: 1 и 2. Њена очекивана вредност је 3/2. Одредити закон верованотће за X.

13. Провери да случајна величина X, која има тзв. ***Кошијеву расподелу*** дату законом вероватноћа

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}\,,$$

нема очекивану вредност.

6. МОДЕЛИ РАСПОДЕЛА СЛУЧАЈНИХ ПРОМЕНЉИВИХ

Између могућих вредности случајне променљиве *X*, и њене одговарајуће расподеле постоји функционална веза која нам је најчешће непозната.

На основу природе појаве коју анализирамо, можемо претпоставити да расподела случајне променљиве *X* припада једној класи расподела, одређеној са једним или више непознатих параметара. Ако је та класа изражена функционалном везом

$$P(x;\theta_1,\theta_2,...\theta_k) \tag{6.1}$$

онда кажемо да функција (6.1) представља модел расподеле случајне променљиве *X*.

Модел расподеле (6.1) садржи све информације о посматраној појави које су изражене преко случајне променљиве *X*.

Постоје модели расподела који се најчешће користе и који сасвим добро описују посматрану случајну променљиву. Разликују се две класе модела: ***Прекидни*** *и* ***Непрекидни модели расподела.***

6.1. МОДЕЛИ ПРЕКИДНИХ РАСПОДЕЛА

"0-1" Расподела (Бернулијева расподела)

<u>Дефиниција 6.1:</u> *Ако случајна променљива X може узети једну од вредности 0 или 1 са вероватноћама* ***p*** *и* ***q*** *респективно, при чему је* $p+q=1$*, кажемо да X има "0-1" расподелу .*

Закон вероватноћа има облик:

X	**0**	**1**
p(x)	***p***	***q***

Параметар расподеле је реалан број p такав да је $0<p<1$.

Очекивана вредност и варијанса ове случајне променљиве су

$$m = E(X) = p$$

$$\sigma^2 = E(X-m)^2 = pq$$

Функција генератрисе има облик

$$g(t) = p + qe^t$$

одакле је лако утврдити моменте расподеле.

Биномна расподела

Дефиниција 6.2: *За случајну променљиву* X *кажемо да има Биномну расподелу, ако може узети једну вредност к из низа ненегативних целих бројева* $\{0, 1, 2, ..., n\}$

са вероватноћама

$$P_n(k) = P(X = k) = \binom{n}{k} p^k q^{n-k} \tag{6.2}$$

при чему је $0 < p < 1$, $p + q = 1$, *а* $\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$ *представља тзв. биномни коефицијент.*

Биномну расподелу означаваћемо са

$$X : \mathrm{B}\,(n; p).$$

Закон вероватноћа Биномне расподеле дат је следећом табелом:

X	0	1	2	...	n
p(x)	$\binom{n}{0} p^0 q^{n-0}$	$\binom{n}{1} pq^{n-1}$	$\binom{n}{2} p^2 q^{n-2}$	...	$\binom{n}{n} p^n q^0$

Бројеви n і p су параметри модела.

На основу биномног обрасца, лако је проверити да је збир биномних вероватноћа једнак јединици. Заиста,

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1\,.$$

Функција генератриса биномне расподеле биће једнака:

$$g(t) = \sum_{k=0}^{n} e^{kt} \binom{n}{k} p^k q^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(pe^t\right)^k q^{n-k} = \left(q + pe^t\right)^n \tag{6.3}$$

Извод функције $g(t)$ у тачки $t=0$, представља очекивану вредност:

$$E(X) = np\,. \tag{6.4}$$

Преко другог извода функције $g(t)$ у тачки $t=0$, добићемо други обичан моменат, па се лако добија вредност варијансе:

$$\sigma^2 = npq\,. \tag{6.5}$$

Може се проверити да је коефицијент асиметрије

$$L_1 = \frac{(q-p)^2}{npq} \tag{6.6}$$

а коефицијент спљоштености је

$$L_2 = 3 + \frac{1-6pq}{npq} \tag{6.7}$$

Из (6.6) і (6.7), види се да Биномна расподела тежи симетричној расподели кад $n \to \infty$. Исто тако, она тежи расподели са „нормалном" спљоштеношћу.

Специјално, када је $p = q = 0{,}5$, биномна расподела је симетрична расподела.

Биномна расподела се користи у тзв. ***Бернулијевој шеми*** независних понављања експеримента обрађеној у поглављу 4.2. Видели смо је одиграла изузетно важну улогу у теорији вероватноће, њеној примени, и развоју статистике. Вероватноће случајних догађаја, везаних за Бернулијеву шему експеримената, дате су биномном расподелом.

Претпоставимо да један експеримент можемо поновити неограничено пута. При томе се мисли да смо у стању неограничено пута остварити један комплекс услова. Сваки пут кад се оствари тај комплекс услова, неки случајни догађај A се може остварити или не остварити. Кад се експеримент понавља више пута, претпостављамо да резултат у једном понављању не утиче на резултате експеримента у наредним понављанима. Кажемо да су понављања експеримента међусобно независна. Вероватноће остваривања случајног догађаја A у више различитих понављања експеримента, добиће се множењем одговарајућих вероватноћа догађаја у појединим понављањима, јер су то независни догађаји.

Ево неколико примера таквих експеримената:

Пример 1: У статистичком скупу од ***N*** елемената $p\%$ њих има одређено својство, док престалих $(1-p)\%$ нема то својство. На случајан начин се из посматраог скупа бира један елеменат, а бележи се да ли он има посматрано својство или не. При томе се сваки извучени елеменат враћа у посматрани скуп, па се поново из целог скупа бира други елеменат. Интересује нас проблем одређивања вероватноћа да ће међу ***n*** извучених елемената бити тачно ***k*** са посматраним својством, такође и вероватноћа да ће се количник ***k/n*** разликовати од ***p*** за не више од једног унапред датог броја. Чему ће тежити количник ***k/n*** кад број понављања ***n*** тежи бесконачности?

Пример 2: Баца се неограичено пута правилан новчић и бележи се број падања „писма”. Интересује нас, чему ће тежити фреквенција појављивања „писма” кад неограничено пута бацимо новчић.

Пример 3: Једна машина аутоматски обрађује неки производ са вероватноћом погрешне обраде једнаком p. Ако се у великој серији од *10.000* комада или *100.000* комада, појави велики број неправилно обрађених комада, да ли то значи да се поузданост машине променила?

Пример 4: Зна се да је међу свим порођајима $p\%$ близанаца. Ако се у једној клиници на коју долазе породиље изабране на случајан начин, нађе 1000 породиља, колики се може очекивати број близанаца?

У свим овим примерима, посматра се један експеримент и случајан догађај A, који се може остварити са вероватноћом p, при чему је $0 < p < 1$, а са вероватноћом $q = 1 - p$, може да се не оствари. При томе је $p + q = 1$.

Пуасонова расподела

Пуасонова расподела добила је име по француском математичару Simeon D. Poisson-u (1781-1840). Она описује вероватноћу да ће случајан догађај наступити у одређеном временском или просторном интервалу под одређеним условима при чему је вероватноћа догађаја веома мала, а број покушаја - временски или просторни интервал - веома велики, тако да се догађај уствари реализује неколико пута. Пуасонова расподела описује, на пример, број грешака по страници књиге, или број жртава аутомобилских удеса месечно, у неком великом граду.

Дефиниција 6.3: *За случајну променљиву X кажемо да има Пуасонову расподелу ако може узети вредност k из низа ненегативних целих бројева* $\{0, 1, 2, 3, ...\}$ *и то са вероватноћом*

$$P(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

при чему је $\lambda > 0$*, реалан број и представља параметар расподеле.*

Пуасонову расподелу означићемо са $X : \mathcal{P}(\lambda)$.

Закон вероватноћа дат је табелом

X	0	1	2	...	k	...
p(x)	$\frac{\lambda^0}{0!} e^{-\lambda}$	$\frac{\lambda^1}{1!} e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$	...	$\frac{\lambda^k}{k!} e^{-\lambda}$	...

Зна се да је

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Зато је збир вероватноћа код Пуасонове расподеле једнак

$$\sum_{k=0}^{\infty} P(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1. \tag{6.8}$$

Очекивана вредност Пуасонове расподеле, биће једнака

$$m = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \cdot \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \,. \tag{6.9}$$

Други обичан моменат је једнак:

$$m_2 = \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \left[k(k-1) + k\right] \frac{\lambda^k}{k!} =$$

$$= e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda$$

Варијанса случајне променљиве X једнака је:

$$\sigma^2 = m_2 - m^2 = \lambda \,. \tag{6.10}$$

Кад се у изразу за трећи обичан моменат m_3 вредност k^3 замени вредношћу

$$k(k-1)(k-2) + 3k(k-1) + k \,,$$

и растави на одговарајуће збирове добиће се

$$m_3 = \lambda^3 + 3\lambda^2 + \lambda \tag{6.11}$$

На сличан начин се може утврдити да је четврти обични моменат једнак

$$m_4 = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda \,. \tag{6.12}$$

Из (6.11) и (6.12) могу се одредити централни моменти:

$$\mu_3 = \lambda$$

$$\mu_4 = 3\lambda^2 + \lambda$$

па је коефицијент асиметрије једнак:

$$\beta_1 = \frac{1}{\lambda},$$

а коефицијент спљоштености

$$\beta_2 = 3 + \frac{1}{\lambda}.$$

Модус Poisson-ове расподеле је вредност k за коју је

$$\begin{aligned} p_k &> p_{k-1} \\ p_k &> p_{k+1} \end{aligned} \tag{6.13}$$

Замењујући p_k, p_{k-1} и p_{k+1} у неједначине (6.13) њиховим вредностима

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}; \qquad p_{k-1} = \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda}; \qquad p_{k+1} = \frac{\lambda^{k+1}}{(k+1)!} e^{-\lambda}$$

добиће се неједначине

$$\lambda - 1 < k < \lambda \tag{6.14}$$

Ако λ није цео број, онда ће у интервалу $(\lambda - 1; \lambda)$ постојати једна вредност која ће задовољити (6.14) и то је модус Пуасонове расподеле

$$M_o = [\lambda].$$

А кад је λ цео број, модуси ће бити тачке $M_o = \lambda - 1$ и $M_o = \lambda$.

Важну групу експеримената, којима одговара модел Пуасонове расподеле, чине експерименти у којима се посматра број остварења неког догађаја у датом временском интервалу дужине Δt. Претпоставља се да су независни бројеви остварења догађаја у временским интервалима који се не преклапају, и да је вероватноћа да се у интервалу дужине Δt оствари тачно један догађај, једнака

$$\lambda \cdot \Delta t + 0(\Delta t),$$

при чему је λ константа, а $0(\Delta t)$ је вероватноћа да ће се у том интервалу догађај остварити два или више пута. За $0(\Delta t)$ претпоставља се да тежи нули.

Ако се цео интервал дужине t, подели на n једнаких делова, можемо те делове посматрати као понављања експеримента, при чему је вероватноћа реализације догађаја у сваком понављању, једнака

$$\frac{\lambda t}{n} + 0\left(\frac{1}{n}\right).$$

Кад $n \to \infty$, број остварења догађаја у интервалу t тежиће Пуасоновој расподели са параметром једнаким $\lambda\, t$.

Геометријска и Паскалова расподела

Нека су реализације експеримента или догађај A или догађај $\overline{A}$, и нека је у сваком понављању експеримента вероватноћа реализације догађаја A иста и једнака $P(A) = p$ $\left[P(\overline{A}) = 1 - p = q\right]$. Док се у Бернулијевој шеми тражила вероватноћа да се у n понављања експеримента (n-фиксиран број), догађај A појави x пута, овде се за случајну променљиву X узима број понављања експеримента до прве реализације догађаја A. Овако дефинисана случајна променљива X može да узме вредности 1, 2, 3, ... Вероватноћа да случајна променљива X узме вредност x дата је изразом:

$$p(x) = P(X = x) = q^{x-1} p \qquad \text{за } x = 1, 2, 3 \ldots \tag{6.15}$$

и представља вероватноћу да се у првих $x-1$ понављања експеримента не реализује догађај A, а да се реализује у x-том понављању.

Формулом (6.15) дефинисан је закон вероватноћа Геометријске расподеле. Назив геометријска расподела проистекао је из чињенице што вероватноће ове расподеле чине геометријски низ. Коришћењем обрасца за збир бесконачно опадајућег геометријског низа показаћемо да је збир свих вероватноћа једнак јединици:

$$\sum_{x=1}^{\infty} p(x) = p\sum_{x=1}^{\infty} q^{x-1} = p\frac{1}{1-q} = \frac{p}{p} = 1$$

Функција генератриса геометријске расподеле једнака је:

$$g(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} pq^{x-1} = \frac{pe^t}{1-qe^t}$$

Математичко очекивање и варијанса су:

$$m = \frac{1}{p}; \quad \sigma^2 = \frac{q}{p^2}$$

Важна особина геометријске расподеле састоји се у следећем: ако је познато да се у n понављања експеримента догађај A није реализовао, онда је вероватноћа да се реализује до окончања $n+k$ понављања екперимента иста као и вероватноћа реализације догађаја A у првих k понављања експеримента, јер вероватноћа реализације догађаја A у сваком понављању експеримента не зависи од резултата у претходним понављањима експеримента.

Пример: Вероватноћа да се до прве реализације догађаја A експеримент понови највише k пута износи:

$$p(X \le k) = p_1 + p_2 + \ldots + p_k = p\sum_{x=1}^{k} q^{x-1} = 1 - q^k$$

Овај резултат можемо добити и нешто другачије: ако се експеримент понавља више од k пута, то значи да се догађај A није реализовао у првих k понављања експеримента; овом случају одговара вероватноћа q^k, па је тада вероватноћа супротног догађаја једнака $1-q^k$.

Геометријска расподела може се уопштити на следећи начин: Нека се експеримент понавља док се догађај A не реализује k пута. Број понављања експеримента је случајна променљива X. Потражимо вероватноћу да се експеримент поновио x пута: $P(x) = P(X = x)$.

Догађај, који се састоји у томе да k-та реализација догађаја A произађе у x-том понављању експеримента, еквивалентан је догађају: У $x-1$ понављања експеримента догађај A појави се $k-1$ пута, а у x-том понављању експеримента догађај A се појави још једном. Како су ови догађаји независни, то се множењем вероватноћа $\binom{x-1}{k-1}p^{k-1}q^{x-k}$ и p добија закон вероватноћа случајне променљиве X:

$$p(x) = P(X = x) = \binom{x-1}{k-1}p^k q^{x-k}$$

$$x = k, k+1, ...$$

Расподела вероватноћа одређена овим законом вероватноћа позната је под називом *Паскалова расподела*. Користи се у економетријским моделима кад се једна експланаторна варијабла посматра као тзв. лагваријабла.

Посебан случај Паскалове расподеле, за $k = 1$, назива се *геометријска расподела.*

6.2. МОДЕЛИ НЕПРЕКИДНИХ РАСПОДЕЛА

Нормална расподела

Приликом решавања проблема граничне функције Биномне расподеле, Д Муавр (De Moivre) је 1733. године открио Нормалну расподелу, што је остало доста незапажено све до радова Гауса (C. F. Gauss-a) 1809. године и Лапласа (P. S. Laplace) 1812. године. Ова два аутора су развили теорију ***грешака опсервација*** и дошли до Нормалне расподеле. Лаплас је доказао тзв. ***централну граничну теорему*** и дао одговоре на различита питања из теорије вероватноће.

Под њиховим утицајем, дуго се сматрало да статистичке расподеле, практично, свих врста података приближавају се Нормалној расподели као идеалном облику само ако се обезбеди довољно велики број тачних опсервација. Девијације сваке случајне варијабле од њене очекиване вредности, посматране су као „грешке”, а на основу „закона грешки” описане су Нормалном расподелом. Показало се да је то тачно у многим физичким, астрономским, демографским и биолошким мерењима, а такође и у техничким и економским наукама.

Треба још напоменути да се велики део статистичког закључивања и доношења одлука, базира на моделима који су генерисани Нормалном расподелом.

Дефиниција 6.5: *За случајну променљиву X кажемо да има* ***Нормалну расподелу*** *ако је њен закон вероватноћа дат функцијом*

$$f(x)=\frac{1}{b\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2b^2}} \quad \text{за } x\in(-\infty,\infty) \tag{6.16}$$

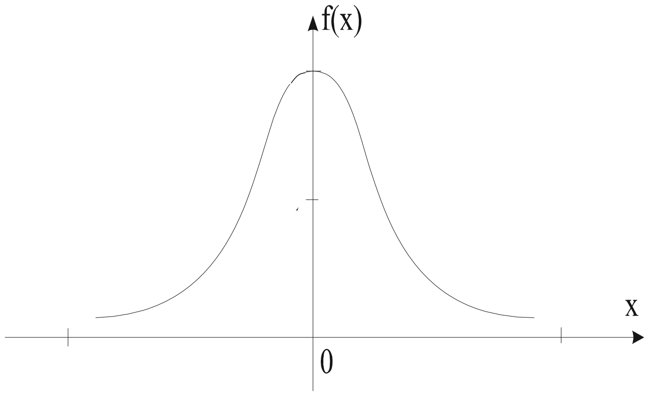
Вредности a и b су параметри модела, a је било који реалан број, а b је такође реалан позитиван број. Нормалну расподелу означаваћемо са

$$X : N(a;b).$$

Функција (6.16) је симетрична у односу на праву $x = a$. Има максимум у тачки $x = a$. Максимална вредност функције је једнака:

$$f(a) = \frac{1}{b\sqrt{2\pi}}.$$

Графички приказ функције (6.16) дат је на слици:



Слика 6.1. Нормална расподела

Очекивана вредност Нормалне расподеле добиће се преко интеграла:

$$m = E(X) = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-a)^2}{2b^2}} dx.$$

Уводећи смену $y = \dfrac{x-a}{b}$ и парцијалном интеграцијом, добиће се да је

$$m = a. \tag{6.17}$$

Варијансу случајне променљиве X са Нормалном расподелом добићемо преко интеграла

$$\sigma^2 = E(X-m)^2 = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-a)^2 e^{-\frac{(x-a)^2}{2b^2}} dx .$$

Уводећи смену $y = \frac{x-a}{b}$, и парцијалном интеграцијом, добиће се

$$\sigma^2 = b . \tag{6.18}$$

Из (6.17) и (6.18) види се да су *a* и *b*, уствари, очекивана вредност и варијанса случајне величине *X*. Зато се и најчешће Нормална расподела означава са

$$X : N\left(m;\sigma^2\right),$$

а њен закон вероватноћа дат је у облику:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

Специјално за $m=0$ *и* $\sigma^2=1$, *добићемо* ***Стандардизовану нормалну расподелу*** *са законом вероватноћа датим тзв. Гаусовом кривом*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

чије вредности су дате у табелама или у статистичким пакетима програма.

Вредности функције расподеле $F(x)$, за стандардизовану Нормалну расподелу, дате су у табелама и означене су са $\Phi(x)$.

За одређивање вероватноћа да ће се *X* наћи у интервалу $(x_1; x_2)$, користимо табеле, односно, функцију расподеле стандардизоване Нормалне расподеле. Наиме, из једначине

$$P\{x_1 < X \le x_2\} = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

после смене

$$y = \frac{x-m}{\sigma},$$

добиће се да је

$$P\{x_1 < X \le x_2\} = \frac{1}{\sqrt{2\pi}} \int_{\frac{x_1-m}{\sigma}}^{\frac{x_2-m}{\sigma}} e^{-\frac{x^2}{2}} dx = \Phi\left(\frac{x_2-m}{\sigma}\right) - \Phi\left(\frac{x_1-m}{\sigma}\right)$$

Функција генератриса Нормалне расподеле је функција

$$g(t) = e^{tm+\frac{t^2\sigma^2}{2}}$$

Налазећи обичне и централне моменте преко функције $g(t)$, лако је проверити да је

$$\beta_1 = 0, \quad \beta_2 = 3$$

из чега се закључује да је то симетрична расподела са нормалном спљоштеношћу.

Гама расподела

Гама функција *или* ***Ојлерова функција друге врсте*** *је функција облика:*

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$$

Гама функција поседује следеће особине:

1. $$\Gamma(p+1) = p\Gamma(p)$$

Ако је p природан број, $p = n$, онда је:

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \ldots = n!$$

при чему се за велике вредности n може користити Stirling-ова формула

$$\Gamma(n+1) = n! \approx \sqrt{2\pi\, n}\, n^n\, e^{-n},$$

2. $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{1}{\sqrt{t}} e^{-t} dt = \sqrt{\pi}$

Из претходних резултата следи:

$$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(\frac{1}{2}+1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$$

$$\Gamma\left(\frac{5}{2}\right) = \Gamma\left(\frac{3}{2}+1\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\cdot\frac{1}{2}\sqrt{\pi}$$

и уопште:

$$\Gamma\left(n+\frac{1}{2}\right) = \frac{2n-1}{2}\cdot\frac{2n-3}{2}\ldots\sqrt{\pi}$$

3. $\Gamma(p)\cdot\Gamma(1-p) = \frac{\pi}{\sin p\pi}, \quad 0 < p < 1$

4. За $n < 0$ Гама функција је једнака:

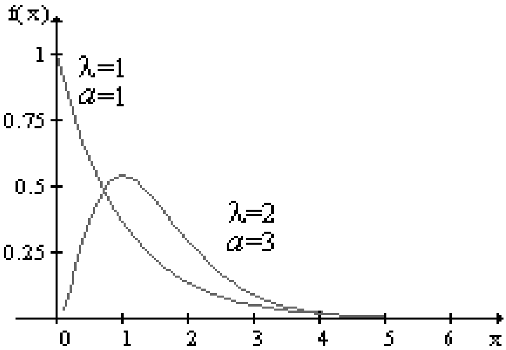
$$\Gamma(n) = \frac{\Gamma(n+1)}{n}.$$

<u>**Дефиниција 6.6**</u>**:** *Случајна промењива X има* ***Гама расподелу****, ако је њен закон вероватноће функција:*

$$f(x,a,\lambda) = \begin{cases} 0, & \text{за x} < 0 \\ \dfrac{a^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-ax}, & \text{за x} > 0 \end{cases}$$

где су параметри $\lambda > 0$ *и* $a > 0$.

За различите вредности параметара закон вероватноћа Гама расподеле има облик:



Слика 6.2. Гама расподела

Функција генератрисе Гама расподеле једнака је:

$$g(t) = E(e^{tx}) = \int_0^{\infty} \frac{a}{\Gamma(\lambda)} e^{tx} (a\,x)^{\lambda-1} e^{-a\,x} dx = \frac{a^{\lambda}}{\Gamma(\lambda)} \int_0^{\infty} x^{\lambda-1} e^{-(a-t)x} dx$$

замењујући $(a-t)x$ са y, за $t < a$, имамо:

$$g(t) = \frac{a^{\lambda}}{\Gamma(\lambda)(a-t)^{\lambda}} \int_0^{\infty} y^{\lambda-1} e^{-y} dy = \frac{a^{\lambda}\Gamma(\lambda)}{\Gamma(\lambda)(a-t)^{\lambda}} = \frac{1}{\left(1-\frac{t}{a}\right)^{\lambda}}$$

Основни параметри Гама расподеле једнаки су :

$$E(X) = g'(0) = \frac{\lambda}{a}$$

Ако је $\lambda > 1$, Гама расподела има модус једнак $M_o = \dfrac{\lambda - 1}{a}$.

$$\sigma^2 = E\left(X^2\right) - E^2\left(X\right) = g''(0) - g'^2(0) = \frac{\lambda}{a^2}$$

Теорема: *Ако независне случајне променљиве* $X_1, X_2, ..., X_n$ *имају Гама расподеле са параметрима* $\lambda_i, \left(i = 1,2,...,n\right)$ *и а, онда случајна променљива* $Y = \sum_{i=1}^{n} X_i$ *има Гама расподелу с параметрима* $\sum_{i=1}^{n} \lambda_i$ *и а.*

Доказ: Како је функција генератриса случајне променљиве Y једнака производу функција генератриса случајних променљивих $X_1, X_2, ..., X_n$ добија се:

$$g_y(t) = \prod_{i=1}^{n} \frac{1}{\left(1 - \frac{t}{a}\right)^{\lambda_i}} = \left(1 - \frac{t}{a}\right)^{-\sum_{i=1}^{n} \lambda_i} \tag{6.19}$$

Добијена функција јесте функција генератриса Гама расподеле. Одавде следи да случајна променљива $Y = \sum_{i=1}^{n} X_i$ има Гама расподелу са параметрима

$$\sum_{i=1}^{n} \lambda_i \ \text{ i } \ a \ .$$

Специјално,

Гама расподела с параметрима $\lambda = \dfrac{n}{2}$ *и* $a = \dfrac{1}{2}$ *биће функција:*

$$f\left(y,\frac{1}{2},\frac{n}{2}\right)=\begin{cases}0, & \text{за y} < 0\\ \dfrac{\frac{1}{2}}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}y\right)^{\frac{n}{2}-1}\cdot e^{-\frac{y}{2}}, & \text{за } 0 < \text{y} < \infty\end{cases}$$

*позната као **Хи-квадрат расподела**.*

Хи-квадрат расподела (χ^2)

Хи-квадрат расподела представља специјалан случај Гама расподеле. Дефинише се као расподела суме квадрата независних случајних променљивих које имају Нормалну расподелу, са средњом вредношћу 0 и стандрадном девијацијом 1. Ова расподела је важна у статистици јер описује расподелу вероватноћа које се користе у неколико најчешћих статистичких процедура, укључујући Пирсонов *Хи*-квадрат тест, а користи се и за испитивање валидности модела.

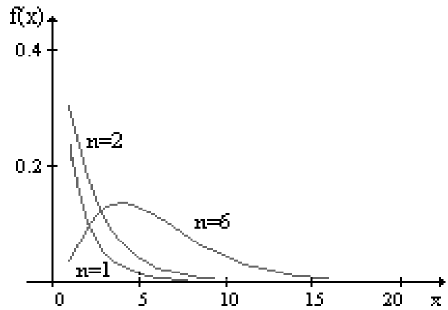
Дефиниција 6.7: *За случајну променљиву X кажемо да има Хи-квадрат расподелу са n степени слободе ако је њен закон вероватноћа дат функцијом*

$$k_n(x)=\begin{cases}c_n x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & \text{за x} > 0\\ 0, & \text{за x} \le 0\end{cases}$$

при чему је n природан број, а константа c_n је дата са

$$c_n=\frac{1}{\int_0^\infty x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx}=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$$

Закон вероватноће за *Хи*-квадрат расподелу приказан је на слици:



Слика 6.7. Hi - квадрат расподела

Лако је видети из Гама расподеле да је

$$f\left(x,\frac{1}{2},\frac{n}{2}\right)=\begin{cases}0, & \text{за x < 0}\\ \dfrac{\frac{1}{2}}{\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}x\right)^{\frac{n}{2}-1}\cdot e^{-\frac{x}{2}}, & \text{за } 0 < x < \infty.\end{cases}$$

Случајну променљиву са *Хи*-квадрат расподелом, означаваћемо са

$$X : \chi_n^2(x).$$

Функција расподеле ове случајне променљиве означава се са

$$K_n(z)=C_n\int_0^z x^{\frac{n}{2}-1}e^{-\frac{x}{2}}dx$$

Њене вредности су дате табелама за степене слободе од 1 до 30. За веће степене слободе могу се користити приближне вредности из Нормалне расподеле.

Очекивана вредност случајне променљиве са *Хи*-квадрат расподелом једнака је

$$E(x)=n,$$

а варијанса је

$$\sigma^2 = E(x-m)^2 = 2n .$$

Обичан моменат реда k дат је изразом

$$m_k = n(n+2)(n+4)\ldots(n+2k-2) ,$$

тако да је

$$\mu_3 = 8n$$
$$\mu_4 = 12n(n+4)$$

Коефицијент асиметрије има вредност

$$\beta_1 = \frac{8}{n},$$

а коефицијент спљоштености је

$$\beta_2 = 3 + \frac{12}{n} .$$

Из формуле за коефицијент асиметрије се види да *Хи*-квадрат расподела тежи симетричној расподели кад $n \to \infty$, а из формуле за коефицијент спљоштености се види да она тежи расподели са нормалном спљоштеношћу.

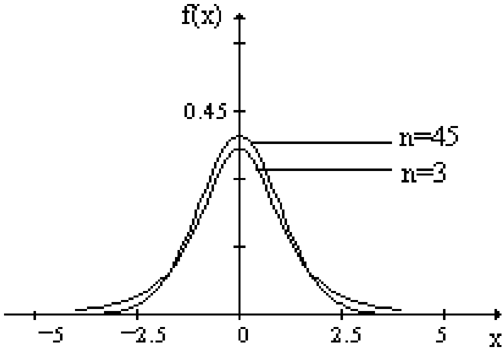
Студентова расподела

Студентову расподелу дефинисао је Госет (W. S. Gosset) у раду објављеном под псеудонимом „Студент”.

<u>Дефиниција 6.8</u>: *За случајну променљиву X кажемо да има* ***Студентову расподелу са n степени слободе*** *ако је њен закон вероватноћа дат функцијом*

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi \cdot n} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty < x < \infty$$

На слици је приказана функција $f_n(x)$, за $n=3$ и $n=45$:



Слика 6.3. Студентова расподела

Закон вероватноћа Студентове расподеле дат је симетричном функцијом са максималном вредношћу за $x=0$. Очекивана вредност је

$$m = E(x) = 0\,.$$

Обични и централни моменти су међусобно једнаки. Варијанса је једнака:

$$\sigma^2 = \frac{n}{n-2},$$

коефицијент асиметрије $\beta_1 = 0$, а коефицијент спљоштености је једнак

$$\beta_2 = 3 + \frac{6}{n-4}\,.$$

Кад $n \to \infty$, Студентова расподела тежи стандардизованој нормалној расподели.

Униформна расподела

<u>Дефиниција 6.9</u>: *Случајна променљива X има* ***Униформну расподелу*** *ако је њен закон вероватноћа дат функцијом:*

$$f(x)=\begin{cases}\dfrac{1}{\beta-\alpha}, & \alpha<x<\beta\\ 0, & \text{за остале } x\end{cases}$$

Вероватноћа да ће се X налазити у интервалу дужине h једнака је

$$\frac{1}{(\beta-\alpha)}h\ .$$

Теорема: *Било коју расподелу непрекидне променљиве X можемо трансформисати у униформну расподелу дефинишући случајну променљиву*

$$Y=\mathrm{F}(X)$$

где је $\mathrm{F}(X)$ *функција расподеле случајне променљиве X, тј.*

$$\mathrm{F}(x)=P\{X\le x\}.$$

Карактеристике ове расподеле су следеће:

Математичко очекивање:

$$E(X)=\frac{a+b}{2}.$$

Варијанса:

$$\sigma^2=\frac{(b-a)^2}{12}\ .$$

Центар симетрије:

$$\frac{a+b}{2}.$$

7. УВОД У СТАТИСТИЧКО ЗАКЉУЧИВАЊЕ

Наша знања, наши ставови и наше активности су у великој мери базирани на **узорцима**. Нешто посматрамо или **«меримо»** делимично (у временским тренуцима, у појединим деловима простора, у одређеним облицима биолошких, друштвених или техничких процеса и сл.) и на основу тога доносимо одлуке или усмеравамо своје активности. Овакав начин рада је заступљен како у свакодневном животу, тако и у научним истраживањима. Наиме, своје активности заснивамо на опсервацијама или мерењима која никад или скоро никад, нису свеобухватна. Идемо тако далеко да на пример, појединц доноси мишљење о некој институцији која се свакодневно бави хиљадама трансакција, на основу једног или два контакта које је имао са том институцијом у току неколико дана, месеци или година.

Путник који проведе 10 дана у некој земљи доноси суд, чак почиње и писати о становништву, индустрији, политичком систему, храни у хотелима и сл. Са друге стране то исто уради неко ко је провео 20 година живећи и проучавајући земљу. Наравно, њихови описи су сасвим различити. Обоје су донели своје закључке на основу **узорака**. Први има много **мањи узорак** па су његови закључци **мање веродостојни**. Код другог је **узорак знатно већи** и његовим закључцима ћемо сигурно више веровати. Другим речима, његови закључци ће бити **веродостојнији.**

Очигледно је да ће ваљаност закључака и предузетих активност зависити од тога како бирамо узорак и како одређујемо његову величину.

До краја двадесетих година овог века мала је пажња била посвећена проблемима како добити ***добре узорке*** и како донети ***ваљане закључке, односно научно валидне закључке.*** Нисмо увек у ситуацији да је посматрана појава на којој вршимо «мерење» (популација из које узимамо узорак), униформна тако да свака врста узорка даје скоро исте резултате. На пример, лабораторијска дијагноза базира се на неколико капи крви. Поступак је базиран на претпоставци да циркулација крви увек добро помеша крв па једна кап говори о крви исто као и друга. Међутим кад је узорак из популације која није тако униформна, као што је најчешћи случај, метод добијања узорка постаје критичан. Проучавање метода и техника које доводе до узорака ***на основу којих можемо веровати да ће закључци бити истинити или су близу истине***, постаје изузетно важно.

7.1. МЕТОДЕ ЛОГИЧКОГ ПРОЦЕСА ЗАКЉУЧИВАЊА

Основне методе логичког процеса закључивања су индуктивна и дедуктивна метода. Ако логички процес доводи до генерализације, а на основу посебних случајева (пропозиција), онда је то **индукција.** Обрнути процес је **дедукција**, а то је логички процес који доводи до специјалних закључака, а на основу општих пропозиција.

Дедуктивни метод закључивања полази од главне и споредне премисе (претпоставке) из којих следи специјалан закључак.

То је метод оваквог облика:

<u>Главна премиса</u>: Сви људи су смртни.

<u>Споредна премиса</u>: Сократ је човек.

<u>Закључак</u>: Сократ је смртан.

или, на пример:

<u>Главна премиса</u>: У складишту су сви производи прве класе.

<u>Споредна премиса</u>: Изабрани производи су из посматраног складишта.

<u>Закључак</u>: Изабрани производи су прве класе.

Валидност дедуктивног начина закључивања заснована је на следећем:

- Постоји потпуна унутрашња конзистентност свих околности садржаних у премисама;

- Закључци следе из премиса, а последице су универзалних пропозиција (истина) садржних у премисама;
- У ланцу закључивања ради се у потпуно затвореном систему. Све релевантне околности су изван нас и садржане су у премисама.

Индуктивни метод закључивања полази од појединачних случајева или премиса, а заклучак се односи на оно што није садржано у премисама или што је изван премиса.

Оно има овакав облик:

Прва премиса:	Из складишта смо узели пет производа;
Друга премиса:	Узети производи су прве класе;
Закључак:	У складишту су сви производи прве класе.

Или, на пример:

Прва премиса:	Сократ, Демокрит и Есхил су људи;
Друга премиса:	Сократ, Демокрит и Есхил су смртни;
Закључак:	Сви људи су смртни.

Разлике код ова два метода закључивања су очигледне. Наиме, закључци из дедуктивних аргументација садржани су већ у премисама. Ако су истините премисе, онда је истинит и закључак. Према томе, ништа ново се не додаје изван оног што је у премисама. Валидност дедуктивног закључивања је у томе, што ланци закључивања имају велику вредност у спознаји истина садржаних у премисама.

Међутим, закључци добијени индуктивним поступком шири су од премиса. Уколико су истинити они ће значити нешто ново у укупном људском сазнању. Али, постоји и једна цена која се мора платити за евентуалну екстензију људског знања. Наиме, могуће је да закључци индуктивног метода не буду истинити. Не постоји потпуна сигурност у тачност оваквих закључака. Они могу бити непотпуни, или чак и погрешни.

Основне карактеристике индуктивног метода закључивања можемо сумирати на следећи начин:

1. Индуктивни закључци морају бити дати преко **пробабилистичких израза**, тј. преко израза који укључују вероватноћу. Они се односе на случајеве изван посматрања.
2. Постоје одређене чињенице које нису садржане у премисама, али су оне релевантне за наш закључак. Тако на пример, метод опсервација (мерења) оног што је у премисама веома је важан за истинитост закључка.
3. Сам метод укључује претпоставку о егзистенцији извесне униформности у систему чињеница садржних у премисама и закључку. Та претпоставка је, у извесном смислу, премиса коју не спомињемо.
4. Верификација индуктивних закључака захтева објективне анализе засноване на посматрањима и мерењима конзистентним са чиниоцима који одређују природу света у коме живимо.

Статистичко закључивање

У статистичком закључивању користи се и индуктивни и дедуктивни метод. Комбинацијом ова два метода, у статистици сазнајемо релевантне истине садржане у премисама које се односе на реалан свет, а затим долазимо до закључака о реалном свету који нису садржани у премисама.

Основу статистике чини **теорија вероватноће** која је базирана на планираним експериментима и могућим резултатима. Користећи систем математичких пропозиција, формулисани су математички модели којима су описане законитости у посматраним експериментима. Зато је први проблем статистике да се утврди исправност математичких модела за опис посматраног експеримента или појаве.

Из теорије вероватноће, дедуктивном методом закључивања формирају се модели статистичких правилности, односно, статистичких законитости везаних за резултате из низа понављања експеримената. Ове статистичке законитости низа понављања експеримента треба да нам омогуће да се да одговор на основни проблем: Да ли су математички модели адекватни за научно валидне описе појаве, експеримента, или процеса, или не?

На пример, када бацимо новчић, математички модел је постојање једнаке вероватноће падања писма, односно грба. Међутим, ако заиста узмемо један новчић и бацамо га више пута, шта ће се десити? Ако је новчић правилног облика (што је укључено у модел), може се очекивати подједнак број падања писма и грба. Из математичког модела произилази то „шта се може очекивати”, ако је новчић правилног облика. Бацањем новчића утврдићемо да ли је произашло то „што смо очекивали” или не. Ако није произашло то што смо очекивали значи да модел није адекватан, тј. не одговара овом експерименту тј. овом новчићу кога смо бацали. Ако је резултат био у складу са „очекивањем”, значи да модел може бити потпуно одговарајући или исправан.

Према томе, дедуктивним методама у статистици се утврђују последице произашле из модела, а које се односе на резултате понављања експеримента везаног за тај модел. Кад се прикупе резултати понављања експеримента, онда се индуктивним закључивањем утврди да ли је модел добар или не.

Други статистички проблем везан је за анализу података који се односе на реалан свет. Наиме, у проучавањима природних и друштвених појава користе се резултати мерења исказани у статистичким подацима. Те резултате је потребно научно оценити и донети одређени закључак о посматраним појавама. Методе статистичког закључивања, индуктивног и дедуктивног, омогућавају научну обраду статистичких података и доношење закључака на бази таквих података.

На пример, произвођач сијалица жели да утврди век трајања својих производа. Измериће век трајања, рецимо, код сто сијалица и добиће резултат. На основу ових сто мерења одредиће просечни век трајања, просечан варијабилитет и слично. Али, његов основни интерес је да утврди век свих сијалица које је произвео. Какав закључак може донети о целокупној производњи на основу ових сто мерења?

Исто тако, при употреби новог лека у лечењу неке болести, посматра се група болесника који користе стари лек и група болесника који користе нови лек. Проценат излечења у обе групе неће бити исти. Да ли су разлике у процентима излечења значајне? Да ли, заиста лекови имају различит ефекат, или су разлике последица избора болесника?

Затим, у једној фабрици се мери просечна температура на радном месту и број грешака радника. На основу десет мерења одређен је **коефицијент корелације** 0.60. Да ли се може закључити да постоји зависност између броја грешака и просечне температуре на радном месту?

Индуктивни и дедуктивни методи закључивања омогућавају да се дају одговори на постављена и слична питања, тј. да се на основу измерених резултата, донесе закључак и о ономе што није измерено.

Проблеми статистичког закључивања могу се поделити у две групе:

1. **Проблеми оцењивања,**
2. **Проблеми тестирања хипотезе**.

Основу решавања ових проблема чине теорија вероватноће и математички модели везани за **популацију, узорак** и расподеле случајних промењљивих и **статистика.**

Зато ћемо у овом поглављу обрадити те основне појмове, а затим ћемо у следећа два поглавља изложити моделе и примену **теорије оцењивања** и **теоерију тестирања хипотеза**.

7.2. ПОПУЛАЦИЈА И УЗОРАК

У статистичким испитивањима срећемо се са следећим проблемима:

а) Посматрамо статистички скуп од N елемената, и на њима извесно обележје X дато расподелом

X	$a_0 - a_1$ x_1	$a_1 - a_2$ x_2	...
p	$p_1 = \frac{f_1}{N}$	$p_2 = \frac{f_2}{N}$	...

Табела 7.1. Расподела статистичког обележја

при чему су $x_1, x_2, \ldots$ могуће вредности обележја X, ако је то прекидно обележје, или средине интервала $a_0 - a_1$, $a_1 - a_2$, ..., ако је то непрекидно обележје.

Особине статистичког скупа одређујемо на основу вредности *параметара* тог обележја, а саме параметре рачунамо преко расподеле обележја X дате Табелом 7.1.

Међутим, у пракси нам је расподела обележја X непозната, па самим тим су нам непознати и параметри статистичког скупа. Наиме, да би одредили расподелу обележја X, морали би на сваком елементу скупа измерити његову вредност. Код већих статистичких скупова то је, практично, немогуће. Једино што можемо урадити, то је следеће: на одређени начин изабрати један мањи подскуп из статистичког скупа и на елементима тог подскупа измерити обележје *X*. На основу ових вредности донећемо одговарајуће закључке о целом статистичком скупу.

У оваквом случају, статистички скуп се зове ***популација***, а изабрани подскуп, на чијим елементима ћемо измерити вредности обележја *X*, зваћемо ***узорак.***

На пример, ако треба утврдити просечну висину становника једне државе, требало би измерити висину сваког становника те државе, а то је практично немогуће. Зато ћемо, изабрати мањи број становника, па на основу њихових висина донети суд о просечној висини свих становника те државе. Становници би представљали *популацију,* а изабрана група би била *узорак*.

Популацију могу да чине, на пример, ТВ гледаоци, студенти, локалне цене неких или свих производа на пијацама, квалитет или број производа у дневној, недељној или годишњој производњи и сл.

б) Посматраћемо експеримент чији резултат је случајна променљива X са датим законом вероватноћа. Да би проверили да ли тај закон вероватноће, заиста, одговара овом експерименту, поновићемо експеримент одређени број пута. На основу резултата у овим понављањима донећемо закључак о расподели случајне величине X.

И у овом случају, расподелу случајне променљиве *X* зваћемо ***популација***, а резултате понављања експеримента, зваћемо ***узорак***.

в) Сваки скуп података којим располажемо, можемо сматрати као подскуп неког ширег (коначног или бесконачног) скупа. Тако се подаци о Сунчевој галаксији могу сматрати као подскуп података о целој васиони.

Подаци о природним и друштвеним кретањима на Земљи у једном периоду, могу се сматрати као део податка о тим кретањима у дужим периодима, и сл.

У сваком случају, имамо следећи проблем: На основу расположивих података о некој појави, потребно је да донесемо закључке о законитостима посматране појаве и у оним областима за које немамо података.

Скуп расположивих података зове се ***узорак***, а појава, за коју доносимо закључак, зове се ***популација***.

г) На крају, сваку природну или друштвену појаву испитујемо тако што формулишемо одређени **модел** којим објашњавамо ту појаву. Наравно поставља се проблем исправности модела. Решићемо га тако што ћемо извршити одређен број мерења, па на основу резултата тих мерења утврдити ваљаност модела.

И у овом случају, сам модел зваћемо ***популација***, а резултате мерења ***узорак***.

У свим овим случајевима потребно је, на основу података из ***узорка, донети одговарајуће закључке о популацији***, тј. потребно је на основу „измереног”, донети суд о оном што нисмо „измерили”.

Наравно, методе закључивања ће зависити од начина добијања података, тј. начина „узимања” узорка. На основу овог критеријума узорке делимо на следеће врсте:

1. ***Прост случајни узорак;***
2. ***Систематски узорак;***
3. ***Стратификовани узорак;***
4. ***Узорак скупина.***

Код ***простог случајног узорка, избор било ког елемента популације не зависи од избора других елемената***, односно вероватноћа избора једног елемента у узорку не зависи од вероватноћа избора других елемената узорка. Ова врста узорка има једноставна теоријска својства и представља основу за изучавање и примену свих других врста узорка у проблемима статистичког закључивања.

При избору елемента у **систематском узорку**, на случајан начин се одабере први елеменат, а затим се из популације узима сваки к-ти елеменат (сваки пети, десети и сл.).

Ако је **популација подељена на стратуме** као хомогеније целине, избор можемо вршити за сваки стратум посебно, тако што се из стратума узима одговарајући број јединица.

Код узорка скупина, елементи популације су подељени у скупине, па се врши избор скупина, а затим се посматрају сви елементи скупине као елементи узорка **(једнофазни узорак)** или се, у оквиру изабране скупине врши избор елемената који ће ући у узорак **(двофазни узорак)** итд..

Комбинацијама ове четири врсте узорака, добијамо и друге врсте узорака. Основно је да се код оваквих врста узорака постиже тзв. ***репрезентативност узорка, која представља основну подлогу ваљаности донетих закључака на основу узорака.***

Иначе, теорија која проучава разне врсте узорака, и методологију избора узорка, зове се ***планирање узорака*** или ***планови узорковања***. Овде се нећемо бавити овом теоријом, јер ћемо проучавати само *просте случајне узорке* и теорију доношења одлука базирану на таквој врсти узорака.

Прост случајан узорак

Посматрајмо један експеримент чији је резултат случајна променљива X. Она ће имати неку расподелу одређену *законом вероватноћа* или *функцијом расподеле*. Понављаћемо експеримент више пута и посматрати резултате тих понављања.

Прво ћемо претпоставити да вршимо *независно* понављање експеримента.

Означићемо са X_1 резултат првог понављања експеримента. Наравно, X_1 ће бити случајна променљива, имаће исту расподелу као и променљива X.

Са X_2 ћемо означити резултат другог понављања експеримента. И X_2 ће бити случајна променљива са истом расподелом као и X. Поред тога, случајне променљиве X_1 и X_2 су независне (због независних понављања експеримената), итд.

На крају, са X_n означићемо резултат n-тог понављања експеримената. То ће, такође, бити случајна променљива са расподелом као и променљива X.

Према томе, резултат n независних понављања експеримената биће скуп случајних променљивих $X_1, X_2, ..., X_n$, које су међусобно независне и све имају исту расподелу као и случајна променљива X.

Случајна променљива X и њена расподела представљају расподелу ***популације***, а $X_1, X_2, ..., X_n$ представља ***прост случајан узорак*** из посматране популације.

Дакле, ***прост случајан узорак величине n, представља скуп од n независних случајних променљивих, које имају исту расподелу и то је расподела популације.***

Кад посматрамо статистички скуп и расподелу обележја X као расподелу популације, онда избор елемената у узорку вршимо тако да на случајан начин одабирамо један по један елемент, и утврђујемо вредности обележја X код изабраних елемената. Претпоставићемо да избор елемената вршимо са понављањем, тако да се сваки изабрани елемент из популације враћа поново у популацију, и врши избор следећег елемента.

Означићемо са $X_1, X_2, ..., X_n$ вредности обележја X на изабраним елементима у узорку. Променљиве $X_1, X_2, ..., X_n$ ће имати исту расподелу као и обележје X на посматраној популацији и чиниће *прост случајан узорак*.

Код изабраних елемената узорка вредности обележја X (или случајне променљиве X) биће конкретни бројеви, тј. биће

$$X_1 = x_1,\ X_2 = x_2, ..., X_n = x_n,$$

и то ће представљати једну реализацију случајних променљивих $X_1, X_2, ..., X_n$. Вредности $x_1, x_2, \ldots, x_n$ биће резултати једног узорка из посматране популације.

Прост случајан узорак се може посматрати као *n*-димензионална случајна променљива ($X_1, X_2, ..., X_n$), а бројеви ($x_1, x_2, \ldots, x_n$) су једна реализација вредност случајног узорка.

Да би искористили резултате узорака за (статистичко) закључивање о особинама популације из које је узорак узет, потребно је проучити основне особине узорка. Зато ћемо, у првом кораку, морати увести неке нове појмове, а затим разрадити одговарајућу теорију везану за прост случајан узорак.

Веродостојност узорка

Пошто прост случајан узорак представља *n*-димензионалну случајну променљиву, онда ће расподела тог узорка бити одређена законом вероватноћа *n*-димензионалне случајне променљиве.

Ако је случајна променљива *X* прекидног типа са законом вероватноћа

X	x_1'	x_2'	...
p(x)	$p(x_1')$	$p(x_2')$	...

онда је закон вероватноћа простог случајног узорка ($X_1, X_2, ..., X_n$), дат функцијом

$$f(x_1, x_2, ..., x_n) = P\{X_1 = x_1\} \cdot \ldots \cdot P\{X_n = x_n\} = p(x_1)p(x_2)...p(x_n)$$

односно

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i) \tag{7.1}$$

При томе је свака вредност у узорку нека вредност из скупа могућих вредности $\left(x_1^{'}, x_2^{'}, ...\right)$ случајне промењљиве *X*. Функција $f\left(x_1, x_2, ..., x_n\right)$ представља ***веродостојност узорка***.

Кад је случајна промењлива *X* непрекидног типа са законом вероватноћа $f\left(x\right)$, онда функција

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

представља закон вероватноће узорка $(X_1, X_2, ..., X_n)$ и то је ***веродостојност узорка***.

Види се да је веродостојност узорка $f\left(x_1, x_2, ..., x_n\right)$ одређена расподелом популације, односно њеним законом вероватноћа. На пример, ако *X* представља **„0-1"** ***расподелу*** са законом вероватноћа,

X	1	0
p(x)	p	1-p

а добијене вредности у узорку су *(1, 1, 0, 1, 0, 1, 1)*, онда је веродостојност овог узорка

$$f(1,1,0,1,0,1,1) = pp(1-p)p(1-p)pp = p^5(1-p)^2.$$

Ово ће бити вероватноћа да ће се добити узорак са вредностима

$$\left(X_1 = 1,\ X_2 = 1,\ X_3 = 0,\ X_4 = 1,\ X_5 = 0,\ X_6 = 1,\ X_7 = 1\right).$$

Ако *X* има експоненцијалну расподелу са законом вероватноћа

$$f(x) = \begin{cases} ae^{-ax} & \text{за } x > 0 \\ 0 & \text{за } x \le 0, \end{cases}$$

онда је веродостојност узорка (X_1, X_2, ..., X_n) дата функцијом

$$f(x_1, x_2, ..., x_n) = a^n e^{-a(x_1 + ... + x_n)}.$$

7.3. СТАТИСТИКЕ И ЊИХОВЕ РАСПОДЕЛЕ

Особине узорка проучавамо преко функција заданих на узорку. Пошто је узорак ($X_1, X_2, ..., X_n$) случајна променљива, онда било каква функција тог узорка, такође, ће бити случајна променљива. Таква функција, уствари, представља известан параметар узорка и зове се ***статистика.***

Функција

$$Z = f\left(X_1, X_2, ..., X_n\right),$$

представља ***статистику****, а за конкретан узорак са вредностима* $\left(x_1, x_2, ..., x_n\right)$*, статистика Z има одређену вредност*

$$z = f\left(x_1, x_2, \ldots, x_n\right)$$

и представљаће једну реализацију те статистике. За неке друге вредности узорка $\left(x_1^{'}, x_2^{'}, ..., x_n^{'}\right)$*, статистика Z ће имати другу вредност*

$$z^{'} = f\left(x_1^{'}, x_2^{'}, \ldots, x_n^{'}\right), \qquad итд.$$

Другим речима, са променом вредности у узорку мењаће се и вредности статистике ***Z***. Вероватноћа појединих вредности узорака одређује се из веродостојности узорка, па према томе и вероватноћа појединих вредности статистике биће одређене преко веродостојности узорка. Дакле, расподелу, односно, закон вероватноћа статистике ***Z***, дефинисане на узорку ($X_1, X_2, ..., X_n$), одређиваћемо из веродостојности узорка. Кад је случајна променљива X на популацији непрекидног типа из

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i),$$

односно из функције

$$f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i)$$

кад је на популацији прекидног типа.

Да би се одредила расподела, односно, закон вероватноћа једне статистике, потребно је утврдити све њене могуће вредности, (за све могуће узорке из посматране популације), а затим вероватноће свих тих вредности. Да то није тако једноставно, показаћемо на најједноставнијем примеру.

Претпоставимо да једна машина „обрађује" неки производ, и да су 3/4 производа добри, а 1/4 лоши. Ако са X означимо случајну променљиву која означава квалитет производа и има вредност 1 кад је производ добар, а нулу кад је производ лош, тада X на популацију има закон вероватноћа

X	0	1
p	$\frac{1}{4}$	$\frac{3}{4}$

Посматраћемо све узорке величина 3. То су узорци (0,0,0), (0,0,1), (0,1,0), (1,0,1), (0,1,1), (1,0,1), (1,1,0) i (1,1,1).

Посматраћемо следеће статистике:

Z_1 : Број лоших комада; Z_2 : Просечан број добрих комада у узорку;

Z_3 : Распон узорка; Z_4 : Варијансу узорка, итд.

За посматране узорке одредићемо веродостојности узорака, а затим вредности појединих статистика. Резултати су у табели 7.2.

		Статистике			
Узорак	Веродостојност узорка	Z_1	Z_2	Z_3	Z_4
(0, 0, 0)	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{64}$	3	0	0	0
(0, 0, 1)	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}=\frac{3}{64}$	2	$\frac{1}{3}$	1	$\frac{2}{9}$
(0, 1, 0)	$\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}=\frac{3}{64}$	2	$\frac{1}{3}$	1	$\frac{2}{9}$
(1, 0, 0)	$\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{1}{4}=\frac{3}{64}$	2	$\frac{1}{3}$	1	$\frac{2}{9}$
(0, 1, 1)	$\frac{1}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}=\frac{9}{64}$	1	$\frac{2}{3}$	1	$\frac{2}{9}$
(1, 0, 1)	$\frac{3}{4}\cdot\frac{1}{4}\cdot\frac{3}{4}=\frac{9}{64}$	1	$\frac{2}{3}$	1	$\frac{2}{9}$
(1, 1, 0)	$\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{1}{4}=\frac{9}{64}$	1	$\frac{2}{3}$	1	$\frac{2}{9}$
(1, 1, 1)	$\frac{3}{4}\cdot\frac{3}{4}\cdot\frac{3}{4}=\frac{27}{64}$	0	1	0	0

Z_1	0	1	2	3
p	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

Z_2	0	$\frac{1}{3}$	$\frac{2}{3}$	1
p	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

Z_3	0	1
p	$\frac{28}{64}$	$\frac{36}{64}$

Z_4	0	$\frac{2}{9}$
p	$\frac{28}{64}$	$\frac{36}{64}$

Табела 7.2. Узорци и статистике из „0-1" расподеле

На овај начин смо у стању да из расподеле популације, односно, закона вероватноћа промењљиве *X*, одредимо закон вероватноћа статистике. Према томе, расподела статистике зависи од расподеле популације и од облика те статистике, па ће једна иста статистика имати различите расподеле (за различите популације), а за једну исту расподелу можемо утврдити неограничен број статистика. Зато је потребно проучавати оне статистике које се најчешће користе у примени, а истовремено посматрати оне популације и њихове расподеле које су, такође, најчешће у примени.

Средина узорка

За одређивање расподеле појединих статистика прво ћемо одредити очекивану вредност и варијансу посматране статистике, а тек онда ћемо одредити и њен закон вероватноћа.

Нека је средина (аритметичка средина) обележја X на популацији једнака m, а његова варијанса има вредност σ^2, тј. нека је

$$m = E(X)$$

$$\sigma^2 = Var(X)$$

при чему је X случајна променљива везана за неки експеримент. Посматрајмо узорак величине n, извучен из ове популације. Означимо елементе узорка са $X_1, X_2, ..., X_n$.

***Аритметичка средина узорака**, или краће **средина узорка** је статистика дата функцијом*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n.$$

а) Статистика $\bar{x}$ је линеарна функција случајних промењљивих X_i, које су међусобно независне и све имају исту расподелу, па је очекивана вредност ове статистике једнака линеарној функцији очекиваних вредности променљивих X_i. Очекивана вредност и варијанса сваке променљиве X_i једнака је очекиваној вредности *m* и варијанси σ^2 популације.

Зато је

$$E(\bar{x}) = \frac{1}{n} nm = m$$

$$Var(\bar{x}) = \frac{1}{n^2} n\sigma^2 = \frac{1}{n}\sigma^2,$$

односно, очекивана вредност средине узорка је једнака очекиваној вредности (средини) популације, а њена варијанса је једнака варијанси популације подељеној са величином узорка, тј.

$$E(\bar{x}) = m$$

$$Var(\bar{x}) = \frac{\sigma^2}{n}.$$

Из ове особине се може закључити да ће се са повећањем узорка, варијанса средине узорка $\bar{x}$ смањивати и тежити нули кад $n \to \infty$. То значи, вероватноћа да ће се $\bar{x}$ наћи у одређеној околини око *m,* тежиће јединици кад $n \to \infty$, тј. за довољно велики обим узорка моћи ћемо, **скоро са сигурношћу, тврдити да ће се средина узорка мало разликовати од средине популације**.

б) Претпоставимо да обележје *X* (односно случајна променљива *X*) на популацији има нормалну расподелу, тј.

$$X : N(m; \sigma^2).$$

Код узорка узетог из овакве популације, средина узорка, као линеарна функција независних случајних променљивих, имаће Нормалну расподелу

$$\bar{x} : N\left(m; \frac{\sigma^2}{n}\right).$$

Одавде се може закључити да ће и статистика

$$Z^* = \frac{\bar{x} - m}{\sigma}\sqrt{n}$$

која представља ***стандардизовану средину узорка***, имати стандардизовану Нормалну расподелу, тј.

$$Z^* = \frac{\bar{x} - m}{\sigma}\sqrt{n} : N(0;1).$$

Овај резултат се може искористити за одређивање вероватноћа појединих догађаја везаних за средину узорка $\bar{X}$. Најчешће ће нас интересовати проблеми везани за вероватноће разлика средине узорка $\bar{X}$ и средине популације ***m***. То ће бити догађај облика

$$\left|\bar{x} - m\right| \le \varepsilon,$$

при чему је $\varepsilon > 0$. Тако ћемо добити да је

$$P\left\{\left|\bar{x} - m\right| \le \varepsilon\right\} = 2\phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - 1$$

Дајући вредности ε, за задани обим узорка n и варијансу популације σ^2, може се одредити тражена вероватноћа из таблица за функцију Нормалне расподеле. Обрнуто, за задану вероватноћу може се одредити околина средине m у којој ће се наћи средина узорка.

Често ће се одређивати обим узорка тако да би се, са задатом вероватноћом, тврдило да ће се средина узорка наћи у датој околини средине популације.

в) Да би се одредила расподела средине узорка $\bar{x}$ за сваку популацију, користи се тзв. ***централна гранична теорема***:

Ако је очекивана вредност (средина) популације m, а варијанса σ^2, тада расподела средине $\bar{x}$ узорка тежи Нормалној расподели са средином m и варијансом σ^2 / n, кад n неограничено

расте, па, за довољно велико n, можемо рећи да ће средина узорка $\overline{X}$ имати приближно Нормалну расподелу, тј.

$$\overline{x} : (\approx) N\left(m; \frac{\sigma^2}{n}\right).$$

Доказ. Мада централна гранична теорема важи и у општијим случајевима, ми ћемо се овде ограничити само на оне расподеле за које *постоји функција генератриса*.

Означићемо са V случајну промењиву

$$V = \frac{X - m}{\sigma},$$

при чему су m и σ^2 средина и варијанса X. Њена функција генератриса је функција

$$g_v(t) = E\left\{ e^{t\frac{X-m}{\sigma}} \right\}. \tag{7.2}$$

Користећи Маклоренов ред за функцију e^x, добија се

$$g_v(t) = E\left\{ 1 + \frac{t}{\sigma}\frac{X-m}{1!} + \frac{t^2}{\sigma^2}\frac{(X-m)^2}{2!} + \ldots \right\},$$

односно,

$$g_v(t) = 1 + \frac{t^2}{\sigma^2}\frac{\sigma^2}{2!} + \frac{t^3}{\sigma^3}\frac{\mu_3}{3!} + \ldots, \tag{7.3}$$

при чему је

$$\mu_k = E(X - m)^k,$$

k-ти моменат случајне промењиве X, *(k = 3, 4, ...)*.

Стандардизована средина узорка је

$$Z = \frac{\overline{x} - m}{\sigma}\sqrt{n},$$

а њена функција генератриса ће имати облик

$$g_z(t) = E\left\{ e^{t\frac{\overline{x}-m}{\sigma}\sqrt{n}} \right\} = E\left\{ e^{\frac{t}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i - m}{\sigma}} \right\} = \prod_{i=1}^{n}\left\{ Ee^{\frac{t}{\sqrt{n}}\frac{X_i - m}{\sigma}} \right\}.$$

Поређењем са (7.2), види се да је

$$g_z(t) = \prod_{i=1}^{n} g_v\left(\frac{t}{\sqrt{n}}\right), \text{ tj.}$$

$$g_z(t) = \left[g_v\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Њен логаритам ће бити

$$\log g_z(t) = n\log\left\{ g_v\left(\frac{t}{\sqrt{n}}\right)\right\},$$

па после замене (7.3), добија се да је

$$\log g_z(t) = n\log\left\{ 1 + \frac{1}{n}\left[\frac{t^2}{2} + \frac{t^3}{\sigma^3\sqrt{n}}\frac{\mu_3}{3!} + \dots\right]\right\}.$$

Користећи Маклоренов ред за функцију

$$\log(1 + x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 + \dots$$

добиће се да је

$$\log g_z(t) = n\left\{ \frac{1}{n}\left[\frac{t^2}{2} + \frac{t^3}{\sigma^3\sqrt{n}}\frac{\mu_3}{3!} + \dots\right] - \frac{1}{2n^2}\left[\frac{t^2}{2} + \frac{t^3}{\sigma^3\sqrt{n}}\frac{\mu_3}{3!} + \dots\right] + \dots\right\},$$

па је

$$\lim_{n\to\infty} \log g_z(t) = \frac{t^2}{2}.$$

То значи да ће функција $g_z(t)$ тежити функцији $g(t) = e^{\frac{t^2}{2}}$, а то је функција генератриса стандардизоване Нормалне расподеле. Према томе, расподела статистике $Z = \frac{\overline{x} - m}{\sigma}\sqrt{n}$ тежи Стандардизованој нормалној расподели. То значи да ће и расподела средине узорка $\overline{x}$ тежити нормалној расподели, а то је и требало доказати.

Јасно је да, на основу централне граничне теореме, нормална расподела постаје, тако рећи, универзално примењива, па зато представља расподелу која се најчешће користи у пракси и у обради статистичких података. За једну групу података, основно што треба рачунати, јесте њихова средина, а тај параметар има приближно нормалну расподелу без обзира какву расподелу има посматрано обележје на популацији. Наравно, уз претпоставку да се подаци могу схватити као узорак са довољно великим бројем елемената.

Користећи ову теорему, можемо лако решавати следеће проблеме:

1. Колика је вероватноћа да ће се средина узорка и средина популације разликовати за мање од датог броја ε? То ће бити следећа вероватноћа

$$P\{|\overline{x} - m| < \varepsilon\} \approx 2\phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - 1;$$

2. Одредити интервал око средине узорка, тако да са заданом вероватноћом, тврдимо да ће средина популације бити у том интервалу. То ће бити интервал са границама

$$(\overline{x} - \varepsilon; \overline{x} + \varepsilon)$$

   при чему ε треба одредити тако да функција

$$2\phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right)-1$$

буде једнака задатој вероватноћи;

3. За који обим узорка n можемо, са задатом вероватноћом, тврдити да ће се средина узорка и средина популације разликовати за мање од датог броја ε? То ће бити она вредност n за коју функција

$$2\phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right)-1$$

има вредност задате вероватноће.

Варијанса узорка

Претпоставимо да су параметри популације средина $\boldsymbol{m}$, варијанса σ^2, обични моменти m_k и централни моменти μ_k, (k=2,3,...).

а) *Ако је ($X_1, X_2, ..., X_n$) прост случајни узорак, тада је варијанса узорка дата функцијом*

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2,$$

а то је статистика чија је очекивана вредност једнака

$$E\left(s^2\right) = \frac{n-1}{n}\sigma^2, \tag{7.4}$$

а варијанса

$$Var\left(s^2\right) = \frac{\mu_4 - \sigma^4}{n} - \frac{2\left(\mu_4 - 2\sigma^4\right)}{n^2} + \frac{\mu_4 - 3\sigma^4}{n^3}. \tag{7.5}$$

Да је тачан израз (7.4), којим је дата очекивана вредност варијансе узорка, провериһемо на следећи начин: Варијанса узорка се може написати у следећем облику

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i\right\}^2 .$$

После квадрирања израза у средњој загради, добиће се

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i^2 + \sum_{i}^{n} \sum_{\neq\, j}^{n} X_i X_j\right\} =$$

$$= \frac{n-1}{n^2}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i}^{n} \sum_{\neq\, j}^{n} X_i X_j ,$$

па је очекивана вредност

$$E\left(s^2\right) = \frac{n-1}{n^2}\sum_{i=1}^{n} E\left(X_i^2\right) - \frac{1}{n^2}\sum_{i}^{n} \sum_{\neq\, j}^{n} E\left(X_i\right)E\left(X_j\right)$$

Последњи члан у горњем изразу, има укупно *n(n-1)* чланова чије су вредности

$$E\left(X_i\right)E\left(X_j\right) = m \cdot m = m^2 ,$$

а први члан на десној страни има *n* чланова који имају вредности

$$E\left(X_i^2\right) = m_2 ,$$

тако да је

$$E\left(s^2\right) = \frac{n-1}{n^2} n \cdot m_2 - \frac{1}{n^2} n(n-1)m^2 ,$$

односно,

$$E\left(s^2\right) = \frac{n-1}{n}\left[m_2 - m^2\right] = \frac{n-1}{n}\sigma^2 ,$$

а то је израз (7.4).

Тачност израза за варијансу варијансе узорка проверићемо на сличан начин. Наиме, варијанса варијансе узрока једнака је

$$Var\left(s^2\right) = E\left[s^2\right]^2 - \left\{E\left(s^2\right)\right\}^2 = E\left(s^4\right) - \left\{\frac{n-1}{n}\sigma^2\right\}^2 \tag{7.6}$$

Квадрат варијансе узорка се може написати у облику

$$s^4 = \frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i^2\right\}^2 - \frac{2}{n}\bar{x}^2\sum_{i=1}^{n} X_i^2 + \bar{x}^4 =$$

$$= \frac{1}{n^2}\left\{\sum_{i=1}^{n} X_i^2\right\}^2 - \frac{2}{n^3}\left\{\sum_{i=1}^{n} X_i\right\}^2 \sum_{i=1}^{n} X_i^2 + \frac{1}{n^4}\left\{\sum_{i=1}^{n} X_i\right\}^4 .$$

Користећи особине очекиваних вредности за збир и производ независних промењљивих, лако је проверити да је

$$E\left\{\sum_{i=1}^{n} X_i^2\right\}^2 = n \cdot m_4 + n(n-1)m_2^2$$

$$E\left\{\sum_{i=1}^{n} X_i\right\}^2 \cdot \sum_{i=1}^{n} X_i^2 = n \cdot m_4 + n(n-1)m_2^2 + 2n(n-1)m_3 m + n(n-1)(n-2)m^3$$

$$E\left\{\sum_{i=1}^{n} X_i\right\}^4 = n \cdot m_4 + n(n-1)m_2^2 + 7n(n-1)m_3 m + 2n(n-1)(n-2)m_2 m^2 +$$
$$+ n(n-1)(n-2)(n-3)m^4$$

Уврштавањем ових вредности у израз (7.6) добиће се израз за варијансу варијансе узорка дат у (7.5).

Из (7.5) види се да се са порастом узорка, $Var\left(s^2\right)$ опада, тако да се, за довољно велики узорак, може сматрати да је

$$Var\left(s^2\right) \approx \frac{\mu_4 - \sigma^4}{n}. \tag{7.7}$$

На сличан начин би могли одредити и моменте вишег реда за варијансу узорка. Сви ће они бити изражени као функција момената популације и обима узорка *n*.

б) Пошто смо посматрали две статистике из узорка: средину и варијансу узорка, онда се поставља проблем зависности ове две статистике. За почетак ћемо одредити њихову коваријансу

$$Cov\left(\bar{x}, s^2\right) = E\left(\bar{x} \cdot s^2\right) - E\left(\bar{x}\right)E\left(s^2\right).$$

Користећи претходне изразе, лако је проверити да је коваријанса између средине и варијансе узорка једнака

$$Cov\left(\bar{x}, s^2\right) = \frac{n-1}{n}\mu_3 \tag{7.8}$$

при чему је μ_3 трећи централни моменат популације.

Из израза (7.8) следи

Закључак*: **Ако је расподела популације симетрична, тада су средина и варијанса узорка некорелиране статистике**.*

в) Да би одредили расподелу варијансе узорка, доказаћемо следећу теорему.

Теорема 7.1. ***Нека су*** $\bar{x}$ ***и*** s^2 ***средина и варијанса узорка из популације са Нормалном расподелом*** $N\left(m;\sigma^2\right)$**.**
Тада

- $\bar{x}$ ***и*** s^2 ***су међусобно независне статистике;***
- ***Средина узорка*** $\bar{x}$ ***има нормалну расподелу;***
- ***Статистика***

$$\frac{ns^2}{\sigma^2}$$

има Хи-квадрат расподелу са (n-1) степени слободе.

Доказ. Ако су елементи узорка $X_1, X_2, ..., X_n$, тада ћемо са $Z_1, Z_2, \ldots, Z_n$ означити нихов стандардизовани облик, тј.

$$Z_i = \frac{X_i - m}{\sigma},\ i=1,2,\ldots,n$$

Из претпоставки теореме, и из ове смене, следи да ће ове промењиве имати стандардизоване Нормалне расподеле. Њихов збир квадрата се може написати у облику

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \frac{\left(X_i - \overline{x} + \overline{x} - m\right)^2}{\sigma^2} = \frac{ns^2}{\sigma^2} + \frac{n}{\sigma^2}\left(\overline{x} - m\right)^2 \qquad (7.9)$$

Њихова аритметичка средина ће бити

$$\overline{z} = \frac{\overline{x} - m}{\sigma},$$

тако де се (7.9) може написати у облику

$$\sum_{i=1}^{n} Z_i^2 = \frac{ns^2}{\sigma^2} + n\overline{z}^2 .$$

Увешћемо нове промењиве $Y_1, Y_2, \ldots, Y_n$ преко линеарне трансформације следећег облика

$$Y_1 = \frac{1}{\sqrt{n}}\left(Z_1 + \ldots + Z_n\right)$$

$$Y_2 = \frac{1}{\sqrt{2 \cdot 1}}\left(Z_1 - Z_2\right)$$

$$\ldots$$

$$Y_i = \frac{1}{\sqrt{i(i-1)}}\left[Z_1 + Z_2 + \ldots + Z_{i-1} - (i-1)Z_i\right]$$

$$\ldots$$

$$Y_n = \frac{1}{\sqrt{n(n-1)}}\left[Z_1 + \ldots + Z_{n-1} - (n-1)Z_n\right]$$

Пошто је очекивана вредност Z_i једнака нули, онда из ове трансформације следи да су очекиване вредности променљивих Y_1, Y_2, …, Y_n, такође једнаке нули, тј.

$$E(Y_i) = 0,$$

за свако i = 1, 2, ..., n.

Због независности променљивих $Z_1, Z_2, \ldots, Z_n$ и њиховог стандардизованог облика, добиће се да је

$$Var(Y_1) = \frac{1}{n}\sum_{i=1}^{n} Var(Z_i) = 1$$

$$Var(Y_i) = \frac{1}{i(i-1)} \cdot \left\{ \sum_{j=1}^{i-1} Var(Z_j) + (i-1)^2 Var(Z_i) \right\} =$$

$$= \frac{1}{i(i-1)} \left\{ (i-1) + (i-1)^2 \right\} = 1,$$

за свако i = 2, ..., n.

Коваријансе између Y_i и Y_j биће једнаке

$$Cov(Y_i, Y_j) = \frac{1}{\sqrt{i(i-1)j(j-1)}} E\left\{ \sum_{k=1}^{i-1} Z_k - (i-1)Z_i \right\}\left\{ \sum_{l=1}^{j-1} Z_l - (j-1)Z_j \right\} =$$

$$= \frac{1}{\sqrt{i(i-1)j(j-1)}} \left\{ E\left(Z_1^2\right) \underbrace{+ 0 + \ldots + 0}_{(1)} + \underbrace{0 + E\left(Z_2^2\right) + 0 + \ldots}_{(2)} + 0 + \ldots + \right.$$

$$\left. + 0 \underbrace{+ \ldots + E\left(Z_{j-1}^2\right)}_{(j-1)} - (j-1)E\left(Z_j^2\right) \right\} = 0$$

Према томе, случајне променљиве $Y_1, Y_2, \ldots, Y_n$ су међусобно независне и имају стандардизовану нормалну расподелу.

Из израза за X_1, види се да је

$$Y_1^2 = n\bar{z}^2 = \frac{n}{\sigma^2}(\bar{x} - m)^2. \tag{7.10}$$

Такође је

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} Z_i^2,$$

па се, на основу (7.9) и (7.10), може написати

$$\sum_{i=2}^{n} Y_i^2 + Y_1^2 = \frac{ns^2}{\sigma^2} + \frac{n}{\sigma^2}(\bar{x} - m)^2 .$$

У горњем изразу, први члан на левој страни једначине има *Хи*-квадрат расподелу са *(n-1)* степени слободе (као збир квадрата независних промењљивих са стандардизованом Нормалном расподелом). То значи да ће статистика

$$\frac{ns^2}{\sigma^2}$$

такође имати ту расподелу, а то је тврђење в).

Због (7.10) и независности промењљивих $Y_1, Y_2, \ldots, Y_n$, следи да ће $\bar{x}$ бити независна од $Y_2, \ldots, Y_n$, односно, од варијансе узорака s^2, а то је тврђење а).

На крају, тврђење б) смо већ проверили при одређивању расподеле за средину узорка.

Заједнички закон вероватноћа за статистике $\bar{x}$ и s^2 биће дат функцијом

$$f\left(\bar{x}, s^2\right) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} e^{-\frac{n}{2\sigma^2}(\bar{x}-m)^2} \cdot C_{n-1} \frac{1}{\sigma^{n-1}} s^{n-2} \cdot e^{-\frac{ns^2}{2\sigma^2}} \tag{7.11}$$

при чему је C_{n-1} константа одређена из *Хи*-квадрат расподеле. Могу се одредити и моменти вишег реда варијансе узорка s^2 извученог из Нормалне расподеле.

Тако ћемо, за моменат реда *k*, имати следећи израз

$$E\left(s^2\right)^k = \frac{(n-1)(n+1)(n+3)...(n+2k-3)}{n^k}\sigma^{2k} \qquad (7.12)$$

Специјално, за *k = 2*, добићемо варијансе узорка

$$Var\left(s^2\right) = \frac{(n-1)(n+1)}{n^2}\sigma^4$$

Преко израза (7.12) могу се одредити и коефицијенти асиметрије и спљоштености, који ће имати вредности

$$\beta_1\left(s^2\right) = \frac{2\sqrt{2}}{\sqrt{n-1}}, \ \beta_2\left(s^2\right) = \frac{12}{n-1}.$$

Студентова t-статистика

Госет (Gosset) [16] је 1908. године објавио свој рад, под псеудонимом „Студент", у коме је одредио закон вероватноћа количника

$$(\overline{x} - m)/s,$$

а Фишер (R. A. Fisher) [13] је 1925. године, извео ригорознији доказ Госетовог резултата. Фишер је дефинисао *t*-статистику изразом

$$t_{n-1} = \frac{\overline{x} - m}{s}\sqrt{n-1}$$

при чему је $\overline{x}$ средина, а s^2 варијанса узорка $X_1, X_2, ..., X_n$ извученог из Нормалне расподеле са параметрима *m* и σ^2.

Доказ. Из теореме 7.1. следи да су статистике

$$\frac{\bar{x}-m}{\sigma}\sqrt{n} \text{ и } \frac{ns^2}{\sigma^2}$$

независне, и да прва има стандардизовану нормалну расподелу, а друга има *Хи*-квадрат расподелу са *(n-1)* степени слободе. Њихов количник

$$\frac{\frac{\bar{x}-m}{\sigma}\sqrt{n}}{\sqrt{\frac{ns^2}{\sigma^2/(n-1)}}}=\frac{\bar{x}-m}{s}\sqrt{n-1},$$

имаће Студентову расподелу. Тако смо дошли до следећег резултата:

Теорема 7. 2. ***Студентова t-статистика***

$$t=\frac{\bar{x}-m}{s}\sqrt{n-1},$$

код узорка из нормалне расподеле, има Студентову t-расподелу са (n-1) степени слободе.

Ова теорема омогућава широку примену t-статистике у решавању разних проблема статистичког закључивања. Њене вредности и расподела не зависе од варијансе популације, која нам је у већини случајева непозната.

Статистике из два назависна узорка

Често се срећемо са два независно изабрана узорка из популације. Нека је $X_1, X_2, ..., X_n$ узорак изабран из популације са Нормалном расподелом и параметрима m и σ_1^2, а $X_1', X_2', ..., X_m'$ узорак из исте популације са параметрима m_2 и σ_2^2.

Средине ових узорака

$$\bar{x}_1=\frac{1}{n}\sum_{i=1}^{n}X_i$$

$$\bar{x}_2 = \frac{1}{m}\sum_{i=1}^{m} X_i^{,}$$

имају Нормалне расподеле, а њихова разлика

$$\bar{x}_1 - \bar{x}_2$$

ће бити статистика која ће имати Нормалну расподелу са параметрима

$$E\{\bar{x}_1 - \bar{x}_2\} = m_1 - m_2$$

$$Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}.$$

Варијансе узорака

$$s_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{x}_1\right)^2$$

$$s_2^2 = \frac{1}{m}\sum_{i=1}^{m}\left(X_i^{©} - \bar{x}_2\right)^2$$

имаће расподеле одређене теоремом 7.1.

Због независности узорака, статистика

$$\frac{ns_1^2}{\sigma_1^2} + \frac{ms_2^2}{\sigma_2^2}$$

имаће Хи-квадрат расподелу са $(n+m-2)$ степени слободе. Поред тога, ова статистика је независна од статистике $(\bar{x}_1 - \bar{x}_2)$.

Студентова t-статистика

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (m_1 - m_2)}{\sqrt{ns_1^2 + ms_2^2}}\sqrt{\frac{nm}{n+m}(n+m-2)}$$

имаће Студентову t-расподелу са $(n+m-2)$ ***степени слободе.***

Посебно значајна у примени је *F-статистика*. То је статистика дефинисана количником

$$F = \frac{(n-1)ms_2^2}{(m-1)ns_1^2}$$

Ако је $\sigma_1^2 = \sigma_2^2$, онда F-статистика има F-расподелу са (m-1) и (n-1) степени слободе.

Ова расподела је последица независности статистика

$$\frac{ms_2^2}{\sigma_2^2} \text{ и } \frac{ns_1^2}{\sigma_1^2}.$$

7.4. ЗАДАЦИ

1. Одреди све узорке обима 5 из Биномне расподеле $B(0.2;9)$.

2. Одреди расподеле статистика $\overline{x}$, s^2, $R = X_{\max} - X_{\min}$, и слично, за узорак обима 5 из расподеле $B(0,2;9)$.

3. Одреди очекиване вредности и варијансе обичних момената узорка $\hat{m}_2, \hat{m}_3$ и $\hat{m}_4$, при чему је

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k .$$

4. Одреди расподелу статистика

$$\overline{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2 ,$$

$$\overline{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2 ,$$

за узорак извучен из популације са Нормалном расподелом.

5. Статистика

$$s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)^2 ,$$

има варијансу једнаку

$$Var\left(s'^2\right) = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\sigma^4\right).$$

Одреди очекивану вредност, варијансу и расподелу статистике s'^2, ако је узет узорак из Нормалне расподеле.

6. Одреди варијансу варијансе узорка из Нормалне расподеле $N(m;\sigma^2)$.

7. Одреди коефицијенте асиметрије и спљоштености средине узорка $\bar{x}$ из популације са Нормалном расподелом $N(m;\sigma^2)$.

8. Код узорка из популације са Нормалном расподелом одреди очекивану вредност и варијансу *медијане узорка*.

9. Показати са код узорка из Пуасонове расподеле $P(1)$, статистика $2\sqrt{\bar{x}}$

   има приближно Нормалну расподелу са параметрима

$$E(2\sqrt{\bar{x}}) = \sqrt{\lambda}$$
$$Var(2\sqrt{\bar{x}}) = \frac{1}{n}.$$

10. Ако популација има коначну очекивану вредност m и варијансу σ^2, тада статистика s^2 тежи по вероватноћи σ^2, а статистика

$$\frac{\sqrt{n}(\bar{x}-m)}{s}$$

   има приближно Нормалну расподелу $N(0;1)$.

11. Нека су $\bar{x}_1$ и s_1^2, средина и варијанса узорка величине n_1 из популације са средином m_1 и варијансом σ_1^2, а $\bar{x}_2$ и σ_2^2 средина и варијанса другог узорка од n_2 елемената из популације са средином m_2 и варијансом σ_2^2. Покажи да ће статистика

$$\frac{(\bar{x}_1-\bar{x}_2)-(m_1-m_2)}{\sqrt{n_2 s_2^2 + n_1 s_2^2}}\sqrt{n_1 n_2}$$

   имати приближно Нормалну расподелу $N(0;1)$.

8. ТЕОРИЈА ОЦЕЊИВАЊА

У деценијама друге половине 20-ог века дошло је до наглог ширења примене статистике у свим, или скоро свим областима истраживања. Тада је развијена и теорија ***статистичког оцењивања***, или краће**,** ***теорија оцењивања***.

Типичан проблем којим се бави теорија оцењивања ћемо приказати кроз неколико примера. Желимо пласирати производ на ново тржиште. Колико ће купаца куповати наш производ, не знамо. Требали би питати све становнике да ли ће то чинити, а то је практично немогуће. Зато ћемо извршити анкету преко узорка из те популације и на основу резултата анкете оценити потенцијални број купаца.

Претпоставимо да смо добили нову врсту ђубрива коју ћемо користити за сејање пшенице. Да би испитали повећање приноса, засејаћемо на одређеним парцелама пшеницу. Употребићемо ново ђубриво на неким парцелама, а на неким га нећемо употребити. Измерићемо приносе и добити приносе по парцелама за парцеле са новим ђубривом и за парцеле на којима нисмо користили ђубриво.

Потребно је утврдити просечно повећање приноса. Кад би имали бесконачно, или велики број мерења, онда би могли утврдити праве просечне приносе при употреби овог ђубрива и без његове употребе. Поставља се питање како ћемо на основу ових мерења, "утврдити" те приносе и повећања. Поступак за "утврђивање" приноса и повећања зваћемо ***оцењивање***. Праве просечне приносе не можемо тачно утврдити, али их можемо на основу мерења, оценити.

Сличан је и следећи проблем везан за посматрање нове методе учења. Њену просечну ефикасност би утврдили кад би посматрали бесконачан број испитаника. То је ипак немогуће. Зато ћемо испитати ефикасност методе на једној групи испитаника. Потребно је да на основу резултата код ове групе *оценимо* праву ефикасност методе.

Или, стабилност производног процеса изражена је варијабилитетом одређених параметара. Нисмо у стању да у сваком тренутку утврђујемо вредности тих параметара. Морамо на основу мерења у одређеним тренуцима оценити стабилност процеса у посматраном периоду.

Дакле, ради се о одређеним параметрима статистичког скупа на великој или бесконачној популацији, који су нам непознати. На основу мерења вредности обележја на узорку, ***оцењиваћемо непознате параметре популације*.**

Општији проблем теорије оцењивања је следећи. На основу резултата мерења обележја X у узорку, треба оценити расподелу посматраног обележја на целој популацији, а затим из те расподеле оцењивати и непознате параметре популације.

У теорији оцењивања полази се од случајне променљиве *X* и њене расподеле. Претпоставља се да расподела обележја *X* припада једној класи расподела, одређеној параметрима $\theta_1, \theta_2, ..., \theta_k$. Другим речима, познато је да је закон вероватноћа случајне променљиве *X* функција $p(x; \theta_1, \theta_2, ..., \theta_k)$ која зависи од непознатих параметара $\theta_1, \theta_2, ..., \theta_k$. Непознате параметре $\theta_1, \theta_2, ..., \theta_k$ **оценићемо** на основу мерења резултата експеримента којим је одређена случајна променљива *X*. Означимо их са $X_1, X_2, ..., X_n$ Дакле, проблем се своди на експериментално утврђивање непознатих параметара.

При решавању тог проблема битна су следећа два момента:

1. Како извести експеримент и измерити резултате његовог извођења?

2. Како на основу добијених резултата мерења, **“утврдити”** вредности, односно, оценити непознате параметре $\theta_1, \theta_2, ..., \theta_k$?

Статистичка теорија која даје одговоре на прво питање, назива се **планирање експеримента**. Ми се нећемо овде бавити том теоријом, већ ћемо се ограничити на независна понављања експеримента и измерене резултате тих понављања. Као што знамо, то ће бити ***прост случајан узорак*** кога смо дефинисали као скуп независних случајних промењивих

$$X_1, X_2, \ldots, X_n$$

које имају исту расподелу као и случајна промењива X на популацији.

Поновљени експеримент n пута, представља једну реализацију простог случајног узорка. То ће бити бројеви $x_1, x_2, \ldots, x_n$, а на основу ових вредности треба да донесемо закључке о вредностима непознатих параметара популације.

Теорија оцењивања се може поделити у више различитих области, а овде ћемо се, углавном, бавити класичном теоријом оцењивања чије основе је дао Р. А. Фишер.

8.1. КРИТЕРИЈУМИ ИЗБОРА ОЦЕНА

Посматраћемо статистички скуп као популацију и обележје X на тој популацији. Означимо са θ непознати параметар посматране популације. Да би оценили вредност тог непознатог параметра, посматраћемо прост случајан узорак

$$(X_1, X_2, \ldots, X_n),$$

изабран из дате популације.

Најопштија оцена параметра θ може бити нека функција дефинисана на овом узорку. То ће бити **статистика**. Означимо је са

$$\hat{\theta} = f(X_1, X_2, \ldots, X_n).$$

Дакле, оцена параметра популације је статистика дефинисана на узорку.

Избором функција f изабрали смо оцену параметра θ, преко посматраног узорка. Како функција f може бити произвољно изабрана, поставља се питање критеријума избора функције f, односно, избора оцене параметра θ.

Критеријум избора оцене зависи од циља који се жели постићи том оценом. Зато и не постоји јединствен критеријум по коме би се за сваки параметар вршио избор оцена. Потребно је утврдити шта ће бити “добра” оцена, шта ћемо подразумевати под “бољом” оценом и шта ће то значити “најбоља оцена”.

Непристрасне оцене

Оцена непознатог параметра θ је статистика $\hat{\theta}$. То је, уствари случајна промењљива која ће узимати разне вредности за разне узорке из исте популације. Та статистика ће имати своју очекивану вредност, а то ће бити она вредност статистике која представља параметар њене централне тенденције. Пожељно је да тај параметар централне тенденције статистике $\hat{\theta}$ буде вредност параметра θ, јер је то она вредност за коју "очекујемо" да ће статистика, на основу једног узорка, бити у њеној околини.

Дакле, очекивана вредност статистике биће њена мера централне тенденције, па је пожељно да буде испуњен услов

$$E(\hat{\theta}) = \theta,$$

тј. да је очекивана вредност оцене $\hat{\theta}$ једнака параметру θ.

Ако је очекивана вредност оцене

$$\hat{\theta} = f(X_1, X_2, ..., X_n)$$

непознатог параметра популације θ, једнака том параметру θ, онда се таква оцена зове непристрасна оцена, тј. када је $E\hat{\theta}) = \theta$, оцена је непристрасна.

Ако је $E(\hat{\theta}) > \theta$, онда је таква оцена позитивно пристрасна, а ако је $E(\hat{\theta}) < \theta$, онда је то негативно пристрасна оцена.

На пример, нека је очекивана вредност (просек) популације *m* непознат, а статистика је средина узорка

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Видели смо да је

$$E(\bar{x}) = m,$$

што значи да средина узорка представља непристрасну оцену очекиване вредности популације.

Такође смо видели да је очекивана вредност статистике

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{x})^2$$

једнака

$$E(s^2) = \frac{n-1}{n}\sigma^2,$$

што значи да је варијанса узорка негативно пристрасна оцена варијансе популације σ^2.

Ако нам је очекивана вредност популације m позната, *онда статистике*

$$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^k$$

представљају непристрасне оцене централних момената, и то за свако $k = 1,2,3....$

Обични моменти узорака, такође, представљају непристрасне оцене обичних момената популације.

Основну предност непристрасних оцена можемо објаснити на следећи начин. Посматрајмо само оне непристрасне оцене $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$, чије су варијансе мање од варијансе популације. То ће бити оцене код којих је

$$E(\hat{\theta}_i) = \theta,\ Var(\hat{\theta}_i) < \sigma^2$$

Тада ће оцена

$$\overline{\theta}_k = \frac{1}{k}\sum_{i=1}^{k}\hat{\theta}_i$$

бити оцена за коју је

$$E\left(\overline{\theta}_k\right) = \theta$$

$$Var\ (\overline{\theta}_k) = \frac{1}{k^2}\sum_{i=1}^{k}\sigma_i^2 < \frac{\sigma^2}{k}$$

Ако k тежи бесконачности, тада варијанса оцене $\overline{\theta}_k$ тежи нули, што значи да ће $\overline{\theta}_k$ бити све ближе и ближе правој вредности параметра θ.

Међутим, ако би оцене $\hat{\theta}_i$ биле пристрасне са пристрасношћу једнаком δ, онда би оцена $\overline{\theta}_k$ била све ближе и ближе вредности $\theta + \delta$, а не вредности параметра θ. И што је пристрасност већа то су и ове разлике веће. Непристрасност оцена је пожељна особина, која обезбеђује да, на основу већег броја узорака из исте популације, у просеку тачно одредимо вредност непознатог параметра. То, ипак, не значи да ћемо на основу једног узорка, тачно одредити ту вредност. Може се десити да је пристрасност једне оцене мала у односу на њену стандардну девијацију, па је не можемо сматрати "лошом" оценом.

Зато је потребно утврдити неке друге критеријуме за избор ваљаности оцена. Поред тога, за један исти параметар популације θ, може постојати и више непристрасних оцена, $\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_k$ па је потребно у класи непристрасних оцена, изабрати **"бољу"** или **"најбољу"** оцену.

Оцене са минималном варијансом

Ограничићемо се на **класу непристрасних оцена** непознатог параметра θ. Потребно је да изградимо критеријум "ваљаности" једне такве оцене.

Интуитивно, оцена $\hat{\theta} = f(X_1, X_2, ..., X_n)$ параметра θ биће добра ако је, у извесном смислу, близу праве вредности тог параметра.

Зато је неопходно да утврдимо меру **“блискости”** статистике (оцене) и параметра.

Посматраћемо неки интервал око праве вредности, тј. интервал

$$\left(\theta-\lambda_1;\theta+\lambda_2\right]$$

при чему су бројеви $\lambda_1,\lambda_2 \in (0;\lambda)$. Статистика $\hat{\theta}$, може узети вредност из овог интервала. Нека је вероватноћа да се то деси једнака

$$P\left\{\theta-\lambda_1<\hat{\theta}\leq\theta+\lambda_2\right\} \tag{8.1}$$

Посматраћемо сада, било какву другу статистику

$$\hat{\theta}_1 = f_1(X_1, X_2, ..., X_n)$$

и одредити вероватноћу

$$P\left\{\theta-\lambda_1<\hat{\theta}_1\leq\theta+\lambda_2\right\} \tag{8.2}$$

Ако је вероватноћа (8.1) увек већа од вероватноће (8.2), и то за свако $\lambda_1,\lambda_2 \in (0;\lambda)$ онда је статистика $\hat{\theta}$ боља оцена од статистике $\hat{\theta}_1$. Да би тај услов био испуњен потребно је да **варијанса статистике** $\hat{\theta}$ буде мања од варијансе статистике $\hat{\theta}_1$.

Варијанса статистике која представља непристрасну оцену, зове се средња квадратна грешка оцене.

Да би оцена $\hat{\theta}$ једног параметра θ, била боља од оцене $\hat{\theta}_1$ тог истог параметра, потребно је да је

$$Var\,(\hat{\theta})=E(\hat{\theta}-\theta)^2 \leq Var\,(\hat{\theta}_1)=E(\hat{\theta}_1-\theta)^2,$$

тј. потребно је да је њена средња квадратна грешка мања од средње квадратне грешке оцене $\hat{\theta}_1$.

Иначе, средња квадратна грешка оцене изражава меру концентрисаности оцене око праве вредности параметра, па је зато пожељно да та концентрисаност буде што већа, односно, да средња квадратна грешка оцене буде што мања.

Да би дошли до "најбоље" непристрасне оцене једног параметра, потребно је да у класи свих непристрасних оцена, нађемо оцену чија варијанса има минималну вредност, тј. *потребно је да нађемо ону статистику* $\hat{\theta}$ *за коју је*

$$Var\,(\hat{\theta}) \leq Var\,(\hat{\theta}_1),$$

при чему је $\hat{\theta}_1$ било каква непристрасна оцена параметра θ.

Непристрасна оцена са минималном варијансом представља *оптималну* оцену параметра θ у класи непристрасних оцена.

На пример, ако се посматра Нормална расподела са непознатом очекиваном вредношћу *m*, може се показати да медијана узорка $\hat{M}_e$ и средина узорка $\bar{x}$ представљају непристрасне оцене *m*, али је $\bar{x}$ ипак боља од $\hat{M}_e$. То је уједно и оцена са минималном варијансом.

Код Нормалне расподеле са познатом очекиваном вредношћу *m* и непознатом варијансом σ^2, оцена

$$\bar{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

биће *оптимална* оцена варијансе σ^2.

Ефикасне оцене

За параметар популације θ потребно је наћи његову оцену $\hat{\theta} = f(X_1, X_2, ..., X_n)$. Видели смо да ће та оцена бити боља ако је њена варијанса мања (кад је то непристрасна оцена). Зато можемо израз

$$E(\hat{\theta} - \theta)^2,$$

тј. **средњу квадратну грешку оцене параметра**, сматрати као меру ефикасности те оцене. Што је та вредност ближа нули, то је оцена ефикаснија. Да би добили праву меру ефикасности једне оцене, потребно је упоредити ову средњу квадратну грешку са сличним показатељима.

Најлогичније је да посматрамо ону оцену $\hat{\theta}_0$ која ће имати најмању средњу квадратну грешку. То ће бити оцена за коју је

$$E(\hat{\theta}_0 - \theta)^2 = \min_{\hat{\theta}} E(\hat{\theta} - \theta)^2$$

Ефикасност оцене $\hat{\theta}$ параметра θ је количник минималне средње квадратне грешке и средње квадратне грешке оцене $\hat{\theta}$, тј.

$$Ef(\hat{\theta}) = \frac{E(\hat{\theta}_0 - \theta)^2}{E(\hat{\theta} - \theta)^2}.$$

Ако се посматрају само непристрасне оцене, онда је ефикасност оцене $\hat{\theta}$ количник варијанси, тј.

$$Ef(\hat{\theta}) = \frac{Var(\hat{\theta}_0)}{Var(\hat{\theta})}$$

при чему је $\hat{\theta}_0$ оцена са минималном варијансом.

Јасно је да је ефикасност оцене број који испуњава услов

$$0 \le Ef\left(\hat{\theta}\right) \le 1.$$

Кад је $Ef(\hat{\theta}) = 1$, кажемо да је $\hat{\theta}$ ефикасна оцена параметра θ.

На пример, код узорка из нормалне популације са познатом очекиваном вредношћу m, средња девијација

$$\hat{e} = \frac{1}{n}\sum_{i=1}^{n}\left|X_i - m\right|$$

има очекивану вредност и варијансу једнаку

$$E(\hat{e}) = \sqrt{\frac{2}{\pi}}\sigma, \qquad \mathrm{Var}\ (\hat{e}) = \frac{\pi - 2}{\pi} \cdot \frac{\sigma^2}{n}.$$

Статистика

$$\sqrt{\frac{\pi}{2}}\hat{e},$$

биће непристрасна оцена стандардне девијације. Минимална вредност средњег квадратног одступања је једнака

$$\min_{\hat{\theta}} \; E(\hat{\theta}-\sigma)=\frac{\sigma^2}{2n}.$$

Ефикасност оцене

$$\sqrt{\frac{\pi}{2}}\hat{e}=\sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^{n}\left|X_i-m\right|$$

једнака је

$$Ef\left(\sqrt{\frac{\pi}{2}}\hat{e}\right)=\frac{1}{\pi-2}=0.876$$

што значи да ова оцена не представља ефикасну оцену стандардне девијације али је њена ефикасност прилично велика.

8.2. ИНТЕРВАЛИ ПОВЕРЕЊА

До сада смо посматрали прост случајан узорак $X_1, X_2, \dots, X_n$ из неке популације. Решавали смо проблем како на основу узорка оценити вредност непознатог параметра θ, тј. како на основу узорка одредити једну тачку као вредност тог параметра.

Међутим, у многим проблемима није нам потребно да утврдимо једну вредност као вредност тог параметра. Биће нам довољно да утврдимо неки интервал у коме би могла да се налази права вредност тог параметра. Кад такав интервал одређујемо преко узорка $X_1, X_2, \dots, X_n$, онда тај интервал зовемо ***интервал поверења***.

Теорија одређивања интервала поверења почела се развијати релативно касно. Први концепт је дао Нојман (J.Neyman) 1935. године. Овај метод оцењивања параметара је мање рестриктиван од метода оцењивања одређивањем тачака.

Из узорка $X_1, X_2, \dots, X_n$ одређиваћемо две статистике

$$Z_1 = f_1(X_1, X_2, ..., X_n),$$

$$Z_2 = f_2(X_1, X_2, \dots, X_n),$$

такве да је увек $Z_1 < Z_2$. Ове две статистике су случајне променљиве које одређују неки интервал

$$(Z_1; Z_2],$$

чије границе се мењају са променом узорка. За неки узорак, овај интервал ће прекрити непознати параметар θ, а за неке узорке, неће. Означимо са $\gamma\%$, проценат узорака који ће садржавати тај параметар, а са $(1-\gamma)\%$, проценат узорака који неће садржавати тај параметар.

Другим речима, нека је γ вероватноћа да интервал $(Z_1; Z_2]$ прекрије непознати параметар θ, а $(1-\gamma)$ је вероватноћа да се то не оствари, тј.

$$P\{Z_1 < \theta \le Z_2\} = \gamma ,$$

$$P\{\theta \notin (Z_1, Z_2]\} = 1 - \gamma .$$

Вероватноћа γ се зове *ниво поверења*, а интервал $(Z_1, Z_2]$, *интервал поверења*.

Ако из исте популације бирамо више узорака, онда ће проценат узорака који садрже параметар θ, бити једнак нивоу поверења $\gamma\%$.

Поступак одређивања интервала поверења

Основни проблем у одређивању интервала поверења је проблем одређивања, односно избора статистика Z_1 и Z_2 за унапред одређени ниво поверења. Решаваћемо га следећим поступком.

1) Одредићемо неку функцију узорка $X_1, X_2, ..., X_n$ и параметра θ, рецимо функцију

$$g(X_1, X_2, ..., X_n; \theta),$$

са следећим особинама:

а) Функција g је дефинисана за сваку вредност θ , непрекидна је и монотоно растућа, или монотоно опадајућа.

б) Расподела функције g не зависи од непознатог параметра θ.

Нека је њен закон вероватноћа функција $\varphi(y)$.

2) За задани ниво поверења γ одредићемо две вредности g_1 и g_2 такве да је

$$\int_{g_1}^{g_2} \varphi(y) dy = \gamma .$$

Ове вредности су такве да је за сваку случајну промењиву Y, која има закон вероватноћа $\varphi(y)$., вероватноћа да се нађе у интервалу $(g_1, g_2]$ једнака γ, тј. такве су да је

$$P\{g_1 < Y \le g_2\} = \gamma$$

3) Решавањем неједначина

$$g(X_1, \ldots, X_n; \theta) \le g_2 ,$$

$$g(X_1, \ldots, X_n; \theta) > g_1 ,$$

по параметру θ добиће се решење

$$f_1(X_1, X_2, \ldots, X_n; g_1) < \theta \le f_2(X_1, X_2, \ldots, X_n; g_2),$$

односно,

$$Z_1 < \theta \le Z_2 .$$

Случајни догађаји

$$g_1 < g(X_1, X_2, \ldots, X_n; \theta) \le g_2 \text{ и } Z_1 < \theta \le Z_2 ,$$

су догађаји са једнаком вероватноћом.

4) Пошто статистика g има закон вероватноћа дат функцијом $\varphi(\gamma)$, онда ће вероватноћа догађаја

$$g_1 < g(X_1, X_2, \ldots, X_n; \theta) \le g_2 ,$$

бити једнака нивоу поверења γ. Зато ћемо и тврдити да је

$$P\{Z_1 < \theta \le Z_2\} = \gamma ,$$

$$P\{\theta \notin (Z_1; Z_2)\} = 1 - \gamma ,$$

тако да је интервал поверења, са заданим нивоом γ, интервал

$$\left(Z_1 = f_1\left(X_1, X_2, ..., X_n; g_1\right);\ Z_2 = f_2\left(X_1, X_2, ..., X_n; g_2\right)\right].$$

Основно за одређивање интервала поверења на овај начин је одређивање функције *g*, у првом кораку и њеног закона вероватноћа.

Код статистичких модела расподела, функцију *g* одређујемо из довољних статистика (ако такве постоје), користећи критеријум факторизације функције веродостојности.

За оцене максималне веродостојности, користимо њихову асимптотску расподелу, тј. Нормалну расподелу. Бирамо ону оцену која има најмању варијансу. На тај начин ћемо, за задани ниво поверења, доћи до интервала поверења са најмањом дужином интервала.

Иначе, дужина интервала поверења зависи од нивоа поверења. Са повећањем нивоа поверења повећаће се дужина интервала поверења, а то је непожељно. Обрнуто, са смањењем дужине интервала поверења, смањиће се и ниво поверења, а то је такође непожељно. Зато се у пракси унапред одреди ниво поверења. Најчешће је то 95%-ни или 99%-ни интервал поверења.

Интервал поверења за очекивану вредност

Нека је очекивана вредност популације *m* непозната. Прво ћемо одредити интервал поверења за овај параметар, у случају кад популација има Нормалну расподелу.

Нека је $X_1, X_2, ..., X_n$ узорак из такве популације. Тада средина узорка

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i ,$$

има Нормалну расподелу са параметрима

$$E(\overline{x}) = m;$$

$$Var(\bar{x}) = \frac{\sigma^2}{n},$$

Статистика

$$Z = \frac{\bar{x} - m}{\sigma}\sqrt{n},$$

има стандардизовану Нормалну расподелу са законом вероватноћа

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Према томе, за функцију $g(X_1,...,X_n;\theta)$ можемо узети баш статистику *Z*. Међутим, статистика *Z* зависи и од варијансе популације σ^2. Зато ћемо посматрати два случаја.

1) Нека је варијанса популације σ^2 позната. Из закона вероватноћа статистике *Z*, одредићемо интервал са границом $(-z_0; z_0)$ тако да је

   $$P\{-z_0 \leq Y \leq z_0\} = \gamma,$$

   Користећи таблице за функцију Нормалне расподеле, одредићемо z_0 тако да је

   $$2\phi(z_0) - 1 = \gamma,$$

   односно

   $$\phi(z_0) = \frac{1+\gamma}{2} \tag{8.3}$$

   У следећем кораку решићемо неједначине

   $$\frac{\bar{x} - m}{\sigma}\sqrt{n} \leq z_0$$

$$\frac{\overline{x}-m}{\sigma}\sqrt{n} \geq -z_0$$

Тако ћемо добити решење

$$\overline{x} - z_0 \frac{\sigma}{\sqrt{n}} < m \leq \overline{x} + z_0 \frac{\sigma}{\sqrt{n}} .$$

Можемо, са вероватноћом γ тврдити да ће се реализовати догађај

$$\overline{x} - z_0 \frac{\sigma}{\sqrt{n}} < m \leq \overline{x} + z_0 \frac{\sigma}{\sqrt{n}} ,$$

а са вероватноћом $(1-\gamma)$ да се неће реализовати.

Зато је интервал поверења за *m*, интервал са границама

$$\left(\overline{x} - z_0 \frac{\sigma}{\sqrt{n}} ; \overline{x} + z_0 \frac{\sigma}{\sqrt{n}} \right],$$

при чему је z_0 одређена једначином (8.3).

2) Нека нам је варијанса популације σ^2 непозната. Сад не можемо користити статистику *Z*, јер њене вредности зависе од σ^2. У овом случају користићемо статистику

$$t = \frac{\overline{x}-m}{s}\sqrt{n-1},$$

која има Студентову расподелу са $(n-1)$ степени слободе. При томе смо са s^2 означили варијансу узорка

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)^2 .$$

Из таблице за Студентову расподелу, одредићемо број t_0 тако да је

$$P\{-t_0 < t_{n-1} \leq t_0\} = \gamma.$$

То ће бити она вредност за коју је

$$2S_{n-1}(t_0)-1=\gamma,$$

односно,

$$S_{n-1}(t_0)=\frac{1+\gamma}{2} \tag{8.4}$$

Решавањем неједначина

$$\frac{\overline{x}-m}{s}\sqrt{n-1}<t_0,$$

$$\frac{\overline{x}-m}{s}\sqrt{n-1}\geq -t_0,$$

добиће се решења

$$\overline{x}-t_0\frac{s}{\sqrt{n-1}}<m\leq \overline{x}+t_0\frac{s}{\sqrt{n-1}},$$

за која тврдимо да је

$$P\left\{\overline{x}-t_0\frac{s}{\sqrt{n-1}}<m\leq \overline{x}+t_0\frac{s}{\sqrt{n-1}}\right\}=\gamma .$$

Интервал поверења за m, у овом случају, је интервал са границом

$$\left(\overline{x}-t_0\frac{s}{\sqrt{n-1}};\ \overline{x}+t_0\frac{s}{\sqrt{n-1}}\right],$$

при чему је t_0 одређено једначином (8.4).

Потребно је још одредити интервал поверења за очекивану вредност m и у случајевима кад популација нема Нормалну расподелу. Користићемо се централном граничном теоремом из које се може закључити да ће статистика

$$Z = \frac{\overline{x} - m}{s}\sqrt{n - 1}\,,$$

имати приближно Нормалну расподелу. За довољно велики узорак, интервал поверења биће интервал са границом

$$\left(\overline{x} - z_0 \frac{s}{\sqrt{n-1}};\ \overline{x} + z_0 \frac{s}{\sqrt{n-1}}\right],$$

при чему је z_0 одређено једначином (8.3).

За задани ниво поверења, дужина интервала поверења за очекивану вредност *m*, обрнуто је пропорционална величини узорка. Зато се са повећањем узорка може доћи до интервала поверења са жељеном дужином.

На пример, код мерења вредности обележја *X* на узорку од 100 елемената, добијена просечна вредност је била $\overline{x} = 58{,}65.$ Ако се зна да је узорак из популације са варијансом $\sigma^2 = 144,$, које ће бити границе 95% интервала поверења за непознату очекивану вредност?

Статистика

$$Z = \frac{\overline{x} - m}{\sigma}\sqrt{n} = \frac{\overline{x} - m}{12} \cdot 10$$

имаће стандардизовану Нормалну расподелу. Из таблица за функцију Нормалне расподеле добићемо да је

$$2\phi\left(z_0\right) - 1 = 0.95,$$

односно,

$$z_0 = 1.96,$$

па тврдимо да ће бити

$$P\left\{-1.96 < Z \leq 1.96\right\} = 0.95\,.$$

На основу узорка, оцењене границе интервала поверења ће бити

$$\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} = 56.30,$$

$$\overline{x} + 1.96\frac{\sigma}{\sqrt{n}} = 61.00.$$

За неко друго испитивање добили смо да је средина узорка $\overline{x} = 24$, а његова varijansa је $s^2 = 36$. Извршено је 17 мерења. Из таблица за функцију Студентове расподеле, из услова

$$P\{-t_0 < t_{16} \le t_0\} = 0.95$$

добиће се да је $t_0 = 2.12$. Оцењене границе 95% интервала поверења за очекивану вредност биће

$$\overline{x} - 2.12\frac{s}{\sqrt{n-1}} = 20.82,$$

$$\overline{x} + 2.12\frac{s}{\sqrt{n-1}} = 27.18.$$

Интервал поверења за варијансу

Посматраћемо узорак $X_1, X_2, \ldots, X_n$ из популације са Нормалном расподелом код које је варијанса σ^2 непозната. Да би одредили интервал поверења за варијансу, користићемо статистику

$$\frac{ns^2}{\sigma^2},$$

која има *Хи*-квадарат расподелу са $(n-1)$ степени слободе.

Прво ћемо одредити *једнострани интервал поверења*. То због тога што нас у пракси најчешће интересује случај кад је варијанса велика.

Из таблица за функцију *Хи*-квадрат расподеле одредићемо број χ_0 тако да је

$$P\left\{\chi^2_{n-1} > \chi_0\right\} = \gamma ,$$

odnosno,

$$K_{n-1}\left(\chi_0\right) = 1 - \gamma , \tag{8.5}$$

Зато можемо тврдити да је γ вероватноћа догађаја

$$\frac{ns^2}{\sigma^2} \geq \chi_0 ,$$

односно, догађаја

$$\sigma^2 \leq \frac{ns^2}{\chi_0}.$$

Пошто варијанса не може бити негативна, онда је

$$P\left\{0 < \sigma^2 \leq \frac{ns^2}{\chi_0}\right\} = \gamma ,$$

тако да је једнострани интервал поверења за варијансу σ^2, интервал са границом

$$\left(0; \frac{ns^2}{\chi_0}\right],$$

при чему је χ_0 одређено једначином (8.5).

За одређивање двостраног интервала поверења варијансе σ^2 користићемо, такође, статистику

$$\frac{ns^2}{\sigma^2}$$

и *Хи*-квадрат расподелу. Одредићемо два броја χ_1 и χ_2 тако да је

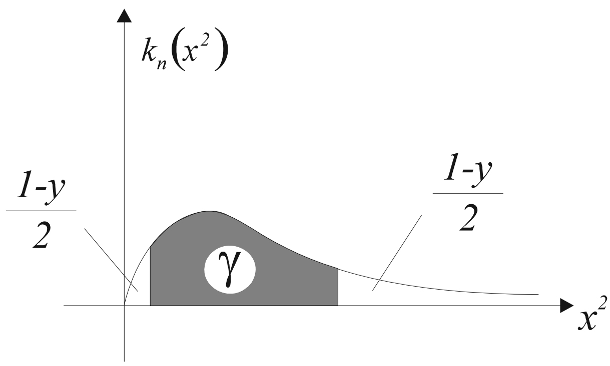
$$P\left\{\chi_1 \le \chi^2_{n-1} \le \chi_2\right\} = \gamma,$$

односно, тако да је

$$P\left\{\chi^2_{n-1} < \chi_1\right\} = \frac{1-\gamma}{2},$$

$$P\left\{\chi^2_{n-1} > \chi_1\right\} = \frac{1-\gamma}{2},$$

као што је приказано на Слици 8.1.



Слика 8.1. Хи-квадрат расподела.

Вредности χ_1 и χ_2 ћемо одредити из функција *Хи*-квадрат расподеле тако да је

$$K_{n-1}(\chi_1) = \frac{1-\gamma}{2}, \tag{8.6}$$

$$K_{n-1}(\chi_2) = \frac{1+\gamma}{2}.$$

Решавањем неједначина

$$\frac{ns^2}{\sigma^2} \le \chi_2,$$

$$\frac{ns^2}{\sigma^2} \ge \chi_1,$$

добићемо неједначину

$$\frac{ns^2}{\chi_2} \leq \sigma^2 \leq \frac{ns^2}{\chi_1},$$

односно, догађај чија вероватноћа је једнака нивоу значајности γ, па је

$$P\left\{\frac{ns^2}{\chi_2} \leq \sigma^2 \leq \frac{ns^2}{\chi_1}\right\} = \gamma,$$

а интервал поверења је интервал са границама

$$\left[\frac{ns^2}{\chi_2};\frac{ns^2}{\chi_1}\right],$$

при чему су χ_1 и χ_2 одређени једначинама (8.6).

На пример, за узорак од *n* елемената, 98% интервал поверења би био интервал поверења

$$\left[\frac{ns^2}{36.191};\frac{ns^2}{7.633}\right].$$

Ако је варијанса узорка $s^2 = 36$, оцењене границе тог интервала би биле

$$\frac{ns^2}{\chi_2} = 19.89 \text{ и } \frac{ns^2}{\chi_1} = 96.32.$$

Интервал поверења коефицијента корелације

Посматраћемо популацију са дводимензионалном Нормалном расподелом код које је непознат коефицијент корелације ρ. Да би одредили интервал поверења за ρ користићемо узорак $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ и коефицијент корелације узорка

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)\left(Y_i - \overline{y}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{y}\right)^2}},$$

који представља оцену коефицијента корелације у популацији.

Познато је да статистика

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r},$$

има приближно Нормалну расподелу са параметрима

$$E(Z) = \frac{1}{2}\ln\frac{1+\rho}{1-\rho},$$

$$Var(Z) = \frac{1}{n-3}.$$

Стандардизацијом статистике Z добићемо статистику која ће имати приближно стандардизовану Нормалну расподелу.

Из функције Нормалне расподеле одредимо z_0 тако да је

$$\phi(z_0) = \frac{1+\gamma}{2}.$$

Решавањем неједначина

$$\left\{\ln\frac{1+r}{1-r} - \ln\frac{1+\rho}{1-\rho}\right\}\cdot\frac{\sqrt{n-3}}{2} < z_0,$$

$$\left\{\ln\frac{1+r}{1-r} - \ln\frac{1+\rho}{1-\rho}\right\}\cdot\frac{\sqrt{n-3}}{3} > -z_0,$$

Добићемо неједначине

$$\frac{e^{2z_1}-1}{e^{2z_1}+1}<\rho<\frac{e^{2z_2}-1}{e^{2z_2}+1},$$

при чему је

$$z_1=\frac{1}{2}\ln\frac{1+r}{1-r}-\frac{z_0}{\sqrt{n-3}},$$

$$z_2=\frac{1}{2}\ln\frac{1+r}{1-r}+\frac{z_0}{\sqrt{n-3}}.$$

Интервал поверења (за ниво поверења γ) коефицијента корелације је интервал са границом

$$\left(\frac{e^{2z_1}-1}{e^{2z_1}+1};\frac{e^{2z_2}-1}{e^{2z_2}+1}\right)$$

На пример, за узорак од 19 елемената и са коефицијентом корелације $r=0.65$, добићемо

$$z_1=0.2853,$$

$$z_2=1.2653,$$

а вредност $z_0=1.96$, што одговара 95% нивоу поверења. Оцењени интервал поверења на основу овог узорка би имао границе

$$(0.2775;0.8524).$$

8.3. ЗАДАЦИ

1) Покажи да су обични моменти узорка непристрасне оцене обичних момената популације.
2) Одреди непристрасну оцену варијансе узорка.
3) За оцену варијансе популације са Нормалном расподелом $X : N(0;\sigma^2)$ користи статистику

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Провери непристраност ове оцене.

4) Одреди интервал поверења за разлику две очекиване вредности m_1 и m_2, на основу два независна узорка обима n_1 и n_2. Претпоставља се да обе популације имају Нормалну расподелу.
5) Случајно је изабрано 25 радника и посматрани су њихови трошкови за исхрану. Добијени су просечни трошкови од 2385 динара, а варијанса је била 309,06. Одреди 95% интервал поверења просечних издатака радника за храну.
6) Из популације са Нормалном расподелом узет је узорак од 200 мерења и добијени су резултати

$$\sum_{i=1}^{200} X_i = 460, \ \sum_{i=1}^{200} X_i^2 = 115.$$

Након одређеног времена узет је нови узорак од 300 мерења и добијени су резултати

$$\sum X_i = 690, \ \sum X_i^2 = 1695$$

За колико се променила дужина 95% интервала поверења непознате очекиване вредности популације у посматраном временском раздобљу?

7) Дужина ексера има Нормалну расподелу. Резултати мерења су следећи

11.8 9.2 9.8 10.6 9.9 10.3 8.6 10.4

Одреди 90% једнострани интервал поверења за варијансу дужине ексера.

8) Ради планирања месечне набавке, продавница је утврдила потражњу за једним производом у току 25 радних дана и добила следеће податке

Потражња	5-7	7-9	9-11	11-13	13-15
Број дана	3	6	7	6	3

Одреди 99% интервал поверења за просечну дневну потрошњу. Шта можете предложити о потребама дневне набавке посматраног производа?

9) Из популације са Нормалном расподелом, код које је варијанса једнака 225, треба одабрати узорак величине n и одредити 99% интервал поверења за очекивану вредност. Колики треба узети узорак, па да дужина интервала поверења не буде већа од 6.

10) Мерења способности, изражена карактеристиком X, дала су следеће резултате

$$\sum_{i=1}^{15} X_i = 0.7$$

$$\sum_{i=1}^{15} X_i^2 = 27.3$$

Одреди параметар који ће изражавати разлике у способностима, а затим оцени границе 80% интервала поверења за тај параметар.

11) Аутоматска машина обрађује полупроизвод. Код карактеристике X варијанса свих производа је 16. Колико треба извршити мерења да би дужина 95% интервала поверења за очекивану вредност карактеристике X била мања од 2.6?

12) На основу следећих података

$$\sum_{i=1}^{20} X_i = 104,$$

$$\sum_{i=1}^{20} X_i^2 = 686,$$

одређен је интервал поверења за варијансу са границама

$$(0;10.5862).$$

Колики је ниво поверења тог интервала?

9. ТЕСТИРАЊЕ ХИПОТЕЗА

При испитивању било каквих појава (друштвених или природних) користимо одређено искуство или одређене логичке закључке, које формулишемо у виду претпоставки о посматраном феномену. Истинитост претпоставки или хипотеза проверавамо на основу мерења, или на основу резултата мерења у једном „експерименту” којим објашњавамо посматрану појаву и доносимо закључак о истинитости или неистинитости хипотеза.

Поступак провере истинитости хипотеза (означавамо је са H_0) о посматраном феномену, представља типичан проблем Теорије одлучивања, а то је проблем утврђивања критеријума за доношење одлуке о прихватању хипотезе H_0 као истините, или о њеном одбацивању. Ако такву одлуку доносимо на основу статистичких података, онда тај поступак називамо **статистичко тестирање хипотеза**. Наравно, овде ћемо се бавити само таквом теоријом, тј. теоријом утврђивања критеријума доношења одлуке на основу низа измерених резултата у једном експерименту, односно на основу статистичких података.

9.1. ОПШТИ ПРОБЛЕМ ТЕОРИЈЕ СТАТИСТИЧКОГ ТЕСТИРАЊА ХИПОТЕЗА

Да би проверили претпоставку о ***правилном облику*** шестостране коцке, потребно је бацити коцку и регистровати фреквенције појављивања појединих страна коцке. На основу резултата мерења у n бацања, хипотезу H_0 (коцка је правилног облика) ћемо прихватити или одбацити.

Нова метода учења може бити боља од неке друге методе. Измерићемо резултате успеха студената који користе нову методу и студената који користе другу. Затим ћемо одредити просечан успех једних и других студената, па на основу разлике у просечном успеху, донети одлуку о тачности (нова метода је боља) или о нетачности H_0.

Предложени систем расподеле дохотка или стимулативна мера, повећаће продуктивност и квалитет производа. Примењујући тај систем или меру стимулације у више фирми, утврдићемо повећање или смањење продуктивности, повећање или смањење квалитета. На основу података прихватићемо или одбацити хипотезе о повећању продуктивности и о повећању квалитета, итд.

Једном речју, хипотезу H_0 ћемо прихватити или одбацити на основу резултата мерења. Сама хипотеза H_0 односиће се на популацију, резултати мерења ће бити резултати узорка, а поступак за проверу хипотезе на основу узорка, зваћемо тестирање хипотезе.

Тестирање хипотезе представља поступак за одлучивање да ли подаци у узорку иду у прилог хипотези H_0 или не. Постоје две класе тестова: ***ПАРАМЕТАРСКИ И НЕПАРАМЕТАРСКИ ТЕСТОВИ***.

Параметарски тестови

Код параметарских тестова се најчешће хипотеза односи на одређене параметре популације. Поступак за тестирање базиран је на некој или неким претпоставкама о расподели популације. На основу особина расподеле популације доносимо закључке о особинама узорака извучених из такве популације. Поређењем резултата у узорку са донетим закључком, доносимо одлуку о прихватању или одбацивању хипотезе. Такве хипотезе се називају *параметарске хипотезе*, а поступак тестирања таквих хипотеза се зове ***параметарски тест***.

Непараметарски тестови

Код непараметарских тестова се закључци о особинама узорака доносе без обзира на расподелу популације, а поређењем резултата у узорку и донетих закључака, доносимо одлуку о хипотези. Такав тест је ***непараметарски тест***.

Са друге стране, хипотеза се може односити на само једну могућу вредност параметра или на само једну тачно одређену расподелу. Таква хипотеза се зове проста хипотеза. На пример, хипотеза о правилној коцки је ***проста хипотеза*** (односи се на претпоставку да је вероватноћа појављивања сваке стране коцке једнака и износи 1/6). Са друге стране, хипотеза да број позива у телефонској централи има Пуасонову расподелу, представља ***сложену хипотезу.***

Претпоставка да је просечан успех студената Београдског универзитета исти као и Скопског универзитета, представља просту хипотезу. Али, претпоставка да је век трајања једног уређаја већи од T, представља сложену хипотезу.

Ако, на било који начин, одлуку о прихватању или одбацивању хипотезе H_0 доносимо на основу статистичких података, могуће су грешке два типа. ***Грешка првог типа*** *је грешка кад хипотезу H_0,* *одбацимо, а она је тачна.* ***Грешка другог типа*** *је грешка ако хипотезу H_0 прихватимо, а она није тачна.*

Главни поступак при тестирању хипотезе састоји у следећем. Посматра се цео скуп могућих вредности у узорку. Нека је то скуп R_n. (То су, у ствари, тачке у *n*-димензионалном простору, који одговара узорку обима *n*.) Скуп R_n поделимо на два дела:

- подскуп *C* (критична област),
- подскуп $R_n - C$.

Одлуку о одбацивању H_0 доносимо ако резултати узорка упадну у област *C*, тј. у критичну област. Ако резултати узорка упадну у другу област $R_n - C$, хипотезу H_0 не одбацујемо, тј. прихватамо да је H_0 тачна.

Кад је хипотеза H_0 тачна, резултати у узорку могу упасти у критичну област C, или не упасти у ту област. У првом случају ми ћемо донети погрешне закључке. Означимо са α вероватноћу да се то деси, тј.

$$\alpha = P\{(X_1, \ldots, X_n) \in C/H_0\},$$

при чему су $\left(X_1, \ldots, X_n\right)$ елементи узорка. Нећемо, у овом случају погрешити ако резултати узорка упадну у област $R_n - C$, а вероватноћа тачности тог закључка је једнака $1-\alpha$, тј.

$$1-\alpha = P\{(X_1, \ldots, X_n) \notin C/H_0\}.$$

У овом случају, погрешан закључак је грешка прве врсте, а α је њена вероватноћа и назива се ниво значајности.

Са друге стране, кад хипотеза H_0 није тачна, означимо то са $\overline{H_0}$, резултати узорка могу бити у области *C*. Тада ћемо, на основу општег критеријума донети исправан закључак, (тј. одбацићемо H_0).

Ако резултати узорка упадну у област $R_n - C$, донећемо погрешан закључак (јер ћемо тад прихватити H_0).

Вероватноћу овог погрешног закључка означимо са β. То ће бити вероватноћа

$$L = P\left\{\left(X_1, \ldots, X_n\right) \notin C/\overline{H}_0\right\},$$

а вероватноћа тачног закључка ће бити

$$1 - L = P\left\{\left(X_1, \ldots, X_n\right) \in C/\overline{H}_0\right\}.$$

Вероватноћа β представља вероватноћу грешке друге врсте, а $1-\beta$ представља моћ теста.

Ради прегледности, приказаћемо овај поступак тестирања хипотеза у следећој табели.

	ЗАКЉУЧАК	
	Ако $(x_1,\ldots,x_n)\in C$	**Ако** $(x_1,\ldots,x_n)\notin C$
ИСТИНА	H_0 **не прихватамо**	H_0 **прихватамо**
H_0 **тачна**	α	$1-\alpha$
H_0 **није тачна**	$1-\beta$	β

Табела 9.1. Табела одлучивања

При овом поступку одлучивања, критична област *C* утиче на вероватноће истинитости донесених закључака (одлука). Ако се мења област *C*, мењаће се и вероватноћа грешака прве и друге врсте. Циљ је да се, за унапред одређену вероватноћу грешке првог типа, тј. за унапред одређени ниво значајности, изабере област *C* која ће довести до најмање могуће вредности грешке другог типа β, тј. да се изабере она област C за коју ће моћ теста $(1-\beta)$ бити највећа.

Ради илустрације овог поступка, посматраћемо пример из контроле квалитета у индустријској производњи. Код великих серија, потребно је утврдити проценат лоших комада у серији. Ако је тај проценат мали (мањи од 5% или 10%, рецимо) тада би целу серију прихватили као добру. У противном би целу серију сматрали лошом.

Међутим, проценат шкарта у серији не знамо јер не можемо извршити сто посто контролу целе серије. Оно што можемо урадити то је да из серије изаберемо један узорак од n производа и да код тог узорка утврдимо проценат шкарта. На основу процента шкарта у узорку одлучићемо да целу серију прихватимо или одбацимо.

Поступак одлучивања је следећи: Ако проценат шкарта у узорку буде мањи од једног броја c, серију ћемо прихватити као добру. Међутим, у серије је могуће да проценат шкарта буде већи од дозвољеног процента. Морамо да одредимо вероватноћу погрешног пријема лоше серије. Та се вероватноћа назива ***ризик купца***.

Са друге стране, ако у узорку буде проценат шкарта већи од броја c, серију ћемо одбацити као лошу. Ипак, у серији је могуће да проценат шкарта буде мањи од дозвољеног, па је потребно одредити вероватноћу погрешног одбацивања добре серије. Та се вероватноћа зове ***ризик произвођача***. Број c утиче на оба ризика. Треба га одабрати тако да се постигне одређено „уједначавање” и смањење ова два ризика.

9.2. ПАРАМЕТАРСКИ ТЕСТОВИ

Кад посматрамо обележје X и његову расподелу у популацији, најчешће нас интересује неки параметар његове расподеле (просечна вредност, варијанса и сл.). Обично располажемо са неким информацијама о вредности тог параметра, а те информације можемо искористити за формулисање одређених претпоставки о параметру. Да би проверили те претпоставке, из популације ћемо узети узорак величине n и на основу вредности у узорку донети закључак о тачности или нетачности претпоставке.

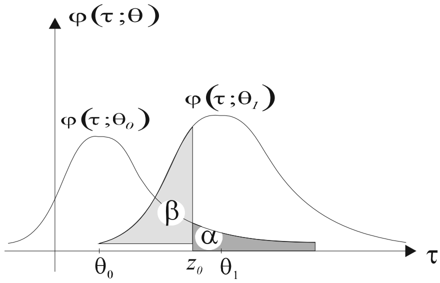
Означимо са θ непозанати параметар популације, а са $X_1, X_2, ..., X_n$ елементе узорка из посматране популације. Најпростији случај би био кад знамо да параметар θ може имати само две вредности, рецимо θ_0 и θ_1, $(\theta_0 < \theta_1)$. Тада хипотезе формулишемо на следећи начин:

- Нулта хипотеза $H_0(\theta = \theta_0)$;
- Алтернативна хипотеза $H_1(\theta = \theta_1)$.

У овом случају, обе хипотезе су просте хипотезе.

Поступак при одлучивању, односно при тестирању хипотезе H_0 састоји се из следећих корака:

1. Одреди се статистика $\tau = f(X_1, ..., X_n)$, дефинисана на узорку, тако да је њена расподела „концентрисана” око вредности θ_0 кад је хипотеза H_0 тачна. Истовремено би требала бити таква да је њена расподела „концентрисана” око вредности θ_1 кад је хипотеза H_1 тачна (Слика 9.1).



Слика 9.1. Расподеле статистике код хипотеза H_0 и H_1.

Означимо са $\varphi(\tau;\theta)$, закон вероватноћа статистике τ. То ће бити функција која ће зависити од вредности статистике τ и од непознатог параметра θ.

2. Могуће вредности статистике τ поделити на два дела избором броја z_0 (критична вредност), који ће бити између θ_0 и θ_1.

3. Критичну вредност z_0 одредити тако да је вероватноћа догађаја $\tau < z_0$, под условом да је $\theta = \theta_0$, једнака прагу значајности α. (Део површине испод криве $\varphi(\tau,\theta_0)$ на Слици 9.1).

4. Са слике се види да ћемо са избором броја z_0, одредити вероватноћу грешака прве и друге врсте. Очито, ако смањимо грешку прве врсте (повећањем z_0) повећаћемо грешку друге врсте, и обрнуто. Зато број z_0 одређујемо тако да α буде унапред одређени број. При томе је вероватноћа грешке друге врсте β једнака површини испод криве $\varphi(\tau,\theta_1)$, на Слици 9.1.

5. Одлуку о прихватању или одбацивању хипотезе H_0 донети на следећи начин: кад из узорка $X_1, X_2, ..., X_n$, вредност статистике τ буде мања од z_0, хипотезу H_0 прихватити, у противном хипотезу H_0 одбацити, тј.

   - ✓ ако је $\tau = f(x_1, x_2, ..., x_n) < z_0$, H_0 усвајамо;
   - ✓ ако је $\tau = f(x_1, x_2, ..., x_n) > z_0$, H_0 не усвајамо.

У овом случају, од свих статистика τ треба изабрати ону статистику која ће имати „најбољу концентрисаност” око θ_0 и θ_1. При томе под „најбољом концентрисаношћу” треба подразумевати следеће:

- ✓ очекивана вредност статистике τ је око вредности θ_0 (или једнака θ_0) кад је хипотеза H_0 тачна, а у противном је око θ_1 (или једнака θ_1);
- ✓ од свих статистика које испуњавају први услов, изабрати ону чија је варијанса најмања.

Пример. У складишту се налазе лежајеви са просечним полупречником 5mm или 10mm. Зна се да полупречник има Нормалну расподелу са варијансом једнаком $\sigma^2 = 4$. Ради утврђивања просечног полупречника свих лежајева, узима се узорак од 16 лежајева и мери њихов полупречник.

а) Одреди критичну вредност z_0 за ниво значајности 0.05 и 0.01.

б) Колика је вероватноћа грешке друге врсте?

в) Ако узорак повећамо на 64 лежаја, колике ће бити грешке прве и друге врсте?

Решење. Означимо са $X_1, X_2, ..., X_n$ вредности полупречника код 16 лежајева у узорку. Просечна вредност у узорку $\bar{x}$ имаће Нормалну расподелу са средином *m* и варијансом.

$$Var(\bar{x}) = \frac{\sigma^2}{n} = 1.$$

а) Хипотезе су следеће:

$$H_0(m = 5);\ H_1(m = 10).$$

Средина узорка има очекиване вредности:

$$E(\overline{x}) = 5 \text{ ако је } H_0 \text{ тачна;}$$

$$E(\overline{x}) = 10 \text{ ако је } H_1 \text{ тачна.}$$

У оба случаја је варијанса 1. За критичну вредност изабраћемо z_0 тако да је

$$P\left\{\overline{x} > z_0 \,\middle|\, m = 5\right\} = 0.05(0.01),$$

тј. тако да вероватноћа, да статистика $\overline{x}$ буде већа од z_0 (под условом да је H_0 тачно), има вредност 0,05 (0,01). Користећи функцију Нормалне расподеле, одредићемо z_0 тако да је

$$1 - \phi\left(\frac{z_0 - 5}{1}\right) = 0.05 \ (0.01),$$

односно, тако да је

$$\phi(z_0 - 5) = 0.95(0.99)$$

одакле је $z_0 = 6.65(7.58)$.

Као тест статистику, могли смо узети било коју линеарну функцију узорка

$$l = a_1 X_1 + a_2 X_2 + \ldots + a_{16} X_{16}.$$

Ако је $a_1 + a_2 + \ldots + a_{16} = 1$, очекиване вредности l су следеће:

$E(l) = 5$ кад је H_0 тачно,

$E(l) = 10$ кад је H_1 тачно.

Варијанса ове функције је

$$Var(l) = \sigma^2\left(a_1^2 + a_2^2 + \ldots + a_{16}^2\right)$$

Може се показати да ће ова варијанса имати најмању вредност ако је

$$a_1 = a_2 = \ldots = a_{16} = \frac{1}{16},$$

тј. кад је l средина узорка. Зато смо и узели средину узорка као најбољу статистику.

б) Вероватноћа грешке друге врсте једнака је

$$\beta = P\{\bar{x} < z_0 | \ m = 10\}.$$

Из функције Нормалне расподеле, добиће се да је

$$\beta = \phi\left(\frac{6,65 - 10}{1}\right) = 0.0004$$

в) Кад се повећа узорак на 64 лежаја, статистика $\bar{x}$ ће имати Нормалну расподелу са варијансом

$$Var(\bar{x}) = \frac{\sigma^2}{n} = \frac{1}{16},$$

па су грешке прве и друге врсте, за $z_0 = 6.65$ једнаке, 0.00004 и 0.000003.

Из овог примера се може закључити да ће се са повећањем узорка смањити вероватноће грешака и прве и друге врсте.

У пракси се ретко појављују овакви случајеви. Најчешће се посматра случај кад је нулта хипотеза проста хипотеза, а алтернативна хипотеза је сложена хипотеза.

Поступак за тестирање хипотезе H_0, кад је алтернативна хипотеза сложена, аналоган је претходном поступку, с тим што се у првом кораку изабере статистика τ, чији је закон вероватноћа функција $\varphi(\tau, \theta)$, па даље закључује као и у претходном случају.

Ипак, чешће ћемо бити у прилици да одредимо статистику облика

$$\tau = f(X_1,...,X_n;\theta),$$

чије вредности зависе од θ, а закон вероватноћа ће бити функција $\varphi(y)$ која неће зависити од параметра θ.

За хипотезу $H_0(\theta = \theta_0)$ одредићемо, из закона вероватноћа $\varphi(y)$, критичну област *C* као интервал, такав да је

$$P\{Y \in C\} = \alpha,$$

а на основу очекиваних вредности статистике τ, кад је хипотеза H_0 тачна и кад хипотеза H_0 није тачна. Даљи поступак у закључивању је исти као и у претходном случају.

Овде нећемо разматрати општији случај кад је нулта хипотеза сложена хипотеза. Посматраћемо поступке за тестирање хипотезе о оним параметрима који се најчешће користе, па кроз њих објаснити и неке елементе опште теорије тестирања хипотеза.

Хипотеза о очекиваној вредности

Прво ћемо посматрати случај кад је узорак $X_1, X_2, ..., X_n$ изабран из популације са Нормалном расподелом. Нека је очекивана вредност m непозанта и желимо да испитамо хипотезу да m има одређену вредност, рецимо m_0. Нулта хипотеза ће бити

$$H_0(m = m_0)$$

Да би тестирали хипотезу H_0, користићемо статистику

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

за коју знамо да има Нормалну расподелу са параметрима

$$E(\overline{x}) = m$$
$$Var(\overline{x}) = \frac{\sigma^2}{n}.$$

Закон вероватноћа ове статистике зависи од варијансе популације σ^2. Зато ћемо претпоставити да нам је варијанса σ^2 позната.

У том случају, статистика

$$\frac{\overline{x} - m}{\sigma}\sqrt{n}$$

има стандардизовану Нормалну расподелу $N(0;1)$. Ова статистика може узети било коју вредност из интервала $(-\infty;\infty)$, са законом вероватноћа

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{\overline{x} - m_0}{\sigma}\sqrt{n}$$

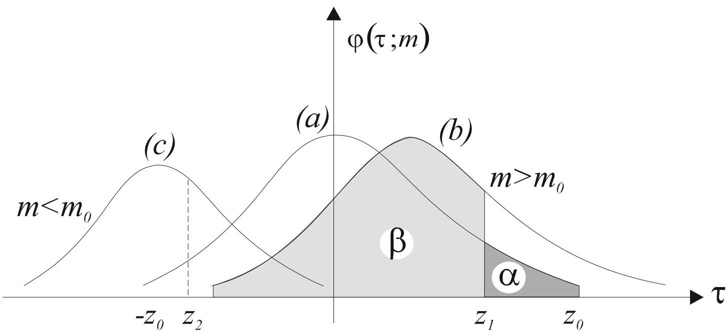
има стандардизовану Нормалну расподелу, а ако H_0 није тачна, ова статистика ће имати Нормалну расподелу са параметрима

$$E(\tau) = \frac{m - m_0}{\sigma}\sqrt{n}$$
$$Var(\tau) = 1.$$

Означимо очекивану вредност статистике τ, кад H_0 није тачна, са

$$d = \frac{m - m_0}{\sigma}\sqrt{n}.$$

Закон вероватноћа статистике τ кад је H_0 тачна, изгледаће као крива (а) на Слици 9.2.



Слика 9.2. Расподела статистике τ .

За одређивање критичне области C разликоваћемо следеће случајеве везане за алтернативну хипотезу.

а) Нека је алтернативна хипотеза

$$H_1(m > m_0).$$

Тада ћемо одредити вредност z_1 тако да, са вероватноћом α , тврдимо да ће статистика τ узети вредност већу од z_1, под условом да је H_0 тачна. То ће бити она вредност z_1 за коју је

$$P\{\tau > z_1 / m = m_0\} = \alpha,$$

односно она вредност за коју је

$$\phi(z_1) = 1 - \alpha . \tag{9.1}$$

Користећи таблице функције Нормалне расподеле одредићмо z_1, а поступак за одлучивање је следећи:

- ✓ ако је $\tau < z_1$, H_0 ћемо усвојити;
- ✓ ако је $\tau > z_1$, H_0 нећемо усвојити.

При томе је вероватноћа грешке друге врсте дата са

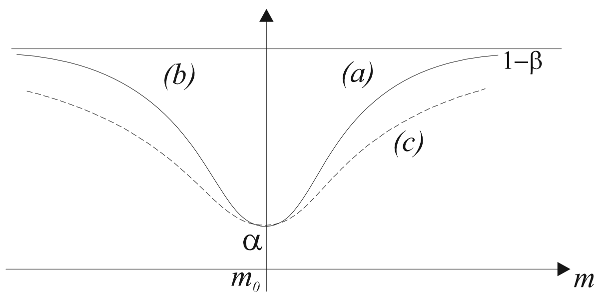
$$\beta = P\{\tau < z_1 / m > m_0\} = P\{\tau - d < z_1 - d / m > m_0\}$$
$$= \phi(z_1 - d) = \phi\left(z_1 - \frac{m - m_0}{\sigma}\sqrt{n}\right).$$

Као што се види из овог израза, β је функција вредности m, и то монотоно опадајуће функције. Види Слику 9.2.

Моћ теста је функција

$$1 - \beta = 1 - \phi(z_1 - d)$$

монотоно растућа за $m > m_0$. Минималну вредност, моћ теста има за $m = m_0$, и та је вредност α. То је крива (а) на Слици 9.3.



Слика 9.3. Моћ теста.

б) Нека је алтернативна хипотеза

$$H_1(m < m_0)$$

Тада ћемо одредити вредност z_2 тако да за задати ниво значајности α буде

$$P\{\tau < -z_2 |\ m = m_0\} = \alpha,$$

односно, тако да је

$$\phi(z_2) = 1 - \alpha \tag{9.2}$$

Одлуку доносимо на следећи начин:

- ✓ ако је $\tau < -z_2$, H_0 одбацујемо;
- ✓ ако је $\tau > -z_2$, H_0 не одбацујемо.

Моћ теста ће бити дата функцијом

$$1-\beta = 1-\phi(z_2 - d)$$

која представља монотоно опадајућу функцију са минималном вредношћу једнаком α. (На Слици 9.3. крива (b)).

в) На крају, нека је алтернативна хипотеза

$$H_1(m \neq m_0).$$

За ниво значајности *a*, изабраћемо вредност z_0 тако да је α вероватноћа да статистика

$$\tau = \frac{\overline{x} - m_0}{\sigma}\sqrt{n}$$

буде по апсолутној вредности, већа од z_0, тј. тако да је

$$P\{|\tau| > z_0 | \ m = m_0\} = \alpha,$$

односно, тако да је

$$\phi(z_0) = 1 - \frac{\alpha}{2} \tag{9.3}$$

Одлуку доносимо на следећи начин:

- ✓ ако је $|\tau| > z_0$, H_0 одбацујемо са нивоом значајности *a*;
- ✓ ако је $|\tau| < z_0$, H_0 не одбацујемо.

Моћ теста је функција

$$1-\beta = 1-P\{|\tau| < z_0 / m \neq m_0\} =$$
$$= 2-\phi(z_0 - d)-\phi(z_0 + d)$$

која има минималну вредност за $m = m_0$, а та вредност је једнака α. (Крива (c) на Слици 9.3.)

До сада смо претпоставили да нам је варијанса популације позната. Ако нам варијанса није позната (што је најчешћи случај) не можемо користити претходни тест. У овом случају корстимо тзв. Студентов *t*-тест. У ствари, ако популација има Нормалну расподелу, нулта хипотеза

$$H_0(m = m_0)$$

је сложена хипотеза, јер се односи и на сваку вредност варијансе популације. Тачније, то је хипотеза

$$H_0(m = m_0, \sigma^2 > 0).$$

За тестирање хипотезе H_0 користимо статистику

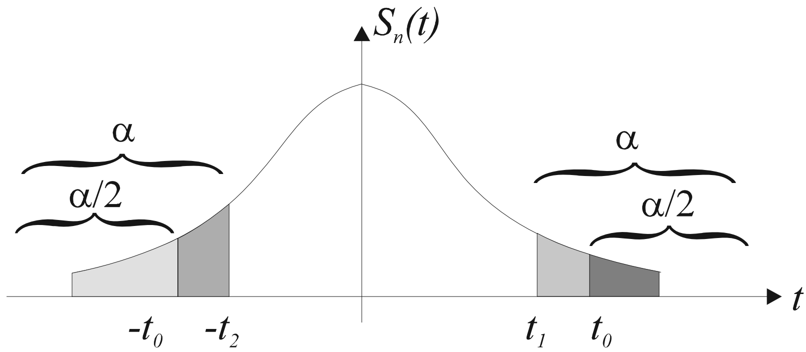
$$t_{n-1} = \frac{\bar{x} - m}{s}\sqrt{n-1}$$

за коју знамо да има Студентову *t*-расподелу са (*n-1*) степени слободе.

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{\bar{x} - m_0}{s}\sqrt{n-1} \tag{9.4}$$

има, такође, Студентову расподелу. У том случају, за задати ниво значајности α, из функције Студентове расподеле, можемо одредити критичну област. Избор области зависиће од алтернативне хипотезе. И овде ћемо посматрати три могућности:



Слика 9.4. Избор критичне области за Студентов t-тест.

а) Алтернативна хипотеза је

$$H_1(m > m_0)$$

Ако је тачна алтернативна хипотеза, статистика τ ће имати очекивану вредност позитивну (зависиће од разлике $m - m_0$), јер се статистика може написати у облику

$$\tau = \frac{\bar{x} - m}{s}\ \sqrt{n-1} + (m - m_0)\ \sqrt{n-1} \cdot \frac{1}{s}$$

Први члан овог израза има Студентову расподелу, па је његова очекивана вредност једнака нули. Пошто је s ненегативна статистика, онда је очекивана вредност

$$E\left(\frac{1}{s}\right) > 0.$$

Зато ћемо одредити број t_1 из функције Студентове расподеле тако да је

$$P\{t_{n-1} > t_1\} = \alpha,$$

односно, тако да је

$$S_{n-1}(t_1) = 1 - \alpha \qquad (9.5)$$

Одлуку о прихватању или одбацивању хипотезе H_0 доносимо на следећи начин:

- ✓ ако је $\tau > t_1$, H_0 одбацујемо;
- ✓ ако је $\tau < t_1$, H_0 не одбацујемо,

при чему је τ одређено из узорка на основу обрасца (9.4)

б) Алтернативна хипотеза је

$$H_1(m < m_0)$$

Кад је алтернативна хипотеза тачна, статистика ће имати негативну очекивану вредност. Из Студентове расподеле бира се t_2 тако да је

$$P\{t_{n-1} < -t_2\} = \alpha,$$

односно, тако да је

$$S_{n-1}(t_2) = 1 - \alpha \tag{9.6}$$

а одлуку доносимо на следећи начин:

- ✓ ако је $\tau < -t_2$, хипотезу H_0 одбацујемо;
- ✓ ако је $\tau > -t_2$, хипотезу H_0 не одбацујемо.

Наравно, статистика τ је одређена из узорка, преко обрасца (9.4).

в) Алтернативна хипотеза је

$$H_1(m \neq m_0)$$

Кад је тачна хипотеза H_1, очекивана вредност статистике τ је или негативна или позитивна, (зависно од разлике $m - m_0$). Зато ћемо овде одредити вредност t_0 из таблица функције Студентове расподеле, тако да је

$$P\left\{\left|t_{n-1}\right| > t_0\right\} = \alpha,$$

тј. тако да је

$$S_{n-1}(t_0) = 1 - \frac{\alpha}{2}$$

Одлуку о прихватању или одбацивању нулте хипотезе доносимо на следећи начин:

- ✓ ако је $|\tau| > t_0$, H_0 одбацујемо;
- ✓ ако је $|\tau| < t_0$, H_0 не одбацујемо.

Вредност статистике τ одређујемо из узорка, на основу обрасца (9.4), а хипотезу H_0 одбацујемо на нивоу значајности α.

На крају, посматрајмо узорак из било какве расподеле са непознатом очекиваном вредношћу и варијансом σ^2.

Ако је узорак довољно велики онда статистика

$$\frac{\bar{x} - m}{\sigma}\sqrt{n}$$

има приближно Нормалну расподелу. Зато можемо користити статистику

$$\tau = \frac{\bar{x} - m_0}{\sigma}\sqrt{n}$$

која има стандардизовану Нормалну расподелу под условом да је хипотеза H_0 тачна.

Избор критичне вредности и поступак одлучивања о прихватању или одбацивању нулте хипотезе $H_0(m = m_0)$, потпуно је аналоган поступку тестирања ове хипотезе кад је узорак из Нормалне расподеле, користећи тест стандардизоване Нормалне расподеле.

У случају да нам је варијанса популације непозната, користимо статистику

$$\tau = \frac{\overline{x} - m_0}{s}\sqrt{n-1}$$

која има приближно Нормалну расподелу.

Пример 1. Зна се да је дужина ексера у великој серији случајна променљива са Нормалном расподелом и стандардном девијацијом 0.7 mm. Прописана просечна дужина у серији је 80 mm. На случајан начин је изабрано 10 ексера и измерене су њихове дужине:

80 81 82 81 81 82 80 82 81 81.

а) Да ли се са нивоом значајности 0.05, може закључити да је у серији просечна дужина ексера била једнака прописаној?

б) Одреди функцију моћи теста и нађи њену вредност у случају кад је просечна вредност дужине порасла за две стандардне девијације.

Решење. Претпоставља се да је прописана и реализована просечна дужина једнака. Потребно је проверити хипотезу

$$H_0(m = 80),$$

а алтернативна хипотеза би била

$$H_1(m \neq 80),$$

где је са m означена просечна дужина ексера у серији.

а) На основу измерених вредности у узорку добићемо оцењену просечну дужину ексера.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} = 81.1.$$

Треба да проверимо да ли је разлика 81.1 и 80 значајна, тј. да ли се та разлика може објаснити као последица случајности узорка, или је та разлика последица одступања просечне дужине у серији од прописане дужине.

Статистика $\bar{x}$ има Нормалну расподелу. Ако је хипотеза H_0 тачна, онда је

$$E(\bar{x}) = 80,$$

$$Var(\bar{x}) = 4.9.$$

Ако је тачна алтернативна хипотеза, онда је

$$E(\bar{x}) = m \neq 80,$$

$$Var(\bar{x}) = 4.9.$$

Из таблица функције Нормалне расподеле одредићемо z_0 тако да је

$$P\{|Z| > z_0\} = \alpha,$$

тј. тако да је

$$\phi(z_0) = 1 - \frac{\alpha}{2} = 0.975$$

Вредност је $z_0 = 1{,}65$.

Ако је H_0 тачно, статистика

$$\tau = \frac{\bar{x} - m_0}{\sigma}\sqrt{n}$$

има стандардизовану Нормалну расподелу, па је

$$P\left\{\left|\frac{\bar{x} - m_0}{\sigma}\sqrt{n}\right| > z_0\right\} = \alpha,$$

односно,

$$P\left\{\left|\frac{\bar{x} - 80}{0.7}\sqrt{10}\right| > 1.65\right\} = 0.05.$$

Из узорка се добија вредност статистике

$$\tau = \frac{81.1 - 80}{0.7}\sqrt{10} = 4.8$$

Пошто је 4.8 > 1.65, закључујемо да хипотеза H_0 није тачна.

б) Ако није тачна хипотеза H_0, онда статистика

$$\tau = \frac{\overline{x} - m_0}{\sigma}\sqrt{n}$$

има Нормалну расподелу са параметрима

$$E(\tau) = \frac{m - m_0}{\sigma}\sqrt{n} = \frac{m - 80}{0.7}\sqrt{10}\ ,$$

$$Var\ \tau) = 1.$$

У том случају ћемо H_0 прихватити, ако је

$$|\tau| < z_0\,,$$

$$|\tau| < 1.65.$$

Наш закључак је погрешан, а вероватноћа те грешке је

$$\beta = P\{|\tau| < z_0 \,/\, m \neq m_0\}.$$

Моћ теста је функција

$$1 - \beta = 1 - P\{|\tau| < z_0 \,/\, m \neq 80\},$$

односно,

$$1 - \beta = 2 - \phi(1.65 - d) - \phi(1.65 + d),$$

при чему је

$$d = \frac{m-80}{0.7} \cdot \sqrt{10}\,.$$

Кад је $m = m_0 + 2\sigma = 81.4$, моћ теста ће имати вредност

$$1-\beta = 2-\phi\left(1.65-2\sqrt{10}\right)-\phi\left(1.65+2\sqrt{10}\right)=0...$$

Пример 2. Стандардна тежина једног лека је 1,2g по јединици. Због аутоматске производње могло је доћи до смањења у тежини. Провери ту могућност на основу измерених вредности:

1.19 1.20 1.19 1.23 1.18 1.21 1.27 1.17 1.15 1.14.

Решење. Претпоставићемо да је тежина случајна променљива са Нормалном расподелом. Потребно је проверити претпоставку

$$H_0\left(m=1.2\right),$$

а супротна претпоставка је

$$H_1\left(m<1.2\right).$$

Користићемо Студентов *t*-тест и посматрати статистику

$$\tau = \frac{\overline{x}-m}{s}\sqrt{n-1}.$$

Кад је хипотеза H_0 тачна, онда ова статистика има Студентову расподелу са 9 степени слободе. Ако је тачна супротна хипотеза, очекивана вредност ове статистике ће бити негативна јер је разлика $m-m_0$ негативна.

Зато ћемо одредити критичну вредност t_2 тако да је

$$P\left\{t_{n-1}<-t_2\right\}=\alpha,$$

односно, тако да је

$$S_9\left(t_2\right)=1-\alpha$$

За $\alpha = 0.05$, критична вредност је $t_2 = 1.83$. Из узорка ћемо одредити средину

$$\overline{x} = \frac{1}{n} \sum X_i = 1.193$$

и варијансу узорка

$$s^2 = \frac{1}{n} \sum \left(X_i - \overline{x} \right)^2 = 0.001 .$$

Посматрана статистика ће имати вредност

$$\tau = \frac{\overline{x} - m_0}{\sigma} \sqrt{n-1} = -0.64.$$

Пошто је -0.64>-1.83, немамо разлога да одбацимо хипотезу H_0. Можемо закључити да није дошло до смањења просечне тежине лека.

Пример 3. Просечни издаци за храну у београдским домаћинствима били су 10.000 динара. Да би проверили да ли је дошло до промене у просечним издацима домаћинства, на случајан начин је одабрано 50 домаћинстава. Добијени су следећи подаци:

Издаци за храну (у хиљадама)	-8	8-9	9-10	10-11	11-12	12-
Број домаћинстава	1	5	9	18	12	5

а) Изврши анализу добијених података.

б) Коју вредност ће имати Студентова *t*-статистика, ако користимо податке изражене у динарима?

Решење. Одредићемо просечну вредност издатака домаћинстава у узорку. То ће бити

$$\overline{x} = \frac{1}{50}\left(7.5 \cdot 1 + 8.5 \cdot 5 + \ldots + 12.5 \cdot 5\right) = 10.5 \text{ хиљада динара.}$$

а) Да ли је разлика 10.5 и 10 хиљада динара значајна?

Из таблица за функцију Студентове расподеле одредићемо t_0 тако да је

$$P\{|t_{49}| > t_0\} = \alpha .$$

Ако за α узмемо вредност 0,01, онда ћемо добити $t_0 = 2.6$.

Потребно је проверити хипотезу

$$H_0(m = 10),$$

а алтернативна хипотезе је

$$H_1(m \neq 10),$$

где смо са m означили просечне издатке свих београдских домаћинстава.

Ако је хипотеза H_0 тачна, онда статистика

$$\tau = \frac{\overline{x} - m_0}{s}\sqrt{n-1} = \frac{\overline{x} - 10}{s}\sqrt{49},$$

има Студентову расподелу. Зато је вероватноћа

$$P\{|\tau| > t_0 / m = 10\} = 0.001 .$$

Из узорка срачуната варијанса је $s^2 = 1180$, а статстика τ има вредност

$$\tau = \frac{10,5 - 10}{\sqrt{1180}}\sqrt{49} = 2.966.$$

Пошто је 2.966>2.6, прихватићемо хипотезу да је дошло до значајног пораста издатака за храну.

б) Кад вредности $X_1, X_2, ..., X_{50}$ у узорку, изражене у хиљадама, заменимо вредностима

$$Y_i = 1000\, X_i,$$

тада ће просечни издаци бити

$$\overline{y} = 1000 \cdot \overline{x},$$

а варијанса у узорку ће бити

$$s_y^2 = 1000^2 \cdot s_x^2$$

Вредност посматране статистике биће

$$\tau_y = \frac{\overline{y} - m_0(y)}{s_y}\sqrt{50} = \frac{\overline{x} - 10}{s}\sqrt{50}, \qquad (9.7)$$

што значи да ће Студентова статистика имати исту вредност као и раније.

Хипотеза о варијанси

Нека је X статистичко обележје или случајна променљива која на популацији има Нормалну расподелу. Потребно је да проверимо неке претпоставке о вредности варијансе у популацији. При томе коритстимо тзв. *Хи*-квадрат тест. Најчешће ће нас интересовати претпоставка да је варијанса већа од неке унапред одређене „жељене” вредности. Зато ћемо посматрати само следећи случај нулте хипотезе

$$H_0\left(\sigma^2 = \sigma_0^2\right),$$

и алтернативне хипотезе

$$H_1\left(\sigma^2 > \sigma_0^2\right).$$

Да би донели закључак о тачности или нетачности хипотезе H_0, посматраћемо узорак из популације $X_1, X_2, ..., X_n$.

Из узорка можемо оценити варијансу преко статистике

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)^2,$$

која представља варијансу узорка. Пошто је s^2 случајна промењљива, она може узети било коју вредност из интервала $(0;\infty)$, са одређеним законом вероватноћа.

Знамо да статистика

$$\frac{ns^2}{\sigma^2}$$

има *Хи*-квадрат расподелу са (*n-1*) степени слободе.

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{ns^2}{\sigma_0^2}$$

има Хи-квадрат расподелу са (n-1) степени слободе и очекиваном вредношћу једнаком (n-1).

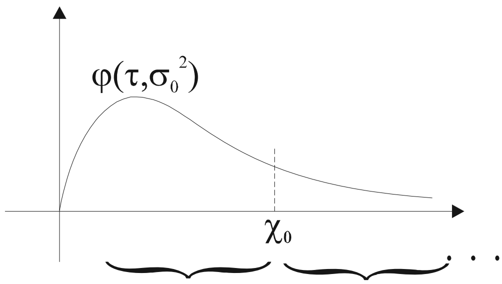
Ако хипотеза H_0 није тачна, тј. тачна је хипотеза да је $\sigma^2 > \sigma_0^2$, тада се ова статистика може написати у облику

$$\tau = \frac{ns^2}{\sigma^2}\frac{\sigma^2}{\sigma_0^2},$$

па ће њена очекивана вредност бити већа од (*n-1*), јер је

$$\frac{\sigma^2}{\sigma_0^2} > 1.$$

Зато ћемо одредити критичну вредност χ_0 тако да за вредности статистике веће од тог броја, одбацујемо H_0, у противном је усвајамо (Слика 9.5).



H_0 усвајамо H_0 не усвајамо

Слика 9.5. Избор критичне области код Хи-квадрат теста.

За задати праг значајности α одредићемо вредности χ_0, из таблица функције *Хи*-квадрат распделе тако да је

$$P\left\{X^2_{n-1} > \chi_0\right\} = \alpha,$$

тј. тако да је

$$K_{n-1}(\chi_0) = 1 - \alpha. \tag{9.8}$$

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{ns^2}{\sigma_0^2} \tag{9.9}$$

има *Хи*-квадрат расподелу, па је вероватноћа

$$P\left\{\tau > \chi_0 / \sigma^2 = \sigma_0^2\right\} = \alpha\,.$$

Одлуку о прихватању или одбацивању хипотезе H_0 доносимо на следећи начин:

- ✓ ако је $\tau > \chi_0$, хипотезу H_0 одбацујемо;
- ✓ ако је $\tau \le \chi_0$, хипотезу H_0 не одбацујемо,

при чему вредност статистике τ срачунамо из узрока, а на основу обрасца (9.9), а критичну вредност одређујемо из услова (9.8).

Вероватноћа грешке прве врсте је α, а вероватноћа грешке друге врсте је функција

$$\beta = P\{\tau < \chi_0 / \sigma^2 > \sigma_0^2\} =$$
$$= P\left\{\frac{ns^2}{\sigma^2} < \chi_0 \frac{\sigma_0^2}{\sigma^2}\right\}.$$

Моћ теста је функција

$$1 - \beta = 1 - K_{n-1}\left(x_0 \frac{\sigma_0^2}{\sigma^2}\right),$$

која представља монотоно растућу функцију чија је минимална вредност једнака α.

Пример. Резултати мерења случајне променљиве X у узорку су следећи:

40 60 60 70 50 40 50 30.

Ако је узорак из Нормалне расподеле, да ли се са нивоом значајности 0.05, може сматрати да је варијанса у популацији једнака 60.

Решење. Потребно је проверити следеће хипотезе

$$H_0(\sigma^2 = 60);$$
$$H_1(\sigma^2 > 60).$$

Из таблица за *Хи*-квадрат расподелу добија се да је $K_7(\chi_0) = 0.95$, за $\chi_0 = 14.067$.

Статистика

$$\tau = \frac{ns^2}{60}$$

има *Хи*-квадрат расподелу са 7 степени слободе (ако је хипотеза H_0 тачна). Зато је $P\{\tau > \chi_0 / \sigma^2 = 60\} = 0.05$

Из узорка се добија

$$\bar{x} = 50, \; s^2 = 150,$$

а вредност статистике је

$$\tau = \frac{ns^2}{\sigma_0^2} = 20.$$

Пошто је 20.000 > 14.067, хипотезу H_0 не можемо прихватити.

Хипотеза о вероватноћи или пропорцији *p*

У серијској производњи, готови или полупроизводи се контролишу и утврђује њихова исправност или неисправност. Проценат лоших комада у великој серији представља број *p* који нам је непознат. У ствари, *p* представља пропорцију лоших производа у серији.

Оваквој контроли одговара следећи статистички модел. Из серије се бирају производи и контролишу. Нека је *A* догађај да је изабрани производ лош, а $\overline{A}$ догађај да је производ добар. Тада је *p* проценат лоших у серији, односно вероватноћа догађаја *A*.

То је, у ствари, модел „0-1” расподеле са непознатим параметром *p*, при чему *X* узима вредност 1 ако се реализовао догађај *A*, а вредност 0 ако се није реализовао догађај *A*.

Да би се утврдила вредност *p*, посматра се узорак од *n* елемената (*n* изабраних производа или *n* пута поновљени експеримент). Нека су елементи узорка $X_1, X_2, ..., X_n$. При томе, свака од променљивих X_i узима вредност 0 или 1.

Статистика

$$Y = \sum_{i=1}^{n} X_i$$

представља број лоших производа у узорку, односно број понављања експеримента код којих се реализовао догађај A.

Средина узорка

$$\bar{x} = \frac{1}{n} Y = \hat{p}$$

представља оцену пропорције (вероватноће) p.

Желимо да проверимо претпоставку

$$H_0\left(p = p_0\right),$$

при чему је p_0 одређена, унапред дата вредност.

На основу граничне теореме Moivre-Laplace, зна се да ће статистика Y имати приближно Нормалну расподелу. Статистика $\bar{x} = \hat{p}$, имаће, такође, Нормалну расподелу са параметрима

$$E\left(\hat{p}\right) = p\ ,$$

$$Var\left(\hat{p}\right) = \frac{p\left(1 - p\right)}{n}.$$

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{\hat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)}} \sqrt{n},$$

има приближно стандардизовану Нормалну расподелу.

Поступак одлучивања о прихватању или одбацивању хипотезе H_0, аналоган је поступку тестирања хипотезе о очекиваној вредности. Из функције Нормалне расподеле за задани ниво значајности, одредимо критичну вредност, а у складу са алтернативном хипотезом. Ако је $H_1\left(p \neq p_0\right)$, одредимо z_0 тако да је

$$P\left\{\left|Z\right| > z_0\right\} = \alpha,$$

односно тако да је

$$\phi(z_0) = 1 - \frac{\alpha}{2} \tag{9.10}$$

Кад је хипотеза H_0 тачна, статистика

$$\tau = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)}}\sqrt{n} \tag{9.11}$$

има приближно стандардизовану Нормалну расподелу. Зато је вероватноћа

$$P\{|\tau| > z_0 \,/\, p = p_0\} = \alpha.$$

Одлуку доносимо на следећи начин

- ✓ ако је $|\tau| > z_0$, хипотезу H_0 одбацујемо;
- ✓ ако је $|\tau| < z_0$, хипотезу H_0 не одбацујемо.

При томе је z_0 одређено из услова (9.10), а статистику τ одређујемо из узорка преко обрасца (9.11).

Пример. Претпоставља се да проценат лица са одређеном крвном групом није већи од 8.28%. Ради провере те претпоставке изабрано је 400.000 лица, и утврђен је проценат 5.48%. Да ли је ова разлика значајна?

Решење. Потребно је проверити претпоставку

$$H_0\left(p = 0.0828\right),$$

а супротна претпоставка је

$$H_1\left(p < 0.0828\right).$$

За ниво значајности 0.01 одредићемо критичну вредност z_1, тако да је вероватноћа

$$P\{Z < -z_1\} = 0.01,$$

односно,

$$\phi(z_1) = 0.99.$$

Та вредност је $z_1 = 2.58$.

Ако је хипотеза H_0 тачна, онда је

$$P\{\tau < -2.58 / p = 0.0828\} = 0.01,$$

при чему је

$$\tau = \frac{\hat{p} - 0.0828}{\sqrt{0.0828 \cdot 0.9172}} \sqrt{400000}.$$

Ако није тачна хипотеза H_0, очекујемо да ће вредност ове статистике бити негативна $(p - p_0 < 0)$. Зато ћемо одлуку донети на следећи начин:

- ✓ ако је $\tau < -2.58$, хипотезу H_0 ћемо одбацити;
- ✓ ако је $\tau > -2.58$, хипотезу H_0 нећемо одбацити.

Из узорка од 400.000 лица добићемо $\hat{p} = 0.0548$, па је вредност статистике

$$\tau = -64.2.$$

Како је $\tau < -2.58$, хипотезу H_0 ћемо одбацити и донети закључак: проценат лица са посматраном крвном групом је мањи од 8.28%.

Хипотеза о параметрима два узорка

Често се у пракси срећемо са проблемима везаним за два независна узорка из неке популације. Параметри таквих узорака ће се међусобно разликовати. Потребно је објаснити те разлике. Оне могу бити последица тога што су параметри популација, које одговарају узорцима, различити, или последица случајности узорака. Посматраћемо таква два случаја.

1. Разлике просечних вредности. Испитивање новог лека, при лечењу једне болести, врши се тако што се на „контролној” групи болесника користи уобичајена терапија, а на другој групи се примењује нови лек. Поређењем разлика просечног ефекта у групама болесника, утврђује се ефикасност новог лека.

У индустрији је потребно поредити резултате две различите технологије. Измере се одређене карактеристике производа из обе технологије и пореде просечне вредности тих карактеристика.

У пољопривреди се испитују разлике у приносу код две сорте, или код две различите врсте ђубрива, и слично. Оваквим проблемима одговара следећи статистички модел:

Обележје X има Нормалну расподелу са параметрима m_1 и σ_1^2 на првој популацији, а на другој популацији параметри су m_2 и σ_2^2. Из прве популације узет је узорак од n_1 елемената.

$$X_1, X_2, \ldots, X_{n_1},$$

а из друге је узет узорак од n_2 елемената

$$X_1', X_2', \ldots, X_{n_2}'.$$

Потребно је, на основу ова два узорка, проверити хипотезу о једнакости очекиваних вредности, тј. хипотезу

$$H_0\left(m_1 = m_2\right).$$

Статистике $\bar{x}_1 \, u \, \bar{x}_2$, као средине узорака

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i-1}^{n_1} X_i ,$$

$$\bar{x}_2 = \frac{1}{n_2}\sum_{i-1}^{n_2} X_i^{'} ,$$

имаће Нормалне расподеле са параметрима

$$E(\bar{x}_1) = m_1$$

$$Var(\bar{x}_1) = \frac{\sigma_1^2}{n_1}$$

$$E(\bar{x}_2) = m_2$$

$$Var(\bar{x}_2) = \frac{\sigma_2^2}{n_2}.$$

Због независности узорака, статистика

$$\bar{x}_1 - \bar{x}_2$$

има Нормалну расподелу са параметрима

$$E(\bar{x}_1 - \bar{x}_2) = m_1 - m_2$$

$$Var(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Претпоставићемо да је $\sigma_1^2 = \sigma_2^2 = \sigma^2$ и да нам је варијанса σ^2 позната. Тада статистика

$$\tau = \frac{\bar{x}_1 - \bar{x}_2 - (m_1 - m_2)}{\sigma}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

има стандардизовану Нормалну расподелу. Из таблице за функцију Нормалне расподеле, одређујемо критичну вредност у складу са алтернативном хипотезом. Ако је алтернативна хипотеза

$$H_1(m_1 \neq m_2),$$

одредићемо z_0 тако да је

$$\phi(z_0) = 1 - \frac{\alpha}{2}.$$

Када је хипотеза H_0 тачна, статистика

$$\tau = \frac{\bar{x}_1 - \bar{x}_2}{\sigma} \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \tag{9.12}$$

има стандардизовану Нормалну расподелу, па је вероватноћа

$$P\{|\tau| > z_0 / m_1 = m_2\} = \alpha .$$

Одлуку доносимо на следећи начин:

- ✓ ако је $|\tau| > z_0$, хипотезу H_0 одбацујемо;
- ✓ ако је $|\tau| < z_0$, хипотезу H_0 не одбацујемо;

а вредност статистике τ срачунамо из узорака преко обрасца (9.12).

У случају када нам је варијанса популација непозната, користимо статистику

$$\tau = \frac{\bar{x}_1 - \bar{x}_2 - (m_1 - m_2)}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)},$$

која има Студентову *t*-расподелу са $(n_1 + n_2 - 2)$ степени слободе.

При томе смо са $s_1^2 \; u \; s_2^2$ означили варијансе узорака

$$s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} \left(X_i - \bar{x}_1\right)^2,$$

$$s_2^2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \left(X_i^{'} - \bar{x}_2\right)^2.$$

Из функције Студентове расподеле, а у складу са алтернативном хипотезом, одређујемо критичну вредност. За алтернативну хипотезу

$$H_1(m_1 \neq m_2),$$

критичну вредност t_0 одређујемо из услова.

$$S_{n_1+n_2-2}(t_0) = 1 - \frac{\alpha}{2}.$$

Ако је хипотеза H_0 тачна, статистика

$$\tau = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}(n_1 + n_2 - 2)} \qquad (9.13)$$

има Студентову расподелу, па је вероватноћа

$$P\{|\tau| > t_0 / m_1 = m_2\} = \alpha.$$

Одлуку доносимо на следећи начин:

- ✓ ако је $|\tau| > t_0$, хипотезу H_0 одбацујемо;
- ✓ ако је $|\tau| < t_0$, хипотезу H_0 не одбацујемо.

Вредност статистике τ срачунавамо из узорака користећи образац (9.13).

На крају, ако посматрамо два узорка великог обима, користићемо статистику

$$\tau = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Ако је хипотеза H_0 тачна, ова статистика има приближно стандардизовану Нормалну расподелу из које ћемо одредити критичну област као и у претходном случају (кад нам је варијанса популација позната). Поступак одбацивања или прихватања хипотезе H_0, биће аналоган поступку кад је варијанса позната.

Пример 1. Ради провере количине метала у руди, посматране су две зоне (*A* и *B*). Из зоне *A*, у 14 мерења добијени су резултати

$$\overline{x}_1 = 2.43\ kg/m^3,$$

$$s_1^2 = 15.07\ \left(kg/m^3\right)^2.$$

Из зоне *B*, у 10 мерења добијени су резултати

$$\overline{x}_2 = 4.90\ kg/m^3,$$

$$s_2^2 = 20.25\ \left(kg/m^3\right)^2.$$

Да ли се са нивоом значајности 0.05, може сматрати да је просечна количина у обе зоне иста?

Решење. Потребно је проверити хипотезу

$$H_0\left(m_1 = m_2\right),$$

где m_1 и m_2 представљају просечне количине метала у зонама *A* и *B*, респективно. Користићемо Студентов *t*-тест. Из таблица функције Студентове расподеле добија се да је $S_{22}\left(t_0\right) = 0.975$ за $t_0 = 2.07$.

Статистика

$$\tau = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{14\,s_1^2 + 10\,s_2^2}}\sqrt{\frac{14\cdot 10}{24}\cdot 22}$$

има Студентову расподелу, под условом да је H_0 тачно. Зато је

$$P\left\{\left|\tau\right| > 2.07 / m_1 = m_2\right\} = 0.05.$$

Вредност статистике из узорка је једнака $\tau = -1.36$. Пошто је $\left|\tau\right| = 1.36 < 2.07$, закључићемо да је хипотеза H_0 тачна, тј. закључићемо да су просечне количине у зонама једнаке.

Пример 2. На основу података о висини регрута у војсци САД, у годинама 1917 и 1943, може ли се закључити да нису настале промене у просечној висини свих Американаца са старошћу која је дозвољавала служење у војсци? Резултати су следећи

Година	Број војника	Просечна висина	Стандардна девијација
1917	868.445	67.49 инча	2.71 инча
1943	67.995	68.11 инча	2.59 инча

Решење. Нека је m_1 просечна висина свих Американаца војничке доби у 1917 години, а m_2 просечна висина за 1943 годину. Потребно је проверити хипотезу

$$H_0\left(m_1 = m_2\right).$$

Оба узорка су велика, па ћемо користити стандардизовани Нормални тест. Из података ћемо добити статистику

$$\tau = \frac{67.49 - 68.11}{\sqrt{\dfrac{2.71^2}{868445} + \dfrac{2.59^2}{67995}}} = -62.0\,.$$

Пошто је $\left|\tau\right| < z_0$ за сваку коначну вредност нивоа значајности a, хипотезу H_0 ћемо одбацити и закључити да је дошло до промене у просечној висини код ове две популације.

2. Количник варијанси. Стабилност процеса у индустрији се мери варијабилитетом одређених карактеристика. Поређењем варијанси карактеристика у одређеним периодима, утврђују се промене у уједначености процеса. Поређењем варијанси карактеристика код два технолошка процеса, утврђује се који процес је стабилнији.

За такве анализе користи се следећи статистички модел. Из популације са Нормалном расподелом узимају се два узорка

$$X_1, X_2, ..., X_{n_1},$$
$$X'_1, X'_2, ..., X'_{n_2},$$

из којих се одређују варијансе. Количник варијанси узорака

$$\frac{s_1^2}{s_2^2},$$

у општем случају, разликоваће се од јединице.

Потребно је проверити да ли су те разлике случајне, или су последица разлике у варијансама популација. Другим речима, потребно је проверити хипотезу

$$H_0\left(\sigma_1^2 = \sigma_2^2\right),$$

при чему смо са $\sigma_1^2 \; и \; \sigma_2^2$ означили варијансе популација из којих су узети узорци.

Одлуку о одбацивању или прихватању хипотезе H_0, доносимо користећи *F*-тест. Наиме, статистике

$$\frac{n_1 s_1^2}{\sigma_1^2} \; и \; \frac{n_2 s_2^2}{\sigma_2^2}$$

имају *Хи*-квадрат расподелу са $(n_1 - 1)$ и $(n_2 - 1)$ степени слободе. Због независности узорака, статистика

$$F = \frac{\dfrac{n_1 s_1^2}{\sigma_1^2(n_1 - 1)}}{\dfrac{n_2 s_2^2}{\sigma_2^2(n_2 - 1)}} = \frac{\sigma_2^2}{\sigma_1^2}\,\frac{n_1(n_2 - 1)}{n_2(n_1 - 1)}\,\frac{s_1^2}{s_2^2}$$

има *F*-расподелу sa $(n_1 - 1)$ и $(n_2 - 1)$ степени слободе.

Из таблице за *F*-расподелу, одређујемо критичне вредности у складу са алтернативном хипотезом. Ако је алтернативна хипотеза

$$H_1\left(\sigma_1^2 > \sigma_2^2\right),$$

онда одређујемо вредност F_0 тако да је

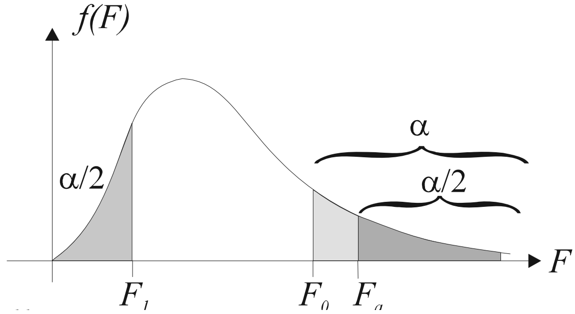
$$P\{F > F_0\} = \alpha. \tag{9.14}$$

Кад је хипотеза H_0 тачна, статистика

$$\tau = \frac{n_1(n_2-1)}{n_2(n_1-1)}\,\frac{s_1^2}{s_2^2} \tag{9.15}$$

има F-расподелу са (n_1-1) и (n_2-1) степени слободе. Зато је вероватноћа

$$P\{\tau > F_0 / \sigma_1^2 = \sigma_2^2\} = \alpha\ .$$



H_0 одбацити **H_0 одбацити**

Слика 9.6. Избор критичне области за количник варијанси.

Одлуку о хипотези H_0 доносимо на уобичајени начин:

- ✓ ако је $\tau > F_0$, хипотезу H_0 одбацујемо;
- ✓ ако је $\tau < F_0$, хипотезу H_0 не одбацујемо.

Вредности статистике τ израчунавамо из узорака, а на основу обрасца (9.15). Критичну вредност F_0 одређујемо из услова (9.14) преко таблица за функцију F-расподеле са (n_1-1) и (n_2-1) степени слободе. (Види Слику 9.6.).

У случају да је алтернативна хипотеза

$$H_1\left(\sigma_1^2 \neq \sigma_2^2\right),$$

одредимо критичне вредности F_1 и F_2 тако да је

$$P\{F < F_1\} = \frac{\alpha}{2}$$
$$P\{F > F_2\} = \frac{\alpha}{2}. \tag{9.16}$$

Одлуку доносимо по следећем поступку:

- ✓ ако је $\tau < F_1$ или $\tau > F_2$, хипотезу H_0 одбацујемо;
- ✓ ко је $F_1 < \tau < F_2$, хипотезу H_0 не одбацујемо.

Вредности статистике τ одређујемо по обрасцу (9.15), а критичне вредности из услова (9.16), користећи таблице функције *F*-расподеле са $(n_1 - 1)$ и $(n_2 - 1)$ степени слободе. (Види Слику 9.6.).

Пример. Ради поређења уједначености рада две машине, мери се карактеристика *X* готових производа са прве и друге машине. У 5 мерења код прве машине, добијена је варијанса $s_1^2 = 20$, а код 10 мерења производа са друге машине добијено је $s_2^2 = 50$.

Са прагом значајности 0,05, проверити да ли је рад прве машине стабилнији.

Решење. Потребно је проверити следеће претпоставке.

$$H_0\left(\sigma_1^2 = \sigma_2^2\right),$$
$$H_1\left(\sigma_1^2 > \sigma_2^2\right).$$

Критична вредност је F_0, одређена тако да је

$$P\{F < F_0\} = 0.05,$$

а то је вредност $F_0 = 3.63$, па је

$$P\left\{\tau > 3.63 / \sigma_1^2 = \sigma_2^2\right\} = 0.05.$$

Статистика има вредност

$$\tau = \frac{5\left(10-1\right)}{10\left(5-1\right)}\ \frac{s_1^2}{s_2^2} = 0.45.$$

Пошто је 0,45<3,63, нема разлога да одбацимо хипотезу H_0. Другим речима, прихватићемо одлуку да је рад машина подједнако стабилан.

9.3. ЗАДАЦИ

1. Одреди моћ Студентовог *t*-теста за проверу хипотезе о математичком очекивању популације.
2. Број часова исправног рада машине има Нормалну расподелу са стандардном девијацијом од 2 сата. Између 8 отказа машине, број сати рада без квара био је следећи

   53 48 50 54 51 50 51.

   Са прагом значајности 0.01 проверити претпоставку да је просечан број сати рада машине између кварова једнак 50 сати.
3. У тексту од 100 случајно изабраних страница једне књиге, број грешака је био следећи:

Број грешака	Број страница
0-10	60
10-20	20
20-30	10
30-40	5
40-	5

   Да ли је просечан број грешака по страници једнак 8, или је већи од 8?
4. При аутоматској обради цилиндра његов полупречник је случајна променљива са Нормалном расподелом. У једном дану измерени су полупречници 10 цилиндара и добијени су следећи подаци:
   - Збир измерених полупречника је 50cm;
   - Збир квадрата измерених полупречника је 400cm^2.

Другог дана, на основу 15 мерења, добијени су следећи резултати:

- Збир измерених полупречника је 100cm;
- Збир квадрата измерених полупречника је 525cm^2.

а) Са прагом значајности 0,01 провери претпоставке да је у оба дана мерења просечни полупречник свих цилиндара једнак 20cm.

б) На основу оба узорка провери претпоставку под а).

в) Да ли су разлике просечних вредности по данима значајне?

5. Мери се издржљивост синтетичког влакна пре и после бојења. Добијени су следећи подаци

Издржљивост пре бојења	4.0	3.5	4.1	5.5	4.6	6.0	5.1	4.3
Издржљивост после бојења	3.0	3.0	3.8	2.1	4.9	5.3	3.1	2.7
Број мерења	1	1	2	10	8	1	1	6

а) Испитај да ли је просечна издржљивост материјала већа пре бојења од просечне издржљивости после бојења.

б) Оцени варијансу разлике издржљивости материјала пре и после бојења.

в) Да ли је стандардна девијација разлике једнака један или је већа од један?

6. Проверава се просечна висина једне мање популације. Претпоставља се да је 165cm. Ради провере, случајно је изабрано 16 особа и добијене су њихове висине

140 150 160 170 180 185 175 170
171 168 150 154 182 180 165 164.

Да ли се може прихватити претпоставка о просечној висини популације од 165cm?

7. У пољопривредном комбинату се испитује утицај новог ђубрива на принос. На 12 парцела употребљено је ново ђубриво, и добијен је просечни принос од 5.22m-центи, са стандардном девијацијом од 0.36. На 10 контролних парцела није употребљено ново ђубриво, а добијен је просечни принос од 4.80m-центи, са стандардном девијацијом од 0.40. Да ли се, са 5% нивоа значајности, може сматрати да је просечан принос већи при употреби новог ђубрива?

8. У три независна узорка добијени су следећи резултати

Узорак	Величина узорка	Средина узорка	Варијанса узорка
1	25	21	17
2	9	22	19
3	16	23	18

Код ког се узорка може прихватити претпоставка да је очекивана вредност у популацији једнака 20? За праг значајности користити 0.0024.

9. Психолог испитује способност код две групе ученика. У првој групи резултати теста су следећи

100 95 90 75,

а у другој групи

90 85 80 80 75 70.

а) Да ли се може сматрати да је прва група „способнија” од друге?
б) С обзиром на мали број података, коментариши закључак под а).

10. Стандардном производњом предвиђено је 10% шкарта. У 100 мерења добијен је проценат шкарта од 15.2%. Да ли треба прихватити хипотезу да се производња креће у оквиру стандарда, или треба производњу сматрати као лошу (изван оквира стандарда)?

11. Проценат шкарта на две машине је једнак p_1 и p_2. Ради провере претпоставке о једнакости процента шкарта на обе

машине, користе се два већа узорка (већа од 30). Нека су $\hat{p}_1$ и $\hat{p}_2$ оцене p_1 и p_2 на основу узорка величине n_1 и n_2.

а) Одреди очекивану вредност и варијансу статистике $\hat{p}_1 - \hat{p}_2$.

б) Одреди, приближно, расподелу за статистику

$$\tau = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{Var(\hat{p}_1 - \hat{p}_2)}}$$

в) Формулиши тест за тестирање хипотезе

$$H_0(p_1 = p_2).$$

12. Нека је у експерименталној серији било 4 неисправна производа од 100 контролисаних, а у првој партији производње од 500 контролисаних производа, било је 12 неисправних. Објасни разлике у експерименталној и пробној серији, нивоом значајности 0.01.

13. На основу података из задатка 5, оцени коефицијент корелације између издржљивости материјала пре и после бојења, а затим, са прагом значајности 0.002, провери независност ова два обележја.

14. Ради провере зависности обележја X и Y, треба изабрати узорак од n елемената. Ако је ниво значајности 0.01, а доња граница коефицијента корелације за одбацивање хипотезе

$$H_0(\rho = 0)$$

једнака 0.50, колики треба да је узорак? Колики ће бити узорак ако је праг значајности 0.05.

10. НЕПАРАМЕТАРСКИ СТАТИСТИЧКИ ТЕСТОВИ

Параметарски статистички тест је поступак провере хипотезе који је базиран на одређеним претпоставкама везаним за параметре и расподеле популације из које је посматрани узорак извучен. Обично се исправност (тачност) ових претпоставки не проверава, па зато закључци параметарских тестова могу бити и погрешни. Параметарски тестови такође захтевају да резултати, добијени "мерењем" у узорку, буду изражени бар, преко интервалне скале.

Непараметарски статистички тест је тест базиран на моделу који не укључује никакве предуслове у вези параметара популације из које је узорак извучен. Одређене претпоставке су карактеристичне за већину непараметарских статистичких тестова, као што је на пример: да су опсервације независне, да је посматрана променљива непрекидна итд. Али оваквих претпоставки је мање и слабије су него оне код параметарских тестова. Шта више, непараметарски тестови не захтевају тако прецизна "мерења" као параметарски тестови. Већина непараметарских тестова је ваљана са подацима у ординалној скали, а неки и са подацима из номиналне скале.

Због тога што се непараметарским тестовима може повећати моћ једноставним повећањем n (величина узорка), и због тога што научници често не могу да спроведу мерења која омогућавају дословно коришћење параметарских тестова, непараметарским статистичким тестовима припада значајна улога у наукама које користе статистику.

Предности непараметарских тестова

1) Искази о вероватноћи, добијени из већине непараметарских тестова су тачне вероватноће (осим у случајевима великих узорака, где се користе одличне апроксимације), без обзира на облик расподеле популације из које је извучен случајни узорак. У извесним случајевима, у непараметарским тестовима се претпоставља непрекидност расподеле, али то је и претпоставка параметарских тестова.

2) Ако је величина узорка мала, као на пример $n = 6$, онда нема алтернативе у коришћењу непараметарских статистичких тестова, сем у случају кад је расподела популације прецизно дефинисана.

3) Постоји неколико подобних непараметарских статистичких тестова за узорке добијене из опсервација неколико различитих популација. Ниједан од параметарских тестова се не може употребити за овакве податке без увођења неких нереалних претпоставки.

4) Непараметарски статистички тестови су погодни за податке који су по својој природи рангирани, као и за податке чије очигледне нумеричке вредности имају снагу ранга. Другим речима истраживач може да каже који од посматраних објеката има вишу или мању посматрану карактеристику, без потребе да прецизира колико мању или вишу. На пример, ако проучавамо нестрпљивост, можемо да кажемо да је A нестрпљивији од B, а да незнамо прецизно колико је A нестрпљивији. Ако су подаци по својој природи рангирани или, чак ако се само могу и категоризовати као плус или минус (више или мање, боље или лошије), на њих се могу применити непараметарске методе. Иначе на такве податке се не може применити ниједна параметарска метода ако се не уведе нека нереалистична претпоставка о њиховој расподели.

5) Непараметарске методе се могу применити и на податке у номиналној скали. Ни једна параметарска техника се не може применити на такве податке.
6) Непараметарски статистички тестови су обично много лакши за учење и примену од параметарских.

Недостаци непараметарских тестова

1) Ако су све претпоставке параметарског статистичког модела испуњене у подацима онда непараметарски статистички тестови често губе информације из података. Степен губитака је изражен преко моћи (ефикасности) непараметарског теста. Ако је, рецимо, моћ непараметарског теста 90%, то значи да тамо где су испуњени сви услови параметарског теста, одговарајући параметарски тест би био подједнако ефикасан као и непараметарски, али са 10% мањим узорком.
2) Још увек не постоји непараметарски метод за тестирање интеракција у моделу анализе варијансе, сем ако се не уведу специјалне претпоставке о адитивности. Овај недостатак би можда требало и занемарити, јер параметарски статистички тестови такође захтевају такве претпоставке. Ипак, са проблемом интеракција вишег реда, се тек треба сусрести у литератури о непараметарским методама.

10.1. НЕПАРАМЕТАРСКИ ТЕСТОВИ ЗА ЈЕДАН УЗОРАК

У овом поглављу обрађени су непараметарски статистички тестови који се користе за тестирање хипотеза на основу "извлачења" само једног узорка. Ови тестови нам говоре да ли је одређени узорак могао доћи из неке познате популације, за разлику од тестова базираних на два узорка, који пореде два узорка и испитују да ли је вероватно да су та два узорка дошла из исте популације.

У тестовима једног узорка извлачимо случајан узорак, а онда тестирамо хипотезу да је тај узорак извучен из популације са одређеном расподелом. Тако тестови једног узорка могу да одговоре на питања као што су:

- да ли постоји значајна разлика у локацији (централна тенденција) између узорка и популације?
- да ли постоји значајна разлика у посматраним фреквенцијама и фреквенцијама које бисмо могли очекивати на бази неког принципа?
- постоји ли значајна разлика између посматраних и очекиваних пропорција?
- да ли је разумно веровати да је овај узорак извучен из популације одређеног облика или форме (као нормална, униформна...)?

У случају једног узорка уобичајена параметарска техника је да се примени *t* тест на разлике између посматране (узорак) и очекиване (популација) средине (енгл. ***mean***). Стриктно говорећи, *t* тест предпоставља да су обсервације у узорку дошле из популације са

Нормалном расподелом. Поред тога, t тест такође захтева да опсервације буду мерене бар у интервалној скали.

Постоји много врста података на које се t тест не може применити. Истраживач може утврдити следеће:

- Претпоставке и захтеви t теста су нереалистични за његове податке;
- Згодније је избећи увођење претпоставки t теста и тако добити генералнији закључак;
- Подаци његовог истраживања по природи су рангирани, па због тога незгодни за анализу t тестом;
- Подаци су једноставно квалитативни или не нумерички, па су због тога незгодни за t тест;
- Заинтересован је за разлике било које врсте, а не само за разлику у локацији.

У тим случајевима истраживач се може одлучити за неки од непараметарских тестова једног узорка који су описани у овом поглављу.

χ^2-тест (Хи–квадрат тест)

Тест сагласности

Проблем сагласности теоријских резултата везаних за вероватноћу и стварних опсервација решава се тестовима сагласности (***Goodness of Fit***). Наиме, у реалности посматрамо узорак од n опсервација вредности неке варијабле или статистичког обележја (било које димензије) и треба да знамо да ли се та варијабла може посматрати као случајна промењљива која има дату расподелу.

Означимо са H_o хипотезу да су подаци о n вредности у узорку генерисани случајном промењљивом са потпуно одређеном расподелом $P(S)$ преко које можемо одредити вероватноћу било ког скупа S.

Потребно је обезбедити метод за проверу да ли су наши подаци конзистентни са хипотезом H_o.

Ако је хипотеза H_o тачна, расподела узорка се може посматрати као *статистичка слика* произашла из расподеле одређене са $P(S)$. Обзиром на утицај случајности, ове две расподеле неће бити исте, али за велико *n*, очекујемо да ће расподела узорка бити добра апроксимација расподеле популације. Зато је природно да уведемо *меру девијације* између ове две расподеле, па на основу њене расподеле формирати одговарајући тест.

Мера девијације или разлика међу овим расподелама, може бити конструисана на разне начине. Најчешће коришћена мера је она која је везана за *Хи*-квадрат тест који је увео K. Pearson.

Идеја је једноставна и прихватљива: Скуп могућих вредности променљиве подели се на коначан број подскупова и посматра се разлика између очекиваних фреквенција из расподеле $P(S)$ ***и реализованих у узорку, па се на основу њих формира*** χ^2 ***статистика.***

Означимо са $S_1, S_2, ..., S_k$ подскупова који немају заједничких тачака и који прекривају скуп могућих вредности случајне променљиве. Нека су $p_1, p_2, ..., p_k$ одговарајуће вероватноће добијене из расподеле $P(S)$, тј

$$p_i = P(S_i), \qquad i = 1,2,...,k.$$

Такође предпостављамо да су ове вероватноће *веће од нуле*, а њихов *збир је једнак јединици.*

Означимо са $m_1, m_2, ..., m_k$ одговарајуће фреквенције у узорку, при чему је m_i = (број елемената узорка чија вредност упада у скуп S_i), за $i = 1,2,...,k$. Јасно је да је

$$\sum_{i=1}^{k} m_i = n \tag{10.1}$$

Кад је хипотеза H_o тачна, статистике $m_1, m_2, ..., m_k$ имају Биномне расподеле:

$$m_i : B(n; p_i), \qquad i = 1,2,...,k.$$

са очекиваним вреностима

$$E(m_i) = np_i, \qquad i = 1,2,...,k.$$

K. Pearson посматра разлике између реализованих вредности статистика $m_1, m_2, ..., m_k$ и њихових очекиваних вредности $np_1, np_2, ..., np_k$ и формира статистику

$$\chi^2 = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} \tag{10.2}$$

Кад величина узорка расте, *расподела статистике* χ^2 *тежи Хи-квадрат расподели са* $(k-1)$ *степени слободе* потпуно независно од облика расподеле популације дате са $P(S)$.

На основу ових резултата можемо утврдити поступак за примену *Хи*-квадрат теста као теста сагласности:

1) *Нулта хипотеза.* $H_o : X$ има дату расподелу $P(S)$.

   Алтернативна хипотеза. $H_1 : X$ нема расподелу $P(S)$.

2) Из $P(S)$ одређујемо $p_1, p_2, ..., p_k$, а из узорка $m_1, m_2, ..., m_k$.
3) Одређујемо вредност статистике χ^2.
4) Утврђујемо критичну област. Из *Хи*-квадрат расподеле одређујући χ^2 тако да је

$$P\left\{\chi_{k-1}^{2} > \chi_0\right\} = \alpha$$

5) Доносимо одлуку:
   - ✓ ако је $\chi^2 > \chi_0$ одбацујемо H_o;
   - ✓ ако је $\chi^2 < \chi_0$ не одбацујемо H_o.

Пример: На основу 50 независних опсервација добијени су следећи подаци

X	0	1	2	3	4	5 и више
f	8	10	11	11	5	5

Да ли се може сматрати да су подаци добијени из Пуасонове расподеле $P(2.36)$?

Решење:

X	m_i	np_i	$(m_i - np_i)^2/np_i$
0	8	4.85	2.05
1	10	11.20	0.13
2	11	13.05	0.32
3	11	10.20	0.06
4	5	6.05	0.18
5 и више	5	4.65	0.03

$$\chi^2 = 2.77$$

Из таблица за Хи-квадрат расподелу са (6-1) степени слободе, проналазимо $\chi_0 = 11.070$, $z = \alpha = 0.05$. Хипотезу H_o не треба одбацити.

Колмогоров-Смирнов тест

Тест Колмогоров-Смирнов за један узорак спада у категорију тестова сагласности. Заснован је на утврђивања степена слагања између расподеле вредности узорка (добијени резултати) и неке одређене теоријску расподеле. Овај тест одређује да ли је разумно веровати да је узорак дошао из популације која има претпостављену теоријску расподелу.

Укратко, тест укључује одређивање кумулативне расподеле фреквенција која би се појавила под претпостављеном теоријском расподелом и поређење те расподеле са добијеном (измереном-узоркованом) кумулативном расподелом фреквенција. Теоријска расподела представља оно шта треба очекивати кад је хипотеза H_o тачна. Затим се одређује тачка у којој ове две расподеле, теоријска и добијена, показују највећу разлику (дивергенцију). Расподелом узорка се показује да ли је тако велика разлика могућа случајно. То јест, расподела узорка показује да ли је толика разлика могућа, ако су опсервације заиста случајни узорак из претпостављене теоријске расподеле.

Нека је $F_o(X)$ потпуно одређена функција расподеле, кад је хипотеза H_o тачна. То јест, за било коју вредност *X*, вредност $F_o(X)$ је пропорција случајева код којих се очекује да имају вредност једнаку или мању од *X*, кад је хипотеза H_o тачна.

Нека је $S_n(X)$ добијена (измерена) функција расподеле из случајног узорка са *n* обсервација. Често се $S_n(X)$ ***назива емпиријска функција расподеле.*** Овде је *X* било која вредност,

$$S_n(X) = k / n\,,$$

где је *k* број опсервација једнаких или мањих од *X*.

Кад је нулта хипотеза тачна, тј. кад је узорак извучен из одређене теоријске расподеле, очекује се да за сваку вредност *X*, $S_n(X)$ треба да буде врло блиска $F_o(X)$. То јест, очекујемо да разлика између $S_n(X)$ и $F_o(X)$ буде мала и да се креће у границама случајних грешака. Колмогоров-Смирнов тест се фокусира на највећу девијацију. Највећа вредност $\left|F_0(X) - S_n(X)\right|$ се зове максимална девијација, *D*:

$$D = \max\left|F_0(X) - S_n(X)\right| \tag{10.3}$$

Расподела статистике *D* је позната, а критичне вредности за *D* се могу пронаћи у табелама критичних вредности за тест Колмогоров-Смирнов за један узорак. Ове табеле постоје у уџбеницима, а имплементиране су и у статистички пакет SPSS. На основу тих критичних вредности одређујемо да ли је D значајно или не.

Пример 1: Претпоставимо да је задатак истраживача у једној фабрици дечијих играчака, да открије да ли деца више воле лопте светлијих или тамнијих тонова. Тест се спроводи на популацији од 10 малишана и свакоме од њих се доноси пет лопти обележених бројевима од 1-5. Лопте се разликују по боји и то тако да је бројем 1 означена лопта на којој преовлађују најтамнији тонови, а бројем 5 лопта са најсветлијим тоновима.

1) *Нулта хипотеза*. H_o: Не постоји разлика у очекиваном броју избора свих пет тоналитета и свака разлика која се појави је случајна варијација у случајном узорку из униформне популације где је $f_1 = f_2 = ... = f_5$.

   Алтернативна хипотеза. H_1 фреквенције $f_1 = f_2 = ... = f_5$ нису све једнаке.

2) *Статистички тест*. Колмогорнов-Смирнов тест за један узорак је одабран јер истраживач жели да упореди добијену расподелу података на ординалној скали са теоријском (претпостављеном) расподелом.

3) *Ниво значајности*. Нека је $\alpha = 0.01$, а n је број малишана који су испитани у студији и износи 10.

4) *Расподела статистике*. Различите критичне вредности за D, из расподеле статистике D, се налазе у табели за Колмогоров-Смирнов тест за један узорак, заједно са придруженим, одговарајућим вероватноћама појављивања.

5) *Област одбацивања*. Област одбацивања се састоји од свих вредности D, израчунатих по формули (10.3), које су довољно велике да су вероватноће, придружене са њиховим појављивањем кад је хипотеза H_o тачна, једнаке или мање од $\alpha = 0.01$.

6) *Одлука*. У овој хипотетичкој студији сваки малишан одабира једну од пет лопти. Претпоставимо да је један од њих одабрао лопту означену бројем 2, петоро њих лопту са бројем 4 и четворо њих лопту са бројем 5. Следећа табела приказује ове податке и прилагођава их за примену Колмогоров-Смирнов теста за један узорак.

	Ранг лопти за бирање				
	I	II	III	IV	V
f	0	1	0	5	4
$F_o(X)$ $S_{10}(X)$	1/5 0/10	2/5 1/10	3/5 1/10	4/5 6/10	5/5 10/10
$\lvert F_0(X)-S_{10}(X)\rvert$	2/10	3/10	5/10	2/10	0

Приметимо да је $F_o(X)$ теоријска кумулативна расподела кад је хипотеза H_o тачна (где је H_o хипотеза да ће свака од пет лопти имати по 1/5 избора). $S_{10}(X)$ је кумулативна расподела посматраних избора малишана. Последњи ред у табели даје апсолутне девијације сваке вредности у узорку од њене очекиване вредности. Ако проверимо последњи ред у табели, лако се утврђује да је *D,* за ове податке 5/10, што је 0.500. Из табела за Колмогоров-Смирнов тест једног узорка, може се видети да је за $n = 10$, придружена вероватноћа *p* да буде D≥0.500, таква да је $p < 0.01$. Пошто је *p* придружено добијеној вредности D, мање или једнако од $\alpha = 0.01$, наша одлука за ову конкретну студију је да се H_o одбаци у корист H_1. Закључујемо да наши субјекти (малишани) показују склоности према светлијим лоптама.

Пример 2: Овде због специфичности, дајемо и пример са непрекидном функцијом расподеле. При мерењу једне карактеристике *X* квалитета добијени су следећи подаци:

X	Број мерења
-30.5	5
30.5-33.5	13
33.5-36.5	23
36.5-39.5	22
39.5-42.5	29
42.5-45.5	29
45.5-48.5	16
48.5-54.5	13

Да ли се може сматрати да X има нормалну расподелу са параметрима $m = 40.48$, $s^2 = 32.59$?

Решење. Из табеле функције нормалне расподеле одредићемо функцију расподеле X, а из узорка емпиријску функцију расподеле. То ће бити следећа табела:

X	$S_n(X)$	$\dfrac{X-m}{\sigma}$	$F_0(X) = \Phi\left(\dfrac{X-m}{\sigma}\right)$	$\lvert F_0(X) - S_n(X)\rvert$
30.5	0.0333	-1.75	0.0401	0.0068
33.5	0.1200	-1.22	0.1112	0.0088
36.5	0.2733	-0.70	0.2420	0.0311
39.5	0.4200	-0.17	0.4325	0.0125
42.5	0.6133	0.35	0.6368	0.0253
45.5	0.8067	0.88	0.8106	0.0039
48.5	0.9113	1.40	0.9192	0.0079
54.5	1.0000	2.46	0.9930	0.0070

Као и до сада, дајемо кратак опис корака, које смо прошли у овом тесту:

1) Одредимо теоријску расподелу фреквенција која се очекује кад је хипотеза H_o тачна;

2) Уредимо добијене резултате (узорак) у кумулативну расподелу, упарујући сваки интервал $S_n(X)$ са одговарајућим интервалом $F_o(X)$;

3) За сваки корак у кумулативним расподелама (добијеној и предвиђеној), одузети $S_n(X)$ од $F_o(X)$;

4) Користећи формулу (10.3), наћи *D*;

5) Из табеле за Колмогоров-Смирнов тест за задани праг значајности α наћи критичну вредност и поредити је са добијеном, па на основу тога одбацити или прихватити H_o.

11. ЛИНЕАРНИ СТАТИСТИЧКИ МОДЕЛИ

На вредности статистичког обележја утиче више фактора. Често смо у прилици да поједине факторе можемо изразити преко показатеља $X_1,...,X_k$. Потребно је утврдити како показатељи $X_1,...,X_k$, утичу на обележје Y. Можемо претпоставити да постоји веза облика

$$Y = f\left(X_1,...,X_k\right) + \varepsilon, \tag{11.1}$$

при чему је $f(\cdot)$ функција одређеног облика, која зависи само од показатеља $X_1,...,X_k$. Величина ε представља ***случајну промењиву*** која је последица свих осталих немерљивих фактора. Израз (11.1) представља **статистички модел**.

Ако је функција $f(\cdot)$ линеарна функција, онда је то линеарни статистички модел.

Разликоваћемо две врсте линеарних статистичких модела.

а) У првој врсти су тзв. ***регресиони*** или ***економетријски модели*** код којих показатељи $X_1,...,X_k$ узимају различите вредности на посматраним елементима статистичког скупа. Основ анализе ових модела чини ***регресиона*** **и** ***корелациона анализа.***

б) Другу групу модела чине модели са фиксним вредностима показатеља, $X_1,...,X_k$, на свим елементима статистичког скупа. То су ***модели са константним ефектима фактора***. Основа анализе ових модела су ***методе анализе варијансе***.

На пример, ако се посматра потрошња домаћинстава, приход домаћинства је сигурно фактор који утиче на потрошњу. Овај фактор, квантитативно изражен, има различите вредности код посматраних домаћинстава. Зато ћемо користити регресиону анализу и економетријски модел за испитивање потрошње домаћинства.

При испитивању тржишта, одређује се тражња за посматраним производом. Фактори који утичу на тражњу су цене и доходак. Тражња, цене и доходак ће се мењати из године у годину. Потребно је утврдити како цене и доходак утичу на тражњу. Јасно, користиће се *регресиони (економетријски) модели*.

Кад утврђујемо утицај ђубрива или сорти житарица на принос, различите сорте и различите врсте ђубрива имаће одговарајуће приносе. На засејаним парцелама једном врстом ђубрива и једном сортом, оба фактора имају константне вредности. На другим парцелама са другом врстом ђубрива и сортом, оба фактора ће имати опет константне вредности. Измерени приноси у оба случаја се разликују. Потребно је утврдити те разлике. Овде се користе методе *анализе варијансе*, а посматрани модели су *модели ǯg константним ефектима*.

У наредним поглављима размотрићемо ове две врсте модела, код којих заједничку основу чини ***метод најмањих квадрата***. Зато ће бити нешто детаљније изложен и објешњен овај метод као основа за оцењивање параметара модела, а затим ће се прећи на статистичко закључивање и одлучивање о ваљаности модела.

11.1. РЕГРЕСИОНИ (ЕКОНОМЕТРИЈСКИ) МОДЕЛИ

Посматраћемо n елемената узорка, измерене вредности обележја Y на елементима узорка и измерене вредности показатеља (фактора) $X_1,...,X_k$, такође на елементима узорка. Обележје Y је ***зависна промењива***, а $X_1,...,X_k$ су ***контролисане промењиве***.

Општи линеарни статистички модел је модел облика

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i \qquad (11.1)$$

где су $\left(Y_i, X_{1i}, X_{2i}, ..., X_{ki}\right)$ измерене вредности зависних и контролисаних промењивих код i-тог елемента узорка ($i=1,2,...,n$). Промењиве ε_i су случајне промењиве (***случајан шум***), које одговарају елементима узорка.

Претпостављамо да су очекиване вредности случајних промењивих ε_i једнаке нули, а варијансе су једнаке σ^2.

Параметри $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ и σ^2 су ***параметри модела*** који су непознати. Потребно је оценити ове параметре на основу узорка. Оценићемо их методом најмањих квадрата.

Метод најмањих квадрата

Ради лакшег схватања самог метода, а и интерпретације добијених резултата, посматраћемо модел са једним фактором (једном контролисаном промењивом), а затим ћемо добијене резултате уопштити и на моделе са више фактора.

Нека је Y зависна промењива, а X контролисана промењива. Линеарни статистички модел ће имати облик

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{11.2}$$

за свако $i=1,2,...,n$.

Нека је

$$E(\varepsilon_i) = 0, \qquad Var(\varepsilon_i) = \sigma^2,$$

а $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ су независне случајне промењиве.

Измерене вредности у узорку су

$$(Y_1, X_1)(Y_2, X_2),...,(Y_n, X_n)$$

и представљају рој тачака у равни XOY, (Слика 11.1).

Права

$$y = \beta_0 + \beta_1 x$$

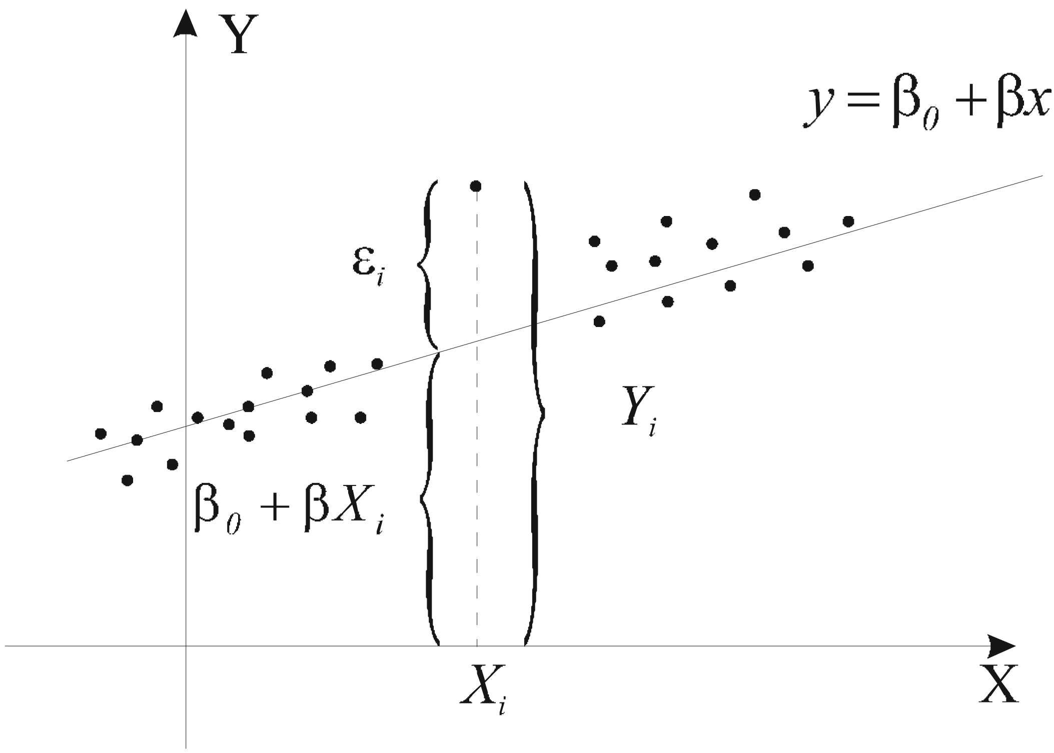
представља очекивану вредност случајне промењиве Y, под условом да је

$$X = x, \qquad \text{и} \qquad E(Y/x) = \beta_0 + \beta_1 x.$$

Коефицијенти β_0 и β_1 су нам непознати. Оценићемо их тако да одступања измерених вредности Y_i у узорку и вредности срачунатих из вредности X_i, буду што мања. Зато ћемо посматрати збир квадрата тих одступања

$$\sum_{i=1}^{n} \left[Y_i - (\beta_0 + \beta X_i)\right]^2 = F(\beta_0, \beta),$$

као функцију $F(\cdot)$ непознатих параметара.



Слика 11.1. Линеарни модел са једним фактором.

Одредићемо вредности β_0 и β_1 тако да ова функција има најмању могућу вредност, тј. оценићемо их тако да средње квадратно одступање буде минимално.

Изједначавањем парцијалних извода функције F са нулом

$$\frac{\partial F(\beta_0, \beta)}{\partial \beta} = 0 \qquad \frac{\partial F(\beta_0, \beta)}{\partial \beta_0} = 0$$

где смо индекс β_1 заменили са β ради краћег писања, добијамо систем једначина.

$$\begin{aligned}\beta\cdot\sum_{i=1}^{n}X_i^2+\beta_0\cdot\sum_{i=1}^{n}X_i&=\sum_{i=1}^{n}X_iY_i\\ \beta\cdot\sum_{i=1}^{n}X_i+\beta_0\cdot n\qquad&=\sum_{i=1}^{n}Y_i,\end{aligned}\tag{11.3}$$

познат под називом систем нормалних једначина.

Детерминанта система нормалних једначина једнака је

$$D=\begin{vmatrix}\sum\limits_{i=1}^{n}X_i^2 & \sum\limits_{i=1}^{n}X_i\\ \sum\limits_{i=1}^{n}X_i & n\end{vmatrix}=n\sum_{i=1}^{n}X_i^2-\left[\sum_{i=1}^{n}X_i\right]^2,$$

односно

$$D=n^2\left\{\frac{1}{n}\sum_{i=1}^{n}X_i^2-\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]^2\right\}.$$

Означимо са $\bar{x}$ и s_x^2, средину и варијансу вредности контролисане промењљиве у узорку. Биће то следеће вредности:

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}X_i;\qquad s_x^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{x}\right).$$

Детерминанта система нормалних једначина је једнака

$$D=n^2s_x^2,$$

Пошто је D>0, систем (11.3) нормалних једначина имаће једно, и само једно решење. То решење ће бити тзв. оцене најмањих квадрата

$$b_0 = \overline{Y} - \frac{\overline{x}}{ns_x^2}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{x}\right) = \overline{Y} - b\overline{x}$$
$$b = \frac{1}{ns_x^2}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{x}\right), \tag{11.4}$$

где смо са $\overline{Y}$ означили средину зависне промењиве Y у узорку

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i .$$

Оцене b_0 и b се могу написати и у следећем облику

$$b_0 = \frac{1}{ns_x^2}\sum_{i=1}^{n}\left[s_x^2 - \overline{x}\left(X_i - \overline{x}\right)\right]Y_i$$
$$b = \frac{1}{ns_x^2}\sum_{i=1}^{n}\left(X_i - \overline{x}\right)Y_i \tag{11.5}$$

1. Особине оцена најмањих квадрата. Коефицијенти b_0 и b су случајне промењиве које зависе од узорка. Из претпоставки модела, параметри зависне промењиве Y у узорку су

$$E\left(Y_i\right) = \beta_0 + \beta X_i \qquad Var\left(Y_i\right) = \sigma^2, \qquad Cov\left(Y_i, Y_j\right) = 0,$$

због независности узорка.

Користећи изразе (11.5), можемо закључити следеће:

а) ***Оцењени регресиони коефицијенти*** b_0 ***и*** b***, као оцене, су линеарне функције вредности зависне промењиве*** $Y_1, Y_2, \ldots, Y_n$***, у узорку. Могу се написати у облику***

$$b_0 = \sum_{i=1}^{n} C_i^{'} Y_i$$
$$b = \sum_{i=1}^{n} C_i Y_i , \tag{11.6}$$

где смо са $C_i^{'}, C_i$ означили вредности

$$C_i^{'} = \frac{s_x^2 - \overline{x}\left(X_i - \overline{x}\right)}{ns_x^2}, \quad i = 1,2,\ldots,n,$$

$$C_i = \frac{\left(X_i - \overline{x}\right)}{ns_x^2}, \qquad i = 1,2,\ldots,n.$$

Лако је проверити да је

$$\sum_{i=1}^{n} C_i^{'} = 1 \qquad \sum_{i=1}^{n} C_i^{'} X_i = 0,$$

$$\sum_{i=1}^{n} C_i = 0 \qquad \sum_{i=1}^{n} C_i X_i = 1.$$

Коефицијенти $C_i^{'}, C_i$ нису случајне променљиве. Њихове вредности зависе само од вредности контролисане променљиве X у узорку.

б) ***Оцене b_0 и b су непристрасне оцене регресионих коефицијената β_0 и β.***

Заиста, из израза (6.6) се добија да је

$$E\left(b_0\right) = \sum_{i=1}^{n} C_i^{'}\left(\beta_0 + \beta X_i\right) =$$

$$= \beta_0 \sum_{i=1}^{n} C_i^{'} + \beta \sum_{i=1}^{n} C_i^{'} X_i =$$

$$= \beta_0 \quad \left(\text{из особине } a\right).$$

На сличан начин се може проверити да је

$$E\left(b\right) = \beta_0 \sum C_i + \beta \sum C_i X_i = \beta.$$

в) ***Варијансе оцена b_0 и b дате су изразима***

$$Var(b_0) = \sigma^2 \sum_{i=1}^{n} C_i^{'2} = \frac{\sigma^2\left(\overline{x}^2 + s_x^2\right)}{ns_x^2},$$

$$Var(b) = \sigma^2 \sum_{i=1}^{n} C_i^2 = \frac{\sigma^2}{ns_x^2}.$$

г) ***Коваријанса ових оцена је једнака***

$$Cov(b_0, b) = \sigma^2 \sum_{i=1}^{n} C_i^{'} C_i = -\frac{\sigma^2 \overline{x}}{ns_x^2}.$$

д) ***Оцене b_0 и b представљају најбоље непристрасне линеарне оцене параметара β_0 и β у следећем смислу: Варијансе било којих других непристрасних линеарних оцена, биће веће од варијанси ових оцена. Другим речима, оне представљају најефикасније оцене у класи линеарних непристрасних оцена, тј. оцене са минималном варијансом.***

Проверићемо то на оцени b. Нека је $b^{'}$ било која оцена линеарног облика

$$b^{'} = \sum_{i=1}^{n} d_i Y_i = \sum_{i=1}^{n} d_i \left(\beta_0 + \beta X_i + \varepsilon_i\right).$$

Да би била непристрасна, мора бити

$$\sum_{i=1}^{n} d_i = 0, \qquad \sum_{i=1}^{n} d_i X_i = 1.$$

Њена варијанса је једнака

$$Var\left(b^{'}\right) = \sigma^2 \sum_{i=1}^{n} d_i^2 . \tag{11.7}$$

Минимална вредност ове варијансе добиће се за вредности које минимизирају Лагранжову функцију

$$F = \sigma^2 \sum_{i=1}^{n} d_i^2 - \lambda \sum_{i=1}^{n} d_i - \lambda_1 \left(\sum_{i=1}^{n} d_i X_i - 1 \right).$$

Решавањем система једначина

$$\frac{\partial F}{\partial d_i} = 0 \qquad i = 1,2,\ldots,n$$

$$\frac{\partial F}{\partial \lambda} = 0$$

$$\frac{\partial F}{\partial \lambda_1} = 0,$$

лако се уверити да је

$$d_i = \frac{X_i - \overline{x}}{n s_x^2} C_i,$$

што је и требало проверити.

е) ***Расподела оцена*** b_0 ***и b. Ако случајна промењива Y има Нормалну расподелу, онда*** b_0 ***и b, као линеарне функције, такође имају Нормалну расподелу.***

Ако, пак, *Y* има било коју расподелу, онда ће b_0 **и *b* имати приближно Нормалну расподелу. То је последица *централне граничне теореме.***

ж) ***Желимо да оценимо вредност случајне промењиве Y, за неку вредност*** $X = x$***. Оцењена вредност је***

$$\hat{Y} = b_0 + bx. \tag{11.8}$$

Наравно, то је линеарна функција њених вредности у узорку. Очекивана вредност ове оцене је

$$E\left(\hat{Y}\right) = \beta_0 + \beta x,$$

а то је тачка на регресионој правој. Можемо сматрати да је оцена $\hat{Y}$ непристрасна оцена очекиване вредности променљиве Y кад је

$$X = x.$$

Варијанса оцене (11.8) биће једнака

$$Var(\hat{Y}) = Var(b_0) + x^2 Var(b) + 2xCov(b_0, b).$$

Замењујући вредности израза на десној страни њиховим вредностима датим у тачки в, добиће се

$$Var(\hat{Y}) = \frac{\sigma^2}{ns_x^2}\left\{s_x^2 + (\bar{x} - x)^2\right\}.$$

Из овог израза се види да ће оцењене вредности случајне променљиве *Y* бити прецизније кад је вредност *x*, за коју вршимо оцену, ближа просечној вредности $\bar{x}$. Такође, са повећањем узорка, побољшаћемо оцену, тј. смањићемо њену варијансу.

з) ***Просечна вредност зависне променљиве Y,***

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i,$$

је статистика са параметрима

$$E(\bar{Y}) = \beta_0 + \beta\bar{x},$$

$$Var(\bar{Y}) = \frac{\sigma^2}{n}.$$

и) ***Варијанса σ^2 у моделу је непознати параметар модела. То је варијанса случајних променљивих ε_i. Такође је то и варијанса зависне променљиве Y, тј.***

$$Var(Y_i) = \sigma^2$$

Потребно је да оценимо овај параметар. Оценићемо га преко разлика

$$Y_i - (b_0 + bX_i),$$

које представљају тзв. резидуале. То су разлике између измерених вредности промењиве Y и срачунатих вредности из линеарне везе преко X.

Оцена параметра σ^2 је статистика

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - (b_0 + bX_i)\right]^2 \tag{11.9}$$

Кад се у изразу (11.9) дода и одузме $\beta_0 + \beta X_i$, а затим се изврши квадрирање и растављање збира на збирове чланова, користећи систем нормалних једначина, оцена варијансе се може написати и у облику

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta_0 + \beta X_i)^2 - (\overline{Y} - \beta_0 - \beta\overline{x})^2 - s_x^2(b-\beta)^2.$$

Одавде је лако проверити да је очекивана вредност ове оцене

$$E\ \hat{\sigma}^2) = \frac{n-2}{n}\sigma^2,$$

што значи да је ово негативно пристрасна оцена.

За непристрасну оцену варијансе σ^2 можемо узети статистику

$$\hat{\sigma}_n^2 = \frac{n}{n-2}\sigma^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left[Y_i - (b_0 + bX_i)\right]^2$$

2. Расподела оцена најмањих квадрата. Испитујемо зависну промењиву Y и контролисану промењиву X преко модела са следећим претпоставкама:

1. Између промењивих X и Y постоји линеарна веза

$$Y = \beta_0 + \beta X + \varepsilon,$$

где је ε случајна променљива.

2. Параметри случајне променљиве су

$$E(\varepsilon) = 0,$$
$$Var(\varepsilon) = \sigma^2.$$

3. Случајна променљива ε има Нормалну расподелу

$$\varepsilon : N(0; \sigma^2)$$

Параметри модела су β_0, β и σ^2 које оцењујемо на основу узорка

$$(Y_1, X_1), (Y_2, X_2), ..., (Y_n, X_n)$$

методом најмањих квадрата. Оцене b_0, b и $\hat{\sigma}^2$ дате су изразима (11.5) и (11.9).

Користећи претпоставке модела, може се закључити да је

а) $E(Y_i) = \beta_0 + \beta X$,

б) $Var(Y_i) = \sigma^2$,

в) Y_i има Нормалну расподелу

$$Y_i : N(\beta_0 + \beta X_i; \sigma^2)$$

г) Променљиве $Y_1, Y_2, ..., Y_n$ су независне и све имају исту расподелу.

Користећи претпоставку 3, лако је закључити да ће оцене максималне веродостојности регресионих параметара, бити оцене добијене методом најмањих квадрата.

Пошто су b_0 и b линеарне оцене облика

$$b_0 = \sum_{i=1}^{n} C_i^{'} Y_i$$

$$b = \sum_{i=1}^{n} C_i Y_i,$$

онда оне имају Нормалну расподелу са одговарајућим параметрима:

$$b_0 : N\left[\beta_0; \frac{\sigma^2\left(s_x^2 + \overline{x}^2\right)}{n s_x^2}\right]$$

$$b : N\left[\beta; \frac{\sigma^2}{n s_x^2}\right].$$

Оцена варијансе σ^2, може се изразити у следећем облику

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 - \left[\overline{Y} - \left(\beta_0 + \beta\overline{x}\right)\right]^2 - \left(b - \beta\right)^2 s_x^2, \tag{11.10}$$

а оцене регресионих коефицијената, такође се могу изразити у следећем облику

$$\begin{aligned} b_0 &= \beta_0 + \sum_{i=1}^{n} C_i^{'} \varepsilon_i \\ b &= \beta + \sum_{i=1}^{n} C_i \varepsilon_i \end{aligned} \tag{11.11}$$

где су $C_i^{'}$, C_i дати изразима

$$C_i^{'} = \frac{s_x^2 - \overline{x}\left(X_i - \overline{x}\right)}{n s_x^2}; \qquad C_i = \frac{X_i - \overline{x}}{n s_x^2}; \quad i = 1,2,\ldots,n.$$

Из (11.10) се може закључити да $\hat{\sigma}^2$, као збир квадрата независних промењљивих, има *Хи*-квадрат расподелу. Други и трећи члан израза на левој страни у (11.10), такође су квадрати промењљивих са Нормалном расподелом, и то су промењљиве које су независне од $\varepsilon_1, \ldots, \varepsilon_n$. Зато се може закључити:

Оцена $\hat{\sigma}^2$ има расподелу као и случајна променљива

$$\sigma^2 \chi^2$$

где χ^2 има Хи-квадрат расподелу са $(n-2)$ степени слободе. Поред тога, статистике b_0 и $\hat{\sigma}^2$ су независне. Такође су независне и статистике b и $\hat{\sigma}^2$.

Тестирање хипотеза и одређивање интервала поверења

За закључивање о оценама коефицијената b_0 и b, користе се статистике

$$t_{b_0} = \frac{\sqrt{s_x^2 - \bar{x}^2}}{s_x} \cdot \frac{b_0 - \beta_0}{\hat{\sigma}^2} \sqrt{n-2},$$

$$t_b = \frac{1}{s_x} \cdot \frac{b - \beta}{\hat{\sigma}} \sqrt{n-2},$$

које имају Студентову t-расподелу са $(n-2)$ степени слободе.

За задати ниво поверења γ, одредићемо вредност t_0 тако да је

$$S_{n-2}(t_0) = \frac{1+\gamma}{2},$$

па на основу статистике t_b, **интервал поверења** за β биће интервал са границама

$$\left[b - t_0 \frac{s_x \hat{\sigma}}{\sqrt{n-2}};\ b + t_0 \frac{s_x \hat{\sigma}}{\sqrt{n-2}} \right].$$

На сличан начин би одредили и интервал поверења за β_0, користећи статистику t_{b_0}. При тестирању хипотеза о регресионим коефицијентима, користимо аналоган поступак као и при тестирању хипотеза о математичком очекивању кад је варијанса популације непозната. Ако желимо да преко регресионог модела оценимо вредност зависне променљиве Y за задану вредност $X = x$, онда је то оцена

$$\hat{Y} = b_0 + bx,$$

која има Нормалну расподелу са параметрима

$$E(\hat{Y}) = \beta_0 + \beta x$$
$$Var(\hat{Y}) = \frac{\sigma^2}{ns_x^2}\left\{s_x^2 + (\overline{x} - x)^2\right\}.$$

Статистика

$$t = \frac{\sqrt{s_x^2 + (\overline{x} - x)^2}}{s_x} \cdot \frac{(b_0 + bx) - (\beta_0 + \beta x)}{\hat{\sigma}}\sqrt{n-2},$$

има Студентову расподелу са $(n-2)$ степени слободе. Ова се статистика користи за одређивање интервала поверења и тестирање хипотезе о оцењеним вредностима зависне променљиве, преко регресионог модела.

Анализа варијансе у једнопараметарском моделу

Резултате зависне променљиве Y у узорку, посматраћемо преко збира квадрата њихових одступања од просека (варијанса узорка помножена са обимом узорка)

$$\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \tag{11.12}$$

Овај збир показује **варијабилитет зависне променљиве**. Њен укупан варијабилитет је последица варијабилитета, односно промена, контролисане променљиве X, а остали део је последица свих других фактора.

Потребно је оценити део тог варијабилитета који је последица промена X. Тај део варијабилитета можемо оценити збиром квадрата

$$\sum_{i=1}^{n}\left[b_0 + bX_i - \frac{1}{n}\sum\left(b_0 + bX_i\right)\right]^2$$

где је у загради разлика између оцењене вредности Y_i преко X_i и просечне вредности тих оцењених вредности. Овај збир квадрата се зове ***регресиони збир квадрата***. Може се изразити у облику

$$b^2 \sum_{i=1}^{n}\left(X_i - \bar{x}\right)^2 . \tag{11.13}$$

На крају, остали део варијабилитета зависне променљиве Y, може се мерити преко тзв***. резидуалног збира квадрата***, који је дат изразом

$$\sum_{i=1}^{n}\left[Y_i - \left(b_0 + bX_i\right)\right]^2 ,$$

а одређен је разликом између измерених вредности зависне променљиве и оцењених њених вредности преко вредности контролисане променљиве. Може се изразити у облику

$$\sum_{i=1}^{n}\left[Y_i - \bar{y} + b\left(X_i - \bar{x}\right)\right]^2 . \tag{11.14}$$

Лако се доказује следећа релација:

$$\sum_{i=1}^{n}\left(Y_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left[Y_i - \left(b_0 + bX_i\right)\right]^2 + b^2 \sum_{i=1}^{n}\left(X_i - \bar{x}\right), \tag{11.15}$$

што значи:

Варијабилитет зависне променљиве Y једнак је збиру објашњеног варијабилитета преко регресије и резидуалног (необјашњеног) дела варијабилитета.

Посматраћемо тзв. *коефицијент детерминације*. Успешност објашњења промена зависне променљиве изражена је односом објашњеног варијабилитета преко регресије, и укупног варијабилитета променљиве *Y*.

Коефицијент детерминације је количник

$$r^2 = \frac{b^2 \sum_{i=1}^{n} \left(X_i - \overline{x}\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{y}\right)^2} \tag{11.16}$$

и представља релативну меру успеха објашњења варијабилитета зависне променљиве Y преко промена променљиве X.

Из релације (11.15) јасно је да је

$$r^2 = 1 - \frac{\sum_{i=1}^{n} \left[Y_i - \left(b_0 + bX_i\right)\right]^2}{\sum_{i=1}^{n} \left(Y_i - \overline{y}\right)^2}, \tag{11.17}$$

што значи: $\left(1 - r^2\right)$ представља однос необјашњеног варијабилитета Y и укупног њеног варијабилитета.

Кад се у (11.16) замени вредност регресионог коефицијента *b*, добиће се израз

$$r^2 = \frac{s_{xy}^2}{s_x^2 s_y^2}$$

где је s_{xy} коваријанса у узорку $\left(X_1, Y_1\right)$, ..., $\left(X_n, Y_n\right)$, s_x^2 и s_y^2 су варијансе променљивих у том узорку.

Зато је коефицијент детерминације, уствари, квадрат коефицијента корелације узорка.

Из релације (11.9) и (11.5) добиће се да је оцењена вредност варијансе дата изразом

$$\hat{\sigma}^2 = s_y^2\left(1 - r^2\right)$$

и представља резидуалну варијансу.

Регресиони збир квадрата може се написати у облику

$$b^2 \sum_{i=1}^{n} \left(X_i - \bar{x}\right)^2 = \left[z_1 + \beta \sqrt{\sum_{i=1}^{n} \left(X_i - \bar{x}\right)} \right]^2,$$

где смо са z_1 означили вредност

$$z_1 = \left(b - \beta\right) \sqrt{\sum_{i=1}^{n} \left(X_i - \bar{x}\right)^2} .$$

Кад је $\beta = 0$, регресиони збир квадрата ће имати расподелу као и случајна променљива

$$\sigma^2 \chi^2,$$

где χ^2 има *Хи*-квадрат расподелу са једним степеном слободе.

Са друге стране, резидуални збир квадрата има расподелу као и случајна променљива

$$\sigma^2 \chi^2,$$

где χ^2 има *Хи*-квадрат расподелу са $\left(n - 2\right)$ степени слободе.

Количник ове две статистике има *F* расподелу са *1* и $\left(n - 2\right)$ степени слободе. Тај се количник може изразити и преко коефицијента детерминације

$$F = \frac{r^2}{1 - r^2}\left(n - 2\right),$$

па се за тестирање хипотезе

$$H_0(\beta = 0),$$

користи F-тест. Из таблица за функцију F-расподеле одредимо критичну вредност F_0 за задати ниво значајности, па, на уобичајени начин, доносимо одлуку о хипотези H_0.

Резултате приказујемо у тзв. ***табели анализе варијансе***:

Извор варијација	Степени слободе	Збир квадрата	Количник F
Регресија	1	$b^2 \sum_{i=1}^{n} \left(X_i - \bar{x}\right)^2$	$F = \frac{r^2}{1-r^2}\sqrt{n-2}$
Резидуал	n-2	$\sum_{i=1}^{n} \left[Y_i - \left(b_0 + bX_i\right)\right]^2 = n\hat{\sigma}^2$	

Пример. Претпоставља се да се са повећањем броја радника предузећа повећава и доходак по раднику. Да би се то проверило, изабрано је 11 предузећа и добијени су следећи резултати:

Доходак (у хиљ.)	10	15	17	16	23	28	30	40	50	60	80
Број радника	10	12	15	20	25	25	30	35	30	40	40

а) Оцени параметре линеарног регресионог модела.

б) Да ли се може сматрати да број радника утиче на доходак по раднику?

в) Одредити коефицијент детерминације и дати његово објашњење.

Решење. Доходак по раднику је зависна променљива Y, а контролисана променљива је број радника X. Модел је облика

$$Y = \beta_0 + \beta X + \varepsilon$$

а) Оцене параметара β_0 и β, добијене методом најмањих квадрата, представљају решење система нормалних једначина

$$\beta \sum_{i=1}^{n} X_i^2 + \beta_0 \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} X_i Y_i,$$

$$\beta \sum_{i=1}^{n} X_i + \beta_0 \cdot b = \sum_{i=1}^{n} Y_i,$$

односно,

$$8344\beta + 282\beta_0 = 12230,$$
$$282\beta + 11\beta_0 = 459.$$

Тако ћемо добити оцењене регресионе коефицијенте

$$b_0 = -14.07$$
$$b = 1.86.$$

Варијансу дохотка по раднику оценићемо са

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{11} \left[Y_i + 14.07 - 1.86 X_i \right]^2 = 87.2.$$

б) Потребно је проверити претпоставку

$$H_0(\beta = 0),$$

а алтернативна претпоставка је

$$H_1(\beta > 0).$$

Користимо статистику

$$t_b = \frac{1}{ns_x^2} \frac{b-0}{\hat{\sigma}} \sqrt{n-2}$$

која има вредност

$$t_b = 6.00\,.$$

Из таблица Студентове *t*-расподеле одредићемо t_0 тако да је

$$S_9\left(t_0\right) = 1 - \alpha$$

За $\alpha = 0{,}01$ добићемо

$$t_0 = 2.821\,.$$

Пошто је $t = 6.00$, $t_0 = 2.821$, хипотезу H_0 ћемо одбацити, и закључити да са порастом броја радника расте и доходак по раднику.

в) Варијанса зависне промењиве *Y* једнака је

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{y}\right)^2 = 130.017\,.$$

Знамо да постоји веза

$$\hat{\sigma}^2 = s_y^2\left(1 - r^2\right).$$

Кад у овај израз заменимо $\hat{\sigma}^2$ и s_y^2, добићемо коефицијент детерминације

$$r^2 = 0.32932\,.$$

То значи да је 33% варијација зависне промењиве, односно дохотка (промена у дохотку), последица промена броја радника. Остале варијације су последица других фактора.

11.2. ЗАДАЦИ

1. У једнопараметарском линеарном моделу

$$Y = \beta_0 + \beta_1 X + \varepsilon ,$$

одреди оцене параметара β_0, β_1 и σ^2 методом максималне веродостојности.

2. На једној раскрсници се мери интензитет саобраћаја. Мерења су вршена свака два сата у периоду од 6 до 14 часова. Добијени су следећи резултати

Време	6	8	10	12	14
Интензитет	26	35	40	52	57

а) Формулиши статистички модел за објашњење интензитета саобраћаја у зависности од времена и оцени одговарајуће параметре.

б) Да ли се са порастом времена током дана у посматраном периоду, интензитет саобраћаја повећава?

3. Број грешака и број часова рада у трећој смени, били су следећи

Број часова рада	3	4	5	6	7	8
Број грешака	5	6	7	8	10	13

а) Оцени линеарну везу између броја грешака и броја часова рада

б) Одреди коефицијент детерминације.

в) Да ли је узорак довољно велики да би се закључци о оцењеном моделу могли прихватити?

4. Степен аутоматизације утиче на зараде радника. На основу узорка од 100 радника, добијени су следећи подаци:

Зараде (у хиљ.)	1.4	1.6	1.8	2.0	2.1	2.2	2.3	2.4	2.5
Степен аутоматизације	0.5	0.6	0.9	0.5	0.6	0.9	0.5	0.6	0.9
Број радника	15	5	0	15	15	5	5	15	25

Извршити статистичку анализу зависности зарада и степена аутоматизације. Објасни добијену вредност коефицијента детерминације.

5. У технолошком ланцу, на машини II, број исправних производа зависи од количине производа обрађених на претходној машини I. У мерењима су добијени резултати

X	5	6	8	5	9	10	7	5	8	7
Y	3	4	5	2	8	7	4	5	6	6

Оцени вредност Y преко X, а затим одреди ону вредност X за коју ће број неисправних производа на другој машини бити 5.

6. Од одређене врсте гуме направљено је 10 цеви различитог пречника. Извршена су мерења максималног унутрашњег притиска који могу да издрже цеви. Добијени су следећи подаци

$$\sum X = 17{,}3 \qquad \sum Y = 56{,}6 \qquad \sum XY = 109{,}53$$
$$\sum X^2 = 33{,}15 \quad \sum Y^2 = 377{,}22.$$

а) Оцени регресиону праву.

б) Одреди 95% интервал поверења за параметар β.

в) Тестирај хипотезу да је $\beta_0 = 1$.

ПРАКТИКУМ

12. ПРАКТИКУМ

Збирка питања за самосталну проверу знања

1. Дескриптивна статистика је део статистике који се бави:

- прикупљањем, обрадом и представљањем постојећих статистичких података
- описом статистичких чињеница
- графичким приказима података
- пописом и сређивањем резултата пописа

2. Статистичка анализа је део статистике коју чини (чине):

- анализирање статистички прикупљених анкета
- скуп статистичких метода за квантитативну анализу међусобних односа између појава које имају масовни карактер и помоћу којих се доносе одређени закључци о законитостима по којима сеодвијају посматране појаве
- математичке методе за сређивање и приказ података
- интерпретација података

3. Статистичка теорија је

- статистичка математичка теорија
- део статистике који се бави проналажењем нових и усавршавањем постојећих статистичких метода
- теоријски део статистике
- временски и просторно ограничена

4. Статистика је

- скуп научних метода које се служе вероватноћом као алатом
- скуп научних метода које се користе за прикупљање, приказивање, анализу и интерпретацију података и доношење статистичких закључака
- економска наука
- грана математике

5. Предмет проучавања математичке статистике су

- нумеричке вредности прикупљене из основног скупа
- описни статистички подаци
- резултати посматрања или експеримената, које је могуће поновити неограничен број пута при истим условима, али није могуће тачно предвиђање резултата за појединачни експеримент
- нумеричке бројне вредности
- математичко-статистички подаци

6. Појаве код којих услови под којима се одвија та појава једнозначно одређују резултат (исход) су

- детерминистичке појаве
- случајне појаве
- сигурне појаве
- неизвесне појаве

7. Вероватноћа проучава

- стохастичке (случајне) појаве
- детерминистичке појаве
- неслучајне појаве
- ништа од наведеног

8. Подаци су

- скуп чињеница који су добијени као резлутат посматрања индивидуалних објеката
- цене производа, продате количине робе, захтеви за поједином робом у пословању предузећа
- имена ученика, оцене и изостанци у школама
- лични подаци спортиста и резултати у спортским клубовима
- подаци о пацијентима и болестима у здравственим картонима

9. Популација (статистички или основни скуп)

- је скуп свих елемената на којима се одређена појава статистички посматра
- је скуп статистичких јединица
- је велики скуп из кога прикупљамо податке
- је скуп људи

10. Статистичка јединица је

- елемент популације или узорка о ком се прикупљају подаци
- једна особа
- предмет
- фирма

11. Опсервација (податак) је

- чињеница о некој особи
- истинита информација о неком објекту
- вредност промењиве (обележја) која се односи на јединицу посматрања
- део опсерваторије

12. Хомогеност популације је

- истоврсност популације
- истоветност популације
- истоветност статистичких јединица узорка
- различитост статистичких јединица узорка

13. У статистици, са N се означава

- број елемената популације
- број елемената узорка
- број људи у скупу
- природан број

14. Узорак је

- део већег скупа људи
- одабрани скуп фирми чији се подаци анализирају
- подскуп популације која је предмет статистичке анализе
- подскуп било ког скупа

15. Репрезентативан узорак

- одражава карактеристике основног скупа
- не одражава карактеристике основног скупа
- делимично одражава карактеристике основног скупа
- је било који узорак

16. Број чланова узорка у статистици се означава са

- n
- m
- k
- s

17. Обележје је

- карактеристика (особина) јединице статистичког скупа или узорка
- број
- опис неке особине
- атрибут

18. Обележје може бити

- квалитативно и квантитативно
- нумеричко
- квалитативно
- атрибутно

19. Код атрибутног (квалитативног) обележја

- вредност обележја се изражава описно
- вредност обележја се изражава искључиво бројевима
- вредност обележја се изражава искључиво кодовима
- вредност обележја се изражава само у процентима

20. Нумеричко обележје може бити

- прекидно или непрекидно
- аритметичко или геометријско
- континуално или непрекидно
- прекидно или дискретно

21. Прекидно нумеричко обележје

- је резултат мерења
- узима вредности из бесконачног скупа
- се изражава природним бројевима
- је резултат анкете

22. Непрекидно (континуално) нумеричко обележје

- се изражава реалним бројевима
- је резултат пребројавања
- узима вредности из коначног скупа
- је резултат пописа

23. Емпиријски подаци су

- стварни подаци прикупљени на терену или резултати експеримента
- подаци добијени из фирме
- подаци о анкетираним особама
- резултат мерења

24. Неуређени статистички подаци су

- неуређени подаци добијени из фирме добијених као резултат посматрања сваке јединице у изабраном скупу
- резултати експеримента поновљеног више пута
- низ неуређених нумеричких вредности добијених као резултат посматрања сваке јединице у изабраном скупу
- подаци о анкетираним особама

25. Број појављивања вредности x_i обележја X је у посматраном скупу података је:

- релативна фреквенција
- релативно учешће
- број појављивања обележја
- апсолутна фреквенција

26. Број појављивања вредности x_i обележја X подељен са укупним бројем података у посматраном скупу (популацији или узорку) је

- апсолутна фреквенција
- релативна фреквенција
- учешће
- број појављивања обележја

27. Основни скуп у статистици је скуп свих

- жена и мушкараца
- статистичких јединица које су предмет проучавања
- људи који живе у једној земљи
- елемената проучавања

28. Узорак у статистици је подскуп

- људи изабраних из основног скупа једне земље
- људи изабраних из основног скупа једне области
- основног скупа који је предмет проучавања
- скупа статистичких јединица

29. Екстремне вредности обележја у скупу су

- вредности из основног скупа
- најмање или највеће вредности обележја у односу на остале вредности обележја у скупу
- најмање вредности обележја у односу на већину вредности у скупу
- највеће вредности обележја у односу на већину вредности у скупу

30. Груписани статистички подаци су

- стварни подаци прикупљени на терену
- подаци приказани у облику расподеле фреквенција
- низ неуређених нумеричких вредности
- резултати експеримента поновљеног више пута
- подаци о анкетираним особама

31. Интервал који укључује све вредности скупа квантитативних података који се налазе између два броја назива се

- низ уређених нумеричких вредности
- групни интервал
- статистички интервал
- *i*-ти интервал

32. Дијаграм који се састоји од низа спојених правоугаоника где су на *x*-оси приказани групни интервали, а на *y*-оси фреквенције или релативне фреквенције назива се

- кумуланта
- полигон
- хистограм
- кружни дијаграм (пита)

33. Крива која представља графички приказ расподеле кумулативних фреквенција назива се

- полигон
- кумуланта
- кружни дијаграм
- кумулативни хистограм

34. Графикон који се састоји од стубића чија висина представља апсолутну или релативну фреквенцију одговарајуће категорије назива се

- хистограм
- кумуланта
- штапићасти дијаграм
- полигон
- кружни дијаграм (пита)

35. Расподела фреквенција је

- табела која приказује све категорије или групне интервале и њима припадајуће фреквенције
- хистограм која приказује фреквенције
- хистограм која приказује релативне фреквенције
- кумуланта која приказује све категорије или групне интервале и њима припадајуће фреквенције

36. Ширина групног интервала је

- разлика између горње и доње граничне вредности једног групног интервала
- величина групног интервала
- збир две граничне вредности једног групног интервала
- количник две граничне вредности једног групног интервала

37. Мере централне тенденције су

- аритметичка средина, модус, медијана и стандардна девијација
- аритметичка средина, медијана и варијанса
- аритметичка средина, модус и стандардна девијација
- аритметичка средина, медијана и модус

38. Мере централне тенденције су

- мере које описују центар расподеле
- мере које описују распршеност расподеле
- мере које описују дисперзију расподеле
- варијанса и стандардна девијација

39. Мере дисперзије су

- мере које описују центар расподеле
- мере које описују распршеност расподеле
- мере којима се израчунава аритметичка средина расподеле
- модус, варијанса и стандардна девијација

40. Вредност средњег члана у рангираном низу (серији) података назива се

- модус
- медијана
- варијанса
- аритметичка средина

41. Медијана дели уређени низ (серију) података на

- три једнака дела
- четири једнака низа са истим бројем елемената
- два низа да истим бројем елемената
- не дели је равномерно

42. Дескриптивна мера која се израчунава за податке из узорка назива се

- параметар
- перцентил
- статистика
- горња спољашња граница

43. Бимодална расподела

- има два модуса
- има три модуса
- има више модуса
- нема модуса

44. Расподела са једним модусом је

- бимодална
- унимодална
- мултимодална
- тримодална

45. Која од следећих мера може имати више од једне вредности?

- аритметичка средина
- модус
- медијана
- интервал варијације

46. Вредности варијансе и стандардне девијације

- никад нису негативне
- увек су позитивне
- никад нису једнаке нули
- су негативне

47. Скуп свих исхода неког експеримента се назива

- простор узорка
- пресек догађаја
- заједничка вероватноћа
- сложени догађај

48. Крајњи исход неког експеримента назива се

- сложени догађај
- елементарни догађај
- комплементарни догађај
- празан догађај

49. Сложени догађај укључује

- све крајње исходе
- тачно два исхода експеримента
- више од једног исхода експеримента
- три исхода експеримента

50. Два једнаковероватна догађаја

- имају исту вероватноћу реализације
- не могу да се реализују истовремено
- ниједан нема утицаја на реализацију другог
- имају исти број елемената

51. Који од следећих приступа вероватноћи може да се примени само на експерименте са једнаковероватним исходима?

- субјективна вероватноћа
- емпиријска вероватноћа
- класична вероватноћа
- заједничка вероватноћа

52. Два међусобно искључива догађаја

- имају исту вероватноћу реализације
- не могу да се реализују истовремено ниједан нема утицаја на реализацију другог
- имају исти број елемената у пресеку

53. Два независна догађаја су они догађаји за које важи да

- имају исту вероватноћу реализације
- не могу да се реализују истовремено
- ниједан нема утицаја на реализацију другог
- имају исти број елемената у унији

54. Вероватноћа реализације неког догађаја је увек реалан број

- мањи од 0
- између 0 и 1
- већи од 0
- већи од 1

55. Сума вероватноћа свих крајњих (елементарних) исхода неког експеримента је увек

- 0
- између 0 и 1
- 100
- 1

56. Два комплементарна догађаја

- су увек међусобно искључива
- никада нису међусобно искључива
- имају непразну унију
- имају више од једног елементарног догађаја

57. Сигуран догађај је догађај који

- се реализује са вероватноћом 1
- има вероватноћу реализације једнаку 0
- је празан скуп
- садржи извесне исходе

58. Закон по коме вероватноћа догађаја добијена на основу релативне фреквенције тежи стварној (теоријској) вероватноћи уколико се неки експеримент понавља велики број пута, назива се

- Студентов закон
- Закон великих бројева
- Закон малих бројева
- Пуасонов закон

59. Нумеричка мера могућности да ће се одређени догађај реализовати зове се

- вероватноћа
- шанса реализације
- постотак
- проценат могућности реализације

60. Вероватноћа реализације било које појединачне вредности x_i дискретне случајне промењљиве X је

- у реалном интервалу од 0 до 1
- 1
- већа од 1
- 100%

61. Збир вероватноћа реализације свих могућих појединачних реализација x_i дискретне случајне промењљиве X је

- у реалном интервалу од 0 до 1
- 1
- већа од 1
- мања од 1
- 0

62. Параметри Биномне расподеле вероватноћа су

- *n, p, q*
- *n, p*
- *n, p, x*
- *n, q, x*

63. Дискретна случајна променљива је она случајна променљива која

- узима непребројиво много вредности
- узима пребројиво много вредности
- узима коначно много вредности
- узима бесконачно много вредности

64. Стандардна девијација дискретне случајне променљиве је

- мера распршености за расподелу вероватноћа дискретне случајне променљиве
- мера вероватноће за расподелу вероватноћа дискретне случајне променљиве
- мера реализације за расподелу вероватноћа дискретне случајне променљиве
- стандардна вредност дискретне случајне променљиве

65. Пермутације скупа од *n* елемената су

- број начина за избор уређених *n*-торки
- број начина на који се може изабрати *k* елемената скупа где је *k* мање од *n*
- број начина за избор уређених *k*-торки где је *k* веће од *n*
- број начина на који се може изабрати *k* елемената скупа где редослед избора није битан

67. Варијације скупа од н елемената k–те ($k < n$) класе су

- број начина на који се може изабрати n елемената тог скупа без понављања елеманата
- број начина на који се може изабрати k елемената из n-точланог скупа а да редослед избора није битан
- број начина на који се може изабрати k елемената из n-точланог скупа а да редослед избора јесте битан

68. Комбинације скупа од н елемената k–те ($k < n$) класе су

- број начина на који се може изабрати n елемената тог скупа без понављања елеманата
- број начина на који се може изабрати k елемената из n-точланог скупа а да редослед избора није битан
- број начина на који се може изабрати k елемената из n-точланог скупа а да редослед избора јесте битан

69. Параметар Пуасонове расподеле вероватноћа је

- укупан број опита, n и вероватноћа успеха, p
- просечан број реализација, λ (ламбда), током интервала за Пуасонову расподелу вероватноћа
- укупан број опита, n и вероватноћа неуспеха, q
- укупан број реализација током интервала за Пуасонову расподелу вероватноћа

70. Непрекидна случајна променљива је она случајна променљива која

- узима непребројиво много вредностиу једном или више интервала
- узима пребројиво много вредности
- узима коначно много вредности
- узима бесконачно много вредности

71. Нормална расподела вероватноћа се примењује у случају

- нијединог од поменутог
- дискретне случајне променљиве
- непрекидне случајне променљиве
- било које случајне променљиве

72. За непрекидну случајну променљиву, вероватноћа да узме једну појединачну вредност x увек је:

- 0
- између 0 и 1
- 1
- већа од 1

73. Која од следећих особина није није карактеристика Нормалне расподеле?

- Укупна површина испод криве је једнака 1.
- Крива је симетрична у односу на аритметичку средину.
- Крива је дефинисана на реалној правој.
- Вредност аритметичке средине је увек позитивна.

74. Параметри Нормалне расподеле

- су z, μ, σ
- су μ, σ
- су μ, σ, x
- су μ, σ, e

75. За стандардизовану Нормалну расподелу је

- $\mu = 0, \sigma = 1$
- $\mu = 0{,}5$, $\sigma = 1$
- $\mu = 1, \sigma = 0$
- $\mu = \sigma$

76. Узорак се сматра великим ако је

- обим узорка већи или једнак 30
- обим узорка мањи или једнак 30
- обим узорка већи или једнак 100
- обим узорка већи или једнак 50

77. Узорачка расподела је расподела вероватноћа

- параметра основног скупа
- статистике узорка
- било које случајне променљиве
- непрекидне случајне променљиве

78. Неслучајне грешке су

- грешке које настају јер је узорак превелики у односу на основни скуп
- грешке које настају приликом прикупљања, бележења и уношења података у табеле
- грешке које настају јер истраживање спроводи недовољно искусна особа
- грешке које настају у рачуну

79. Случајна грешка је

- разлика између вредности статистике узорка и вредности одговарајућег параметра основног скупа
- грешке које настају приликом прикупљања, бележења и уношења података у табеле
- грешке које настају јер је узорак мали
- грешке које настају у рачуну 80. Статистика узорка која се користи за оцењивање параметара основног скупа назива се
- непристрасна оцена
- оцена

- конзистентна оцена
- позитивна оцена

81. Оцена чија је очекивана вредност (средња вредност) једнака вредности одговарајућег параметра основног скупа назива се

- непристрасна оцена
- оцена
- конзистентна оцена
- позитивна оцена

82. Статистика узорка чија се стандардна грешка смањује са порастом величине узорка назива се

- непристрасна оцена
- оцена
- конзистентна оцена
- позитивна оцена

83. Аритметичка средина статистике узорка је увек једнака

- σ
- μ
- μ - 10
- $\sigma(x)$

84. Параметар (параметри) Студентове t расподеле је (су)

- μ
- n - 5
- број степени слободе, *df*
- n

85. Ниво поузданости се означава са

- $(1-\alpha)100\%$
- $\alpha\,100\%$
- α
- β

86. Вредност статистике узорка која се користи за одређивање одговарајућег параметра основног скупа назива се

- непристрасна оцена
- оцена
- овењена вредност
- позитивна оцењена вредност

87. Аритметичка средина Студентове *t* расподеле је једнака

- μ
- 5
- 0
- n-1

88. Тачкаста оцењена вредност је

- непристрасна оцењена вредност
- вредност статистике узорка додељена одговарајућем параметру скупа
- бројна вредност
- позитивна бројна вредност додељена одговарајућем параметру скупа

89. Интервал конструисан око вредности статистике узорка који се користи за оцењивање одговарајућег параметра основног скупа је

- интервал неповерења
- интервал поверења
- отворени интервал
- затворени интервал

90. Поступак у коме се нумеричка вредност додељује параметру скупа на основу података из узорка је

- процењивање
- оцењивање
- додељивање
- узорковање

91. Тестирање хипотезе се увек односи на

- статистику узорка
- параметар основног скупа
- статистички тест
- аритметичку средину

92. Грешка I (прве) врсте прави се када

- се не одбацује нулта хипотеза иако је стварно нетачна
- се одбацује нулта хипотеза иако је стварно тачна
- се одбацује алтернативна хипотеза иако је стварно тачна
- се не одбацује алтернативна хипотеза иако је стварно нетачна

93. Грешка II (друге) врсте прави се када

- се одбацује нулта хипотеза иако је стварно тачна
- ниједно од понуђеног
- се не одбацује алтернативна хипотеза иако је стварно нетачна
- се не одбацује нулта хипотеза иако је стварно нетачна

94. Ниво значајности (α) показује

- вероватноћу јављања грешке прве врсте
- вероватноћу јављања грешке друге врсте
- вероватноћу одбацивања истините алтернативне хипотезе
- вероватноћу јављања грешке треће врсте

95. Вредност β даје

- вероватноћу јављања грешке прве врсте
- вероватноћу јављања грешке друге врсте
- вероватноћу јављања грешке треће врсте
- јачину теста

96. Вредност $1-\beta$ даје

- вероватноћу јављања грешке прве врсте
- вероватноћу јављања грешке друге врсте
- вероватноћу јављања грешке треће врсте
- јачину теста

96. Двострани тест је тест са

- две области неодбацивања нулте хипотезе
- једном облашћу неодбацивања и две области одбацивања нулте хипотезе
- два статистичка теста
- две тест статистике

97. Једнострани тест има

- једну област неодбацивања нулте хипотезе
- једну област одбацивања и једну област неодбацивања нулте хипотезе
- једну област одбацивања нулте хипотезе
- једну тест статистику

98. Знак у алтернативној хипотези у двостраном тесту је увек

- $<$
- $>$
- $\neq$
- $\leq$

99. Знак у алтернативној хипотези у левостраном тесту је увек

- $<$
- $>$
- $\neq$
- $\leq$

100. Знак у алтернативној хипотези у десностраном тесту је увек

- $<$
- $>$
- $\neq$
- $\leq$

101. Нулта хипотеза је

- Тврђење или исказ о неком параметру основног скупа који се сматра неистинитим све док се супротно не докаже
- Тврђење или исказ о неком параметру основног скупа који ће сматрати истинитим ако је почетна претпоставка неистинита

- Тврђење или исказ о неком параметру основног скупа који ће сматрати неистинитим ако је нулта хипотеза неистинита
- Тврђење или исказ о неком параметру основног скупа који се сматра истинитим све док се супротно не докаже

102. Алтернативна хипотеза је

- Тврђење или исказ о неком параметру основног скупа који се сматра неистинитим све док се супротно не докаже
- Тврђење или исказ о неком параметру основног скупа који ће сматрати истинитим ако је нулта хипотеза неистинита
- Тврђење или исказ о неком параметру основног скупа који ће сматрати неистинитим ако је почетна претпоставка неистинита
- Тврђење или исказ о неком параметру основног скупа који се сматра истинитим све док се супротно не докаже

103. Да бисте тестирали хипотезу да је крвни притисак универзитетских професора у просеку нижи од крвног притиска извршних директора компанија, који од следећих тестова бисте користили?

- двострани тест
- левострани тест
- деснострани тест
- једнострани тест

104. Два узорка изабрана из два основна скупа тако да избор једног не утиче на избор другог називају се

- зависни узорци
- независни узорци
- парни узорци
- узорци усклађених парова

105. Проста регресија је регресиони модел који садржи

- само једну објашњавајућу променљиву
- само једну зависну променљиву
- више од једне објашњавајуће променљиве
- једну објашњавајућу и једну зависну променљиву

106. Веза између једне објашњавајуће и једне зависне променљиве приказана линеарном регресијом је

- линеарна функција
- нелинеарна функција
- полином другог степена
- степена функција

107. Детерминистички регресиони модел

- садржи случајну грешку
- приказује егзактну везу
- приказује нелинеарну регресиону везу
- је полином другог степена

108. Стохастички регресиони модел

- садржи случајну грешку
- приказује егзактну везу
- приказује нелинеарну регресиону везу
- је полином другог степена

109. Метод најмањих квадрата минимизира суму

- резидуала
- квадрата резидуала
- предвиђања
- остатака

110. Вредност коефицијента просте линеарне регресије је увек у интервалу од

- 0 до 1
- -1 до 1
- 1 до 2
- -1 до 0

111. Ако између две променљиве не постоји линеарна корелација и тада је вредност коефицијента просте линеарне регресије једнака

- 1
- -1
- 2
- 0

112. Да би сте тестирали хипотезу да је крвни притисак универзитетских професора у просеку нижи од крвног притиска директора компанија, који од следећих тестова би сте користили?

- левострани тест
- двосмерни тест
- деснострани тест

113. Фреквенција неке одређене вредности обележја X је

- атрибут обележја
- број појављивања те вредности у посматраном скупу података подељен са величином тог скупа
- број појављивања те вредности у посматраном скупу података

114. Процедура или поступак избора одређених статистичких јединица које ће ући у узорак назива се:

- анализирање статистичких јединица
- математичке методе за сређивање и приказ података
- узорковање
- интерпретација података

115. Узорак који одражава карактеристике основног скупа назива се

- репрезентативан узорак
- добар узорак
- пожељан узорак
- велики узорак
- мали узорак

116. Релативна фреквенција неке одређене вредности обележја X је

- атрибут обележја
- број појављивања те вредности у посматраном скупу података подељен са бројем елемената тог скупа
- број појављивања те вредности у посматраном скупу података

117. Груписани квалитативни подаци се могу графички приказати помоћу:

- хистограма
- полигона
- кумуланте
- релативне кумуланте
- штапићастог дијаграма и структурног круга (пите)

118. Структурни круг или пита је

- врста графикона у облику круга који је подељен на кружне исечке
- затворена полигонална линија
- врста графикона који се још назива и штапићасти дијаграм

119. Позиционе мере су

- аритметичка средина, медијана и варијанса
- аритметичка средина, модус и стандардна девијација
- аритметичка средина, модус, квартили и стандардна девијација
- медијана, модус, квартили, перцентили

120. Одредити модус(е) серије података 1, 22, 23, 45, 44, 45, 34, 23, 47

- 1
- 45
- 23 и 45
- 23
- 47

121. Одредити модус(е) серије података 1, 22, 45, 44, 45, 34, 23, 47

- 1
- 45
- 23 и 45
- 23
- 47

122. Аритметичка средина нумеричког обележја *X* посматраног над популацијом је

- променљива
- случајна променљива
- статистика
- константа

123. Аритметичка средина нумеричког обележја *X* посматраног над популацијом је

- променљива
- параметар расподеле тог обележја
- случајна променљива
- статистика

124. Аритметичка средина нумеричког обележја *X* посматраног над узорком је

- променљива
- параметар расподеле тог обележја
- случајна променљива
- константа

125. Израчунати аритметичку средину података 10, 20, 30, 40

- 40
- 20
- 25
- 10

126. Збир одступања аритметичке средине од вредности обележја једнак је

- 0
- 1
- између 0 и 1
- n

127. Означи тачну тврдњу.

- Геометријска средина изравнава апсолутне промене између вредности података.
- Геометријска средина изравнава релативне или пропорционалне промене између вредности података.
- Геометријска средина је геометријски појам и односи се на троугао.

128. Ако вредности обележја X, од n елемената, поређамо у неопадајући низ и n је паран број. Медијана обележја X

- се не може одредити
- припада овом скупу
- не припада овом скупу
- се може одредити али није јединствена

129. Ако вредности обележја X, од n елемената, поређамо у неопадајући низ и n је непаран број. Медијана обележја X

- се не може одредити
- припада овом скупу
- не припада овом скупу
- се може одредити али није јединствена

130. Вредности аритметичке средине, медијане и модуса за симетричан хистограм и симетричну криву расподеле фреквенција се

- поклапају
- не поклапају и аритметичка средина је већа од медијане и модуса
- не поклапају и аритметичка средина је мања од медијане и модуса

131. Однос вредности аритметичке средине, медијане и модуса за хистограм расподеле фреквенција асиметричне удесно је следећи од мањег ка већем:

- поклапају се
- модус, медијана и аритметичка средина
- не поклапају и аритметичка средина је мања од медијане и модуса

132. Модел вишеструке регресије садржи

- само једну објашњавајућу променљиву
- само једну зависну променљиву
- више објашњавајућих променљивих и једну зависну променљиву
- једну објашњавајућу и једну зависну променљиву

133. Модел вишеструке линеарне регресије представља

- линеарну функцију више објашњавајућих променљивих
- линеарну функцију једне објашњавајуће променљиве
- само једну независну променљиву
- једну објашњавајућу и једну зависну променљиву

134. Да би сте тестирали хипотезу да је крвни притисак универзитетских професора у просеку нижи од крвног притиска директора компанија, који од следећих тестова би сте користили?

- левострани тест
- двосмерни тест
- деснострани тест

135. Да би сте тестирали хипотезу да је садржај активне супстанце у једном леку различит од садржаја те супстанце у другом леку, који од следећих тестова би сте користили?

- левострани тест
- двосмерни тест
- деснострани тест

136. Да би сте тестирали хипотезу да запослени једне компаније имају у просеку веће зараде од просека зарада у другој компанији, који од следећих тестова би сте користили?

- левострани тест
- двосмерни тест
- деснострани тест

137. Да би сте тестирали хипотезу да је просечна оцена студената економије различита од просечна оцена студената информатике, који од следећих тестова би сте користили?

- левострани тест
- деснострани тест
- двосмерни тест

138. Предмет изучавања статистике су промењиве појаве које се дешавају:

- у реалном простору и времену
- у малом броју случајева
- у великом броју случајева
- у ограниченом простору и коначном временском интервалу

139. Метод узорка је посматрање:

- случајно одабраних типичних статистичких јединица основног скупа на одређеном простору и у одређеном моменту
- одређеног броја случајно изабраних статистичких јединица основног скупа
- свих статистичких јединица основног скупа на одређеном простору и у одређеном моменту

140. Број деце и број запослених у домаћинству су:

- прекидна нумеричка обележја
- непрекидна нумеричка обележја
- атрибутивна обележја

141. Апсолутна фреквенција је:

- учесталост изражена у релативном броју
- учесталост изражена у апсолутном броју
- учесталост изражена у јединицама мере

142. Структура једног хомогеног скупа, који се посматра независно од других, најчешће се графички приказује:

- кружним дијаграмом
- хистограмом
- полигоном

143. Средње вредности се могу поделити у две групе:

- израчунате и позиционе средње вредности
- израчунате и апсолутне средње вредности
- позиционе и релативне средње вредности

144. Медијана је:

- израчуната средња вредност
- релативна средња вредност
- позициона средња вредност

145. Сума квадрата одступања појединих вредности од аритметичке средине неког скупа даје:

- нулу
- минималну вредност квадрата одступања од аритметичке средине
- максималну вредност квадрата одступања од аритметичке средине
- не може се унапред одредити јер зависи од скупа

146. Хармонијска средина се користи у случајевима где се посматране величине међусобно налазе у:

- обрнуто пропорционалном односу
- инверзном виду
- директном пропорционалном односу

147. Стандардна девијација је:

- релативна мера корелације
- релативна мера коваријације
- апсолутна мера дисперзије

148. Коефицијент варијације нарочито је погодан за:

- упоређивање дисперзије двеју или више серија
- испитивање просечног одступања
- испитивање просечног раста двеју или више серија

149. Структура скупа графички се може приказати помоћу штапићастог дијаграма:

- вертикално и хоризонтално
- само вертикално
- само хоризонтално

150. Индекс већи од 100 показује релативно:

- повећање појаве за онолико процената колико је мањи од 100
- повећање појаве за онолико процената колико је већи од 100
- смањење појаве за онолико процената колико је мањи од 100

151. Проста линеарна корелација може бити:

- само негативна
- позитивна или негативна
- само позитивна

152. Ако појава X расте а појава Y опада тада је:

- коефицијент корелације између ових појава позитиван
- коефицијент корелације између ових појава једнак нули
- коефицијент корелације између ових појава негативан

153. Параметри линеарног тренда одређују се:

- методом најмањих одступања
- методом најмањих квадрата
- методом најмањих разлика

154. Параматар *a* код линеарног тренда показује

- просек података временске серије
- средњи апсолутни раст
- темпо развоја

155. Алгебарска сума одступања стварних података од одговарајућих вредности функције тренда:

- једнака је нули
- већа је од нуле
- мања је од нуле

156. Производња једног рудника у тонама је 1994. износила 188.173 тона док је 2011. године износила 347.232. Геометријска стопа раста је 6.32%. То значи да се производња у периоду 1994-2011. године:

- повећавала сваке године у просеку за 6.32%
- повећавала само за 0.632 %
- повећала за цео период само за 6.32%

Решени мешовити задаци за вежбу

Груписање података и нумеричке дескриптивне мере

1. Висина играча једног кошаркашког тима износи у центриметрима 195, 180, 190, 202, 205, 201, 198, 199, 200, 204. Наћи просечну висину играча.

Решење:

$$\overline{X} = \frac{195+180+190+202+205+201+198+199+200+204}{10}$$

2. У једном испитном року добијене су следеће оцене на испиту из једног предмета. Израчунати просечну оцену из тог предмета.

Оцена	6	7	8	9	10
Број студената	32	14	11	8	5

Решење:

Оцене (x_i)	Број студената (f_i)	Групни производ ($x_i f_i$)
6	32	192
7	14	89
8	11	88
9	8	72
10	5	50
	$\sum f_i = 70$	$\sum x_i f_i = 500$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{500}{70} = 7{,}14$$

3. На једном колоквијуму 150 студената освојило је следећи број бодова који су дати у табели. Израчунати просечан број освојених бодова.

Број бодова	0 - 5	6 - 10	11 - 15	16 - 20	21 - 25
Број студената	28	21	52	34	15

Решење:

Број бодова (x_i)	Број студената (f_i)	Средина интервала	Групни производ ($x_i f_i$)
0 - 5	28	2,5	70
6 - 10	21	8	168
11 - 15	52	13	676
16 - 20	34	18	612
21 - 25	15	23	345
	$\sum f_i = 150$		$\sum x_i f_i = 1871$

$$\overline{X} = \frac{\sum_{i=1}^{5} x_i f_i}{\sum_{i=1}^{5} f_i} = \frac{1872}{150} = 12,47$$

Подсећање

Проста геометријска средина израчунава се на следећи начин:

$$G = \sqrt[n]{x_1 x_2 \ldots x_n}$$

Применом логаритамског рачуна добија се проста геометријска средина у логаритамском облику:

$$\log G = \frac{\log x_1 + \log x_2 + \ldots + \log x_n}{n} = \frac{\sum_{i=1}^{n} \log x_i}{n}$$

Пондерисана геометријска средина се израчунава на следећи начин:

$$G = \sqrt[\sum f_i]{x_1^{f_1} x_2^{f_2} \ldots x_n^{f_n}}$$

Применом логаритамског рачуна добија се пондерисана геометријска средина у логаритамском облику:

$$\log G = \frac{f_1 \log x_1 + f_2 \log x_2 + \ldots + f_n \log x_n}{f_1 + f_2 + \ldots + f_n} = \frac{\sum_{i=1}^{n} f_i \log x_i}{\sum_{i=1}^{n} f_i}$$

Геометријска средина се не може израчунати ако нумеричка серија има неку вредност која је нула или негативан број.

4. Распоред радника према радном стажу у једној фабрици дат је табелом. Израчунати просечан радни стаж запослених.

Године стажа	5 - 10	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
Број запослених	28	21	38	24	15	8

Решење:

Године стажа (x_i)	Број запослених (f_i)	Средина интервала	log x_i	f_i log x_i
5-10	28	7,5	0.87506	24,50168
10-15	21	12,5	1,09691	23,03511
15-20	38	17,5	1,24303	47,23514
20-25	24	22,5	1.35218	32,45232
25-30	15	27,5	1,43933	21,58995
30-35	8	32,5	1,51188	12,09504
	$\sum f_i = 134$			160,90924

$$\log G = \frac{160,90024}{134} = 1,20082$$

G = 15,87

Просечни радни стаж у фабрици био је 15,87 година.

5. Продаја предузећа је расла на начин као што је приказано у табели. Проценат раста је изражен у односу на преходну годину, а индекс је израчунат тако што се на 1 дода проценат раста подељен са 100. Израчунати просечну стопу раста продаје.

Година	Проценат раста	Индекс (x_i)
2001.	10%	1,10
2002.	8%	1,08
2003.	5%	1,05
2004.	15%	1,15
2005.	11%	1,11
2006.	9%	1,09
2007.	1%	1,01
2008.	-10%	0,9
2009.	-15%	0,85
2010.	-8%	0,92

Решење:

Колона индекс, заправо, садржи ланчане индексе, па због тога рачунамо геометријску средину. У овом случају

$x_1 \cdot x_2 \cdot x_3 \cdot ... \cdot x_{10} = 1{,}2337$ где је n = 10.

Примењујући формулу за геометријску средину, добијамо да она износи $G = \sqrt[10]{1{,}2337} = 1{,}0212$. Дакле, просечна стопа раста продаје овог предузећа за период од 2001. до 2010. године износи 1.0212, односно 2.12%.

Подсећање

Хармонијска средина употребљава се у оним случајевима када нумеричка вредност обележја и обим појаве стоје у обрнутој сразмери и када су вредности обележја за које треба израчунати средину изражене у виду реципрочних односа.

Проста хармонијска средина се израчунава на следећи начин:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Пондерисана хармонијска средина се израчунава на следећи начин:

$$H = \frac{f_1 + f_2 + ... + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + ... + \frac{f_n}{x_n}} = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

6. Аутомобил пређе 150 km брзиним од 60 km/h и врати се брзином 50 km/h. Израчунати средњу брзину кретања аутомобила.

Решење:

$$H=\frac{2}{\frac{1}{60}+\frac{1}{50}}=54,54$$

7. За израду једног производа први радник утроши 10 min, други радник потроши 6 min, а трећи 8,5 min. Израчунати просечно време утрошено за израду производа.

Решење:

$$H=\frac{3}{\frac{1}{10}+\frac{1}{6}+\frac{1}{8,5}}=7,8054$$

Подсећање

Модус је онај податак који се најчешће јавља у низу тј. има највећу фреквенцију. Нумерички низ може имати један, два или више модуса, а може се десити да модус не постоји.

Модус код интервалних низова података - Постоји модални интервал са највећом фреквенцијом који садржи вредност модуса и та вредност се израчунава по обрасцу:

$$M_o=x+i\frac{(f_2-f_1)}{(f_2-f_1)+(f_2-f_3)}$$

x – доња граница модалнох интервала

i – ширина модалног интервала

f_1 – фреквенција интервала који претходи модалном

f_2 – фреквенција модалног интервала

f_3 - фреквенција интервала који следи иза модалног

8. Одредити модус сђедећег низа података: 10, 7, 8, 13, 8, 17, 19, 12, 12, 10, 6, 7, 12, 10, 15, 12.

Решење: $M_o = 12$

9. У следећем низу података одредити модус: 14, 19, 19, 19, 23, 23, 23, 25, 25, 25, 27, 29, 30.

Решење: Овде се ради омултимодалном низу података са модусима $M_o=19$, $M_o=23$, $M_o=25$.

10. У табели је дата месечна потрошња неког артикла по домаћинствима. Израчунати најчешћу тј. Најфреквентнију потрошњу по домаћинству.

Потрошња	2 - 4	4 - 6	6 - 8	8 - 10	10 - 12	12 - 14
Број домаћинстава	8	10	21	19	15	6

Решење:

У овом низу модални интервал са највећом фреквенцијом је од 6 до 8, па су вредности следеће: $x=6$, $i=2$, $f_1=10$, $f_2=21$, $f_3=19$

$$M_o = 6 + 2\frac{(21-10)}{(21-10)+(21-19)}$$

$M_o = 7{,}692$

Најчешћа потрошња прехрамбеног производа по домаћинству је 7,692 килограма.

Подсећање

Медијана је средња вредност свих вредности обележја $x_1, x_2, x_3, \ldots, x_k$ уређених по величини. Код одређивања медијане морамо разликовати случајеве када је број вредности обележја паран и непаран.

Ако је број вредности обележја непаран број, онда постоји једна вредност која је средња вредност обележја. Ако је број вредности обележја паран број, тада постоје два средња члана и за медијану се узима њихова аритмеричка средина.

Медијана код интервалних серија са непарним бројем података израчунава се по обрасцу:

$$M_e = x_1 + \left(\frac{x_2 - x_1}{W_2 - W_1}\right)\left(\frac{\sum_{i=1}^{n} f_i}{2} - W_1\right)$$

Медијана код интервалних серија са парним бројем података израчунава се по обрасцу:

$$M_e = x_1 + \left(\frac{x_2 - x_1}{W_2 - W_1}\right)\left(\frac{\sum_{i=1}^{n} f_i + 1}{2} - W_1\right)$$

x_1 - доња граница медијалног интервала

x_2 – горња граница медијалног интервала

W_2 – збирна фреквенција медијалног интервала из растуће кумуланте

W_1 - збирна фреквенција пре медијалног интервала из растуће кумуланте

11. Одредити медијану следећег низа података: 18, 25, 19, 22, 26, 28, 30, 16, 36.

Решење:

Низ података се најпре уреди у растућем (неопадајућем) поретку:

16, 18, 19, 22, 25, 26, 28, 30, 36.

Позиција медијане је $\frac{N+1}{2} = \frac{9+1}{2} = 5$, што значи да је медијана вредност петог члана серије $M_e = 25$.

12. На случајан начин изабрано је 85 студената једног факултета и измерена је њихова висина. Израчунати медијану.

Висина	150-160	160-170	170-180	180-190	190-200	200-210
Број студената	5	15	30	24	8	3

Решење:

Висина	Број студената	Растућа кумуланта
150 - 160	5	5
160 - 170	15	20
170 - 180	30	50
180 - 190	24	74
190 - 200	8	82
200 - 210	3	85
Ukupno	$\sum f_i = 85$	

Место медијане је: $M_e = \frac{\sum_{i=1}^{n} f_i}{2} = \frac{85}{2} = 42,5$

Медијални интервал је (170 - 180), одакле је $x_1 = 170$, $x_2 = 180$, $W_2 = 50$, $W_1 = 20$

и најзад медијана је

$$M_e = 170 + (\frac{180-170}{50-20})(42,5-20) = 177,5 .$$

Половина студената има висину мању од 177,5 cm а друга половинаима висину већу од 177,5 cm.

Подсећање

Варијанса ($\boldsymbol{\sigma^2}$) представља просечно квадратно одступање података у низу од аритметичке средине тог низа. Њена вредност се креће од нуле до бесконачно. Варијанса се код простих серија израчунава по обрасцу

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n}$$

Варијанса се код низова датих расподелом фреквенција израчунава по обрасцу

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i(x_i - \overline{X})^2}{\sum_{i=1}^{n} f_i}$$

Стандардна девијација (σ) представља квадратни корен из варијансе.

За просте низове података:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n}},$$

а за расподеле фреквенција:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i - \overline{X})^2}{\sum_{i=1}^{n} f_i}} .$$

Коефицијент варијације (K_v) израчунава се по обрасцу obrascu

$$K_v = \frac{\sigma}{\overline{X}} .$$

13. На 40 парцела принос жита (*x_i*) у тонама по хектару (*t/ha)* дат је табелом расподеле фреквенција. Израчунати варијансу серије података.

Принос жита $t/_{ha}$	3.1 – 5	5.1 – 7	7.1 – 9	9.1 – 11	11.1 – 13	13.1 – 15
Број парцела	5	8	11	9	4	3

Решење:

Принос жита	Број парцела (f_i)	Средина интервала	$x_i\ f_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$f_i(x_i - \overline{x})^2$
3.1-5	5	4.05	20.25	-4.4	19.36	96.8
5.1-7	8	6.05	48.4	-2.4	5.76	46.08
7.1-9	11	8.05	88.55	-0.4	0.16	1.76
9.1-11	9	10.05	90.45	1.6	2.56	23.04
11.1-13	4	12.05	48.2	3.6	12.96	51.84
13.1-15	3	14.05	42.15	5.6	31.36	94.08
Укупно	$\sum f_i = 40$	/	338	/	/	313.6

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{338}{40} = 8.45$$

Просечан принос жита на 40 парцела износи 8.45 $t/_{ha}$.

$$\sigma^2 = \frac{\sum_{i=1}^{n} f_i(x_i-\overline{X})^2}{\sum_{i=1}^{n} f_i} = \frac{313.6}{40} = 7.84$$

Квадратно одступање приноса жиза од просечног приноса жита на 40 парцела је 7,84 .

14. Број угоститељских објеката по општинама је дат табелом. Груписати податке у облику интервалне нумеричке серије (низа).

5	20	15	35	48	17	38	39
7	19	20	40	46	28	35	41
9	17	19	38	39	31	42	29
10	16	30	42	50	42	25	19

Решење:

На који начин ћемо организовани овај низ података у групне интервале? Одговор на ово питање даје Стургесово правило или формула. Стургесова формула се користи за одређивање оптималног броја интервала *k* у које треба груписати *n* података:

$$k = 1+3,3 \log n$$

<u>Број интервала: Ширина интервала: Доња граница првог интервала:</u>

$k = 1+3.3\log 32$ $i = \frac{x_{\max} - x_{\min}}{k}$ $x_0 = x_{\min} - \frac{i}{2}$

$k = 1+3,3*1.50515$ $i = \frac{50-5}{5.96699}$ $x_0 = 5 - \frac{7.54}{2}$

$k = 1+4{,}96699$ $i = 7.54$ $x_0 = 1.23$

$k = 5.96699$ $i \approx 8$ $x_0 \approx 1$

$k \approx 6$

Број групних интервала је 6. Максималан број угоститељских радњи x_{max} је 50, а минималан број је x_{min} је 5. Ширина групног интервала је i ≈ 8 . Доња граница групног интервала је $x_0 \approx 1$.

На основу израчунатих вредности k, i, x_0 можемо да формирамо табелу груписања општина према броју угоститељских радњи.

Број угост. објеката	Број општина f_i	Средина интерв. x_i	Растућа кумуланта	Опадајућа кумуланта	Релативна фрекв. %	Растућа кумуланта у %
1 - 9	3	5	3	32	9.38	9.38
10 - 18	5	14	8	29	15.62	25
19 - 27	6	23	14	24	18.75	43.75
28 - 36	6	32	20	18	18.75	62.5
37 - 45	9	41	29	12	28.13	90.62
46 - 54	3	20	32	3	9.37	100
Укупно	$\sum fi$ =32	/	/	/	100	/

15. За груписане податке у интервалној нумеручкој серији у задатку бр. 14 израчунати сритметичку средину, модус, медијану и стандардну девијацију.

16. Потрошња млека у 24 домаћинства у литрима износила је:

17 9 9 20 24 30 30 13

9 9 10 12 12 14 15 13

7 7 9 9 14 15 15 17

Груписати податке у облику интервалне нумеричке серије. Израчунати медијану и варијансу.

17. Анкетирано је 30 купаца у продавницама о просечној дневној потрошњи хлеба у kg. Добијени су следећи подаци:

1 2 3 2 1 1 6 3 2 3 1 5 4 2 2

3 1 4 2 6 3 1 5 2 4 2 2 4 2 2

Груписати податке у облику интервалне нумеричке серије. Израчунати модус, варијансу и стандардну девијацију.

18. Компанија врши контролу свог производног процеса пратећи број произведених сатова у једном производном погону. Контрола производних линија једног погона дала је следеће резултате:

73 81 75 74 79 80 77 76 79 81

77 74 77 73 76 80 79 78 75 75

Груписати податке у облику интервалне нумеричке серије. Израчунати аритметичку средину, медијану, варијансу и стандардну девијацију.

19. Принос овса у t/ha кретао се према следећим подацима:

4,05 5,56 6,28 7,03 8,24 5,61 9,52 10,00

7,42 4,59 5,89 6,97 7,24 7,95 8,26 8,08

7,82 8,10 8,95 9,90 6,04 4,99 8,40 6,64

Груписати податке у облику интервалне нумеричке серије. Израчунати модус, варијансу и стандардну девијацију.

20. Величина газдинстава у хектарима у једној месној заједници приказана је следећим подацима:

1,10 2,10 1,15 3,00 2,59 3,15 4,72 4,83 5,00

4,87 4,70 3,15 5,05 2,,19 5,15 7,00 6,85 7,03

8,90 9,50 9,80 1,20 2,50 7,52 5,25 3,15 3,51

6,60 4,69 1,20 2,25 3,59 3,64 6,05 5,37 8,60

Груписати податке у облику интервалне нумеричке серије. Израчунати модус, медијану, варијансу и стандардну девијацију.

21. На писменом испиту из статистике дата су три задатка и сваки је максимално оцењен са 25 поена. Расподела 60 ученика према броју освојених бодова дата је табелом:

Поени	15 - 25	25 - 35	35 - 45	45 - 55	55 - 65	65 -75
Број ученика	8	12	20	10	6	4

а) Конструисати полигон и хистограм расподеле фреквенција и растућу кумуланту.

б) Одредити средњу вредност, модус, медијану и стандардну девијацију.

22. На пријемном испиту из информатике расподела 150 пријављених ученика према броју укупно освојених поена дата је табелом:

Поени	0 – 20	20 - 40	40 - 60	60 - 80	80 - 100	Укупно
Број ученика	20	35	60	25	10	150

а) Конструисати полигон и хистограм расподеле фреквенција и опадајућу кумуланту.

б) Одредити средњу вредност, модус, медијану, интервал варијације и варијансу.

23. Компанија врши контролу свог производње у једном производном погону аутомобилских делова. Контрола производних линија дала је следеће резултате:

74 81 75 74 79 80 77 76 79 81 76 85 74 84 76

77 74 77 74 76 80 79 78 75 75 81 82 83 77 82

а) Груписати податке у облику интервалне нумеричке серије.

б) Израчунати аритметичку средину, медијану, варијансу и стандардну девијацију.

24. На пријемном испиту на информационим технологијама дато је 20 задатака и сваки је максимално оцењен са 5 поена. Расподела 60 ученика према броју освојених бодова дата је табелом:

Поени	0-20	20-40	40-60	60-80	80-100
Број ученика	15	10	34	20	11

а) Конструисати полигон и хистограм расподеле фреквенција и кумуланту.
б) Израчунати аритметичку средину, варијансу и стандардну девијацију.

25. У евиденцији бироа за запошљавање дата је следећа табела расподеле незапослених по годинама старости.

Класе	15-25	25-35	35-45	45-55	55-65
Број незапослених	50	10	34	76	110

а) Конструисати полигон и хистограм расподеле фреквенција у процентима и растућу кумуланту.

б) Израчунати аритметичку средину, модус, варијансу и стандардну девијацију.

26. Дат је статистички скуп табелом:

Класе	54	55	56	57	58
Фреквенције	2	5	6	4	3

Нацртати хистограме и полигоне фреквенција (апсолутне и релативне), растућу и опадајућу кумуланту. Израчунати аритметичку средину података, дисперзију и стандардну девијацију.

Комбинаторика

Подсећање

Нека је дат скуп $A=\{a_1,a_2,...,a_n\}$.

Варијација без понављања *k*-те класе од *n* елемената је било која *k*-торка различитих елемената скупа *A*. Број варијација без понављања се израчунава формулом

$$V_k^n = n(n-1)\cdots(n-k+1)$$

Варијација са понављањем *k*-те класе од *n* елемената је било која *k*-торка елемената скупа А. Број варијација са понављањем дат је формулом

$$\overline{V}_k^n = n^k$$

27. Колико има двоцифрених бројева који се могу написати помоћу цифара 1, 2, 3?

Решење: Тражени бројеви су: 11, 12, 13, 21, 22, 23, 31, 32, 33 и има их укупно 3^2=9.

28. На колико се начина могу изабрати четири особе на четири различите дужности, од девет пријављених кандидата?

Решење: V_4^9= 9·8·7·6 = 3024

29. Колико се различитих четвороцифрених бројева може формирати ако се цифре у броју не понављају и на првом месту не може бити 0?

Решење: $V_4^{10} - V_3^9 =$ 10·9·8·7 - 9·8·7 = 4536

30. Инвеститор ће случајно изабрати 6 тржишта од 20 за инвестирање. На колико начина је то могуће да уради ако је редослед којим се бирају тржишта битан?

Решење: $V_6^{20} = 20\cdot19\cdot18\cdot17\cdot16\cdot15 = 27907200$

31. Колико у граду има телефона са петоцифреним бројевима:

а) ако су све цифре различите

б) ако се цифре понављају

Решење:

а) $V_5^{10} = 10\cdot9\cdot8\cdot7\cdot6 = 30240$

б) $\overline{V}_5^{10} = 10^5 = 100000$

32. Дат је скуп $A = \{1,2,3,4\}$.

а) Формирати све двоцифрене бројеве од елемената овог скупа, код којих се цифре не понављају и одредити њихов број.

б) Формирати све двоцифрене бројеве од елемената овог скупа и одредити њихов број.

Решење:

а) Тражени бројеви су: 12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42, 43 и њихов број је

$V_2^4 = 4\cdot3 = 12$

б) Тражени бројеви су: 11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43, 44 и њихов број је $\overline{V}_2^4 = 4^2 = 16$.

33. Одредити број речи од 5 слова које се могу написати помоћу азбуке од 30 слова без обзира да ли се у речима понављају слова и да ли се добијају речи без значења.

Решење: $\overline{V}_5^{30} = 30^5 = 24300000$

34. Колико се петоцифрених бројева може образовати од цифара 0,1,3,5,7,9, ако се 0 не налази на последњем месту и ако се ниједна цифра не понавља?

Решење: $V_5^6 - V_4^5 = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 - 5 \cdot 4 \cdot 3 \cdot 2 = 720 - 120 = 600$

35. Колико се Морзеових речи може формирати из основних знакова · и –, ако се једна реч састоји из највише 5 елементарних знакова?

Решење: $\overline{V}_1^2 + \overline{V}_2^2 + \overline{V}_3^2 + \overline{V}_4^2 + \overline{V}_5^2 = 2^1 + 2^2 + 2^3 + 2^4 + 2^5 = 62$

36. Број елемената четврте класе варијација са понављањем од x елемената износи 50625. Одредити број елемената x.

Решење:

$$\overline{V_4^x} = x^4 = 50625$$
$$x = \sqrt[4]{50625} = 15$$

37. Број варијација друге класе са понављањем од n елемената већи је за 15 од броја варијација друге класе без понављања од истог броја елемената. Одредити број елемената n.

Решење:

$$\overline{V_2^n} = V_2^n + 15$$
$$x^2 = \frac{x(x-1)}{2} + 15$$
$$2x^2 - x^2 + x = 30$$
$$x^2 + x = 30$$
$$x(x+1) = 30$$
$$x = 5$$

38. Доказати да је $V_{n-1}^n = V_n^n = n!$.

Решење: $V_{n-1}^n = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 2 = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 2\cdot 1 = V_n^n = n!$

39. Колико се различитих четвороцифрених бројева може формирати од десет цифара ако се цифре у броју понављају и на првом месту не може бити 0?

Решење: $\overline{V}_4^{10} - \overline{V}_3^{9} =$ 10·10·10·10 - 10·10·10 = 9·10·10·10 = 9000

40. Одредити број различитих шестоцифрених бројева који се могу формирати од елемента скупа E={1, 2, 3, 4, 5, 6, 7} тако да се цифре не понављају, а да крајње две цифре буду парне.

Решење: $V_4^5 \cdot V_2^3 = 5\cdot 4\cdot 3\cdot 2\cdot 3\cdot 2 = 720$

41. Одредити број различитих петоцифрених бројева који се могу формирати од елемента скупа E={1, 2, 3, 4, 5, 6, 7} тако да се цифре не понављају, а да је збир крајњих цифара паран број.

Решење: Збир крајњих цифара је паран ако су сабране две парне или две непарне цифре. У скупу E има 4 непарних и три парне цифре. Дакле,

$$V_3^5(V_2^3 + V_2^4) = 5\cdot 4\cdot 3\cdot(3\cdot 2 + 4\cdot 3) = 1080$$

42. а) Колико има петоцифрених бројева код којих су све цифре различите? (Број не може почети нулом!)

б) Колико има петоцифрених бројева код којих цифре не морају бити различите? (Број не може почети нулом!)

Решење:

а) Варијација без понављања пете класе од 10 елемената таквих да прва цифра броја не може бити нула има:

$$V_5^{10} - V_4^9 = 9 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 27216$$

б) Варијација са понављањем пете класе од 10 елемената таквих да прва цифра броја не може бити нула има:

$$\overline{V_5^{10}} - \overline{V_4^9} = 10^5 - 10^4 = 100000 - 10000 = 9 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 90000$$

Подсећање

Нека је дат скуп међусобно различитих елемената $A = \{a_1, a_2, ..., a_n\}$.

Пермутација је било који распоред свих *n* елемената скупа *A*. Број пермутација скупа од *n* различитих елемената износи:

$$P(n) = n(n-1)...1 = n!$$

Симбол *n*! је скраћеница за записивање производа првих *n* узастопних природних бројева и чита се “ен факторијел”. По дефиницији је 0!=1.

Нека је дат скуп $A = \{a_1, a_2, ..., a_n\}$ од којих се неки понављају и то $k_1, k_2, ..., k_j$ пута. **Пермутација са понављањем** је било који распоред свих *n* елемената скупа *A*. Број пермутација скупа од *n* елемената са понављањем износи:

$$P_{k_1,k_2,\dots,k_j}(n) = \frac{n!}{k_1!k_2!\cdots k_j!}$$

43. На колико начина је могуће распоредити 5 особа на пет столица?

Решење: $P(5) = 5! = 5\cdot4\cdot3\cdot2\cdot1=12$

44. Колико различитих петоцифрених бројева се могу написати помоћу цифара 0, 1, 2, 3, 4, а да се цифре не понављају?

Решење: $P(5) – P(4) = 5! - 4! = 120 - 24 = 96$

45. Дат је скуп $A = \{1,2,3,4,5,6\}$.

а) Колико шестоцифрених бројева (над овим скупом без понављања цифара) почиње цифрама 1, 2 у датом поретку?

б) У колико шестоцифрених бројева (над овим скупом без понављања цифара) цифре 1, 2 стоје једна поред друге у датом поретку?

Решење:

а) Цифре 12 посматрамо као један сегмент који се налази на првом месту, па остале четири цифре треба разместити на четири места, дакле решење је

$P(4) = 4! = 24$

б) Цифре 12 посматрамо као један сегмент али се он сада може може наћи на пет места у броју тако да је одговор

$P(5) = 5! = 120$

46. Формирати све пермутације од елемената *a, b, b, c* и одредити њихов број.

Решење:

Пермутације су:
abbc,abcb,acbb,babc,bbac,bbca,bcba,bacb,bcab,cabb,cbab,cabb

а има их $P_2(4) \equiv \frac{4!}{2!} = 12$.

47. Дате су цифре 0, 0, 0, 0, 1, 1, 1. Колико има могућих пермутација од ових елемената?

Решење: $P_{4,3}(7) \equiv \frac{7!}{4!3!} = 35$

48. Колико пермутација од елемената $\{a, a, a, a, a, b, b, b, c\}$ почиње словом *a*, колико словом *b*, колико словом *c*, а колико словима *cab*?

Решење:

Словом *a* почиње: $P_{4,3}(8) \equiv \frac{8!}{4!3!} = 280$

Словом *b* почиње: $P_{5,2}(8) = \frac{(9-1)!}{5!\cdot(3-1)!} = \frac{8!}{5!\cdot 2!} = 168$

Словом *c* почиње: $P_{5,3}(8) = \frac{8!}{5!\cdot 3!} = 56$

Словима *cab* почиње: $P_{5,2}(8) = \frac{(9-3)!}{2!\cdot(3-1)!} = \frac{6!}{2!\cdot 4!} = 15$

49. Колико пермутација од елемената 1,2,2,2,3,3,3,3,4,4,4 почиње са 22?

Решење: $P_{4,3}(9)=\frac{(11-2)!}{4!\cdot 3!}=2520$

50. Колико пермутација од елемената 1,2,2,2,3,3,3,3,4,4,4 почиње са 313?

Решење: $P_{3,2,3}(8)=\frac{(11-3)!}{3!\cdot 2!\cdot 3!}=560$

51. У колико пермутација цифара 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 стоје непарне цифре 1, 3, 5, 7, 9 једна до друге

а) у датом поретку на почетку пермутације,

б) у датом поретку,

ц) у произвољном поретку на почетку пермутације?

д) у произвољном поретку?

Решење:

а) $P(5) = 5! = 120$

б) $P(6) = 6! = 720$

ц) $P(5) \cdot P(5) = 5!\cdot 5! = 14400$

д) $P(6) \cdot P(5) = 6!\cdot 5! = 86400$

52. Одредити број пермутација од латиничних слова речи MENADŽMENT?

Решење: Реч је о пермутацијама са понављањем и њихов број је $P_{2,2,2}(10)=\frac{10!}{2!\cdot 2!\cdot 2!}=453600$

53. Дате су цифре 0, 0, 1, 1, 1, 1, 2. Колико има могућих пермутација од ових елемената?

Решење: Реч је о пермутацијама са понављањем и њихов број је $P_{2,4}(7)=\frac{7!}{2!\cdot 4!}=105$

Подсећање

Нека је дат скуп међусобно различитих елемената $A=\{a_1,a_2,...,a_n\}$

Комбинација без понављања *k*-те класе од *n* елемената је било која неуређена *k*-торка различитих елемената скупа *A* где је $k<n$. Број комбинација износи

$$C_k^n=\binom{n}{k}=\frac{n!}{k!(n-k)!}$$

Израз $\binom{n}{k}$ чита се "н над к" и представља број свих подскупова датог скупа *A* који садржи *k* елемената.

Комбинација са понављањем *k*-те класе од *n* елемената је било која неуређена *k*-торка елемената скупа *A* где је $k<n$ од којих се неки елементи понављају. Број таквих комбинација износи

$$\bar{C}_k^n=\binom{n+k-1}{k}$$

54. На колико се начина од 20 ученика једног одељења може изабрати трочлана делегација?

Решење: $\binom{20}{3}$=1140

55. На колико се начина од 12 играча једног кошаркашког тима може изабрати прва петорка?

Решење: $\binom{12}{5}$=792

56. Из групе од 8 мушкараца и 5 жена треба изабрати 5 особа тако да међу њима буду бар 2 жене. На колико начина се то може учинити?

Решење: $\binom{5}{2}\binom{8}{3}+\binom{5}{3}\binom{8}{2}+\binom{5}{4}\binom{8}{1}+\binom{5}{5}$=881

57. Кошаркашки тим сачињавају 5 бекова, 4 центра и 3 крила. На колико начина се може формирати петорка ако у њој морају да играју бар два бека и бар један центар?

Решење: $\binom{5}{2}\binom{4}{1}\binom{3}{2}+\binom{5}{2}\binom{4}{2}\binom{3}{1}+\binom{5}{2}\binom{4}{3}+\binom{5}{3}\binom{4}{1}\binom{3}{1}+\binom{5}{3}\binom{4}{2}+\binom{5}{4}\binom{4}{1}=540$

58. На школској забави налази се 12 девојака и 15 младића. На колико начина је могуће од њих изабрати 4 пара за плес?

Решење: $C_4^{12}\cdot C_4^{15}=\binom{12}{4}\binom{15}{4}=\frac{12\cdot 11\cdot 10\cdot 9}{1\cdot 2\cdot 3\cdot 4}\cdot\frac{15\cdot 14\cdot 13\cdot 12}{1\cdot 2\cdot 3\cdot 4}=495\cdot 1365=675675$

59. Три члана пороте ће случајно бити изабрана од пет људи. Колико различитих комбинација је могуће?

Решење: $C_3^5=\binom{5}{3}=\frac{5\cdot 4\cdot 3}{1\cdot 2\cdot 3}=10$

60. У игри лото извлачи се 7 од могућих 39 бројева. Редослед извучених бројева не утиче на остваривање награде. Колико има могућих комбинација? Која је вероватноћа да добијете „седмицу“ на лотоу ако сте уплатили један листић?

Решење: $C_7^{39} = \binom{39}{7} = 15380937$

$$P(A) = \frac{1}{C_7^{39}} = \frac{1}{\binom{39}{7}} = \frac{1}{15380937} = 0{,}000000065$$

61. Инвеститор ће случајно изабрати 6 тржишта од 20 за инвестирање. На колико начина је то могуће да уради ако редослед којим се бирају тржишта није битан?

Решење: $C_6^{20} = \binom{20}{6} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 38760$

62. На шаховском турниру учествује 15 шахиста. Сваки треба да одигра партију са сваким. Колико ће партија бити одиграно?

Решење: $C_2^{15} = \binom{15}{2} = \frac{15 \cdot 14}{1 \cdot 2} = 105$

Биномна формула

Подсећање

Биномна формула је формула помоћу које се израчунава израз

(a+b)^n, где је $n \in N$.

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-1}b^2 + \ldots + \binom{n}{n-1}ab^n + \binom{n}{n}b^n = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$$

$$\binom{n}{k} = \frac{n(n-1)(n-2)...(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

$$\binom{n}{0} = \binom{n}{n} = 1$$

63. Одредити пети члан у развијеном облику бинома $\left(x^{\frac{1}{2}} - x^{\frac{2}{3}}\right)^{12}$.

Решење:

$$T_5 = \binom{12}{4}\left(x^{\frac{1}{2}}\right)^{12-4} \cdot \left(-x^{\frac{2}{3}}\right)^4 = 495x^{\frac{20}{3}}$$

64. Збир коефицијената првог, другог и трећег члана бинома $\left(x^2 + \frac{1}{x}\right)^n$ је 46. Одредити члан који не садржи x.

Решење:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} = 46$$

$$1 + n + \frac{n(n-1)}{2} = 46$$

$$n = 9$$

Бином гласи $\left(x^2 + \frac{1}{x}\right)^9$.

$$T_{k+1} = \binom{9}{k}\left(x^2\right)^{9-k} \cdot \left(\frac{1}{x}\right)^k = \binom{9}{k} x^{18-3k}$$

$$18 - 3k = 0$$

$$k = 6$$

Тражени члан је:

$$T_7 = \binom{9}{6} = \binom{9}{3} = \frac{9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3} = 84$$

65. Предузеће запошљава 20 радника. Запослени су одлучили да случајно изаберу три радника за одлазак на прославу, али тако да је редослед избора битан. Колико избора је могуће?

66. На забави се налази 25 девојака и 30 мушкараца. На колико начина можемо изабрати три пара за плес.

67. Доказати да у месту у нашој земљи са хиљаду становника живе бар две особе са истим иницијалима.

68. Колико има различитих четвороцифрених бројева дељивих са 5 ако ниједан број не садржи једнаке цифре?

69. Научно друштво има 25 чланова. На колико начина се могу изабрати председник, подпредседник и секретар тог друштва ако сваки члан друштва може имати највише једну функцију?

70. Ако се регистарске таблице на аутомобилима састоје од 2 слова азбуке и иза њих шестоцифреног броја (од 000000 до 999999), колики је број различитих таблица?

71. Колико има могућих комбинација у игри лото (од 39 бројева извлачи се 7)?

72. На колико начина се из комплета који садржи 52 различите карте може изабрати 6 карата, тако да међу њима буде бар једна карта из сваке од 4 боје?

73. У ормару се налази 10 различитих пари ципела. На колик о начина можемо изабрати 4 ципеле тако да међу њима буде бар један пар?

74. Из комплета од 52 карте извучено је 10 карата. Карте су извлачене без враћања, а редослед извлачења се не сматра битним. Колико има различитих избора? У колико случајева се међу извученим картама налази тачно један кец?

75. У кутији се налазе куглице нумерисане бројевима 1, 2, 3, . . ., 10. Из кутије се истовремено извлаче три куглице. У колико случајева ће збир бројева на извученим куглицама бити једнак 9 .

76. Колико у граду има телефона са петоцифреним бројевима:
а) ако сваки број чине различите цифре,
б) ако се цифре не понављају?

77. На шаховском турниру одиграно је 210 партија. Одредити број учесника, ако је сваки учесник одиграо партију са сваким.

78. У кампањи за изборе председнички кандидат мора да обиђе 7 од 15 градова у Србији. Да би постигао што бољи резултат он кампању мора да заврши у Београду. На колико различитих начина он то може да уради?

79. Одредити члан који у развијеном облику бинома не садржи x.

$$\left(x+x^{-2}\right)^{12}$$

80. Одредити члан који у развијеном облику наведеног бинома има промењљиву x на пети степен.

$$\left(x^{\frac{1}{3}}+x^{\frac{1}{2}}\right)^{11}$$

81. Коефицијенти четвртог и шестог члана у развијеном облику бинома односе се као 5:18. Одредити члан који не зависи од x.

$$\left(\frac{1}{x}+\sqrt{x}\right)^{n}$$

82. Одредити десети члан у развијеном облику бинома , ако је биномни коефицијент трећег члана једнак 105.

$$\left(9x+\frac{1}{\sqrt{3x}}\right)^{n}$$

83. Биномни коефицијент трећег члана у развијеном облику датог бинома је 78. Наћи члан који не садржи x.

$$\left(x\sqrt{x}+x^{-5}\right)^{n}$$

Вероватноћа

Подсећање

Елементарни догађај ω је сваки могући исход експеримента. Скуп свих елементарних догађаја односно простор исхода експеримента је сигуран догађај $\Omega=\{\omega_1, \omega_2, ..., \omega_n\}$.
Сваки подскуп скупа Ω назива се сложеним догађајем или догађајем. Догађаје обележавамо великим словима A, B, C,...
Догађај који се никада не може појавити при реализацији експеримента назива се немогућ догађај.

Класична дефиниција вероватноће: Нека скуп Ω садржи n елементарних догађаја који су дисјунктни и једнако могући. Ако скуп A садржи m елементарних догађаја, тада је вероватноћа $P(A)$ једнака $P(\mathbf{A})=\frac{m}{n}$, где је m број свих повољних исхода, а n број свих могућих исхода догађаја A.

84. Десет картица нумерисано је бројевима од 1 до 10. Извлаче се две картице истовремено.Наћи вероватноћу да је збир бројева на извученим картицама једнак 10.

Решење: $P(A) = \frac{4}{\binom{10}{2}} = \frac{4}{45}$

85. У кутији се налази 10 црвених и 6 белих куглица. Насумице извлачимо две куглице. Колика је вероватноћа да ће

а) извучене куглице бити различитих боја,

б) обе куглице бити беле.

Решење: а) $P(A) = \frac{\binom{10}{1}\binom{6}{1}}{\binom{16}{2}} = 0{,}5$ б) $P(B) = \frac{\binom{6}{2}}{\binom{16}{2}} = 0{,}125$

86. У кутији се налази 8 црвених и 6 плавих куглица. Насумице извлачимо 5 куглица. Колика је вероватноћа да ће међу њима бити тачно 3 плаве?

Решење: $P(A) = \frac{\binom{6}{3}\binom{8}{2}}{\binom{14}{5}} = 0{,}2797$

87. Бацамо 3 новчића један за другим. Наћи вероватноћу да ћемо добити два писма и један грб.

Решење: Број свих елементарних исхода је $2^3 = 8$. Повољни исходи су {ППГ, ПГП, ГПП}, дакле има их 3, па је тражена вероватноћа P(A) $= \frac{3}{8}$.

Могуће је и решење преко Биномне расподеле вероватноћа *X*: B(8; 0,5)

$$P\{x = 2\} = \binom{3}{2}\frac{1}{2^2}\cdot\frac{1}{2} = \frac{3}{8}$$

88. Телефонски број се састоји од 6 цифара. Ако се претпостави да постоје сви телефонски бројеви од 000 000 до

999 999, која је вероватноћа да у произвољно изабраном броју све цифре буду различите?

Решење: $P(A) = \frac{V_6^{10}}{\overline{V}_6^{10}} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5}{10^6} = 0{,}1512$

89. Студент је од 30 испитнх питања научио 24. На испиту је добио 3 питања. Колика је вероватноћа да је одговорио тачно на најмање 2 питања.

Решење: $P(A) = \frac{\binom{24}{2}\binom{6}{1}+\binom{24}{3}}{\binom{30}{3}} = 0{,}9064$

90. Од седмоцифрених бројева случајно се бира један. Одредити вероватноћу да се цифра 2 у његовом запису јавља тачно два пута, а на осталим местима су 1 или 3.

Решење: $P(A) = \frac{\binom{7}{2} \cdot 2^5}{9 \cdot 10^6} = 0{,}00007$

91. Из скупа четвороцифрених бројева у којима се не јављају 0 и 9 случајно се бира један. Одредити вероватноћу да се у том броју цифра 1 јавља тачно једном.

Решење: $P(A) = \frac{\binom{4}{1} \cdot 7^3}{8^4} = 0{,}3350$

92. Коцка је бачена 6 пута. Колика је вероватноћа да:

а) неће пасти ниједна шестица,

б) да падне тачно једна шестица?

Решење:

а) $P(A) = \frac{5^6}{6^6} = 0{,}3349$ б) $P(B) = \frac{\binom{6}{1}5^5}{6^6} = 0{,}4019$

93. Имамо 7 пертли различитих боја, од којих је једна црвена, а једна зелена. Наћи вероватноћу да ће црвена и зелена пертла бити једна поред друге, ако се пертле ређају на случајан начин на прав конац једна до друге.

Решење: Црвену и зелену пертлу посматрамо као једну, па отуд 6! Осим тога, две су могућности да стоје црвена па зелена И обрнуто, зато је 2 пута 6! То све се дели са бројем пермутација свих пертли 7!

$$P(A) = \frac{2 \cdot 6!}{7!} = 0{,}2857$$

94. У серији од пет сијалица једна је неисправна. Колика је вероватноћа да између 3 насумично изабране сијалице буде једна неисправна сијалица?

Решење: При решавању овог задатка користимо Биномну расподелу B(3; 0,2). Дакле, број понављања експеримента је 3, вероватноћа да је сијалица неисправна је 0,2 а да је исправна је 0,8. Одатле вероватноћа догађаја *A* да у 3 насумично изабране сијалице једна неисправна дата је са

$$P(A) = \binom{3}{1} 0{,}2 \cdot 0{,}8^2$$

95. Наћи вероватноћу да се при истовременом бацању три коцке добије збир мањи од 5.

Решење: Повољни исходи су {111,121,211,112}.

$$P(A) = \frac{4}{216} = 0{,}0185$$

96. Шта је вероватније добити при истовременом бацању три коцке: збир 11 или 12?

Решење: Укупан број могућности при бацању три коцке је 216.

За збир 11 повољне моућности су: (1,4,6), (1,5,5), (1,6,4), (2,3,6), (2,4,5), (2,5,4), (2,6,3), (3,2,6), (3,3,5), (3,4,4), (3,5,3), (3,6,2), (4,1,6), (4,2,5), (4,3,4), (4,4,3), (4,5,2), (4,6,1), (5,1,5), (5,2,4), (5,3,3), (5,4,2), (5,5,1), (6,1,4), (6,2,3),(6,3,2), (6,4,1).

Укупно је 27 могућности па је вероватноћа: $P(A)=\frac{27}{216}=\frac{1}{8}$

За збир 12 повољне моућности су: (1,5,6), (1,6,5), (2,4,6), (2,5,5), (2,6,4), (3,3,6), (3,4,5), (3,5,4), (3,6,3), (4,2,6), (4,3,5), (4,4,4), (4,5,3), (4,6,2), (5,1,6), (5,2,5), (5,3,4), (5,4,3), (5,5,2), (5,6,1), (6,1,5), (6,2,4), (6,3,3), (6,4,2), (6,5,1). Укупно је 25 могућности па је вероватноћа: $P(B) = \frac{25}{216}$.

Дакле, јасно је да је већа вероватноћа да падне збир 11.

97. Један играч са вероватноћом 0,5 добија у једној игри на срећу. Шта је вероватније да добије 4 од 5 игара или 6 од 8 игара?

Решење:

Број добијених игара представља променљиву која има Биномну расподелу:

$$p4=\binom{5}{4}0,5^5=\frac{5}{32},\quad p6=\binom{8}{6}0,5^8=\frac{7}{64}$$

Вероватније је да играч добије 4 од 5 игара, него 6 од 8 игара.

98. Бацамо динар 10 пута. Која је вероватноћа да се свих 10 пута појави грб?

Решење:

Број грбова у 10 бацања динара представља случајну променљиву *X* која има Биномну расподелу:

$$P\{x = 10\} = \binom{10}{10}\frac{1}{2^{10}} \cdot \frac{1}{2^{0}} = \frac{1}{1024} = 0{,}000977$$

Условна вероватноћа

Подсетник

Вероватноћа догађаја *A*, уколико знамо да се догађај *B* већ реализовао или уколико претпостављамо да ће се реализовати назива се **условна вероватноћа**.

Вероватноћа P(A/B) зове се условна вероватноћа догађаја *A* под условом и дефинише се са

$$P(A/\mathrm{B}) = \frac{P(AB)}{P(B)}, \quad \text{за } P(B)>0.$$

Ако су догађаји ***A* и *B* међусобно зависни** тада је:

$$P(AB)=P(A)\cdot P(B/A)=P(B)\cdot P(A/B)$$

Догађаји ***A* и *B* су међусобно независни** ако је

$$P(A/B)=P(A),\ P(B/A)=\mathrm{P}(B) \text{ или } P(AB)=P(A)\cdot P(B)$$

99. Колика је вероватноћа да ће се на коцки приликом бацања појавити паран број, под условом да је тај број мањи од 4?

Решење: Нека је *A* догађај појаве парног броја, а *B* појава броја мањег од 4.

$A=\{2,4,6\}\quad B=\{1,2,3\}\quad A\cap B=\{2\}\quad P(AB)=\frac{1}{6}\quad P(B)=\frac{3}{6}$

$$\boldsymbol{P(A/B)}=\frac{P(AB)}{P(B)}=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{1}{3}$$

100. У једном одељењу од 30 ученика, 12 носи наочаре, 8 пише левом руком, а 6 има обе особине. Колика је вероватноћа да случајно изабрани ученик пише левом руком, ако знамо да носи наочаре?

Решење: Нека је *A* догађај да ученик пише левом руком, а *B* догађај да ученик носи наочаре.

$$\boldsymbol{P(A/B)}=\frac{P(AB)}{P(B)}=\frac{\frac{6}{30}}{\frac{12}{30}}=\frac{1}{2}$$

101. У једној кутији налазе се 4 беле и 8 црних куглица, а у другој 3 беле и 9 црних. Извлачимо из сваке кутије по једну куглицу. Одредити вероватноћу да је из обе кутије извучена бела?

Решење: Нека је *A* догађај да је бела куглица из прве кутије, *B* догађај да је бела куглица из друге кутије. Догађаји *A* и *B* су независни.

$\boldsymbol{P(AB) = P(A) \cdot P(B)} = \frac{4}{12} \cdot \frac{3}{12} = \frac{1}{12} = 0{,}08$

102. При бацању две коцке посматрамо збир који се појављује на њима. Колика је вероватноћа да је збир 6, ако се зна да је збир паран број?

Решење: Нека се бацају две коцкице одједном и нека је A догађај да је збир на коцкицама 6 и B догађај да је збир паран број.

Повољни исходи: $A = \{15, 24, 33\}$, а $B = \{11, 13, 15, 22, 24, 26, 33, 35, 44, 46, 55, 66\}$.

Пресек догађаја A и B чини $\{15, 24, 33\}$.

$$P(AB) = \frac{3}{21} = \frac{1}{7}, \qquad P(B) = \frac{12}{21}, \quad P(A/B) = 0{,}25$$

Ако би се бацала једна коцкица два пута заредом онда би тражена вероватноћа била:

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{\frac{5}{36}}{\frac{18}{36}} = \frac{5}{18} = 0{,}28$$

103. Структура запослених у једној фирми по полу и степену образовања дата је следећом табелом:

Пол	Средње образовање	Више образовање	Високо образовање	Укупно
Женски	132	75	43	250
Мушки	88	70	42	200
Укупно	220	145	85	450

а) Колика је вероватноћа да ће случајно изабрана особа бити женског пола (A)?

б) Колика је вероватноћа да ће случајно изабрана особа имати средње образовање (B)?

в) Колика је вероватноћа да ће случајно изабрана особа бити женског пола (A) или имати средње образовање?

Решење:

а) $P(A) = \frac{250}{450} = 0,55$

б) $P(B) = \frac{220}{450} = 0,49$

в) $P(A+B) = P(A) + P(B) - P(AB) = \frac{250}{450} + \frac{220}{450} - \frac{132}{450} = 0,75$

104. При бацању две коцке наћи вероватноћу да падну:

а) два иста броја,

б) бројеви чији је збир 7,

в) бар једна јединица.

105. У кутији се налази 5 плавих, 6 црвених и 7 белих куглица. Из кутије случајно извлачимо три куглице оједном. Одредити вероватноћу да међу извученим куглицама буду:

а) све три куглице различитих боја,

б) две беле и једна плава.

106. Од 10 истоветних производа једен фабрике 6 је исправних. Насумоце се бира 5 производа. Колика је вероватноћа да ће међу извученим производима бити 3 исправна?

107. У кутији се налази 10 плавих и 6 белих куглица. Извлачимо 2 куглице. Колика је вероватноћа да ће:

а) извучене куглице бити различитих боја,

б) да ће обе куглице бити плаве,

в) да ће обе куглице бити беле?

108. Колика је вероватноћа да ћемо извлачењем 5 бројева на лутрији од 50 бројева извући бројеве 10, 21, 35?

109. Један студент је од 30 испитних питања научио 25, а други 15. На испиту су добили по три питања. Колика је вероватноћа да ће први, односно други одговорити на најмање 2 питања?

110. Бацамо новчић три пута. Нека је X случајна променљива која представља број писама.

а) Написати закон расподеле.

б) Наћи математичко очекивање.

111. Једна продавница продаје дневно 5 фрижидера. Ако је 0,7 вероватноћа да је један фрижидер овог типа буде исправан и после пет година употребе, наћи вероватноћу да после пет година буде 2 неисправних фрижидера.

112. На основу анализе 100 возила на техничком прегледу добијени су следећи резултати:

	Неисправне кочнице	Исправне кочнице	Укупно
Излизане гуме	5	25	30

У добром стању	10	60	70
Укупно	15	85	100

Под претпоставком да је овај узорак репрезентитативан за сва возила на друмовима, израчунајте вероватноћу да ће случајно изабрано возило имати:

а) Излизане гуме

б) Неисправне кочнице

в) Неисправне кочнице или излизане гуме

г) Излизане гуме под условом дас у кочнице неисправне

д) Добре гуме под условом да су кочнице исправне.

113. Узет је случајан узорак од 80 адвоката и постављено им је питање да ли су за или против смртне казне. У следећој табели су дате две класификације њихових одговора:

	За смртну казну	Против смртне казне
Мушко	32	24
Женско	13	11

Одредити вероватноћу да је случајно изабран адвокат:

а) за смртну казну

б) женско

в) против смртне казне под условом да је женско

г) женско и за смртну казну

д). против смртне казне или је мушко.

114. Агенција за заштиту потрошача је случајно изабрала 1700 летова две велике авио компаније А и Б. У следећој табели су дате две класификације ових летова.

	Касни мање од 30 минута	Касни д 30 минута до 1 сата	Касни више од 1 сата
Компанија А	429	390	92
Компанија Б	393	316	80

Ако је један лет изабаран случајним путем од ових 1700 летова, одредити вероватноћу да тај лет:

а) касни више од једног сата

б) касни мање од 30 минута

в) касни више од једног сата под условом да је то лет компаније Б.

115. У једном шеширу налази се пет зелених, осам црвених ис седам плавих кликера. Ако случајно извалчимо један кликер из шешира, наћи вероватноћу догађаја А да је извучен црвени кликер. Који је комплементарни догађај за А и колика је његова вероватноћа?

Тачкасте оцене и интервали поверења за аритметичку средину

116. Пре него што одлучи по којој цени ће издавачка кућа продавати нову књигу за факултете, компанија жели да зна просечну цену свих књига таквог типа у продаји. Истраживачко одељење фирме је узело узорак од 25 одговарајућих уџбеника и прикупило податке о њиховим ценама. Ово истраживање је показало да је просечна цена у узорку 90,5 Еура. Познато је да је стандардна девијација цена свих оваквих књига 7,5 Еура и да основни скуп таквих цена има Нормалну расподелу. Конструишите 90% интервал поверења за просечну цену свих оваквих уџбеника за факултете.

Решење:

Како нам је σ познато и величина узорка мања од 30, а основни скуп нормално расподељен, за конструисање интервала поверења користимо Нормалну расподелу.

$n = 25$, $\overline{X} = 90,5$, $\sigma = 7,5$

Ниво поузданости је 90%. Прво нађимо z вредност за ниво поузданости од 90%. Површина на сваком крају нормалне криве је $\alpha/2 = (1 - 0,90)/2 = 0,05$. У таблици Нормалне расподеле потражимо површину 0,05 и пронађемо вредност за $z = 1,65$.

90% интервал поверења за μ је:

$$\left[\overline{X} - z\frac{\sigma}{\sqrt{n}}, \overline{X} + z\frac{\sigma}{\sqrt{n}}\right] = \left[90,50 - 1,65(1,50), 90,50 + 1,65(1,50)\right] = \left[88,02; 92,98\right]$$

Тако смо 90% сигурни да је просечна цена свих уџбеника између 88,02 и 92,98 Еура.

117. Удружење бивших студената жели да оцени просечан дуг овогодишњих дипломаца. Зна се да је стандардна девијација скупа дуговања овогодишњих дипломаца 11 800 динара. Колико велики узорак би требало изабрати да би оцена била до 800 динара од аритметичке средине скупа уз ниво поузданости 99%?

Решење:

Удружење тражи да 99% интервал поверења за просечни дуг овогодишњих дипломаца буде $\left[\overline{X}-800, \overline{X}+800\right]$. Максимална вредност маргиналне грешке је 800 динара, тј. Е = 800.

Вредност з за ниво поузданости од 99% је -2,58. Вредност σ је дата износи 11 800 динара. Тражену величину узорка добијамо из формуле

$$n=\left(\frac{z\sigma}{E}\right)^2=\left(\frac{(2,58)(11.800)}{800}\right)^2=1.448,18\approx 1.449$$

Дакле, неопходна величина узорка је 1449. Ако удржење бивших студената узме узорак од 1449 студената који су дипломирали ове године, израчуна аритметичку средину дуга у узорку, затим формира 99% интервал поверења око аритметичке средине овог узорка, маргинална грешка оцене ће бити приближно 800 динара.

118. Узет је узорак од 25 људи једног града како би се оценио просечан ниво холестерола. Показано је да је просечан ниво холестерола у овом узорку 186 једуница са стандардном девијацијом 12. Претпоставимо да је ниво холестерола код свих људи тог града приближно нормално расподељен. Конструишите 95 % интервал поверења за аритметичку средину скупа μ.

Решење:

Како нам σ није познато и величина узорка мања од 30, а основни скуп нормално расподељен, за конструисање интервала поверења користимо t -расподелу.

$n = 25$, $\overline{X}=186$, $s = 12$

Ниво поузданости је 95%. Да бисмо пронашли вредност t, потребно је да знамо број степени слободе и површину на сваком крају испод криве t расподеле.

Број степени слободе је $df=n-1=24$. Површина на сваком крају је 0,025. Из таблице t расподеле добијамо да је t = 2,064. 95% интервал поверења за μ је:

$$\left[\overline{X}-t\frac{s}{\sqrt{n}},\overline{X}+t\frac{s}{\sqrt{n}}\right]=\left[186-2,064(2,40),186+2,064(2,40)\right]=\left[181,05;190,95\right]$$

Тако можемо тврдити са поузданошћу 95% да је просечни ниво холестерола код свих људи у датом месту између 181,05 и 190,95.

119. Маркетиншки истраживач жели да нађе 98% интервал поверења за просечан износ који посетиоци неког забавног парка потроше дневно по особи. Стандардна девијација дневно потрошених износа по особи за све посетиоце овог парка је 15 долара. Израчунати интервалну оцену аритметичке средине скупа за узорак обима 52, ако је просечан износ који су потрошили посетиоци из узорка 75$?

Решење:

Како нам је σ познато и величина узорка већа од 30 за конструисање интервала поверења користимо Нормалну расподелу, n=52 и σ=15.

Ниво поузданости је 98%. Прво нађимо z вредност за ниво поузданости од 98%. Површина на сваком крају нормалне криве је $\alpha/2 = (1-0,98)/2 = 0,01$. У таблици Нормалне расподеле потражимо површину 0,01 и пронађемо вредност за z = -2,33.

98% интервал поверења за μ је:

$$\left[\overline{x_{52}}-z\frac{\sigma}{\sqrt{52}},\overline{x_{52}}+z\frac{\sigma}{\sqrt{52}}\right]=[75-2{,}33\frac{15}{7{,}21},75+2{,}33\frac{15}{7{,}21}]=[75-4{,}85;75+4{,}85]=[70{,}15;79{,}85]$$

Тако смо 98% сигурни да је просечан износ који су потрошили сви посетиоци имеђу 70,15 и 79,85. Тачкаста оцењена вредност просечног износа је 75$, а маргинална грешка је 4,85$.

120. Служба маркетинга жели да нађе 95% интервал поверења за просечан износ који посетиоци једног тржног центра потроше дневно по особи. Стандардна девијација дневно потрошених износа по особи за све посетиоце овог тржног центра је 11 Евра. Израчунати интервалну оцену аритметичке средине скупа за узорак обима 36, ако је просечан износ који су потрошили посетиоци из узорка 55 Евра?

Решење:

Како нам је σ познато и величина узорка већа од 30 за конструисање интервала поверења користимо Nормалну расподелу, n=36 и σ=11. Ниво поузданости је 95%. Прво нађимо z вредност за ниво поузданости од 95%. Површина на сваком крају нормалне криве је α/2 = (1-0,95)/2 = 0,025. У таблици Нормалне расподеле потражимо површину 0,025 и пронађемо вредност за z = -1,96. 95% интервал поверења за μ је:

$$\left[\overline{x_{36}}-z\frac{\sigma}{\sqrt{36}},\overline{x_{36}}+z\frac{\sigma}{\sqrt{36}}\right]=\left[55-1{,}96\frac{11}{6},55+1{,}96\frac{11}{6}\right]$$
$$=[55-3{,}59;\ 55+3{,}59]=[51{,}41;\ 58{,}59]$$

Тако смо 95% сигурни да је просечан износ који су потрошили сви посетиоци имеђу 51,41 и 58,59. Тачкаста оцењена вредност просечног износа је 55 Евра, а маргинална грешка је 3,59 Евра.

121. У једном чланку је писало да су играчи бејзбола били плаћени до 150 долара по аутограму на зимском сајму. Претпоставимо да је у случајном узорку од 80 таквих аутограма аритметичка средина била 135 долара по аутограму, а да стандардна девијација основног скупа износи 22 долара. Конструишите 95% интервал поверења за одговарајућу аритметичку средину скупа да би сте проценили да ли је напис у чланку био истинит.

122. Узорак од 25 поруџбина је показао да је просечно време за испоруку производа неке фирме 70 сати. Претпоставимо да је стандардна девијација основног скупа 16 часова и да је расподела основног скупа Нормална. Конструишите 95 % интервал поверења за просечно време неопходно за испоруку свих поруџбина које су пристигле у ову фирму.

123. Менаџер одељења у маркету жели да оцени, са нивоом поузданости од 90%, просечну количину новца коју потроше сви потрошачи у овој продавници. Стандардна девијација потрошеног новца свих муштерија у овој радњи је 31 долар. Коју величину узорка би требало да изабере да би оцена била до 3 долара од аритметичке средине скупа?

Решење:

Максимална вредност маргиналне грешке је 3 долара, тј. E = 3.

Вредност z за ниво поузданости од 90% је 1,64. Вредност σ је дата износи 31 долар. Тражену величину узорка добијамо из формуле $n = \left(\frac{z\,\sigma}{E}\right)^2$.

За $\alpha/2=0{,}05$ у таблицама Нормалне расподеле налазимо вредност за z=1,64. Одатле, $= \left(\frac{1{,}64\cdot 31}{3}\right)^2 = 287{,}19$.

Дакле,неопходна величина узорка је 288.

124. У узорку од 100 опсервација изабраном из неког скупа аритметичка средина је 55,32 , а стандардна девијација основног скупа је 8,4. Формирајте 90 % интервал поверења за аритметичку средину.

Решење:

Како нам је σ познато и величина узорка већа од 30 за конструисање интервала поверења користимо Нормалну расподелу, n = 100, σ = 8,4. Ниво поузданости је 90%. Прво нађимо z вредност за ниво поузданости од 90%. Површина на сваком крају нормалне криве је α/2 = (1-0,90)/2 = 0,05. У таблици ормалне расподеле потражимо површину 0,05 и пронађемо вредност за z = 1,64.

90% интервал поверења за μ је:

$$\left[\overline{x_{100}} - z\frac{\sigma}{\sqrt{n}}, \overline{x_{100}} + z\frac{\sigma}{\sqrt{n}}\right] = \left[55{,}32 - 1{,}64\frac{8{,}4}{\sqrt{100}},\ 55{,}32 + 1{,}64\frac{8{,}4}{\sqrt{100}}\right] = [55{,}32 - 1{,}3776,\ 55{,}32 + 1{,}3776] = [53{,}94\ ;\ \ 56{,}70]$$

125. Узет је узорак од 15 боца сока да би се оценила просечна нето маса једне боце целе производње. Ако се зна да је расподела нормална, са стандардном девијацијом 1,95, оценити просечну нето масу боце сока са поузданошћу 95% на основу добијених резултата мерења нето масе боца сока у грамима:

995,2	996,3	996,9	997,5	997,9
998,3	998,5	998,7	999,1	999,4
999,8	1000,0	1001,3	1001,5	1003,4

Колика је тачност добијене оцене и колика је дужина интервала поверења?

Решење:

Како нам је σ познато и обим узорка је мањи од 30 за конструисање интервала поверења користимо Нормалну расподелу. Ниво поузданости је 95%. Прво нађимо z вредност за ниво поузданости од 95%. Површина на сваком крају нормалне криве је $\alpha/2 = (1\text{-}0{,}95)/2 = 0{,}025$. У таблици нормалне расподеле потражимо површину 0,025 и пронађемо вредност за $z = -1{,}96$.

Из узорка израчунавамо да је $\overline{x_{15}} = 998{,}92$.

$$\left[\overline{x_{15}} - z\frac{\sigma}{\sqrt{n}}, \overline{x_{15}} + z\frac{\sigma}{\sqrt{n}}\right]$$
$$= \left[998{,}92 - 1{,}96\frac{1{,}95}{\sqrt{15}},\ \ 998{,}92 + 1{,}96\frac{1{,}95}{\sqrt{15}}\right]$$
$$= [998{,}92 - 0{,}99\ ,\ \ 998{,}92 + 0{,}99]$$
$$= [997{,}93\ ,\ \ 999{,}91]$$

Тачкаста оцењена вредност је 998,92 док је тачност добијене оцене заправо маргинална грешка и она износи 0,99. Дужина интервала поверења је 2E = 1,98.

126. У једном чланку је писало о плати статиста по емисији на телевизији. Претпоставимо да је у случајном узорку од 60 таквих емисија аритметичка средина била 600 динара по емисији, са стандардном девијацијом 30 динара. Конструишите 98% интервал поверења за одговарајућу аритметичку средину скупа.

Решење:

Како нам је σ непознато и величина узорка већа од 30 за конструисање интервала поверења користимо Студентову t-расподелу, n = 60, S = 30. Ниво поузданости је 98%. Прво нађимо t вредност за ниво поузданости од 98%. Површина на сваком крају криве t расподеле је α/2 = (1-0,98)/2 = 0,01. У таблици Студентове t расподеле потражимо површину 0,01 и пронађемо вредност за t = 2,39. 98% интервал поверења за μ је:

$$\left[\overline{x_{60}} - t\frac{S}{\sqrt{n}}, \overline{x_{60}} + t\frac{s}{\sqrt{n}}\right] = \left[600 - 2{,}39\frac{30}{\sqrt{60}},\ \ 600 + 2{,}39\frac{30}{\sqrt{60}}\right]$$
$$= [590{,}74\ ;\ 609{,}26]$$

Маргинална грешка је E=9,25, а тачкаста оцењена вредност је 600 дин.

127. Узорак од 25 поруџбина у пицерији показао је да просечно време за испоруку пице на тражену адресу износи 30 минута. Претпоставимо да је стандардна девијација скупа 10 минута и да је расподела основног скупа Нормална. Конструишите 95 % интервал поверења за просечно време неопходно за испоруку свих поруџбина које су пристигле у ову пицерију.

Решење:

Како нам је σ познато и величина узорка мања од 30 потребно је да основни скуп буде Нормално расподељен што је испуњено. Тако за конструисање интервала поверења користимо Нормалну расподелу, n=25, σ=10. Ниво поузданости је 95%. Прво нађимо z вредност за ниво поузданости од 95%. Површина на сваком крају нормалне криве је α/2 = (1-0,95)/2 = 0,025. У таблици Нормалне расподеле потражимо површину 0,025 и пронађемо вредност за z = -1,96. Тачкаста оцењена вредност просечног времена испоруке је 30 минута. Треба израчунати маргиналну грешку.

95% интервал поверења за μ је:

$$\left[\overline{x_{25}} - z\frac{\sigma}{\sqrt{n}}, \overline{x_{25}} + z\frac{\sigma}{\sqrt{n}}\right] = \left[30 - 1{,}96\frac{10}{\sqrt{25}},\ \ 30 + 1{,}95\frac{10}{\sqrt{25}}\right]$$
$$= [30 - 3{,}92,\ \ 30 + 3{,}92] = [26{,}08\,;\ \ 33{,}92]$$

Тако смо 95% сигурни да је просечно време неопходно за испоруку свих поруџбина које су пристигле у ову пицерију између 26,08 и 33,92 а маргинална грешка је 3,92 минута.

128. У узорку од 125 опсервација изабраном из неког скупа аритметичка средина је 35,32 , а стандардна девијација 7,4. Формирајте 95% интервал поверења за аритметичку средину.

Решење:

Како нам је σ непознато и величина узорка већа од 30 за конструисање интервала поверења користимо Студентову *t* расподелу, *n*=125, *s*=7,4. Ниво поузданости је 95%. Прво нађимо *t* вредност за ниво поузданости од 95%. Површина на сваком крају криве *t* расподеле је α/2=(1-0,95)/2=0,025. У таблици Студентове *t* расподеле потражимо површину 0,025 и пронађемо вредност за *t*=1,98. 95% интервал поверења за μ је:

$$\left[\overline{x_{125}} - t\frac{S}{\sqrt{n}}, \overline{x_{125}} + t\frac{s}{\sqrt{n}}\right]$$
$$= \left[35{,}32 - 1{,}98\frac{7{,}4}{\sqrt{125}},\ \ 35{,}32 + 1{,}98\frac{7{,}4}{\sqrt{125}}\right] =$$

$$[35{,}32 - 1{,}31\,; 35{,}32 + 1{,}31] = [34{,}01\,; 36{,}63]$$

Маргинална грешка је Е=1,31, а тачкаста оцењена вредност је 35,32.

129. Просечни месечни рачун за струју у јденом граду износио је 77 Еура. Овај просек израчунат је на основу случајног узорка од 500 таквих рачуна, при чему је стандардна девијација овог узорка 26 еура. Формирајте 99 % интервал поверења за просек свих месечних трошкова за струју у том граду.

Решење:

Како нам је σ непознато и величина узорка већа од 30 за конструисање интервала поверења користимо Студентову t расподелу, n=500, S=26. Ниво поузданости је 99%. Прво нађимо z вредност за ниво поузданости од 99%. Површина на сваком крају нормалне криве је α/2=(1-0,99)/2=0,005. У таблици Студентове расподеле потражимо површину 0,005 и пронађемо вредност за t=2,576. Тачкаста оцењена вредност рачуна је 77 долара. Треба израчунати маргиналну грешку. 95% интервал поверења за μ је:

$$\left[\overline{x_{500}} - t\frac{s}{\sqrt{n}}, \overline{x_{500}} + t\frac{s}{\sqrt{n}}\right] = \left[77 - 2{,}576\frac{26}{\sqrt{500}}, \quad 77 + 2{,}576\frac{26}{\sqrt{500}}\right] = [77 - 2{,}99\,; 77 + 2{,}99] = [74{,}00\,; 79{,}99]$$

Напомена: Како је узорак веома велики, а у том случају Студентова t расподела приближно је једнака стандардизованој Нормалној расподели, у овом случају се може користити и Нормална расподела за одређивање интервала поверења.

130. Фирма која производи детерџенте жели да оцени просечну количину детерџента у врећама од 3 kg уз ниво поузданости од 99%. Фирма зна да стандардна девијација количина детерџента у свим таквим врећама износи 0,01 kg. Колико велики узорак би требало да узме фирма да би оцена била до 0,002 kg од аритметичке средине скупа?

131. Случајан узорак од 36 аутомобила средње класе тестираних на потрошњу горива дао је просечну вредност од

26,4 km по литру са стандардном девијацијом узорка од 2,3 km по литру. Конструишите 99% интервал поверења за аритметичку средину скупа.

132. Фирма која продаје компјутерске делове поштом гарантује својим купцима да испоруке врше одмах по пристизању налога. Узорак од 25 последњих поруџбина је показао да је просечно време за испоруку било 70 сати. Претпоставимо да је стандрдна девијација скупа 16 часова и да је расподела основног скупа нормална. Конструишите 95% интервал поверења за одговарајућу аритметчку средину скупа.

133. Директор једне школе је забринут због времена које његови ђаци троше радећи и хтео би да оцени просечан број сати које ђаци проведу на послу недељно. Зна се да је стандардна девијација времена потрошеног недељно на овим пословима израчуната на основу узорка 2,5 сати. Коју величину узорка би требао да изабере да би оцена била до 0,75 сати од аритметичке средине скупа. Директор жели да користи ниво поузданости од 98%.

134. Просечни месечни рачун за услуге мобилне телефоније износио је 1083 динара. Овај просек израчунат је на основу
случајног узорка од 500 таквих рачуна, при чему је стандардна девијација овог узорка 26 динара. Формирајте 99% интервал поверења за просек свих месечних трошкова за мобилно телефонирање.

Тестирање хипотезе о аритметичкој средини једног основног скупа

135. Утврђено је да је просечна дужина међународних разговора који су преко телефонске компаније обављени у једној години била 12,44 минута. Управа компаније је желела да провери да ли се просечна дужина разговора разликује од

12,44 мин. У узорку од 150 таквих разговора просечна дужина је износила 13,71 минут. Стандардна девијација свих разговора је 2,65 минута. Да ли на нивоу значајности од 2 % може да се закључи да се просечна дужина разговора разликује од 12,44 минута?

Решење: Нека је μ просечна дужина разговора који се обављају преко ове компаније, а $\overline{x_n}$ просечна дужина разговора у узорку.

$n = 150$, $\overline{x_n} = 13{,}71$, $\sigma = 2{,}65$, $\alpha = 0{,}02$

Формирамо нулту и алтернативну хипотезу:

H_0: $\mu = 12{,}44$

H_1: $\mu \neq 12{,}44$

Како је σ познато и величина узорка већа од 30, за тестирање хипотезе користимо Нормалну расподелу. Како је знак у алтернативној хипотези $\neq$, за тестирање користимо двострани тест. Из таблице Нормалне расподеле, проналазимо критичне вредности -2,33 и 2,33.

Израчунавамо статистику теста $t = \frac{\overline{x_n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{13{,}71 - 12{,}44}{\frac{2{,}65}{\sqrt{150}}} = 5{,}87$

Како се 5,87 налази у области одбацивања нулте хипотезе, закључак је да се нулта хипотеза одбацује и прихвата алтернативна. Дакле, на нивоу значајности од 2% можемо да закључимо да се просечна дужина разговора разликује од 12,44 минута.

136. Градоначелник једног великог града тврди да просечна нето вредност имовине породица које живе у том граду износи најмање 300 000$. У изабраном узорку од 25 породица, просечна нето вредност имовине износи 288 000$. Претпоставимо да нето вредности имовине свих породица у овом граду имају нормалну расподелу, са стандардном девијацијом узорка 80 000 $. Може ли се закључити на нивоу значајности од 2,5 % да је градоначелникова тврдња неистинита?

Решење:

Формирамо нулту и алтернативну хипотезу:

H_0: $\mu \geq 300\ 000$

H_1: $\mu < 300\ 000$

Како је σ није познато и величина узорка мања од 30, а основни скуп има нормалну расподелу, за тестирање хипотезе користимо т расподелу. Како је знак у алтернативној хипотези < за тестирање користимо левострани тест. Из таблице *t* расподеле, проналазимо критичну вредност -2.064.

$$t = \frac{\overline{x_n} - \mu}{\frac{s}{\sqrt{n}}} = \frac{288000 - 300000}{\frac{80000}{\sqrt{25}}} = -0{,}75$$

Како се - 0,75 > - 2.064 налази у области неодбацивања нулте хипотезе, закључак је да се нулта хипотеза не одбацује. Дакле, на нивоу значајности од 2,5 % можемо да закључимо да је градоначелникова тврдња тачна.

137. Постоји сумња да машина која пакује детерџент, чија је прописана маса 1 кг, није више прецизна, па би требало извршити њен ремонт. Да бисмо донели одлуку о ремонту, изабрали смо узорак од 16 паковања и добили следећи резултат:

1,020 1,010 1,050 1,015 1,002 1,008 1,025 0,998

1,012 1,033 1,017 1,001 1,008 1,011 1,024 1,006

На основу добијених података проверићемо да ли је сумња у прецизност рада машине оправдана. Из искуства знамо да је маса паковања нормално распоређена. Шта се мозе закључити када је ниво значајности теста α=0,05?

Решење:

$$\mu = 1\,kg,\ n = 16,\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{16{,}240}{16} = 1{,}015\ kg$$

Формулисање нулте и алтернативне хипотезе:

H_0: $\mu = 1$ (Просечна маса паковања је 1кг.)

H_1: $\mu \neq 1$ (Просечна маса паковања је различита од 1кг.)

Избор расподеле која ће се користити се врши на основу критеријума: σ непозната, n<30, па ћемо за тестирање хипотезе користити t расподелу са n-1 степени слободе. Одређивање области одбацивања и неодбацивања нулте хипотезе се врши на основу знака ≠ у алтернативној хипотези који указује да је тест двострани, са две области одбацивања које су симетрично распоређене на крајевима криве т расподеле. Ниво значајности је 0,05 па су површине које одговарају областима одбацивања су на сваком крају α/2=0,05/2=0,025.

Критичне вредности су $t_{15;\,0{,}025} = \pm\, 2{,}131$, za α=0,025 i df =16 - 1=15.

Следећи корак је израчунавање вредности статистике теста. Сада израчунавамо реализовану вредност статистике t теста

$$s_{\overline{X}} = \sqrt{\frac{\sum x_i^2 - n\overline{x}^2}{n(n-1)}} = \sqrt{\frac{16{,}486 - 16 \cdot 1{,}015^2}{16 \cdot 15}} = 0{,}0032$$

$$t = \frac{\overline{x_n} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1{,}015 - 1{,}0}{0{,}0032}\sqrt{16} = 18{,}75$$

Реализована вредност статистике је $t = 18{,}75$

На основу реализоване вредности статистике t теста доносимо одлуку о одбацивању или неодбацивању нулте хипотезе.

Критеријум је: H_0 не треба одбацити ако је $|t| \leq 2{,}1315$

H_0 треба одбацити ако је $|t| > 2{,}1315$

Ова вредност t=18,75 је већа од критичне вредности $t = 2{,}1315$ и налази се у области одбацивања нулте хипотезе. Следи да нулту хипотезу одбацујемо и закључујемо да, на основу података у узорку, просечна маса кутије детерџента није једнака 1кг, при ризику од 0,05 да смо одбацили истиниту нулту хипотезу.

138. Исти задатак као претходни, 137. али сумњамо да је маса кутије детерџента већа од 1 кг. Тестирати хипотезу на нивоу значајности од 5%.

Решење:

Формулисање нулте и алтернативне хипотезе:

H_0: $\mu \leq 1$ (Просечна маса паковања није већа од 1кг.)

H_1: $\mu > 1$ (Просечна маса паковања је већа од 1кг.)

Избор расподеле која ће се користити: σ непозната, n<30, па ћемо за тестирање хипотезе користити t расподелу са n-1 степени слободе.

Одређивање области одбацивања и неодбацивања: Знак > у алтернативној хипотези указује да је тест деснострани, са једном облашћу одбацивања нулте хипотезе која је на десном крају криве т расподеле. Ниво значајности је 0,05. Површина која одговара области одбацивања је α=0,05. Критична вредност је $t_{15;\,0,05}$ =1,753, и за α=0,05 и df =16-1=15 је очитана из таблица t расподеле.

Сада израчунавамо реализовану вредност статистике t теста

$$\mu = 1\,kg,\ n = 16,\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{16{,}240}{16} = 1{,}015\ kg$$

$$s_{\overline{X}} = \sqrt{\frac{\sum x_i^2 - n\overline{x}^2}{n(n-1)}} = \sqrt{\frac{16{,}486 - 16 \cdot 1{,}015^2}{16 \cdot 15}} = 0{,}0032$$

$$t = \frac{\overline{x_n} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1{,}015 - 1{,}0}{0{,}0032}\sqrt{16} = 18{,}75$$

Реализована вредност статистике је t = 18,75. На основу реализоване вредности t статистике t-теста доносимо одлуку о одбацивању или неодбацивању нулте хипотезе.

Критеријум је: H_0 не треба одбацити ако је $t \leq 1{,}7531$

H_0 треба одбацити ако је $t > 1{,}7531$

Ова вредност t=18,75 је већа од критичне вредности t = 1,7531 и налази се у области одбацивања нулте хипотезе.

Следи да нулту хипотезу одбацујемо и закључујемо да је, на основу података у узорку, просечна маса кутије детерџента већа од 1кг, при ризику од 0,05 да смо одбацили истиниту нулту хипотезу.

139. Време трајања ремонта једне врсте машина је у просеку 10 дана. Компанија која поправља те машине жели да тестира претпоставку да ће поправка трајати више од 10 дана, са нивоом значајности од 0,01 , на узорку од обима 121, ако се зна да аритметичка средина узорка 10,6 дана, а стандардна девијација основног скупа је 2,2.

Решење:

Формулисање нулте и алтернативне хипотезе:

H_0: $\mu \leq 10$ (Просечно трајање поправке машине није веће од 10 дана)

H_1: $\mu > 10$ (Просечно трајање поправке машине је веће од 10 дана)

Избор расподеле која ће се користити: σ позната, n>30, па ћемо за тестирање хипотезе користити Нормалну расподелу.

Одређивање области одбацивања и неодбацивања: Знак > у алтернативној хипотези указује да је тест деснострани, са облашћу одбацивања десно од критичне вредности на крају криве приближно Nормалне узорачке расподеле. Ниво значајности је 0,01. Површина која одговара области одбацивања је α=0,01.

Критична вредност је $z_{0,99}$ = 2,33 и за α=0,01 и је очитана из таблица Нормалне расподеле.

Сада израчунавамо реализовану вредност статистике z теста

$\mu = 10$, $n = 121$, $\bar{x} = 10{,}6$, $\sigma = 2{,}2$

$$z = \frac{\overline{x_n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{10{,}6 - 10}{0{,}02} = 3$$

Реализована вредност статистике је z=3,0. На основу реализоване вредности z статистике z теста доносимо одлуку о одбацивању или неодбацивању нулте хипотезе.

Критеријум је: H_0 не треба одбацити ако је $z \leq 2{,}33$

H_0 треба одбацити ако је $z > 2{,}33$

Ова вредност $z = 3{,}0$ је већа од критичне вредности $z_{0,99} = 2{,}33$ и налази се у области одбацивања нулте хипотезе.

Следи да нулту хипотезу одбацујемо и закључујемо да је, на основу података из узорка, просечно време потребно за ремонт машине веће од 10 дана, при ризику од 1% да смо одбацили истиниту нулту хипотезу.

140. Психолог тврди да је аритметичка средина старости деце која почињу да ходају 12,5 месеци. Изабран је случајан узорак од 18 деце и утврђено је да аритметичка средина старости деце која почињу да ходају 12,9 месеци са стандардном девијацијом од 0,8 месеци. Да ли можете користећи ниво значајности од 1 % да закључите да се аритметичка средина старости деце која почињу да ходају разликује од 12,5 месеци, под претпоставком да њихова старост има приближно Нормалну расподелу?

Решење:

Формулисање нулте и алтернативне хипотезе:

H_0: μ = 12,5 (Просечна старост деце која почињу да ходају је 12,5 месеци.)

H_1: $\mu \neq$ 12,5 (Просечна старост деце која почињу да ходају није 12,5 месеци.)

Избор расподеле која ће се користити: σ непозната, n=18<30, а основни скуп има Нормалну расподелу па ћемо за тестирање хипотезе користити t расподелу са df= n-1=17 степени слободе.

Одређивање области одбацивања и неодбацивања: Знак ≠ у алтернативној хипотези указује да је тест двострани, са две области одбацивања нулте хипотезе која је се налазе на оба краја криве t расподеле. Ниво значајности је 0,01. Површина која одговара области одбацивања је α/2=0,01/2=0.005. Критична вредност је $t_{17;\ 0,005}$ = 2,898 и за α=0,005 и df =17 је очитана из таблица t расподеле.

Сада израчунавамо реализовану вредност статистике t теста

$$t = \frac{\overline{x_n} - \mu}{\frac{s}{\sqrt{n}}} = \frac{12{,}9 - 12{,}5}{\frac{0{,}8}{\sqrt{18}}} = 2{,}12$$

Реализована вредност статистике теста је t = 2,12.

На основу реализоване вредности t статистике t-теста доносимо одлуку о одбацивању или неодбацивању нулте хипотезе.

Критеријум је: H_0 не треба одбацити ако је $|\ t\ | \leq 2{,}898$

H_0 треба одбацити ако је $|\ t\ | > 2{,}898$

Реализована вредност статистике теста је t=2,12 је мања од друге критичне вредности $t_{17;\ 0,005}$ = 2,898 и већа од прве критичне вредности $t_{17;\ 0,005}$ = - 2,898 налази се у области неодбацивања нулте хипотезе.

Следи да нулту хипотезу не одбацујемо и закључујемо на основу података у узорку да се аритметичка средина старости деце која почињу да ходају не разликује од 12,5 месеци при нивоу значајности од 1%.

141. На основу једног проучавања показано је да је аритметичка средина новца који становници једног европског града троше на одећу 675 Евра годишње. Истраживач је желео да провери да ли је тај резултат и даље тачан. Недавно је изабран случајан узорак од 39 грађана и добијени су следећи подаци о износима које троше на одећу сваке године:

671	1284	328	1698	827	921	725	304	382	539
1070	854	669	328	537	849	930	1234	1195	738
341	189	867	923	721	125	298	473	876	932
973	931	460	1430	391	887	958	674	1482	

Проверите на нивоу значајности од 1% да ли се аритметичка средина износа које становници тог европског града троше на одећу током последње године разликује од 675 Еура. Претпоставимо да је стандардна девијација основног скупа 132 Еура.

142. Новинарка тврди да одрасли становници у њеном граду проводе у просеку 30 и више сати месечно у читању новина, часописа, романа итд. Недавно изабрани узорак од 25 одраслих особа из овог града је показао да су оне у просеку провеле 27 сати месечно у читању. Познато је да основни скуп времена проведеног у читању има нормалну расподелу, са стандардном девијацијом од 7 сати. Можете ли на нивоу значајности од 1% да закључите да је новинаркина тврдња тачна?

143. Компанија која производи боје тврди да аритметичка средина времена сушења њихових боја није дужа од 45 мин. Изабран је случајан узорак од 25 канти боја и утврђено је да је аритметичка средина времена сушења у овом узорку била 49,5 мин са стандардном девијацијом од 3 мин. Претпоставимо да време сушења ових фарби има Нормалну расподелу. Можете ли да закључите, на нивоу значајности од 1%, да је тврђење ове компаније истинито?

Решење:

Формулисање нулте и алтернативне хипотезе:

H_0: $\mu \leq 45$ (Време сушења њихових боја није дуже од 45 мин.)

H_1: $\mu > 45$ (Време сушења њихових боја дуже од 45 мин.)

Избор расподеле која ће се користити: σ непозната, n=25<30, а основни скуп има Нормалну расподелу па ћемо за тестирање хипотезе користити t расподелу са $df = n$-1=24 степена слободе. Користи се деснострани тест. Одређујемо t критично за df=24 и за α=0,01. Критична вредност је $t_{24;\,0,01} = 2{,}492$.

$$t = \frac{\overline{x_n} - \mu}{\frac{s}{\sqrt{n}}} = \frac{49{,}5 - 45}{\frac{3}{\sqrt{25}}} = 7{,}5$$

Реализована вредност статистике теста је t=7,5 је већа од критичне вредности $t_{24;\,0,01} = 2{,}492$ и налази се у области одбацивања нулте хипотезе. Следи да нулту хипотезу одбацујемо и закључујемо на основу података у узорку да је аритметичка средина врема сушења боја је већа 45 дана при нивоу значајности од 1 %.

Тестирање хипотезе о аритметичкој средини два основна скупа

144. Из два независна узорка изабрана из два основна скупа са нормалним расподелама добијене су следеће информације:

$$n_1 = 300,\ \bar{x}_1 = 22{,}0,\ \sigma_1 = 4{,}9\ \ n_2 = 250,\ \bar{x}_2 = 27{,}6,\ \sigma_2 = 4{,}5$$

Тестирати хипотезу да је аритметичка средина другог основног скупа већа од аритметичке средине првог основног скупа са нивоом значајности од 1%.

Решење:

Формулисање нулте и алтернативне хипотезе:

H_0: $\mu_1 = \mu_2$ (Аритметичке средине основних скупова се не разликују)

H_1: $\mu_1 > \mu_2$ (Аритметичка средина првог основног скупа је већа од аритметичке средине другог)

Избор расподеле која ће се користити: Како су σ_1 и σ_2 познате, а оба узорка су велика n_1, $n_2 > 30$, следи да је узорачка расподела статистике апроксимативно Nормална па ћемо за тестирање хипотезе користити Нормалну расподелу.

Одређивање области одбацивања и неодбацивања: Знак > у алтернативној хипотези указује да је тест деснострани, са десном облашћу одбацивања на крају криве приближно нормалне узорачке расподеле. Ниво значајности је 0,01. Површина која одговара области одбацивања је на десном крају $\alpha = 0{,}01$.

Из таблице за стандардизовану Нормалну расподелу налазимо да је критична вредност функције за вредности функције расподеле 0,99 приближно једнака $z_{0,99}=2,33$.

Вредност статистике z теста израчунавамо на следећи начин:

$$\sigma_{\bar{X}_1-\bar{X}_2}=\sqrt{\frac{4{,}9^2}{300}+\frac{4{,}5^2}{250}}=0{,}40$$

$$z=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}}=\frac{(22{,}0-27{,}6)-0}{0{,}40}=-14{,}0$$

Доношење одлуке: Како се реализована вредност статистике теста z = -14,0 налази у области неодбацивања нулте хипотезе, не одбацујемо нулту хипотезу и закључујемо да се аритметичке средине ова два основна скупа не разликују.

145. Истраживач је желео да испита да ли мушкарци и жене који раде у једном граду прелазе исто растојање од куће до посла. У узорку од 25 мушкараца просечно растојање је износило 21 km, док је у узорку од 30 жена оно износило 20 km . Претпоставимо да ова два основна скупа имају Nормалну расподелу и да су њихове стандардне девијације једнаке 5,2 km и 4,4 km, респективно.

а. Нека су μ_1 и μ_2 аритметичке средине два основна скупа (просечна растојања од куће до посла свих мушкараца и жена у том граду). Израчунајте тачкасту оцењену вредност за μ_1 - μ_2.

б. Да ли, на нивоу значајности од 5% можете закључити да се растојања која сви запослени мушкарци и жене у овом граду прелазе од куће до посла, у просеку разликују?

146. Један пословни консултант је желео да испита да ли постојање вртића у кругу фирме смањује одсуства са посла њихових мајки. У узорку од 45 мајки деце предшколског узраста, изабраном из компанија које у кругу фирме имају вртић, просечан број дана одсуства у току прошле године је био 6,4 дана. У другом узорку од 50 мајки, запослених у компанијама без вртића, просечан број дана одсуства у прошлој години је био 9,3 дана. Претпставимо да су стандардне девијације ова два основна скупа 1,20 и 1,85 дана, респективно.

а) Формирајте 98% интервал поверења за разлику аритметичких средина два основна скупа.

б) Да ли на нивоу значајности од 2,5% можете да закључите да постоји разлика у просечном броју дана одсуства са посла мајки запослених у компанијама са и без вртића у оквиру фирме?

СТАТИСТИЧКЕ ТАБЛИЦЕ

ТАБЛИЦА I. ВЕРОВАТНОЋЕ БИНОМНЕ РАСПОДЕЛЕ

$P_{n,x,p} = \binom{n}{x} p^x q^{n-x}$ $\quad$ $q = 1 - p$; $P_{n,x,p} = P_{n,n-x,1-p}$ (за $p > 0,5$)

x/p	0,1	0,2	0,3	0,4	0,5
			n = 2		
0	0,810	0,640	0,490	0,360	0,250
1	0,180	0,320	0,420	0,480	0,500
2	0,010	0,040	0,090	0,160	0,250
			n = 3		
0	0,729	0,512	0,343	0,216	0,125
1	0,243	0,384	0,441	0,432	0,375
2	0,027	0,096	0,189	0,288	0,375
3	0,001	0,008	0,027	0,064	0,125
			n = 4		
0	0,6561	0,4096	0,2401	0,1296	0,0625
1	0,2916	0,4096	0,4116	0,3456	0,2500
2	0,0486	0,1536	0,2646	0,3456	0,3750
3	0,0036	0,0256	0,0756	0,1536	0,2500
4	0,0001	0,0016	0,0081	0,0256	0,0625
			n = 5		
0	0,5905	0,3277	0,1681	0,0778	0,0313
1	0,3280	0,4096	0,3601	0,2592	0,1562
2	0,0729	0,2048	0,3087	0,3456	0,3125
3	0,0081	0,0512	0,1323	0,2314	0,3125
4	0,0005	0,0064	0,0284	0,0768	0,1562
5	0,0000	0,0003	0,0024	0,0102	0,0313
			n = 6		
0	0,5314	0,2621	0,1176	0,0467	0,0156
1	0,3543	0,3932	0,3025	0,1866	0,0938
2	0,0984	0,2458	0,3241	0,3110	0,2344
3	0,0146	0,0819	0,1852	0,2765	0,3125
4	0,0012	0,0154	0,0595	0,1382	0,2344
5	0,0001	0,0015	0,0102	0,0369	0,0938
6	0,0000	0,0001	0,0007	0,0041	0,0156
			n = 7		
0	0,4783	0,2097	0,0824	0,0279	0,0078
1	0,3720	0,3670	0,2471	0,1306	0,0547
2	0,1240	0,2753	0,3177	0,2613	0,1641
3	0,0229	0,1147	0,2269	0,2903	0,2734
4	0,0025	0,0287	0,0972	0,1935	0,2734
5	0,0001	0,0043	0,0250	0,0774	0,1641
6	0,0000	0,0004	0,0036	0,0172	0,0547
7	0,0000	0,0000	0,0002	0,0016	0,0078

Таблица I. наставак

x/p	0,1	0,2	0,3	0,4	0,5
			n = 8		
0	0,4305	0,1678	0,0576	0,0168	0,0039
1	0,3826	0,3355	0,1977	0,0896	0,0313
2	0,1488	0,2936	0,2065	0,2090	0,1094
3	0,0331	0,1468	0,2541	0,2787	0,2188
4	0,0046	0,0459	0,1361	0,2322	0,2734
5	0,0004	0,0092	0,0467	0,1239	0,2188
6	0,0000	0,0011	0,0100	0,0413	0,1094
7	0,0000	0,0001	0,0012	0,0079	0,0313
8	0,0000	0,0000	0,0001	0,0007	0,0039
			n = 9		
0	0,3874	0,1342	0,0403	0,0101	0,0019
1	0,3874	0,3019	0,1556	0,0605	0,0176
2	0,1722	0,3019	0,2668	0,1612	0,0703
3	0,0446	0,1762	0,2668	0,2508	0,1641
4	0,0074	0,0661	0,1715	0,2508	0,2461
5	0,0008	0,1165	0,0735	0,1672	0,2461
6	0,0001	0,0028	0,0210	0,0743	0,1641
7	0,0000	0,0003	0,0039	0,0212	0,0703
8		0,0000	0,0004	0,0035	0,0176
9			0,0000	0,0003	0,0019
			n = 10		
0	0,3487	0,1074	0,0282	0,0060	0,0009
1	0,3874	0,2684	0,1211	0,0403	0,0098
2	0,1937	0,3019	0,2335	0,1209	0,0439
3	0,0574	0,2013	0,2668	0,2149	0,1172
4	0,0112	0,0881	0,2001	0,2508	0,2051
5	0,0015	0,0264	0,1029	0,2007	0,2461
6	0,0001	0,0055	0,0368	0,1115	0,2051
7	0,0000	0,0000	0,0090	0,0425	0,1172
8		0,0001	0,0014	0,0106	0,0439
9		0,0000	0,0001,	0,0016	0,0098
10			0,0000	0,0001	0,0009

Таблица I. наставак

x/p	0,1	0,2	0,3	0,4	0,5
			n = 15		
0	0,2059	0,0352	0,0047	0,0005	0,0000
1	0,3431	0,1319	0,0306	0,0047	0,0005
2	0,2669	0,2309	0,0915	0,0219	0,0032
3	0,1285	0,2502	0,1701	0,0634	0,0139
4	0,0429	0,1876	0,2186	0,1268	0,0416
5	0,0105	0,1031	0,2061	0,1859	0,0917
6	0,0019	0,0430	0,1473	0,2066	0,1527
7	0,0003	0,0139	0,0811	0,1771	0,1964
8	0,0000	0,0034	0,0348	0,1181	0,1964
9		0,0001	0,0115	0,0612	0,1527
10		0,0000	0,0030	0,0245	0,0917
11			0,0006	0,0074	0,0416
12			0,0001	0,0016	0,0139
13			0,0000	0,0003	0,0032
14				0,0000	0,0005
15					0,0000
			n = 20		
0	0,1216	0,0115	0,0008	0,0000	0,0000
1	0,2701	0,0577	0,0068	0,0005	0,0000
2	0,2852	0,1369	0,0279	0,0031	0,0002
3	0,1901	0,2053	0,0716	0,0124	0,0011
4	0,0898	0,2182	0,1307	0,0350	0,0046
5	0,0319	0,1746	0,1789	0,0746	0,0148
6	0,0089	0,1081	0,1916	0,1244	0,0370
7	0,0020	0,0546	0,1643	0,1659	0,0739
8	0,0003	0,0221	0,1144	0,1797	0,1201
9	0,0001	0,0074	0,0653	0,1597	0,1602
10	0,0000	0,0020	0,0309	0,1172	0,1762
11	0,0000	0,0005	0,0120	0,0710	0,1602
12		0,0001	0,0038	0,0355	0,1201
13		0,0000	0,0003	0,0049	0,0370
14			0,0000	0,0013	0,0148
15				0,0003	0,0046
16				0,0000	0,0011
17					0,0002
18					0,0000

Таблица I. наставак

x/p	0,1	0,2	0,3	0,4	0,5
			n = 30		
0	0,0424	0,0012	0,0000	0,0000	0,0000
1	0,1413	0,0093	0,0003	0,0000	0,0000
2	0,2277	0,0337	0,0018	0,0000	0,0000
3	0,2360	0,0785	0,0072	0,0003	0,0000
4	0,1771	0,1325	0,0208	0,0012	0,0000
5	0,1023	0,1723	0,0464	0,0042	0,0002
6	0,0474	0,1795	0,0829	0,0115	0,0005
7	0,0180	0,1538	0,1219	0,0263	0,0019
8	0,0058	0,1105	0,1501	0,0505	0,0055
9	0,0015	0,0676	0,1573	0,0823	0,0823
10	0,0004	0,0255	0,1416	0,1152	0,0280
11	0,0001	0,0161	0,1103	0,1396	0,0508
12	0,0000	0,0064	0,0748	0,1474	0,0806
13		0,0022	0,0444	0,1360	0,1115
14		0,0007	0,0232	0,1101	0,1355
15		0,0002	0,0105	0,0783	0,1444
16		0,0000	0,0043	0,0490	0,1355
17			0,0015	0,0279	0,1115
18			0,0004	0,0119	0,0806
19			0,0002	0,0054	0,0508
20			0,0000	0,0020	0,0280
21				0,0007	0,0133
22				0,0002	0,0055
23				0,0000	0,0019
24					0,0005
25					0,0002
26					0,0000

Таблица II. ВРЕДНОСТИ ФУНКЦИЈА e^x и e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
0,00	1,0000	1,0000	0,59	1,8040	0,55433
0,01	1,0101	0,99005	0,60	1,8221	0,54881
0,02	1,0202	0,98020	0,61	1,8404	0,54335
0,03	1,0305	0,97045	0,62	1,8589	0,53794
0,04	1,0408	0,96079	0,63	1,8776	0,53259
0,05	1,0513	0,95123	0,64	1,8965	0,52729
0,06	1.0618	0,94176	0,65	1,9155	0,52205
0,07	1,0725	0,93239	0,66	1,9348	0,51685
0,08	1,0833	0,92312	0,67	1,9542	0,51171
0,09	1,0942	0,91393	0,68	1,9739	0,50662
0,10	1,1052	0,90484	0,69	1,9937	0,50158
0,11	1,1163	0,89583	0,70	2,0138	0,49659
0,12	1,1275	0,88692	0,71	2.0340	0,49164
0,13	1,1388	0,87810	0,72	2,0544	0,48675
0,14	1,1503	0,86936	0,73	2.0751	0,48191
0,15	1,1618	0,86071	0,74	2,0959	0,47711
0,16	1,1735	0,85214	0,75	2,1170	0,47237
0,17	1,1853	0,84366	0,76	2,1383	0,46767
0,18	1,1972	0,83527	0,77	2,1598	0,46301
0,19	1,2092	0,82696	0,78	2,1815	0,45841
0,20	1,2214	0,81873	0,79	2,2034	0,45384
0,21	1,2337	0,81058	0,80	2,2255	0,44933
0,22	1,2461	0,80252	0,81	2,2479	0,44486
0,23	1,2586	0,79453	0,82	2,2705	0,44043
0,24	1,2712	0,78663	0,83	2,2933	0,43605
0,25	1,2840	0,77880	0,84	2,3164	0,43171
0,26	1,2969	0,77105	0,85	2,3396	0,42741
0,27	1,3100	0,76338	0,86	2,3632	0,42316
0,28	1,3231	0,75578	0,87	2,3869	0,41895
0,29	1,3364	0,74826	0,88	2,4109	0,41478
0,30	1,3499	0,74082	0,89	2,4351	0,41066
0,31	1,3634	0,73345	0,90	2,4596	0,40657
0,32	1,3771	0,72615	0,91	2,4843	0,40252
0,33	1,3910	0,71892	0,92	2,5093	0,39851
0,34	1,4049	0,71177	0,93	2,5345	0,39455
0,35	1,4191	0,70469	0,94	2,5600	0,39063

Таблица II. Вредности функција e^x и e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
0,36	1,4333	0,69768	0,95	2,5857	0,38674
0,37	1,4477	0,69073	0,96	2,6117	0,38389
0,38	1,4623	0,68386	0,97	2,6379	0,37908
0,39	1,4770	0,67706	0,98	2,6645	0,37531
0,40	1,4918	0,67032	0,99	2,6912	0,37158
0,41	1,5068	0,66365	1,00	2,7183	0,36788
0,42	1,5220	0,65705	1,01	2,7456	0,36422
0,43	1,5373	0,65051	1,02	2,7732	0,36060
0,44	1,5527	0,64404	1,03	2,8011	0,35701
0,45	1,5683	0,63763	1,04	2,8292	0,35345
0,46	1,5841	0,63128	1,05	2,8577	0,34994
0,47	1,6000	0,62500	1,06	2,8864	0,34646
0,48	1,6161	0,61878	1,07	2,9154	0,34301
0,49	1,6323	0,61263	1,08	2,9447	0,33960
0,50	1,6487	0,60653	1,09	2,9743	0,33622
0,51	1,6653	0,60050	1,10	3,0042	0,33287
0,52	1,6820	0,59452	1,11	3,0344	0,32956
0,53	1,6989	0,58860	1,12	3,0649	0,32628
0,54	1,7160	0,58275	1,13	3,0957	0,32303
0,55	1,7333	0,57695	1,14	3,1268	0,31982
0,56	1,7507	0,57121	1,15	3,1582	0,31664
0,57	1,7683	0,56553	1,16	3,1899	0,31349
0,58	1,7860	0,55990	1,17	3,2220	0,31037
1,18	2,3544	0,30728	1,77	5,8709	0,17033
1,19	3,2871	0,30422	1,78	5,9299	0,16864
1,20	3,3201	0,30119	1,79	5,9895	0,16696
1,21	3,3535	0,29820	1,80	6,0496	0,16530
1,22	3,3872	0,29523	1,81	6,1104	0,16365
1,23	3,4212	0,29229	1,82	6,1719	0,16203
1,24	3,4556	0,28938	1,83	6,2339	0,16041
1,25	3,4903	0,28650	1,84	6,2965	0,15882
1,26	3,5254	0,28365	1,85	6,3598	0,15724
1,27	3,5609	0,28033	1,86	6,4237	0,15567
1,28	3,5966	0,27804	1,87	6,4883	0,15412
1;29	3,6328	0,27527	1,88	6,5535	0,15259
1,30	3,6693	0,27253	1,89	6,6194	0,15107

Таблица II. Вредности функција e^x и e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
1,31	3,7062	0,26982	1,90	6,6859	0,14957
1,32	3,7434	0,26714	1,91	6,7531	0,14808
1,33	3,7810	0,26448	1,92	6,8210	0,14661
1,34	3,8190	0,26185	1,93	6,8895	0,14515
1,35	3,8574	0,25924	1,94	6,9588	0,14370
1,36	3,8962	0,25666	1,95	7,0287	0,14227
1,37	3,9354	0,25411	1,96	7,0993	0,14086
1,38	3,9749	0,25158	1,97	7,1707	0,13946
1,39	4,0149	0,24908	1,98	7,2427	0,13807
1,40	4,0552	0,24660	1,99	7,3155	0,13670
1,41	4,0960	0,24414	2,00	7,3891	0,13534
1,42	4,1371	0,24171	2,01	7,4633	0,13399
1,43	4,1787	0,23931	2,02	7,5383	0,13266
1,44	4,2207	0,23693	2,03	7,6141	0,13134
1,45	4,2631	0,23457	2,04	7,6906	0,13003
1,46	4,3060	0,23244	20,5	7,7679	0,12873
1,47	4,3492	0,22993	2,06	7,8460	0,12745
1,48	4,3929	0,22764	2,07	7,9248	0,12619
1,49	4,4371	0,22537	2,08	8,0045	0,12493
1,50	4,4817	0,22313	2,09	8,0849	0,12369
1,51	4,5267	0,22091	2,10	8,1662	0,12246
1,52	4,5722	0,21871	2,11	8,2482	0,12124
1,53	4,6182	0,21654	2,12	8,3311	0,12003
1,54	4,6646	0,21438	2,13	8,4149	0,11884
1,55	4,7115	0,21225	2,14	8,4994	0,11765
1,56	4,7588	0,21014	2,15	8,5849	0,11648
1,57	4,8066	0,20805	2,16	8,6711	0,11533
1,58	4,8550	0,20598	2,17	8,7583	0,11418
1,59	4,9037	0,20393	2,18	8,8463	0,11304
1,60	4,9530	0,20190	2,19	8,9352	0,11192
1,61	5,0028	0,19989	2,20	9,0250	0,11080
1,62	5,0531	0,19790	2,21	9,1157	0,10970
1,63	5,1039	0,19593	2,22	9,2073	0,10861
1,64	5,1552	0,19398	2,23	9,2999	0,10753
1,65	5,2070	0,19205	2,24	9,3933	0,10646
1,66	5,2593	0,19014	2,25	9,4877	0,10540

Таблица II. Вредности функција e^x и e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
1,67	5,3122	0,18820	2,26	9,5831	0,10435
1,68	5,3656	0,18637	2,27	9,6794	0,10331
1,69	5,4195	0,18452	2,28	9,7767	0,10228
1,70	5,4793	0,18268	2,29	9,8749	0,10127
1,71	5,5290	0,18087	2,30	9,9742	0,10026
1,72	5,5845	0,17907	2,31	10,074	0,09926
1,73	5,6407	0,17728	2,32	10,176	0,09827
1,74	5,6973	0,17552	2,33	10,278	0,09730
1,75	5,7546	0,17377	2,34	10,381	0,09633
1,76	5,8124	0,17204	2,35	10,486	0,09537
2,36	10,591	0,09442	2,94	18,916	0,05287
2,37	10,697	0,09348	2,95	19,106	0,05234
2,38	10,805	0,09255	2,96	19,298	0,05182
2,39	10,913	0,09163	2,97	19,492	0,05130
2,40	11,023	0,09072	2,98	19,688	0,05079
2,41	11,134	0,08982	2,99	19,886	0,05029
2,42	11,246	0,08892	3,00	20,086	0,04979
2,43	11,359	0,08804	3,05	21,115	0,04736
2,44	11,473	0,08716	3,10	22,198	0,04505
2,45	11,588	0,08629	3,15	23,336	0,04285
2,46	11,705	0,08543	3,20	24,533	0,04076
2,47	11,822	0,08458	3,25	25,790	0,03877
2,48	11,941	0,08374	3,30	27,113	0,03688
2,49	12,061	0,08291	3,35	28,503	0,03508
2,50	12,182	0,08208	3,40	29,964	0,03337
2,51	12,305	0,08127	3,45	31,500	0,03175
2,52	12,429	0,08046	3,50	33,115	0,03020
2,53	12,554	0,07966	3,55	34,813	0,02872
2,54	12,680	0,07887	3,60	36,598	0,02732
2,55	12,807	0,07808	3,65	38,475	0,02599
2,56	12,936	0,07730	3,70	40,447	0,02472
2,57	13,066	0,07654	3,75	42,521	0,02352
2,58	13,197	0,07577	3,80	44,701	0,02237
2,59	13,330	0,07502	3,85	46,993	0,02128
2,60	13,464	0,07427	3,90	49,402	0,02024
2,61	13,599	0,07353	3,95	51,935	0,01925

Таблица II. Вредности функција e^x и e^{-x}

x	e^x	e^{-x}	x	e^x	e^{-x}
2,62	13,736	0,07280	4,00	54,598	0,01832
2,63	13,874	0,07208	4,10	60,340	0,01657
2,64	14,013	0,07136	4,20	66,686	0,01500
2,65	14,154	0,07065	4,30	73,700	0,01357
2,66	14,296	0,06995	4,40	81,451	0,01228
2,67	14,440	0,06925	4,50	90,017	0,01111
2,68	14,585	0,06856	4,60	99,484	0,01005
2,69	14,732	0,06788	4,68	107,77	0,00928
2,70	14,880	0,06721	4,70	109,95	0,00910
2,71	15,029	0,06654	4,80	121,51	0,00823
2,72	15,180	0,06587	4,90	134,29	0,00745
2,73	15,333	0,06522	5,00	148,41	0,00674
2,74	15,487	0,06457	5,10	164,02	0,00610
2,75	15,643	0,06393	5,20	181,27	0,00552
2,76	15,800	0,06329	5,30	200,34	0,00499
2,77	15,959	0,06266	5,40	221,41	0,00452
2,78	16,119	0,06204	5,50	244,69	0,00409
2,79	16,281	0,06142	5,60	270,43	0,00370
2,80	16,445	0,06081	5,70	298,87	0,00335
2,81	16,610	0,06020	5,80	330,30	0,00303
2,82	16,777	0,05961	5,90	365,04	0,00274
2,83	16,945	0,05901	6,00	403,43	0,00248
2,84	17,116	0,05843	6,25	518,01	0,00193
2,85	17,288	0,05784	6,50	665,14	0,00150
2,86	17,462	0,05727	6,75	854,06	0,00117
2,87	17,637	0,05670	7,00	1.096,6	0,00091
2,88	17,814	0,05613	7,50	1.808,0	0,00055
2,89	17,993	0,05558	8,00	2.981,0	0,00034
2,90	18,174	0,05502	8,50	4.914,8	0,00020
2,91	18,357	0,05448	9,00	8.103,1	0,00012
2,92	18,541	0,05393	9,50	13.360	0,00007
2,93	18,728	0,05340	10,00	22.226	0,00005

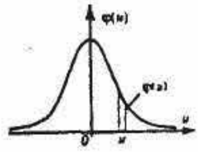
Таблица III. ВЕРОВАТНОЋЕ ПО ПУАСОНОВОМ РАСПОРЕДУ: $p(x)=\frac{m^x}{x!}\cdot e^{-m}$.

	Вредност параметра $np=m$									
x	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	90481	81873	74082	67032	60653	54881	49659	44933	40657	36788
1	09048	16374	22225	26813	30327	32929	34761	35946	36591	36788
2	00452	01638	03334	05363	07582	09879	12166	14379	16466	18394
3	00015	00109	00333	00715	01264	01976	02839	03834	04940	06131
4	00000	00006	00025	00072	00158	00296	00497	00767	01112	01533
5		00000	00002	00006	00016	00036	00070	00123	00200	00307
6			00000	00000	00001	00004	00008	00015	00030	00051
7					00000	00000	00001	00002	00004	00007
8							00000	00000	00000	00001
9										00000

	Вредност параметра $np=m$									
x	2	3	4	5	6	7	8	9	10	11
0	13534	04979	01832	00674	00248	00091	00034	00012	00005	00002
1	27068	14936	07326	03369	01487	00638	00268	00111	00045	00018
2	27068	22404	14653	08422	04462	02234	01074	00500	00227	00101
3	18045	22404	19537	14037	08924	05213	02863	01499	00757	00371
4	09022	16803	19537	17547	13385	09123	05725	03374	01892	01019
5	03609	10082	15629	17547	16062	12772	09160	06073	03783	02242
6	01203	05041	10420	14622	16062	14900	12214	09109	06306	04110
7	00344	02160	05954	10445	13768	14900	13959	11712	09008	06458
8	00086	00810	02977	06528	10326	13038	13959	13176	11260	08879
9	00019	00270	01323	03627	06884	10141	12408	13176	12511	10853
10	00004	00081	00529	01813	04130	07098	09926	11858	12511	11938
11	00001	00022	00193	00824	02253	04517	07219	09702	11374	11938
12	00000	00006	00064	00343	01126	02635	04813	07277	09478	10943
13		00001	00020	00132	00520	01419	02962	05038	07291	09260
14		00000	0006	00047	00223	00709	01692	03238	05208	07475
15			00002	00016	00089	00331	00903	01943	03472	05335
16			00000	00005	00033	00145	00451	01093	02170	03668
17				00001	00012	00060	00212	00579	01276	02373
18				00000	00004	00023	00094	00289	00709	01450
19					00001	00009	00040	00137	00373	00840
20					00000	00003	00016	00062	00187	00462
21						00001	00006	00026	00089	00242
22						00000	00002	00011	00040	00121
23							00001	00004	00017	00058
24							00000	00002	00007	00027
25								00001	00003	00012
26								00000	00001	00005
27									00000	00002
28										00001
29										00000

Напомена: Испред сваке вредности у овој таблици треба ставити децимални зарез.

Таблица IV. НОРМАЛАН РАСПОРЕД – ЗАКОНА ВЕРОВАТНОЋЕ
(Ординате нормалне криве)



$$\varphi(u)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\cdot u^{2}}$$

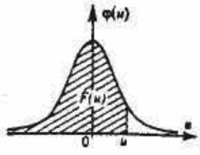
|u|	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38361	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32918	32713	32506	32297	32086	31974	31659
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09560
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06439	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03391	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01889	01842	01797
2,5	01753	01709	01625	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00936	00909	00885	00861	00837	00814
2,8	00792	00770	00727	00748	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00471	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00038	00037	00035	00034	00033	00032	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00020
3,9	00020	00019	00018	00018	00017	00016	00015	00014	00014	00014

Напомена: Испред сваке вредности у овој таблици треба ставити децимални зарез.

Таблица V.

НОРМАЛАН РАСПОРЕД – ФУНКЦИЈА РАСПОРЕДА

(Површине испод нормалне криве у границама од $-\infty$ до u)



$$F(u)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u} e^{-\frac{1}{2}v^2}\,dv$$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
	0,	0	0	0	0	0	0	0	0	0
0,0	5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9292	9279	9306	9319
1,5	9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9713	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	9772	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9980	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
3,0	9987	9987	9987	9988	9988	9989	9989	9990	9990	9990
3,1	9990	9991	9991	9991	9992	9992	9992	9992	9993	9993
3,2	9993	9993	9994	9994	9994	9994	9994	9995	9995	9995
3,3	9995	9995	9995	9996	9996	9996	9996	9996	9996	9997
3,4	9997	9997	9997	9997	9997	9997	9997	9997	9997	9998

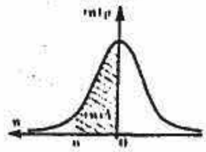
Напомена: Испред сваке вредности у овој таблици треба ставити децимални зарез.

Таблица VI.

ПОВРШИНЕ ИСПОД НОРМАЛНЕ КРИВЕ У ГРАНИЦАМА ОД 0 ДО „u“

(једна страна површине испод криве)

$$\Phi_1(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{1}{2}v^2}\, dv$$



u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0.3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4393	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,5957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4978	0,4979	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,4999									

Таблица VII.

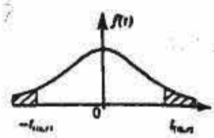


ТАБЛИЦА t-ДИСТРИБУЦИЈЕ ИЛИ „СТУДЕНТОВОГ" РАСПОРЕДА

(вредности t за одговарајуће нивое сигнификантности α и степене слободе r)

Степен слободе (r)	НИВО ГРЕШКЕ (ниво сигнификантности) α								
	0,50	0,40	0,20	0,10	0,05	0,025	0,01	0,005	0,001
1	1,000	1,376	3,078	6,314	12,706	25,452	63,657	127,32	636,619
2	0,816	1,061	1,886	2,920	4,303	6,205	9,925	14,089	31,598
3	0,765	0,978	1,638	2,353	3,182	4,186	5,841	7,453	12,941
4	0,741	0,941	1,533	2,132	2,776	3,495	4,604	5,598	8,610
5	0,727	0,920	1,476	2,015	2,571	3,163	4,032	4,773	6,859
6	0,718	0,906	1,440	1,943	2,447	2,969	3,707	4,317	5,959
7	0,711	0,896	1,415	1,895	2,365	2,841	3,499	3,929	5,405
8	0,706	0,889	1,397	1,860	2,306	2,752	3,355	3,832	5,041
9	0,703	0,883	1,383	1,833	2,262	2,685	3,250	3,690	4,781
10	0,700	0,879	1,372	1,812	2,228	2,634	3,169	3,581	4,587
11	0,697	0,876	1,363	1,796	2,201	2,593	3,106	3,497	4,437
12	0,695	0,873	1,356	1,782	2,179	2,560	3,055	3,428	4,318
13	0,694	0,870	1,350	1,771	2,160	2,533	3,012	3,372	4,221
14	0,692	0,868	1,345	1,761	2,145	2,510	2,977	3,326	4,140
15	0,691	0,866	1,341	1,753	2,131	2,490	2,947	3,286	4,073
16	0,690	0,865	1,337	1,746	2,120	2,473	2,921	3,252	4,015
17	0,689	0,863	1,333	1,740	2,110	2,458	2,898	3,222	3,965
18	0,688	0,862	1,330	1,734	2,101	2,445	2,878	3,197	3,922
19	0,688	0,861	1,328	1,729	2,093	2,435	2,861	3,174	3,883
20	0,687	0,860	1,325	1,725	2,086	2,423	2,845	3,153	3,850
21	0,686	0,859	1,323	1,721	2,080	2,414	2,831	3,135	3,829
22	0,686	0,858	1,321	1,717	2,074	2,406	2,819	3,119	3,792
23	0,685	0,858	1,319	1,714	2,069	2,398	2,807	3,104	3,767
24	0,685	0,857	1,318	1,711	2,064	2,391	2,797	3,090	3,745
25	0,685	0,856	1,316	1,708	2,060	2,385	2,787	3,078	3,725
26	0,684	0,856	1,315	1,706	2,056	2,379	2,779	3,067	3,707
27	0,684	0,855	1,314	1,703	2,052	2,373	2,771	3,056	3,690
28	0,683	0,855	1,313	1,701	2,048	2,368	2,763	3,047	3,674
29	0,683	0,854	1,311	1,699	2,045	2,364	2,756	3,038	3,659
30	0,683	0,854	1,310	1,697	2,042	2,360	2,750	3,030	3,646
35	0,682	0,852	1,306	1,690	2,030	2,342	2,724	2,996	3,591
40	0,681	0,851	1,303	1,684	2,021	2,329	2,704	2,971	3,551
45	0,680	0,850	1,301	1,680	2,041	2,319	2,690	2,952	3,520
50	0,680	0,849	1,299	1,676	2,008	2,310	2,678	2,937	3,496
55	0,679	0,849	1,298	1,673	2,004	2,304	2,669	2,925	3,476
60	0,679	0,848	1,296	1,671	2,000	2,299	2,660	2,915	3,460
70	0,678	0,847	1,294	1,667	1,994	2,290	2,648	2,899	3,435
80	0,678	0,847	1,293	1,665	1,989	2,284	2,638	2,887	3,416
90	0,678	0,846	1,292	1,663	1,986	2,279	2,631	2,878	3,402
100	0,677	0,846	1,290	1,661	1,982	2,276	2,625	2,871	3,390
120	0,677	0,845	1,289	1,658	1,980	2,270	2,617	2,860	3,373
∞	0,6745	0,841	1,281	1,6448	1,9600	2,2414	2,5758	2,8070	3,2905

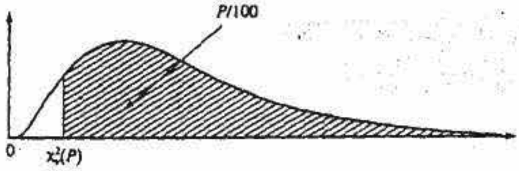
TABLE 8. PERCENTAGE POINTS OF THE χ^2-DISTRIBUTION

This table gives percentage points $\chi^2_\nu(P)$ defined by the equation

$$\frac{P}{100}=\frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})}\int_{\chi^2_\nu(P)}^{\infty}\chi^{\frac{1}{2}\nu-1}e^{-\frac{1}{2}x}\,dx.$$

If X is a variable distributed as χ^2 with ν degrees of freedom, $P/100$ is the probability that $X \geqslant \chi^2_\nu(P)$.

For $\nu > 100$, $\sqrt{2X}$ is approximately normally distributed with mean $\sqrt{2\nu-1}$ and unit variance.



(The above shape applies for $\nu \geqslant 3$ only. When $\nu < 3$ the mode is at the origin.)

P	99.95	99.9	99.5	99	97.5	95	90	80	70	60
$\nu=$ 1	$0.0^6 3927$	$0.0^5 1571$	$0.0^4 3927$	$0.0^3 1571$	$0.0^3 9821$	0.003932	0.01579	0.06418	0.1485	0.2750
2	0.001000	0.002001	0.01003	0.02010	0.05064	0.1026	0.2107	0.4463	0.7133	1.022
3	0.01528	0.02430	0.07172	0.1148	0.2158	0.3518	0.5844	1.005	1.424	1.869
4	0.06392	0.09080	0.2070	0.2971	0.4844	0.7107	1.064	1.649	2.195	2.753
5	0.1581	0.2102	0.4117	0.5543	0.8312	1.145	1.610	2.343	3.000	3.655
6	0.2994	0.3811	0.6757	0.8721	1.237	1.635	2.204	3.070	3.828	4.570
7	0.4849	0.5985	0.9893	1.239	1.690	2.167	2.833	3.822	4.671	5.493
8	0.7104	0.8571	1.344	1.646	2.180	2.733	3.490	4.594	5.527	6.423
9	0.9717	1.152	1.735	2.088	2.700	3.325	4.168	5.380	6.393	7.357
10	1.265	1.479	2.156	2.558	3.247	3.940	4.865	6.179	7.267	8.295
11	1.587	1.834	2.603	3.053	3.816	4.575	5.578	6.989	8.148	9.237
12	1.934	2.214	3.074	3.571	4.404	5.226	6.304	7.807	9.034	10.18
13	2.305	2.617	3.565	4.107	5.009	5.892	7.042	8.634	9.926	11.13
14	2.697	3.041	4.075	4.660	5.629	6.571	7.790	9.467	10.82	12.08
15	3.108	3.483	4.601	5.229	6.262	7.261	8.547	10.31	11.72	13.03
16	3.536	3.942	5.142	5.812	6.908	7.962	9.312	11.15	12.62	13.98
17	3.980	4.416	5.697	6.408	7.564	8.672	10.09	12.00	13.53	14.94
18	4.439	4.905	6.265	7.015	8.231	9.390	10.86	12.86	14.44	15.89
19	4.912	5.407	6.844	7.633	8.907	10.12	11.65	13.72	15.35	16.85
20	5.398	5.921	7.434	8.260	9.591	10.85	12.44	14.58	16.27	17.81
21	5.896	6.447	8.034	8.897	10.28	11.59	13.24	15.44	17.18	18.77
22	6.404	6.983	8.643	9.542	10.98	12.34	14.04	16.31	18.10	19.73
23	6.924	7.529	9.260	10.20	11.69	13.09	14.85	17.19	19.02	20.69
24	7.453	8.085	9.886	10.86	12.40	13.85	15.66	18.06	19.94	21.65
25	7.991	8.649	10.52	11.52	13.12	14.61	16.47	18.94	20.87	22.62
26	8.538	9.222	11.16	12.20	13.84	15.38	17.29	19.82	21.79	23.58
27	9.093	9.803	11.81	12.88	14.57	16.15	18.11	20.70	22.72	24.54
28	9.656	10.39	12.46	13.56	15.31	16.93	18.94	21.59	23.65	25.51
29	10.23	10.99	13.12	14.26	16.05	17.71	19.77	22.48	24.58	26.48
30	10.80	11.59	13.79	14.95	16.79	18.49	20.60	23.36	25.51	27.44
32	11.98	12.81	15.13	16.36	18.29	20.07	22.27	25.15	27.37	29.38
34	13.18	14.06	16.50	17.79	19.81	21.66	23.95	26.94	29.24	31.31
36	14.40	15.32	17.89	19.23	21.34	23.27	25.64	28.73	31.12	33.25
38	15.64	16.61	19.29	20.69	22.88	24.88	27.34	30.54	32.99	35.19
40	16.91	17.92	20.71	22.16	24.43	26.51	29.05	32.34	34.87	37.13
50	23.46	24.67	27.99	29.71	32.36	34.76	37.69	41.45	44.31	46.86
60	30.34	31.74	35.53	37.48	40.48	43.19	46.46	50.64	53.81	56.62
70	37.47	39.04	43.28	45.44	48.76	51.74	55.33	59.90	63.35	66.40
80	44.79	46.52	51.17	53.54	5715	60.39	64.28	69.21	72.92	76.19
90	52.28	54.16	59.20	61.75	65.65	69.13	73.29	78.56	82.51	85.99
100	59.90	61.92	67.33	70.06	74.22	77.93	82.36	87.95	92.13	95.81

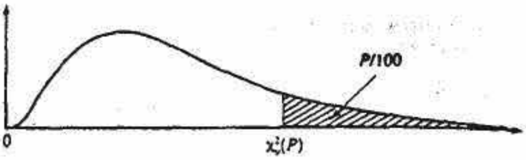
TABLE 8. PERCENTAGE POINTS OF THE χ^2-DISTRIBUTION

This table gives percentage points $\chi^2_\nu(P)$ defined by the equation

$$\frac{P}{100} = \frac{1}{2^{\nu/2}\Gamma(\frac{1}{2}\nu)}\int_{\chi^2_\nu(P)}^{\infty} x^{\frac{1}{2}\nu-1} e^{-\frac{1}{2}x}\,dx.$$

If X is a variable distributed as χ^2 with ν degrees of freedom, $P/100$ is the probability that $X \geqslant \chi^2_\nu(P)$.

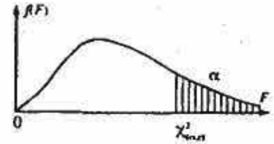
For $\nu > 100$, $\sqrt{2X}$ is approximately normally distributed with mean $\sqrt{2\nu - 1}$ and unit variance.



(The above shape applies for $\nu \geqslant 3$ only. When $\nu < 3$ the mode is at the origin.)

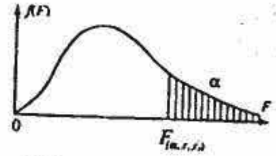
P	50	40	30	20	10	5	2.5	1	0.5	0.1	0.05
$\nu =$ 1	0.4549	0.7083	1.074	1.642	2.706	3.841	5.024	6.635	7.879	10.83	12.12
2	1.386	1.833	2.408	3.219	4.605	5.991	7.378	9.210	10.60	13.82	15.20
3	2.366	2.946	3.665	4.642	6.251	7.815	9.348	11.34	12.84	16.27	17.73
4	3.357	4.045	4.878	5.989	7.779	9.488	11.14	13.28	14.86	18.47	20.00
5	4.351	5.132	6.064	7.289	9.236	11.07	12.83	15.09	16.75	20.52	22.11
6	5.348	6.211	7.231	8.558	10.64	12.59	14.45	16.81	18.55	22.46	24.10
7	6.346	7.283	8.383	9.803	12.02	14.07	16.01	18.48	20.28	24.32	26.02
8	7.344	8.351	9.524	11.03	13.36	15.51	17.53	20.09	21.95	26.12	27.87
9	8.343	9.414	10.66	12.24	14.68	16.92	19.02	21.67	23.59	27.88	29.67
10	9.342	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19	29.59	31.42
11	10.34	11.53	12.90	14.63	17.28	19.68	21.92	24.72	26.76	31.26	33.14
12	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.30	32.91	34.82
13	12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82	34.53	36.48
14	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32	36.12	38.11
15	14.34	15.73	17.32	19.31	22.31	25.00	27.49	30.58	32.80	37.70	39.72
16	15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	34.27	39.25	41.31
17	16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72	40.79	42.88
18	17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	37.16	42.31	44.43
19	18.34	19.91	21.69	23.90	27.20	30.14	32.85	36.19	38.58	43.82	45.97
20	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40.00	45.31	47.50
21	20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.40	46.80	49.01
22	21.34	23.03	24.94	27.30	30.81	33.92	36.78	40.29	42.80	48.27	50.51
23	22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18	49.73	52.00
24	23.34	25.11	27.10	29.55	33.20	36.42	39.36	42.98	45.56	51.18	53.48
25	24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93	52.62	54:95
26	25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29	54.05	56.41
27	26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64	55.48	57.86
28	27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99	56.89	59.30
29	28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34	58.30	60.73
30	29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67	59.70	62.16
32	31.34	33.38	35.66	38.47	42.58	46.19	49.48	53.49	56.33	62.49	65.00
34	33.34	35.44	37.80	40.68	44.90	48.60	51.97	56.06	58.96	65.25	67.80
36	35.34	37.50	39.92	42.88	47.21	51.00	54.44	58.62	61.58	67.99	70.59
38	37.34	39.56	42.05	45.08	49.51	53.38	56.90	61.16	64.18	70.70	73.35
40	39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77	73.40	76.09
50	49.35	51.89	54.72	58.16	63.17	67.50	71.42	76.15	79.49	86.66	89.56
60	59.33	62.13	65.23	68.97	74.40	79.08	83.30	88.38	91.95	99.61	102.7
70	69.33	72.36	75.69	79.71	85.53	90.53	95.02	100.4	104.2	112.3	115.6
80	79.33	82.57	86.12	90.41	96.58	101.9	106.6	112.3	116.3	124.8	128.3
90	89.33	92.76	96.52	101.1	107.6	113.1	118.1	124.1	128.3	137.2	140.8
100	99.33	102.9	106.9	111.7	118.5	124.3	129.6	135.8	140.2	149.4	153.2

Таблица VIII. ТАБЛИЦА χ^2-Распореда (хи-квадрат)
(вредности χ^2 за одговарајуће нивое сигнификантности α и степене слободе r)



Степ. слоб. (r)	Ниво грешке α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,000157	0,000628	0,000393	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,0201	0,0404	0,103	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,115	0,185	0,352	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,341
4	0,297	0,429	0,711	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	0,554	0,752	1,145	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,383	15,082
6	0,872	1,134	1,635	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	1,239	1,564	2,167	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,622	18,475
8	1,646	2,032	2,733	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	2,088	2,532	3,325	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	2,558	3,059	3,940	4,865	6,179	7,267	9,324	11,781	13,441	15,987	18,307	21,161	23,209
11	3,053	3,609	4,575	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	3,571	4,178	5,226	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	4,107	4,765	5,892	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	4,660	5,368	6,571	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	5,229	5,985	7,261	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	5,812	6,614	7,962	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	6,408	7,255	8,672	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	7,015	7,906	9,390	10,865	12,875	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	7,633	8,567	10,117	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	8,260	9,237	10,851	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	8,897	9,915	11,591	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	9,542	10,600	12,338	14,014	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	10,196	11,293	13,091	14,848	17,187	19,021	22,337	26.018	28,429	32,007	35.172	38,968	41,638
24	10,856	11,992	13,848	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	11,524	12,697	14,611	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	12,198	13,409	15,379	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	12,879	14,125	16,151	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,133	44,140	46,963
28	13,565	14,847	16,928	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	14,256	15,574	17,708	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	14,953	16,306	18,493	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

Таблица IX. СНЕДЕКОРОВ F-РАСПОРЕД



Вредности F за ниво сигнификантности $\alpha = 5\%$ $[F_{(0,05; r_1; r_2)}]$
При томе је r_1 степен слободе бројитеља, а r_2 степен слободе именитеља

Степени слободе (r_2)	Степени слободе (r_1)														
	1	2	3	4	5	6	7	8	9	10	12	20	30	50	∞
1	161,4	199,5	215,7	224,5	230,1	233,9	236,7	238,9	240,5	241,5	243,9	248,0	250,1	252,2	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,44	19,46	19,48	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,66	8,62	8,58	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,80	5,74	5,70	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,56	4,60	4,44	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,08	4,00	3,97	3,81	3,75	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3.79	3,73	3,68	3,63	3,57	3,44	3,38	3,32	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,15	3,08	3,03	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,93	2,86	2,80	2,71
10	4,96	4,10	3,71	3,48	3,33	322	3,14	3,07	3,02	2,97	2,91	2,77	2,70	2,64	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,65	2,57	2,50	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,95	2,85	2,80	2,76	2,69	2,54	2,46	2,40	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,46	2,38	2,32	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,39	2,31	2,24	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,33	2,25	2,18	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,28	2,20	2,13	2,01
17	4,45	3,59	3,20	2,96	2,82	2,70	2,62	2,55	2,50	2,45	2,38	2,23	2,15	2,08	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,19	2,11	2,04	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,15	2,07	2,00	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,12	2,04	1,96	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,09	2,00	1,93	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,07	1,98	1,91	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,20	2,04	1,96	1,88	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	2,02	1,94	1,86	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,28	2,24	2,16	200	1,92	1,84	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,16	2,09	1,93	1,84	1,76	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,84	1,74	1,66	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,95	1,78	1,69	1,60	1,44
70	3,98	3,19	2,74	2,50	2,35	2,23	2,14	2,07	2,01	1,97	1,89	1,72	1,62	1,53	1,35
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,82	1,68	1,57	1,43	1,28
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,75	1,57	1,46	1,35	1,00

Таблица IX. СНЕДЕКОРОВ *F*-РАСПОРЕД

Вредности F за ниво сигнификантности $\alpha = 1\%$ $[F_{(0,01;\, r_1;\, r_2)}]$

При томе је r_1 степен слободе бројитеља, а r_2 степен слободе именитеља

Степ. слоб. (r_2)	Степени слободе (r_1)														
	1	2	3	4	5	6	7	8	9	10	12	20	30	50	∞
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,6	6022,5	60,55,8	61,06,3	6208,7	6260,7	6302,0	6366,0
2	98,48	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,38	99,40	99,42	99,45	99,47	99,48	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,13	27,05	26,69	26,50	26,35	26,12
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,37	14,02	13,83	13,69	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,89	9,55	9,38	9,24	9,02
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,39	7,23	7,09	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,47	6,15	5,98	5,85	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,83	5,67	5,36	5,20	5,06	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,11	4,80	4,64	4,51	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,71	4,41	4,25	4,12	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,10	3,94	3,80	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,16	3,86	3,70	3,56	3,36
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,67	3,51	3,37	3,16
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,51	3,34	3,21	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,36	3,20	3,07	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,25	3,10	2,96	2,73
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,16	3,00	2,86	3,65
18	8,28	6,01	5,09	4,58	4,25	4,01	3,85	3,71	3,60	3,51	3,37	3,07	2,91	2,78	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,25	3,43	3,30	3,00	2,84	2,70	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,71	3,56	3,45	3,37	3,23	2,94	2,77	2,63	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,65	3,51	3,40	3,31	3,17	2,88	2,72	2,58	2,36
22	7,94	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,83	2,67	2,53	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,78	2,62	2,48	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,25	3,17	3,03	2,74	2,58	2,44	2,21
25	7,77	5,57	4,66	4,18	3,86	3,63	3,43	3,32	3,21	3,13	2,99	2,70	2,54	2,40	2,17
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,06	2,98	2,84	2,55	2,38	2,24	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,88	2,80	2,66	2,37	2,20	2,05	1,81
50	7,17	5,06	4,20	3,72	3,41	3,18	3,02	2,88	2,78	2,70	2,56	2,26	2,10	1,94	1,68
70	7,01	4,92	4,08	3,60	3,29	2,07	2,91	2,77	2,67	2,59	2,45	2,15	1,99	1,82	1,53
100	6,90	4,82	3,98	3,51	3,28	2,99	2,89	2,69	2,59	2,51	2,36	2,06	1,89	1,73	1,43
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	1,87	1,69	1,52	1,00

Таблица X

ГРАНИЧНЕ ВРЕДНОСТИ БРОЈА НИЗОВА ЗА ТЕСТ ХОМОГЕНОГ НИЗА К.

$p = 0,05$

n_1 \ n^2	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2										2	2	2	2	2	2	2	2	2	2
3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Таблица XI

ДОЊА И ГОРЊА КРИТИЧНА ВРЕДНОСТ W ТЕСТА СУМЕ РАНГОВА

n_2	α		n_1(мањи узорак)						
	Једносмерни	Двосмерни	4	5	6	7	8	9	10
4	.05	.10	11,25						
	.025	.05	10,26						
	.01	.02							
	.005	.01							
5	.05	.10	12,28	19,36					
	.025	.05	11,29	17,38					
	.01	.02	10,30	15,39					
	.005	.01		15,40					
6	.05	.10	13,31	20,40	28,50				
	.025	.05	12,32	18,42	16,52				
	.01	.02	11,33	17,43	24,54				
	.005	.01	10,34	16,44	23,55				
7	.05	.10	14,34	21,44	29,55	39,66			
	.025	.05	13,35	20,45	27,57	36,69			
	.01	.02	11,37	18,47	25,59	34,71			
	.005	.01	10,38	16,49	24,60	32,73			
8	.05	.10	15,37	23,47	31,59	41,71	51,85		
	.025	.05	14,38	21,49	29,61	38,74	49,87		
	.01	.02	12,40	19,51	27,63	35,77	45,91		
	.005	.01	11,41	17,53	25,65	34,78	43,93		
9	.05	.10	16,40	24,51	33,63	43,76	54,90	66,105	
	.025	.05	14,42	22,53	31,65	40,79	51,93	62,109	
	.01	.02	13,43	20,55	28,68	37,82	47,97	59,112	
	.005	.01	11,45	18,57	26,70	35,84	45,99	56,115	
10	.05	.10	17,43	26,54	35,67	45,81	56,96	69,111	82,128
	.025	.05	15,45	23,57	32,70	42,84	53,99	65,115	78,132
	.01	.02	13,47	21,59	29,73	39,87	49,103	61,119	74,136
	.005	.01	12,48	19,61	27,75	37,89	47,105	58,122	71,139

Таблица XII. ГРАНИЧНЕ ВРЕДНОСТИ ЗА ТЕСТ ПРЕДЗНАКА

N	1%	5%	10%	25%	N	1%	5%	10%	25%
1					46	13	15	16	18
2					47	14	16	17	19
3				0	48	14	16	17	19
4				0	49	15	17	18	19
5			0	0	50	15	17	18	20
6		0	0	1			18	19	20
7		0	0	1	52	16	18	19	21
8	0	0	1	1	53	16	18	20	21
9	0	1	1	2	54	17	19	20	22
10	0	1	1	2	55	17	19	20	22
11	0	1	2	3	56	17	20	21	23
12	1	2	2	3	57	18	20	21	23
13	1	2	3	3	58		21	22	24
14	1	2	3	4	59	19	21	22	24
15	2	3	3	4	60	19	21	23	25
16	2	3	4	5	61	20	22	23	25
17	2	4	4	5	62	20	22	24	25
18	3	4	5	6	63	20	23	24	26
19	3	4	5	6	64	21	23	24	26
20	3	5	5	6	65	21	24	25	27
21	4	5	6	7	66	22	24	25	27
22	4	5	6	7	67	22	25	26	28
23	4	6	7	8	68	22	25	26	28
24	5	6	7	8	69	23	25	27	29
25	5	7	7	9	70	23	26	27	29
26	6	7	8	9	71	24	26	28	30
27	6	7	8	10	72	24	27	28	30
28	6	8	9	10	73	25	27	28	31
29	7	8	9	10	74	25	28	29	31
30	7	9	10	11	75	25	28	29	32
31	7	9	10	11	76	26	28	30	32
32	8	9	10	12	77	26	29	30	32
33	8	10	11	12	78	27	29	31	33
34	9	10	11	13	79	27	30	31	33
35	9	11	12	13		28	30	32	34
36	9	11	12	14	81	28	31	32	34
37	10	12	13	14	82	28	31	33	35
38	10	12	13	14	83	29	32	33	35
39	11	12	13	15	84	29	32	33	36
40	11	13	14	15	85	30	32	34	36
41	11	13	14	16	86	30	33	34	37
42	12	14	15	16	87	31	33	35	37
43	12	14	15	17	88	31	34	35	38
44	13	15	16	17	89	31	34	36	38
45	13	15	16	18	90	32	35	36	39

Таблица XIII

ДОЊА И ГОРЊА КРИТИЧНА ВРЕДНОСТ W^+ WILCOXON-овог ТЕСТА РАНГА СА ЗНАКОМ

Једносмерни: Двосмерни:	$\alpha = .05$ $\alpha = .10$	$\alpha = .025$ $\alpha = .05$	$\alpha = .01$ $\alpha = .02$	$\alpha = 0{,}005$ $\alpha = .01$
		Доња, Горња		
5	0,15			
6	2,19	0,21		
7	3,25	2,26	0,28	
8	5,31	3,33	1,35	0,36
9	8,37	5,50	3,42	1,44
10	10,45	8,47	5,50	3,52
11	13,53	10,56	7,59	5,61
12	17,61	13,65	10,68	7,71
13	21,70	17,74	12,79	10,81
14	25,80	21,84	16,89	13,92
15	30,90	25,95	19,101	16,104
16	35,101	29,107	23,113	19,117
17	41,112	34,119	27,126	23,130
18	47,124	40,131	32,139	27,144
19	53,137	46,144	37,153	32.158
20	60,150	52,158	43,167	37,173
21	67,164	58,173	49,182	43,188
22	75,178	66,187	55,198	48,205
23	83,193	73,203	62,214	54.222
24	91,209	81,219	69,231	61,239
25	100,225	89,236	76,249	68,257
26	110,241	98,253	84,267	75,276
27	119,259	107,271	93,285	83,295
28	130,276	116,290	101,305	91,315
29	140,295	126,309	110,325	100,335
30	151,314	137,328	120,345	109,356

Таблица XIV.

КРИТИЧНЕ ВРЕДНОСТИ H KRUSKAL-WALLIS- овог ТЕСТА У СЛУЧАЈУ 3 НЕЗАВИСНА УЗОРКА

			Ниво значајности α						
n_1	n_2	n_3	0.200	0.100	0.050	0.025	0.010	0.005	0.001
2	2	1	3.60						
2	2	2	3.71	4.57					
3	2	1	3.52	4.29					
3	2	2	3.93	4.50	4.71				
3	3	1	3.29	4.57	5.14				
3	3	2	3.78	4.56	5.36	5.56			
3	3	3	3.47	4.62	5.60	5.96	7.20	7.20	
4	1	1	3.57						
4	2	1	3.16	4.50					
4	2	2	3.67	4.46	5.33	5.50			
4	3	1	3.21	4.06	5.21	5.83			
4	3	2	3.44	4.51	5.44	6.00	6.44	7.00	
4	3	3	3.39	4.71	5.73	6.15	6.75	7.32	8.02
4	4	1	3.27	4.17	4.97	6.17	6.67		
4	4	2	3.46	4.55	5.45	6.08	7.04	7.28	
4	4	3	3.42	4.55	5.60	6.39	7.14	7.60	8.33
4	4	4	3.50	4.65	5.69	6.62	7.65	8.00	8.65
5	1	1	3.86						
5	2	1	3.33	4.20	5.00				
5	2	2	3.36	4.37	5.16	6.00	6.53		
5	3	1	3.22	4.02	4.96	6.04			
5	3	2	3.41	4.65	5.25	6.00	6.82	7.19	
5	3	3	3.44	4.53	5.65	6.32	7.08	7.52	8.24
5	4	1	3.09	3.99	4.99	5.78	6.95	7.36	
5	4	2	3.36	4.54	5.27	6.04	7.12	7.57	8.11
5	4	3	3.32	4.55	5.63	6.41	7.45	7.91	8.50
5	4	4	3:33	4.62	5.62	6.67	7.76	8.14	9.00
5	5	1	3.24	4.11	5.13	6.00	7.31	7.75	
5	5	2	3.39	4.51	5.34	6.35	7.27	8.13	8.69
5	5	3	3.43	4.55	5.71	6.49	7.54	8.24	9.06
5	5	4	3.31	4.52	5.64	6.67	7.77	8.37	9.32
5	5	5	3.42	4.56	5.78	6.74	8.00	8.72	9.68

Таблица XV.
КРИТИЧНЕ ВРЕДНОСТИ Q FRIEDMAN-овог ТЕСТА У СЛУЧАЈУ 3 И 4 ЗАВИСНА УЗОРКА

	Ниво значајности α						
	0.200	0.100	0.050	0.025	0.010	0.005	0.001
				k=3			
n							
2	4.00						
3	4.67	6.00	6.00				
4	4.50	6.00	6.50	8.00	8.00	8.00	
5	3.60	5.20	6.40	7.60	8.40	10.00	10.00
6	4.00	5.33	7.00	8.33	9.00	10.33	12.00
7	3.71	5.43	7.14	7.71	8.86	10.29	12.29
8	4.00	5.25	6.25	7.75	9.00	9.75	12.25
9	3.56	5.56	6.22	8.00	9.56	10.67	12.67
10	3.80	5.00	6.20	7.80	9.60	10.40	12.60
11	3.82	5.09	6.55	7.82	9.46	10.36	13.27
12	3.50	5.17	6.50	8.00	9.50	10.17	12.50
13	3.85	4.77	6.62	7.54	9.39	10.31	12.92
				k=4			
2	5.40	6.00	6.00				
3	5.40	6.60	7.40	8.20	9.00	9.00	
4	4.80	6.30	7.80	8.40	9.60	10.20	11.10
5	5.16	6.36	7.80	8.76	9.96	10.92	12.60
6	4.80	6.40	7.60	8.80	10.20	11.40	13.00
7	4.89	6.43	7.80	9.00	10.37	11.40	13.80
8	4.80	6.30	7.65	9.00	10.50	11.85	13.95

Таблица XVI

РАСПОДЕЛА λ КОЛМОГОРОВА – СМИРНОВА

Табела даје вредности функције Q(λ) дате обрасцем

$$Q(\lambda)=\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2\lambda^2}$$

λ	Q(λ)	λ	Q(λ)	λ	Q(λ)	λ	Q(λ)	λ	Q(λ)	λ	Q(λ)
0,32	0,0000	0,66	0,2236	1,00	0,7300	1,34	0,9449	1,68	0,9929	2,00	0,9993
0,33	0,0001	0,67	0,2396	1,01	0,7406	1,35	0,9478	1,69	0,9934	2,01	0,9994
0,34	0,0002	0,68	0,2558	1,02	0,7508	1,36	0,9505	1,70	0,9938	2,02	0,9994
0,35	0,0003	0,69	0,2722	1,03	0,7608	1,37	0,9531	1,71	0,9942	2,03	0,9995
0,36	0,0005	0,70	0,2888	1,04	0,7704	1,38	0,9556	1,72	0,9946	2,04	0,9995
0,37	0,0008	0,71	0,3055	1,05	0,7798	1,39	0,9580	1,73	0,9950	2,05	0,9996
0,38	0,0013	0,72	0,3223	1,06	0,7889	1,40	0,9603	1,74	0,9953	2,06	0,9996
0,39	0,0019	0,73	0,3391	1,07	0,7976	1,41	0,9625	1,75	0,9956	2,07	0,9996
0,40	0,0028	0,74	0,3560	1,08	0,8061	1,42	0,9646	1,76	0,9959	2,08	0,9996
0,41	0,0040	0,75	0,3728	1,09	0,8143	1,43	0,9665	1,77	0,9962	2,09	0,9997
0,42	0,0055	0,76	0,3896	1,10	0,8223	1,44	0,9684	1,78	0,9965	2,10	0,9997
0,43	0,0074	0,77	0,4064	1,11	0,8299	1,45	0,9702	1,79	0,9967	2,11	0,9997
0,44	0,0097	0,78	0,4230	1,12	0,8374	1,46	0,9718	1,80	0,9969	2,12	0,9997
0,45	0,0126	0,79	0,4395	1,13	0,8445	1,47	0,9734	1,81	0,9971	2,13	0,9998
0,46	0,0160	0,80	0,4559	1,14	0,8514	1,48	0,9750	1,82	0,9973	2,14	0,9998
0,47	0,0200	0,81	0,4720	1,15	0,8580	1,49	0,9764	1,83	0,9975	2,15	0,9998
0,48	0,0247	0,82	0,4880	1,16	0,8644	1,50	0,9778	1,84	0,9977	2,16	0,9998
0,49	0,0300	0,83	0,5038	1,17	0,8706	1,51	0,9791	1,85	0,9979	2,17	0,9998
0,50	0,0361	0,84	0,5194	1,18	0,8765	1,52	0,9803	1,86	0,9980	2,18	0,9999
0,51	0,0428	0,85	0,5347	1,19	0,8823	1,53	0,9815	1,87	0,9981	2,19	0,9999
0,52	0,0503	0,86	0,5497	1,20	0,8877	1,54	0,9826	1,88	0,9983	2,20	0,9999
0,53	0,0585	0,87	0,5645	1,21	0,8930	1,55	0,9836	1,89	0,9984	2,21	0,9999
0,54	0,0675	0,88	0,5791	1,22	0,8981	1,56	0,9846	1,90	0,9985	2,22	0,9999
0,55	0,0772	0,89	0,5933	1,23	0,9030	1,57	0,9855	1,91	0,9986	2,23	0,9999
056	0,0876	0,90	0,6073	1,24	0,9706	1,58	0,9864	1,92	0,9987	2,24	0,9999
0,57	0,0987	0,91	0,6209	1,25	0,9121	1,59	0,9873	1,93	0,9988	2,25	0,9999
0,58	0,1104	0,92	0,6343	1,26	0,9164	1,60	0,9880	1,94	0,9989	2,26	0,9999
0,59	0,1228	0,93	0,6473	1,27	0,9206	1,61	0,9888	1,95	0,9990	2,27	0,9999
0,60	0,1357	0,94	0,6601	1,28	0,9245	1,62	0,9895	1,96	0,9991	2,28	0,9999
0,61	0,1492	0,95	0,6725	1,29	0,9283	1,63	0,9902	1,97	0,9991	2,29	0,9999
0,62	0,1632	0,96	0,6846	1,30	0,9319	1,64	0,9908	1,98	0,9992	2,30	0,9999
0,63	0,1778	0,97	0,6964	1,31	0,9354	1,65	0,9914	1,99	0,9993	2,31	1,0000
0,64	0,1927	0,98	0,7079	1,32	0,9387	1,66	0,9919				
0,65	0,2080	0,99	0,7191	1,33	0,9418	1,67	0,9924				

Таблица XVII

МАКСИМАЛНИ БРОЈ ИНВЕРЗИЈА MW ПРИ КОЈЕМ ЈЕ РАЗЛИКА ЈОШ УВЕК ЗНАЧАЈНА

(U – тест Wilcoxon-Mann Whitney)

За оштрији критеријум $\alpha = 0{,}01$

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n_2														
5	–	–	0	1										
6	–	–	1	2	3									
7	–	0	1	2	4	6								
8	–	0	2	4	6	7	9							
9	–	1	3	5	7	9	11	14						
10	–	1	3	6	8	11	13	16	19					
11	–	1	4	7	9	12	15	18	22	25				
12	–	2	5	8	11	14	17	21	24	28	31			
13	0	2	5	9	12	16	20	23	27	31	35	38		
14	0	2	6	10	13	17	22	26	30	34	38	43	47	
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56
16	0	3	7	12	16	21	26	31	36	41	46	51	55	61

за блажи критеријум $\alpha = 0{,}05$

n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	–	0												
4	–	0	1											
5	0	1	2	4										
6	0	2	3	5	7									
7	0	2	4	0	8	11								
8	1	3	5	8	10	13	15							
9	1	4	6	9	12	15	18	21						
10	1	4	7	11	14	17	20	24	27					
11	1	5	8	12	16	19	3	27	31	34				
12	2	5	9	13	17	21	26	30	34	38	43			
13	2	6	10	15	19	24	28	33	37	42	47	51		
14	3	7	11	16	21	26	31	36	41	46	51	56	61	
15	3	7	12	18	23	28	33	39	41	40	55	61	68	72
16	3	8	14	19	25	30	36	42	48	54	60	65	71	77

Таблица XVIII. КРИТИЧКЕ ВРЕДНОСТИ Q_{α} TUKEY-евог ТЕСТА

	Број аритметичких средина које се пореде ($\alpha = 0,05$)								
SSRV*	2	3	4	5	6	7	8	9	10
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11
18	2.97	3,61	4.00	4.28	4.49	4.67	4.82	4.96	5.07
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47

–	11	12	13	14	15	16	17	18	19	20
1	50.59	51.96	53.20	54.33	55.36	56.32	57.22	58.04	58.83	59.56
2	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	9.72	9.95	10.15	10.35	10.52	10.69	10.84	10.98	11.11	11.24
4	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

* Степен слободе резидуалне варијансе

Наставак Таблице XVIII

	Број аритметичких средина које се пореде (α=0,01)								
SSRV*	2	3	4	5	6	7	8	9	10
1	90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6
2	14.04	19.02	22.29	24.72	26.63	28.20	29.53	30.68	31.69
3	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69
4	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44
16	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16

—	11	12	13	14	15	16	17	18	19	20
1	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
2	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
3	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
4	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
5	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	9.30	9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
10	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
14	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39
15	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
18	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
24	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
60	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
120	5.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

* Степен слободе резидуалне варијансе

Таблица XIX СЛУЧАЈНИ БРОЈЕВИ

Ред	Колона									
No	0	1	2	3	4	5	6	7	8	9
0	87331	82442	28104	26432	83640	17323	68764	84728	37995	96106
1	33628	17364	01409	87803	65641	33433	48944	64299	79066	31777
2	54680	13427	72496	16967	16195	95593	55040	53729	62035	66717
3	51199	49794	49407	10774	98140	83891	37195	24066	61140	65144
4	78702	98067	61313	91661	59861	54437	77739	19892	54817	88645
5	55672	16014	24892	13089	00410	81458	76156	28189	40595	21500
6	18880	58497	03862	32368	59320	24807	63392	79793	63043	09425
7	10242	62548	62330	05703	33535	49128	66298	16193	55301	01306
8	54993	17182	94168	23228	83895	73251	68199	64639	83178	70521
9	22686	50885	16006	04041	08077	33065	35065	35237	02502	72062
10	42349	03145	15770	70665	53291	32288	41568	66079	98705	31029
11	18093	09553	39428	75464	71329	86344	80729	40916	18860	51780
12	11535	03924	84252	74795	40193	84597	42497	21918	91384	84721
13	35066	73848	65351	53270	67341	70177	92373	17604	42204	60476
14	57477	22809	73558	96182	96779	01604	25748	59553	64876	94611
15	48647	33850	52956	45410	88212	05120	99391	32276	55961	41775
16	86857	81154	22223	74950	53296	67767	55866	49061	66937	81818
17	20182	36907	94644	99122	09774	29189	27212	79000	50217	71077
18	83687	31231	01133	41432	54542	60204	81618	09586	34438	87683
19	81315	12390	46074	47810	90171	36313	95440	77583	28506	38808
20	87026	52826	58341	76549	04105	66191	12914	55348	07907	06978
21	34301	76733	07251	90524	21931	83695	41340	53581	64582	60210
22	70734	24337	32674	49508	49751	90489	63202	24380	77943	09942
23	94710	31527	73445	32839	68176	53580	85250	53243	03350	00128
24	76462	16987	07775	43162	11777	16810	75158	13894	88945	15539
25	14348	28403	79245	69023	34196	46398	05964	64715	11330	17515
26	74618	89317	30146	25606	94507	98104	04239	44239	37636	88866
27	99442	19200	85406	45358	86253	60638	38858	44964	54103	57287
28	26869	44399	89452	06652	31271	00647	46551	83050	92058	83814
29	80988	08149	50499	98584	28385	63680	44638	91864	96002	87802
30	07511	79047	89289	17774	67194	37362	85684	55505	97809	67056
31	49779	12138	05048	03535	27502	63308	10218	53296	48687	61340
32	47938	55945	24003	19635	17471	65997	85906	98694	56420	78347
33	15604	06626	14360	79542	13512	87595	08542	03800	35443	52823
34	12307	27726	21864	00045	16075	03770	86978	86978	02693	09096
35	02450	28053	66134	99445	91316	25727	89399	85272	67148	78358
36	57623	54382	35236	89244	27245	90500	75430	96762	71968	65838
37	91762	78849	93105	40481	99431	03304	21079	86459	21287	76566
38	87373	31137	31128	67050	34309	44914	80711	61738	61498	24288
39	67094	41485	54149	86088	10192	21174	39948	67268	29938	32476
40	94456	66747	76922	87627	71834	57688	04878	78348	68970	60048
41	68359	75292	27710	86889	81678	79798	58360	39175	75667	65782
42	52393	31404	32584	06837	79762	13168	76055	54833	22841	98889
43	59565	91254	11847	20672	37625	41454	86861	55824	79793	75475
44	48185	11066	20162	38230	16043	48409	47421	21195	98008	57305

Таблица XIX Наставак

	0	1	2	3	4	5	6	7	8	9
45	19230	12187	86659	12971	52207	76546	63272	19312	81662	96557
46	84327	21942	81727	68735	89190	58491	55329	96875	19465	89687
47	77430	71210	00591	50124	12030	50280	12358	76174	48353	09682
48	12462	19108	70512	53926	25595	97085	03833	59806	12351	64253
49	11684	06644	57816	10078	45021	47751	38285	73520	08434	65627
50	12896	36576	68686	08462	65662	76571	70891	09007	04581	01684
51	59090	05111	27587	90349	30789	50304	70650	06646	70126	15284
52	42486	67483	65282	19037	80588	73076	41820	46651	40442	40718
53	88662	03928	03249	85910	97533	88643	29829	21557	47328	36724
54	69403	03626	92678	53460	15465	83516	54012	80509	55976	46115
55	56434	70543	38696	98502	32092	95505	62091	39549	30117	98209
56	58227	62694	42837	29183	11393	68463	25150	86338	95620	39836
57	41272	94927	15413	40505	33123	63218	72940	98349	57249	40170
58	36819	01162	30425	15546	16065	68459	35776	64276	92868	07372
59	31700	66711	26115	55755	33584	18091	38709	57276	74660	90392
60	69855	63699	36839	90531	97125	87875	62824	03889	12538	24740
61	44322	17569	45439	41455	34324	90902	07978	26268	04279	76816
62	62226	36661	87011	66267	78777	78044	40819	49496	39814	73867
63	27284	19737	98741	72531	52741	26699	98755	19657	08665	16818
64	88341	21652	94743	77268	79525	44769	66583	30621	90534	62050
65	53266	18783	51903	56711	38060	69513	61963	80460	88018	86510
66	50527	49330	24839	42529	03944	95219	88724	37247	84166	23023
67	15655	07852	77206	35944	71446	30573	19405	57824	23576	23301
68	62057	22206	03314	83465	57466	10465	19891	32308	01900	67484
69	41769	56091	19892	96253	92808	45785	52774	49674	68103	65032
70	25993	72416	44473	41299	93095	17338	69802	98548	02429	85238
71	22842	57871	04470	37373	34516	04042	04078	35336	34393	97573
72	55704	31982	05234	22664	22181	40358	28089	15790	33340	18852
73	94258	18706	09437	96041	90052	80862	20420	24323	11635	91677
74	74145	20453	29657	98868	56695	53483	87449	35060	98942	62697
75	88881	12673	73961	89884	73247	97670	69570	88888	58560	72580
76	01508	56780	52223	35632	73347	71317	46541	88023	36656	76332
77	92069	43000	23233	06058	82527	25250	27555	20426	60361	63525
78	53366	35249	02117	68620	39388	69795	73215	01846	16983	78560
79	88057	54097	45511	74867	32192	90071	04147	46094	63519	07199
80	85492	82238	02668	91854	86149	28590	77853	81035	45561	16032
81	39453	62123	69611	53017	34964	09786	24614	49514	01056	18700
82	82627	98111	93870	56969	69566	62662	07353	84838	14570	14508
83	61142	51743	38209	31474	96095	15163	54380	77849	20465	03142
84	12031	32528	61311	53730	89032	16124	58844	35386	45521	59368
85	31313	59838	29147	76882	74328	09955	63673	96651	53264	29871
86	50767	41056	97409	44376	62219	35439	70102	99248	71179	26052
87	30522	95669	84966	26554	24768	72247	84993	85375	92518	16334
88	74176	19870	89874	64799	03792	57006	57225	36677	46825	14087
89	17114	93248	37065	91346	04657	93763	92210	43676	44944	75798
90	53005	11825	64608	87587	05742	31914	55044	41818	29667	77424
91	31985	81539	79942	49471	46200	27639	94099	42085	79231	03932
92	63499	60508	77522	15624	15088	78519	52279	79214	43623	69166
93	30506	42444	99047	66010	91657	37160	37408	85714	21420	80996
94	78248	16841	92357	10130	68990	38307	61022	56806	81016	38511
95	64996	84789	50185	32200	64382	29752	11876	00664	54547	62597
96	11963	13157	09136	01769	30117	71486	80111	09161	08371	71749
97	44335	91450	43456	90449	18338	19797	31339	50473	06606	89788
98	42477	11868	44520	01113	11341	11743	97949	49718	99176	42006
99	77562	18863	58515	90166	78508	14864	19111	57183	85808	59385

ЛИТЕРАТУРА

ЛИТЕРАТУРА

Andersosn, T. W. [1958] An Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Ltd., NewYork.

Bartlet S. [1933] On the Theory of Statistical Regresion, Proc. R. Soc., 53, Edingburgh

Blalock H. M. [1979] Social Statistics, McGraw-Hill, New York

Cochran, W. G. [1972] Sampling Techniques, Secondedition, WilleyEastern Private Limited, NewDelhi.

Cramer H. [1946] Mathematical Metods of Statistics, University Press, Princenton.

Cramer H. [1955] The Elements of Probability Theory and Some of its Applications, JohnWiley & Sons, Ltd., New York.

Cuthbertson, K., S. G. Hall, M. P. Taylor [1992] Applied Econometrics Techniques, Phillip Allan, University of Michigan Press.

Dillon, W. R., M. Goldstein [1984] Multivariate Analysis: Methods and Applications, JohnWiley & Sons, Ltd., New York.

Draper N. R., Smith H. [1996] Applied Regression Analysis, J. Wiley., New York

Dugue D. [1973] Appliation des propriete de la limite an sens du calcul des probabilites a l'etude de diverses questions d' estimation, Journ. De l'Ec. Polytechn., Paris.

Enders, W. [1995] Applied Econometric Time Series, A User's Guide, John Wiley & Sons, Ltd., New York.

Feller W. [1968] An Introduction to Probability Theory and its Applications, Vol 1, J. Wiley & Sons, Ltd., New York.

Fisher A. R. [1954] Statistical Methods for Research Workers, Oliver & Bojd, Edinburg.

Gibbons J. D. [1992] Nonparametric Statistical Inference, Marcel Dekker.

Gnedenko V. B. [1969] Kurs teorii veroathnosteii, Moskva.

Gosset S. W. [1908] “Student”: The Probable Error of a Mean, Biometrika, V. 6.

Grais, B. [1983] Exercices de statistique descrptive, Dunod, Paris.

Grđić, G., M. Eremić, D. Mladenović [1980] Ekonomsko-statistička analiza, Ekonomski fakultet, Beograd.

Griffiiths, W., R. Carter Hill, G. Judge [1993] Learning and Practicing Econometrics, JohnWiley & SonsLtd., New York.

Group of Authors [2006] e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook

Hartwig F, B. E. Dearing [1985] Exploratory Data Analysis, SagePublications, Beverly Hills.

Harvey, A. [1990] The Econometric Analysis of Time Series, 2nd ed., MIT Press.

Ivanović B. [1959] Teorija korelacije sa primenom u ekonomskim istraživanjima, Beograd.

Ivanović B., [1973] Teorijska statistika, Naučna knjiga, Beograd.

Ivković, Z [1980] Matematička statistika, Naučna knjiga, Beograd.

Ivković, Z [1988] Verovatnoća i matematička statistika, Naučna knjiga, Beograd

Johnston, J. [1988] Econometrics Methods, McGrawHill.

Juran M. J. [1962]Quality Control Handbook, New York

Kish L. [1965] Survey Sampling, JohnWiley andSons Inc., New York, London, Sydney.

Kish, L. [1987] Statistical Design for Research, John Wiley & Sons, New York.

Klaemer H. C. [1982]Kappa Coefficient, In: Encyclopedic ofStatistical Sciences, S. Kotz andN. L. Jonson, John Willyand Sons, New York.

Klainbaum D. G., Kupper L. L., Mogensferin H. [1982] Epidemiological Research Principles and Quantitative Methods, Lifetime Learnings Publishing.

Kovačić, Z. [1995] Analiza vremenskih serija, Ekonomski fakultet, Beograd.

Kraft H. C., Eeden VanC. [1968] A Nonparametric Introduction to Statistics, New York.

Krčevinac, S. [1993] Ekonometrijske metode, skripta, Fakultet organizacionih nauka, Beograd.

Lehmann L. E. [1959] Testing Statistical Hypotheses, New York.

Mann, P. S. [1995] Statistics for Business and Economics, JohnWiley & Sons, Ltd.

Mann, P. S. [2020] Introductory Statistics, 10th Edition, Wiley

Markland, A. [1988] Quantitative Methods for Management Decisions, JohnWiley & Sons, Ltd.

Markus, G. B. [1984] Analyzing Panel Data, Sage Publications, Bevery Hills.

Merkle, M. [2001] Verovatnoća i statistika, Akademska misao, Beograd

Mills C. F. [1956] Introduction to Statistics, New York.

Milošević, V.[1983] Teoijska statistika, Naučna knjiga, Beograd.

Milošević, V. [2000] Teorija verovatnoće, Zavod za udžbenike, 2000.

Mises VonR. Mathematical Theory of Probability and Statistics, NewYork.

Mladenović, P. [2002] Verovatnoća i statistika, Matematički fakultet, 2002.

Norušis M. J. [1993] SPSS for Windows Base System User's Guide, SPSS Inc., Chicago

Norusis M.J. [1990] SPSS/PC+ v4.0 Advanced Statistica, SPSS Inc, Chicago.

Norušis, M.J. [1990] SPSS/PC+ 4.0 Base Manual, SPSS Inc., Chicago.

Ostrom, C. W., Jr. [1978] Time Series Anaylis: Regression Tecniques, Sage Univeristy Paper serieson Quantitative Applications inthe SocialSciences, SageUniveristy Paper, Beverly Hills/London.

Pearson K. [1900] On A Criterian That A System Of Deviations From The Probable In The Case Of A Correlated System Of Variables Is Such That It Can Ve Reasonable Supposed To Have Arisen In Random Sampling, Phil. Mag. V. 50.

Pindyck, R. S., D. L. Rubinfeld [1976] Econometric Models and Ecnonomic Forecasts, McGraw-Hill, New York.

Prohorov V. A., Rozanov A. [1973] Teoria veroathnosteii, Osnoviponotio, Predillnie teoremi, Slucaiinie processi, Moskva.

Rao R. C. [1968] Linear Statistical Inference and its Applications, New York, Prevod "Nauka", Moskva.

Searle S. R. [1971]Linear Models, John Wiley & Sons, New York

Sharma, S. [1996] Applied Multivariate techniques, JohnWiley & Sons, Ltd., New York

Siegel S. [1956] Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York.

Storm R. [1970] Teoria verovathnosteii matematičeskaa statistika statističeskii kontrola kačestva, Prevodsa nemačkog, Moskva.

Theil H. [1971] Principals of Econometrics, JohnWiley & Sons, Ltd., New York.

Vuković N. [1978 i 1981] Verovatnoća i statistika, Beograd.

Vuković N. [1987] Statistička analiza, Naučna knjiga, Beograd.

Vuković, N. [2000 i 2002] PC Verovatnoća i statistika, FON, Beograd

Vuković, N., Spasić, S. [2020] Statistika za inžinjere, Treće izdanje, Univerzitet Singidunum

Wilks S. [1967] Matematičeskaa statistika, Prevodsa engleskog, Moskva.

Wonnacott, R.J., T. H. Wonnacott [1979] Econometrics, 2 nd Edition, John Wiley & Sons, Ltd., NewYork

Zacks S. [1971] The Theory of Statistical Inference, NewYork, Moskva.

Žarković, S. S. [1976] Upotreba statističkih podataka, Savezni zavodza statistiku, Beograd.

Zečević, T., M. Kovačević, M. Kovačević [1991] Teorija uzoraka i planiranje eksperimenata, Ekonomski fakultet, Beograd.

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

311(075.8)

ВУКОВИЋ, Наход, 1941-
Статистика са практикумом / Наход Вуковић, Слађана Спасић. - 1. изд. - Београд : Универзитет Сингидунум, 2022 (Београд : Бирограф). - VI, 460 стр. : илустр. ; 24 cm

Тираж 300. - Библиографија: стр. 456-460.

ISBN 978-86-7912-776-1

1. Спасић, Слађана, 1966- [аутор]

а) Статистика

COBISS.SR-ID 60090633

www.singidunum.ac.rs

9 788679 127761

Находⅾ Вуковић
Слађана Спасић

СТАТИСТИКА СА ПРАКТИКУМОМ

Ова књига обрађује основе вероватноће и методе статистичке анализе обухваћене програмом предмета Статистика за инжењере који се изучава на првој години Факултета за инжењерски менаџмент Универзитета Сингидунум. Зато је ова књига намењена пре свега студентима Универзитета Сингидунум, али и свима који желе да буду успешни у својим активностима, било да су менаџери, предузетници, инжењери. Стечено и усвојено знање биће им од помоћи у проналажењу решења у свакодневним пословним одлукама које су базиране на статистичким подацима.

Универзитет
Сингидунум