1. Појам, значај и области статистике

ЗНАЧЕЊА СТАТИСТИКЕ

Статистика је део примењене математике, тј. представља тзв. „науку о подацима“. Постоје два основна значења статистике:

1. Статистика као нумерички подаци

Статистику у свакодневном животу доживљавамо као **нумеричке** (бројчане) **податке** које користимо и којих смо сведоци. На пример, то су износ плате, број студената који су уписали Економски факултет, број заражених особа вирусом COVID-19 итд.

2. Статистика као научна дисциплина

У академском животу, статистику доживљавамо као **научну дисциплину**, а не само пуке нумеричке податке из свакодневног живота. У том смислу, статистику схватамо као **научни метод** који се користи за:

- прикупљање и сређивање података
- приказивање података
- анализу и интерпретацију података
- доношење статистичких и економских одлука (тј. за оцењивање, тестирање и прогнозирање)

ПРЕДМЕТ ИЗУЧАВАЊА СТАТИСТИЧКЕ АНАЛИЗЕ

Предмет изучавања статистичке анализе су **варијабилне појаве** (променљиве појаве). Ово су појаве које имају различите вредности или појавне облике од случаја до случаја. Ове појаве се вредносно разликују по:

- **јединици посматрања**: На пример, по земљама имамо БДП Србије, БДП Хрватске, БДП Словеније итд. У овом примеру Србија, Хрватска, Словенија су јединице посматрања. БДП је варијабилна појава која се вредносно разликује у зависности од *јединице* посматрања.
- **времену посматрања**: На пример, различите су вредности за БДП Србије у 2016. години, БДП Србије у 2017. години, БДП Србије у 2018. години, БДП Србије у 2019. години итд. БДП је варијабилна појава која се вредносно разликује у зависности од *времена* посматрања.

Најмањи варијабилитет показују елементарне појаве у физици и хемији (нпр. капи кише). Највећи варијабилитет имају друштвено економске појаве (нпр. успех студента, место становања).

ЦИЉ СТАТИСТИЧКЕ АНАЛИЗЕ

Наравно, постоје разлози због чега се спроводи статистичка анализа. Два основна циља статистичке анализе су:

- откривање статистичких законитости
- анализа одступања од статистичких законитости

На пример, као неке статистичке законитости изводимо:

- *очекивано трајање живота*: жене генерално живе дуже (статистичка законитост), али постоје одређена одступања по земљама од ове законитости

- *бацање новчића*: што више пута бацамо новчић, број понављања „главе" и „писма" биће приближнији - веровали или не! Ово је тзв. **закон великих бројева**. Вероватноћа да бачени новчић покаже писмо или главу износи ½. Што се више понавља овај експеримент, то ће бити вероватније да ће број исхода када „падне глава" (релативна вероватноћа исхода „глава"), бити близак вредности ½, као и релативна вероватноћа исхода „писмо".

ОБЛАСТИ СТАТИСТИКЕ

Статистика као наука обухвата две области – теоријску статистику и примењену статистику.

1. Теоријска (математичка) статистика

Теоријска (математичка) статистика је област статистике која се бави развијањем статистичких модела и теорема, као и доказивањем истих. Пример веома важне статистичке теореме је централна гранична теорема (ЦГТ). Више о овоме ћемо учити касније. За случајну променљиву $\bar{X}$ она гласи:

1) Уколико случајна променљива X има нормалну расподелу, случајна променљива $\bar{X}$ је такође нормално расподељена (без обзира на величину узорка), што записујемо као:

$$X \sim N(\mu\,;\sigma^2) \rightarrow \bar{X} \sim N(\mu_{\bar{X}}\,;{\sigma_{\bar{X}}}^2)$$

2) Уколико случајна променљива X нема нормалну расподелу, случајна променљива $\bar{X}$ је *приближно* нормално расподељена ако је величина узорка бар 30:

$$X \text{ није } N(\mu\,;\sigma^2) \rightarrow \bar{X} \text{ приближно } N(\mu_{\bar{X}}\,;{\sigma_{\bar{X}}}^2)$$

под условом да је $n \geq 30$

У оквиру теоријске статистике, статистичари су под одређеним претпоставкама извели ове закључке, које ми употребљавамо у решавању проблема. Решавање проблема је део примењене статистике, о чему ћемо више научити у наставку.

2. Примењена статистика

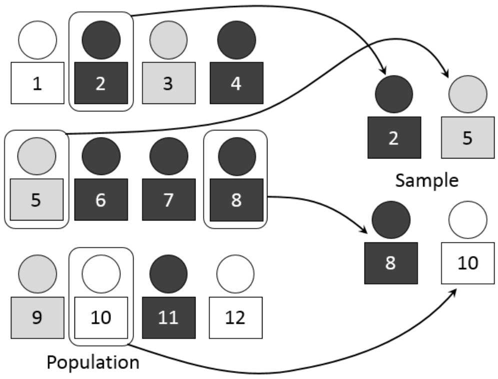
Ово је област статистике која се бави применом статистичких модела и теорема у решавању проблема. У оквиру примењене статистике разликујемо:

- **дескриптивну статистику** – она описује податке помоћу табела, графикона и сумарних показатеља (мера централне тенденције, дисперзије и облика расподеле)
- **инференцијалну (индуктивну) статистику** – она обухвата статистичке методе које примењујемо да бисмо на основу резултата из *узорка* дошли до статистичких закључака о *основном скупу.**

* Као што знамо, основни скуп је скуп свих елемената чије особине статистички испитујемо, а узорак је део основног скупа који је изабран у сврхе статистичке анализе. На пример, од 10.000.000 становника једне земље (основни скуп) бирамо 5.000 људи за анкету о животној средини (узорак).

2. Основни скуп и узорак (врсте узорака)

ШТА ПРЕДСТАВЉА ОСНОВНИ СКУП, А ШТА УЗОРАК?



Основни скуп је скуп свих елемената чије особине статистички испитујемо, а **узорак** је део основног скупа који је изабран у сврхе статистичке анализе. На пример, од 10.000.000 становника једне земље (основни скуп) бирамо 5.000 људи за анкету о животној средини (узорак). Треба напоменути да основни скуп могу чинити:

- појединци (бића) – као што је случај за пример који смо изнад навели;
- ствари – на пример, од 10.000 производа бирамо 100 производа за тестирање исправности;
- догађаји (феномени) – на пример, бирамо инфлацију у претходних пет година за анализу.

КАКО СЕ ВРШИ ПРИКУПЉАЊЕ ПОДАТАКА?

Прикупљање података у сврхе статистичке анализе врши се на основу **анкете**, која може бити попис и узорачка анкета.

1. Попис

Попис је анкета која се спроводи на **целом основном скупу**. Ретко се користи јер је скуп и дуготрајан, а и често је чак и немогуће идентификовати све елементе основног скупа.

2. Узорачка анкета

Узорачка анкета је анкета која се спроводи на **делу основног скупа**, тј. изабраном узорку. Њена сврха је доношење закључака о карактеристикама основног скупа из којег је узорак изабран. Да би ово било поуздано, потребно је да узорак буде *репрезентативан* за основни скуп, што значи да адекватно осликава карактеристике основног скупа.

ВРСТЕ УЗОРАКА

Према томе на који начин се узорак бира из основног скупа, разликујемо случајан и неслучајан (намеран) узорак.

1. Случајан узорак

Случајан узорак је онај где **сви елементи** основног скупа улазе у обзир за избор у одређени узорак. Математички, за све елементе основног скупа важи да је вероватноћа избора у узорак позитивна. Уколико би се десило да било који елемент основног скупа нема могућност да буде изабран у узорак, вероватноћа избора у узорак би била нула, и такав узорак не би био случајан.

2. Неслучајан (намеран) узорак

Неслучајан узорак је онај где **неки елементи** основног скупа **не** улазе у обзир за избор у одређени узорак, тј. неки елементи основног скупа имају вероватноћа избора у узорак која је једнака нули. У том контексту, неслучајне узорке можемо класификовати на:

- **погодан узорак** – У погодан узорак бирају се *најдоступнији* елементи основног скупа, у циљу добијања брзог резултата истраживања. Мање доступни елементи основног скупа се не узимају у обзир за избор у узорак.
- **узорак заснован на субјективном суду истраживача** – У овај узорак бирају се они елементи основног скупа за које истраживач субјективно процени да су највише одговарајући за узорак.
- **квота узорак** – Код овог узорка, основни скуп делимо у различите подскупове на основу одређених карактеристика. Потом из подскупа бирамо подузорак, тако да сваки подузорак има учешће у подскупу пропорционално учешћу тог подскупа у основном скупу. На пример, бирамо 1000 особа из града од којих је 49% мушкараца и 51% жена. Да бисмо изабрали квота узорак бирамо 490 мушкараца и 510 жена, како бисмо задржали пропорцију основног скупа.

Оно што су предности случајних узорака у односу на неслучајне узорке су следеће ставке:

- унапред је позната вероватноћа избора сваког елемента основног скупа у узорак
- може се утврдити поузданост закључка о основном скупу

Неслучајни (намерни) узорци не поседују ове предности.

3. Технике избора случајног узорка

ТЕХНИКЕ ИЗБОРА СЛУЧАЈНОГ УЗОРКА

Рекли смо да је случајан узорак онај где сви елементи основног скупа улазе у обзир за избор у одређени узорак. Постоје две основне технике избора случајног узорка – прост случајан узорак и контролисани случајан узорак.

1. Прост случајан узорак

Прост случајан узорак је техника у којој сваки елемент основног скупа **има исту вероватноћу** да буде изабран. Реч „случајан“ се односи на то да је вероватноћа избора сваког елемента у основни скуп позната, док се реч „прост“ односи на то да су те вероватноће за сваки елемент основног скупа једнаке.

Поступци за избор простог случајног узорка су:

- путем лутрије (извлачењем) – нпр. извлачење из кутије/шешира
- коришћење таблица случајних бројева – све се ређе користи, јер постоје статистички пакети попут Excel-a, Minitab-a, Ti-84, SPSS-a итд.

Прост случајан узорак може бити:

1. Прост случајан узорак са понављањем	**2. Прост случајан узорак без понављања**
➢ вероватноћа избора у извлачењима се не мења ➢ избори елемената у узорак су међусобно независни ➢ ређе се користе у пракси Нпр. извлачење куглица са враћањем	➢ вероватноћа избора у извлачењима се мења ➢ избори елемената у узорак су међусобно зависни ➢ чешће се користе у пракси Нпр. извлачење куглица без враћања – као што је извлачење „лото“ куглица

2. Контролисани случајан узорак

Контролисани случајан узорак је техника у којој сваки елемент основног скупа **нема исту вероватноћу** да буде изабран. Реч „случајан“ се односи на то да је вероватноћа избора сваког елемента у основни скуп позната, док се реч „контролисани“ односи на то да те вероватноће за сваки елемент нису једнаке.

Код оваквог узорка, основни скуп садржи међусобно јасно разграничене подскупове. Контролисани случајан узорак може бити:

- систематски узорак
- стратификовани узорак
- кластер узорак

ВРСТЕ КОНТРОЛИСАНОГ СЛУЧАЈНОГ УЗОРКА

Као што смо већ рекли, контролисани случајан узорак може бити:

1. Систематски узорак

Код систематског узорка, избор елемената вршимо по систематском реду полазећи од случајног изабраног почетка (другим речима, случајно бирамо први елемент, а потом бирамо сваки k-ти елемент). Кораци за формирање систематског узорка су следећи:

- из интервала елемената $\left(0, k = \frac{N}{n}\right)$ коришћењем таблице случајних бројева бирамо први елемент узорка (N је величина основног скупа, а n величина узорка коју желимо)
- почев од првог елемента, сваки следећи k-ти елемент основног скупа бирамо као елемент систематског узорка

На пример, са листе 1000 домаћинстава желимо да изаберемо 50. Први корак је да из интервала елемената $\left(0, k = \frac{N}{n} = \frac{1000}{50} = 20\right)$ изаберемо први елемент узорка случајним путем. Значи, од првих 20 домаћинстава са списка, бирамо једно домаћинство случајним путем (користећи таблице случајних бројева или софтвер). Претпоставимо да смо овим методом одабрали 11. домаћинство. Сада прелазимо на други корак где бирамо сваки следећи k-ти (двадесети) елемент основног скупа као елемент узорка. Значи, бирамо 11. домаћинство, па 31. домаћинство, па 51. домаћинство итд. у наш систематски узорак.

2. Стратификовани узорак

Стратификовани узорак је унија простих случајних узорака од којих је сваки изабран из по једног „стратума“. Кораци за формирање стратификованог узорка су следећи:

- основни скуп делимо на хомогене подскупове, тј. „стратуме“, према одређеној карактеристици
- из сваког стратума бирамо прост случајан узорак одређене величине (важно је напоменути да величине простих случајних узорака треба да буду пропорционалне величинама стратума у основном скупу)
- унија ових скупова представља стратификовани узорак

На пример, од 100 испитаника формирамо 3 групе: са ниским, средњим и високим приходима. Из сваког стратума бирамо прост случајан узорак и унија ових скупова је стратификовани узорак.

3. Кластер узорак

Кластер узорак је сличан стратификованом узорку, али случајно бирамо и кластере и елементе основног скупа. Кораци за формирање кластер узорка су следећи:

- основни скуп прво делимо на (географске) групе, тзв. „кластере“ (примарне целине). Кластери треба да буду што сличнији, јер је сваки кластер репрезентативан за основни скуп
- случајним путем бирамо кластере за узорак
- из сваког кластера случајним путем бирамо елементе који ће заједно чинити кластер узорак

На пример, можемо поделити целу државу Њујорк на 40 региона (кластера), тј. сличних примарних целина. Потом, на случајан начин бирамо 5 кластера, из којих случајно бирамо домаћинства за спровођење узорачке анкете.

4. Основни појмови статистике

С обзиром да се бавимо статистиком, веома је важно да знамо основне појмове статистике. Основне појмове статистике чине:

- основни скуп (циљна популација)
- јединица посматрања
- променљива (варијабла, обележје)
- податак (опсервација)
- серија података

1. Основни скуп (циљна популација)

Основни скуп је скуп свих елемената чије особине статистички испитујемо. Разликујемо га од физичког скупа, који није значајан за статистичку анализу (физички скуп је нпр. скуп честица у хемијском раствору). Основни скуп још називамо и *циљна популација* или само *популација*.

Основна карактеристика основног скупа *релативна хомогеност*. **Релативна хомогеност значи да су елементи основног скупа истоврсни, али не и истоветни** – сви елементи имају бар једну заједничку особину, али се по нечему и разликују.

Основни скуп је одређен:

- просторно (нпр. студенти прве године Економског факултета *у Београду* 2019. године)
- временски (нпр. студенти прве године Економског факултета у Београду *2019. године*)
- појмовно (нпр. *студенти прве године Економског факултета у* Београду *2019.* године)

2. Јединица посматрања

Јединица посматрања је одређени субјекат или објекат на којем се одређена појава статистички посматра тј. субјекат или објекат о којем се прикупљају подаци. Другим речима, то су појединачни елементи основног скупа, односно узорка. Јединице посматрања могу бити:

- појединци (бића) – као што је случај за пример који смо изнад навели – јединице посматрања су студенти прве године Економског факултета у Београду 2019. године;
- ствари – на пример, од 10.000 производа бирамо 100 производа за тестирање исправности – јединице посматрања су производи;
- догађаји (феномени) – на пример, бирамо инфлацију у претходних пет година за анализу – јединице посматрања су бројчане вредности инфлације.

3. Променљива (варијабла, обележје)

Променљиве су особине по којима се јединице посматрања основног скупа разликују. На пример, студенти Економског факултета разликују се по полу, просечној оцени итд. У овом примеру, пол и просечна оцена су променљиве. Променљиве се још називају и *варијабле*, односно *обележја.*

Променљиве имају одређене *модалитете*, тј. различите видове у којима се променљива јавља. На пример, пол студента може бити мушки или женски. „Мушки“ и „женски“ су модалитети променљиве „пол студента“.

Постоје различите врсте променљивих, од којих је за нас најзначајнија подела на квалитативне и квантитативне променљиве.

1. Квалитатна (атрибутивна, категоријска) променљива	**2. Квантитативна (нумеричка) променљива**
Квалитативна променљива је променљива која *не може* узети нумеричке вредности, али се може класификовати у категорије. Ове променљиве се још називају *атрибутивне* и *категоријске* променљиве. **Модалитети квалитативне променљиве** су описни. На пример, квалитативна променљива је пол студента, а модалитети су „мушки“ и „женски“.	**Квантитативна променљива** је променљива која *може* узети нумеричке вредности. Ове променљиве се још називају *нумеричке* променљиве. **Модалитети квантитативне променљиве** су нумерички. На пример, квантитативна променљива је висина студента, а модалитети су нумерички (нпр. 175 cm, 179 cm, 181 cm). **Квантитативне променљиве можемо даље поделити на:** а) *прекидне (дискретне)* – модалитети могу бити само цели бројеви (нпр. број чланова домаћинства, број положених испита…) б) *непрекидне (континуалне)* – модалитети могу бити децимални бројеви (нпр. висина студента)

4. Податак (опсервација)

Податак (који се још назива и опсервација) је вредност променљиве која се односи на једну јединицу посматрања из основног скупа, односно узорка. На пример, за скуп 1000 студената (јединице посматрања основног скупа) анализирамо висину студената (променљива) и знамо да 25. студент на списку у анкети има висину 180 cm (податак, тј. опсервација за ову јединицу посматрања).

5. Серија података

Серија података је скуп података који се односи на једну или више променљивих. На пример, из основног скупа од 1000 студената бирамо 50 студената за спровођење анкете о њиховој телесној тежини. Добијамо резултате: за првог студента 75 килограма, за другог студента 63 килограма, за трећег студента 86 килограма итд. Ови подаци заједно чине серију података.

5. Врсте промењивих и мерне скале

ВРСТЕ ПРОМЕНЉИВИХ

Променљиве су особине по којима се јединице посматрања основног скупа разликују. На пример, студенти Економског факултета разликују се по полу, просечној оцени итд. Пол и просечна оцена су овде променљиве. Променљиве се још називају и *варијабле*, односно *обележја.*

Променљиве имају одређене *модалитете*, што су различити видови у којима се променљива јавља. На пример, пол студента може бити мушки или женски. „Мушки“ и „женски“ су модалитети ове променљиве.

Постоје различите врсте променљивих, од којих је за нас најзначајнија следећа подела:

1. Квалитатна (атрибутивна, категоријска) променљива	2. Квантитативна (нумеричка) променљива
Квалитативна променљива је променљива која *не може* узети нумеричке вредности, али се може класификовати у категорије. Ове променљиве се још називају *атрибутивне* и *категоријске* променљиве. **Модалитети квалитативне променљиве** су описни. На пример, квалитативна променљива је пол студента, а модалитети су „мушки“ и „женски“.	**Квантитативна променљива** је променљива која *може* узети нумеричке вредности. Ове променљиве се још називају *нумеричке* променљиве. **Модалитети квантитативне променљиве** су нумерички. На пример, квантитативна променљива је висина студента, а модалитети су нумерички (нпр. 175 cm). **Квантитативне променљиве можемо даље поделити на:** а) *прекидне (дискретне)* – модалитети могу бити само цели бројеви (нпр. броја чланова домаћинства, број положених испита…) б) *непрекидне (континуалне)* – модалитети могу бити децимални бројеви (нпр. висина студента)

Подаци који су прикупљени о квалитативној променљивој називају се **квалитативни подаци.**
Подаци који су прикупљени о квантитативној променљивој називају се **квантитативни подаци.**

МЕРНЕ СКАЛЕ

Вредности променљивих меримо на основу одређених мерних скала. Битно је разликовати коју мерну скалу користимо за мерење променљивих, јер од нивоа мерљивости података зависи избор статистичких метода које ћемо применити у статистичким истраживањима. Постоје **четири основне врсте мерних скала**, **од којих се прве две односе на квалитативне податке, а следеће две на квантитативне податке**:

1. Номинална скала

Код номиналне скале **модалитети променљиве се *не* рангирају по значају**. На пример, уколико је променљива пол студента, модалитети су „мушки" и „женски". Примењујемо номиналну скалу, јер не рангирамо ове модалитете по значају. Код номиналне скале симболи само раздвајају модалитете (нпр. можемо означити модалитете као „1 мушки" и „2 женски", али нема разлике у значају модалитета, модалитети се не рангирају по значају). Номинална скала је најнепрецизнија скала.

2. Ординална скала

Код ординалне скале **модалитети променљиве се рангирају по значају**, по утврђеним критеријумима. На пример, уколико је променљива стручна спрема, модалитети су „виша школа, „средња школа", „основна школа". Примењујемо ординалну скалу, јер рангирамо ове модалитете по значају (нпр. виша школа је значајнија од основне школе). Код ординалне скале симболи означавају и ранг модалитета (на пример, можемо означити модалитете као „1 виша школа", „2 средња школа" и „3 основна школа"). Важно је пак напоменути да **не можемо одредити апсолутну разлику између модалитета** код ове скале (не можемо тачно рећи колико је виша школа значајнија од средње, односно колико је средња школа значајнија од основне). Наравно, логика налаже да је ординална скала прецизнија од номиналне скале.

3. Интервална скала

Код интервалне скале **нулта тачка није права**, тј. нула не значи одсуство појаве. На пример, за променљиву „температура ваздуха" користимо интервалну скалу, јер нула не означава одсуство температуре, већ одређену вредност температуре (0 °C). Интервалну скалу карактерише одређена јединица мере, која нам **омогућава да нумерички одредимо апсолутну разлику између модалитета променљиве**. На пример, за температуру ваздуха јединица мере су степени Целзијуса. На основу тога, знамо да је апсолутна разлика између модалитета +15 °C и –4°C тачно 19 °C. Међутим, не можемо одредити ***релативну*** **разлику између модалитета променљиве** – нпр. не можемо рећи за колико је процената +15 °C веће од –4°C.

4. Скала односа

Код скале односа **нулта тачка јесте права**, тј. нула значи одсуство појаве. На пример, за променљиву телесна тежина користимо скалу односа, јер 0 kg заиста представља одсуство телесне тежине. Скалу односа карактерише одређена јединица мере, па можемо одредити апсолутну разлику између модалитета променљиве, али **можемо одредити и релативну разлику између модалитета**. На пример, када поредимо 100 kg и 125 kg, знамо да је 125 kg за 25% веће од 100 kg. Скала односа је најпрецизнија мерна скала.

6. Структурне серије и временске серије

Серија података је скуп података који се односи на једну или више промењљивих. На пример, из основног скупа од 1000 студената бирамо 50 студената за спровођење анкете о њиховој телесној тежини. Добијамо резултате: за првог студента 75 килограма, за другог студента 63 килограма, за трећег студента 86 килограма итд. Ови подаци заједно чине серију података.

Према начину формирања и аналитичком садржају можемо класификовати серије података на:

- **структурне** серије података
- **временске** серије података

Ове серије података често називамо и скраћено – „структурне серије“ и „временске серије“.

1. СТРУКТУРНЕ СЕРИЈЕ

Структурне серије чине подаци прикупљени **о различитим јединицама посматрања у истом временском тренутку, тј. периоду**. На пример, уколико је промењљива (обележје) број гледалаца, једна структурна серија података су ТВ програми по броју гледалаца:

ТВ програм (јединица посматрања)	Број гледалаца у милионима (модалитети промењљиве)
РТС 1	1,564
Пинк	1,946
Студио Б	1,346
History	1,167

Структурне серије такође се називају и *подаци пресека*, тј. *упоредни подаци*.

Подела структурних серија података

Структурне серије података можемо поделити у две основне групе, на основу врсте промењљиве која је коришћена за груписање података:

- **структурне серије података са атрибутивним обележјем (квалитативном промењљивом)** – код ових серија модалитети промењљиве се не могу изразити бројчано. На пример, ТВ програме можемо приказати по оцени квалитета од стране гледалаца:

ТВ програм (јединица посматрања)	Оцена квалитета (модалитети промењљиве)
РТС 1	„висок квалитет“
Пинк	„средњи квалитет“
Студио Б	„средњи квалитет“
History	„веома висок квалитет“

- **структурне серије података са нумеричким обележјем (квантитативном промењљивом)** – код ових серија модалитети промењљиве се могу изразити бројчано. Пример овакве серије су ТВ програми по броју гледалаца, што смо приказали раније изнад.

Особине структурних серија података

Структурна серија података има одређене битне карактеристике:

- уређена је по модалитетима променљиве (који могу бити нумерички или атрибутивни, у зависности да ли је реч о нумеричкој или атрибутивној променљивој)
- представља низ података о свим елементима основног скупа или узорка

2. ВРЕМЕНСКЕ СЕРИЈЕ

Временске серије чине подаци прикупљени **о истим јединицама посматрања у различитим временским тренуцима, тј. периодима** (и притом су уређене хронолошки). На пример, уколико је променљива (обележје) бруто домаћи производ по глави становника, једна временска серија података је БДП по глави становника по годинама:

Година (време)	БДП по глави становника (вредност променљиве)
2016.	10.584
2017.	11.687
2018.	12.446
2019.	13.645

Подела временских серија података

Временске серије података можемо поделити у две основне групе, на основу природе података које садрже:

- **моментне временске серије** – показују **ниво појаве** у тачно одређеним сукцесивним *моментима* времена. Битна особина на основу које знамо да је серија моментна је то што *нема смисла сабирати податке.* Горњи пример БДП-а по глави становника кроз време је пример моментне временске серије. Уколико бисмо сабрали вредности БДП по глави становника, не бисмо добили никакав смислени резултат који можемо интерпретирати.
- **интервалне временске серије** – показују **ток (кретање појаве)** у тачно одређеним сукцесивним *интервалима* времена. Битна особина на основу које знамо да је серија интервална је то што *има смисла сабирати податке.* На пример, уколико је променљива „издаци за исхрану у динарима“, имамо једну интервалну временску серију:

Година (време)	Издаци за исхрану у динарима (вредност променљиве)
2016.	200.000
2017.	180.000
2018.	215.000
Укупно (Σ)	595.000

Уколико бисмо сабрали вредности издатака за исхрану за све године (као што смо учинили у последњем реду табеле), добијамо смислени резултат – можемо закључити да су укупни издаци за исхрану у периоду 2016-2018. године износили 595.000 динара.

7. Извори података, попис и узорачка анкета, случајне и неслучајне грешке

ИЗВОРИ ПОДАТАКА

У сврхе статистичких истраживања потребно је да користимо бројне изворе података. Сходно томе, користимо:

- **секундарне изворе података** (интерне и екстерне)
- **примарне изворе података** (непосредно посматрање, статистички експеримент, анкетно испитивање)

1. Секундарни извори података

Секундарни подаци су подаци који већ негде постоје, а који су прикупљени за неку другу сврху – независно од истраживачке акције. Они имају основну предност у погледу нижих трошкова и брже расположивости. У зависности од тога да ли су прикупљени унутар предузећа који је наручилац истраживања или изван предузећа, секундарни подаци се деле на:

- **интерне податке** – то су подаци које предузеће већ поседује (нпр. подаци о запосленима, књиговодствени подаци и сл.)
- **екстерне податке** – то су подаци који су добијени ван предузећа (нпр. новине, стручни часописи, приручници, каталози, адресари, лексикони, књиге, панели, базе података и сл.)

2. Примарни извори података

Примарни подаци су оригинални подаци, тј. подаци које истраживач тржишта прикупља по први пут за потребе конкретног истраживачког пројекта. У ове сврхе се у статистичким истраживањима користе:

- **метод непосредног посматрања** – нпр. мерење, пребројавање итд.
- **статистички експеримент** – подразумева прикупљање података о основном скупу или узорку, *уз контролу фактора* који могу да утичу на карактеристике елемената или резултате. На пример, када се тестира нова вакцина против корона вируса, да бисмо утврдили ефикасност вакцине формирамо две групе добровољаца за тестирање:
    - *експериментална група* – ово је група елемената која се подвргава третману. Чланови ове групе добијају праву вакцину.
    - *контролна група* – ово је група елемената која не добија третман. Чланови ове групе не добијају праву вакцину већ „плацебо“ (који не узрокује било какве ефекте)

  Битно је да се ове групе формирају тако да добровољци у групама буду међусобно слични, што се постиже случајним распоређивањем у групе (тзв. двоструки експеримент „на слепо“).
- **анкетно испитивање** – подразумева прикупљање података о основном скупу или узорку, *без контроле фактора* који могу да утичу на карактеристике елемената или резултате. Анкета може бити:
    - *попис* – попис је анкета која се спроводи на целом основном скупу. Ретко се користи јер је скуп, дуготрајан, а и често је чак и немогуће идентификовати све елементе основног скупа.

- *узорачка анкета* – узорачка анкета је анкета која се спроводи на делу основног скупа, тј. изабраном узорку. Њена сврха је доношење закључака о карактеристикама основног скупа из којег је узорак изабран. Да би ово било поуздано, потребно је да узорак буде репрезентативан за основни скуп, што значи да адекватно осликава карактеристике основног скупа.

Основни начини спровођења анкете су:

- *интервју* – предност овог начина је што постоји велики одзив испитаника и квалитет одговора, али недостатак је што је веома скуп;
- *телефонско анкетирање* – предност овог начина је што постоји велики одзив испитаника и јефтинији је од интервјуа, али недостатак је што квалитет одговора није на високом нивоу из разлога што људи генерално не воле да их зову телефоном;
- *анкетирањем путем поште* – предност овог метода је што је најјефтинији од наведених, али недостатак је што постоји мали одзив испитаника.

СЛУЧАЈНЕ И НЕСЛУЧАЈНЕ ГРЕШКЕ

Резултати који се добију путем узорачке анкете могу садржати две врсте грешака:

1. Случајне (узорачке) грешке

Случајна (узорачка) грешка дефинише се као разлика између резултата који се добијају узорачком анкетом и резултата које бисмо добили да је узорачком анкетом обухваћен цео основни скуп. Она настаје јер сваки узорак који можемо да изаберемо из основног скупа садржи различите елементе основног скупа који се разликују по особини којом се бавимо у истраживању, тако да сваки узорак даје другачије резултате. Када бисмо изабрали све елементе основног скупа, добили бисмо јединствени резултат и случајна грешка не би постојала. Сходно томе, при спровођењу узорачке анкете постоји случајна грешка, док при попису она не постоји.

На пример, уколико се бавимо просечном тежином 100 испитаника, добићемо другачије $\bar{x}$ за сваки узорак који изаберемо. Када бисмо изабрали свих 100 испитаника у узорак, добили бисмо јединствену вредност μ. Апсолутна разлика између ових вредности представља случајну грешку.

2. Неслучајне (систематске) грешке

Неслучајна (систематска) грешка дефинише се као грешка која се јавља при прикупљању, бележењу и формирању табела података, графикона и сумарних показатеља. Она може да се јави при спровођењу и пописа и узорачке анкете. Јављају се услед људског пропуста, те се могу свести на минимум уз пажљиво планирање, организовање и реализацију истраживања. Неке врсте неслучајних грешака су следеће:

- **погрешно изабран „оквир“ из којег се бира узорак** – „оквир“ представља листу елемената основног скупа. Може се десити да је ова листа непотпуна, због чега долази до неслучајне грешке при избору узорка из те листе елемената основног скупа. На пример, уколико бирамо испитанике из телефонског именика који је непотпун, долази до неслучајне грешке.
- **грешка због одсуства одговора** – при анкети, може се десити да неки испитаници не одговоре на питање, чиме долази до неслучајне грешке
- **грешка у одговору** – при анкети, може се десити да неки испитаници не дају истинит одговор, јер нису добро разумели питање или не желе да дају одговор услед одређених предрасуда
- **грешка анкетара приликом евидентирања** – административна грешка при евиденцији
- **грешка јер узорак није случајан** – самим тим, узорак није ни репрезентативан за осн. скуп

8. Сређивање и графичко приказивање квантитативних и квалитативних података

ГРУПИСАНИ, НЕГРУПИСАНИ ПОДАЦИ И ФРЕКВЕНЦИЈЕ

Када анализирамо неку серију података, пре свега је битно да препознамо да ли је то серија *негруписаних* или *груписаних* података.

1. Серија негруписаних података

Ово су сирови подаци добијени статистичким истраживањем. Једноставно речено, нисмо пребројали колико се пута одређени модалитети понављају или их груписали на било који начин.

Пример серије негруписаних података: У узорку од 10 студената пол студената је следећи:

Мушки, Мушки, Женски, Мушки, Женски, Мушки, Мушки, Мушки, Мушки, Мушки

2. Серија груписаних података

Ово су подаци који су на неки начин груписани. Уколико су у питању дискретни подаци онда можемо да их групишемо бројањем, или ако су у питању континуелни подаци можемо их груписати у одређене интервале.

Пример серије груписаних података: У узорку од 10 студената пол студената је следећи:

Пол (x)	Број студената (f)
Мушки	8
Женски	2
Укупно (Σ)	10

Број студената представља **апсолутну фреквенцију (f)** – колико се истих модалитета променљиве налази у узорку или основном скупу.

Уколико поделимо ове апсолутне фреквенције са укупним бројем студената у нашем узорку, добијамо релативне фреквенције ($\frac{f}{\sum f}$):

Пол (x)	Апсолутна фреквенција (f)	Релативна фреквенција ($\frac{f}{\sum f}$)
Мушки	8	8/10 = 0,8
Женски	2	2/10 = 0,2
Укупно (Σ)	10	1

Релативне фреквенције показују релативну учесталост модалитета у узорку или основном скупу, у односу на цео узорак или основни скуп.

Уколико релативне фреквенције помножимо са 100, добијамо **учешћа**:

Пол (x)	Број студената (f)	Релативна фреквенција ($\frac{f}{\sum f}$)	Учешће ($\frac{f}{\sum f} \cdot 100$)
Мушки	8	8/10 = 0,8	80
Женски	2	2/10 = 0,2	20
Укупно (Σ)	10	1	100

Учешће мушкараца у овом узорку студената је 80%, док је учешће женских особа у овом узорку студената 20%.

Битне напомене:

- Уколико изучавамо узорак, збир апсолутних фреквенција је величина узорка n:

$$\sum f = n$$

- Уколико изучавамо основни скуп, збир апсолутних фреквенција је величина тог скупа N:

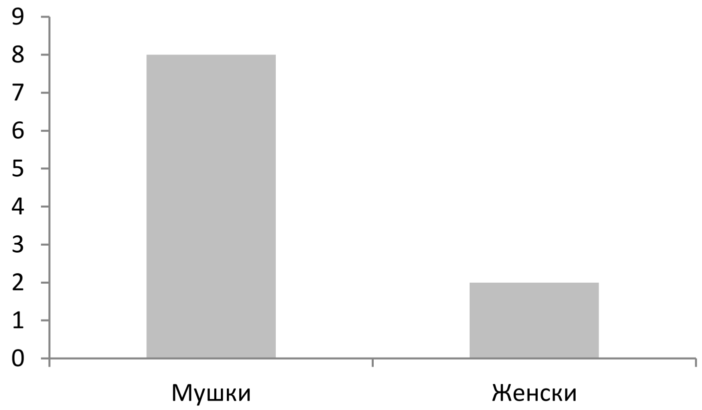
$$\sum f = N$$

ГРАФИЧКО ПРИКАЗИВАЊЕ КВАЛИТАТИВНИХ ПОДАТАКА

Најчешће квалитативне податке приказујемо путем следећих графичких приказа:

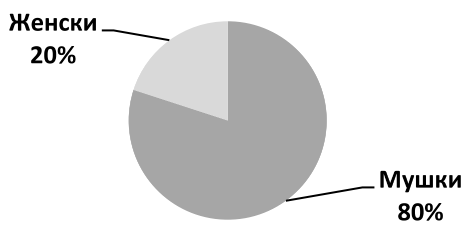
1. Штапићасти дијаграм

На хоризонталној оси означавамо одређене модалитете, док на вертикалној оси означавамо најчешће апсолутне фреквенције (мада је могуће користити и релативне фреквенције или учешћа).



2. Структурни круг (енгл. "pie chart")

Цео круг представља 100%. Одговарајући делови круга представљају учешћа или релативне фреквенције одређених модалитета.

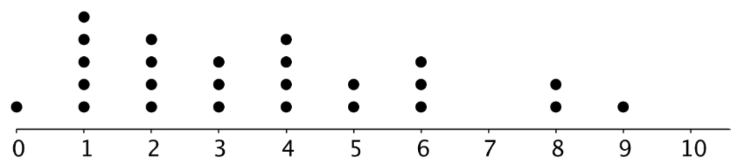


ГРАФИЧКО ПРИКАЗИВАЊЕ КВАНТИТАТИВНИХ ПОДАТАКА

Најчешће квантитативне податке приказујемо путем следећих графичких приказа:

1. Тачкасти дијаграм

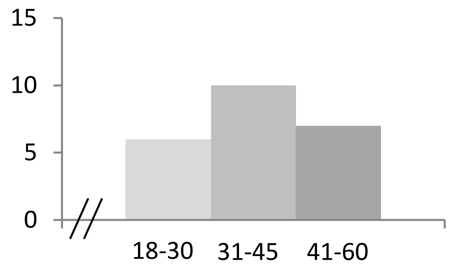
На хоризонталној оси означавамо одређене (нумеричке) модалитете, док на вертикалној оси означавамо апсолутне фреквенције. Из тачкастог дијаграма можемо видети и екстремне вредности.



број сати проведених радећи домаћи задатак недељно

2. Хистограм

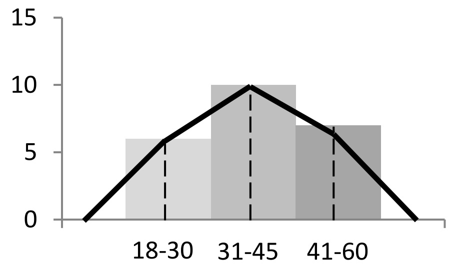
На хоризонталној оси означавамо одређене (нумеричке) модалитете (у оквиру интервала или граничних вредности), док на вертикалној оси означавамо фреквенције. Обавезно цртамо и знак // као *знак за приближавање координатном почетку.*



број година живота

3. Полигон

Конструише се на основу хистограма. Утврђујемо половине интервала и означавамо их на хоризонталној оси. Потом спајамо врхове „стубића“.



број година живота

ИНТЕРВАЛНЕ СЕРИЈЕ ПОДАТАКА

Када анализирамо интервалне серије података, битно је да знамо шта представљају следећи појмови. Узмимо пример следеће интервалне серије података:

Старост (x)	Број запослених (f)
18-30	12
31-43	19

1. Доња и горња граница интервала

За први интервал, доња граница је 18, а горња граница је 30 година.

За други интервал, доња граница је 31, а горња граница 43 године.

2. Средина интервала

За први интервал, средина интервала је $\frac{18+30}{2} = 24$

За други интервал, средина интервала је $\frac{31+43}{2} = 37$

3. Доња и горња гранична вредност интервала

За први интервал, доња гранична вредност интервала је $\frac{18+17}{2} = 17{,}5$

За први интервал, горња гранична вредност интервала је $\frac{30+31}{2} = 30{,}5$

За други интервал, доња гранична вредност интервала је $\frac{31+30}{2} = 30{,}5$

За други интервал, горња гранична вредност интервала је $\frac{43+44}{2} = 43{,}5$

4. Ширина интервала

За први интервал, ширина интервала је $30{,}5 - 17{,}5 = 13$

За други интервал, ширина интервала је $43{,}5 - 30{,}5 = 13$

ФОРМИРАЊЕ ИНТЕРВАЛА

При табеларном приказивању расподеле, потребно је да изаберемо следеће вредности:

1. Број групних интервала

Колико групних интервала ћемо формирати зависи од броја података. Најчешће формирамо између 5 и 20 групних интервала. За утврђивање тачног броја интервала за одређених број података често се користи **Старџесово правило**:

$$C = 1 + 3{,}3\log_{10} N$$

где је C број групних интервала, а N број података.

На пример, уколико имамо 100 испитаника ($N = 100$), Старџесово правило нам говори да треба да формирамо 8 групних интервала:

$$C = 1 + 3{,}3\log_{10} 100 = 1 + 3{,}3 \cdot 2 = 7{,}6 \approx 8$$

2. Ширина групних интервала

Најчешће се узима да су ширине свих групних интервала једнаке. Често се водимо следећим правилом:

$$\text{ширина интервала} = \frac{x_{max} - x_{min}}{C}$$

где је C број групних интервала, x_{min} најмања вредност обележја, а x_{max} највећа вредност обележја. Сходно томе, уколико на пример имамо испитанике старости од 18 до 65 година, а Старџесовим правилом смо утврдили да бирамо 8 групних интервала, ширина интервала коју апроксимативно бирамо је:

$$\text{ширина интервала} = \frac{65 - 18}{8} = 5{,}875 \approx 6$$

3. Доња граница првог групног интервала

Као доња граница првог групног интервала узима се било који адекватан број који је мањи или једнак најмањој вредности обележја, x_{min}. Надовезујући се на претходни пример, доња граница првог групног интервала може бити 18 или мање година, али не и више од 18 година.

КУМУЛАТИВНЕ ФРЕКВЕНЦИЈЕ И КУМУЛАНТА

Поред апсолутних и релативних фреквенција, такође можемо да формирамо и расподелу *кумулативних* фреквенција на следећи начин (узимамо исти пример као раније):

Старост (x)	Број запослених (f)	Кумулативна фреквенција
18-30	12	0+12=12
31-43	19	12+19=31
44-56	14	31+14=45
57-69	5	45+5=50
Укупно (Σ)	50	–

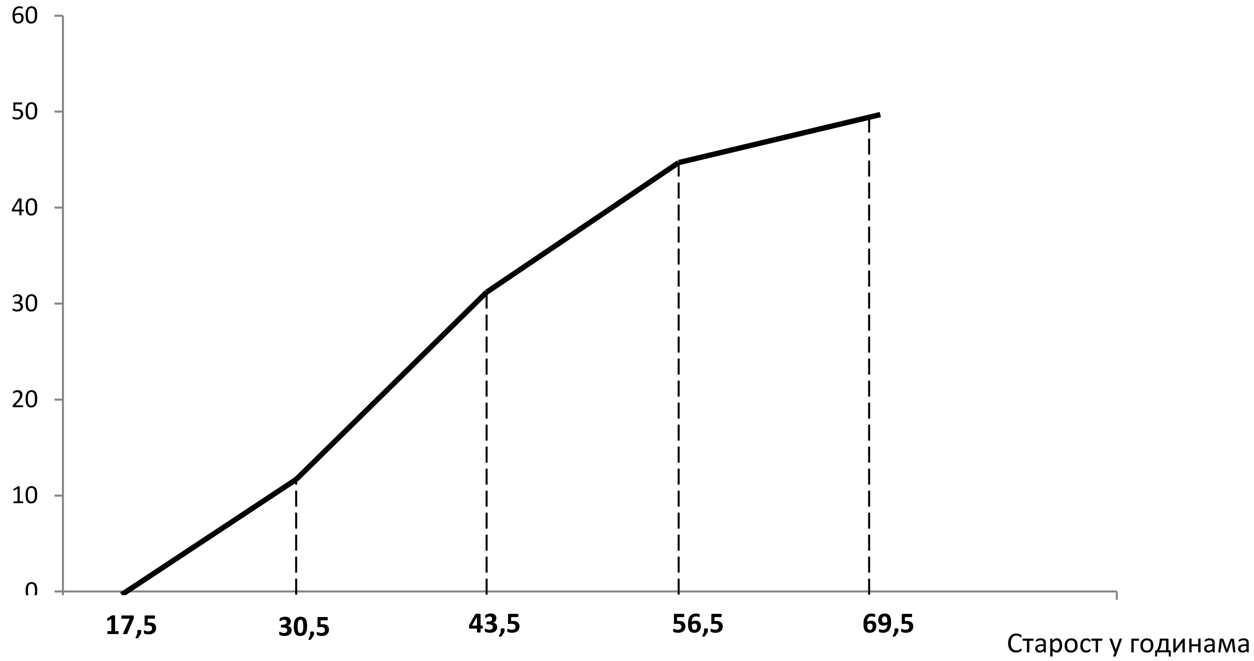
Интерпретација кумулативних фреквенција у овом примеру је следећа:

- 12 особа је старости 30 година и мање.
- 31 особа је старости 43 године и мање.
- 45 особа је старости 56 година и мање.
- 50 особа је старости 69 година и мање.

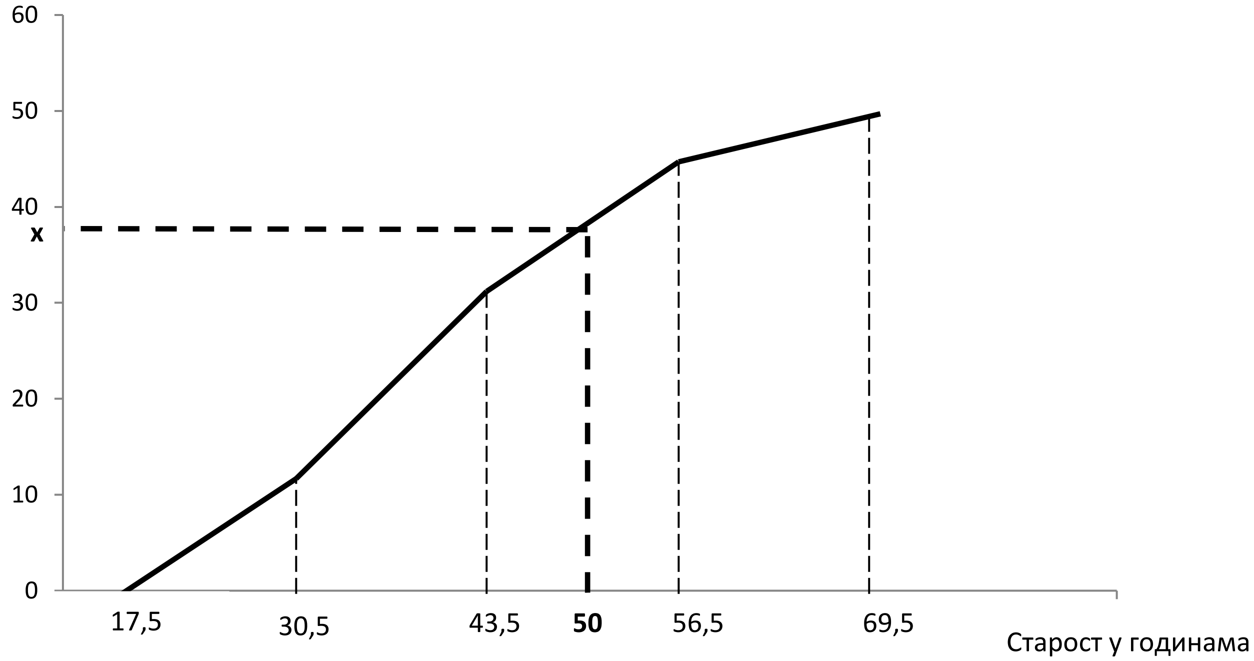
Битно је напоменути да кумулативне фреквенције можемо формирати **само за нумеричке податке**. Такође, кумулативну фреквенцију можемо приказати графички **кумулантом (огивом)** уколико је у питању интервална серија података, слично полигону. Уместо средина интервала користимо **граничне вредности интервала.** У нашем примеру додајмо колону за граничну вредност интервала:

Старост (x)	Број запослених (f)	Кумулативна фреквенција	Гранична вредност интервала
18-30	12	12	од 17,5 до мање од 30,5
31-43	19	31	од 17,5 до мање од 43,5
44-56	14	45	од 17,5 до мање од 56,5
57-69	5	50	од 17,5 до мање од 69,5
Укупно (Σ)	50	-	-

Кумуланта изгледа овако (искључиво је цртамо са граничним вредностима интервала):



Оригинал функције (хоризонтална оса) јесте старост у годинама. Слика функције (вертикална оса) у одређеној старости у годинама даје информацију колико постоји особа не старијих од те вредности. На пример:



Интерпретација: Постоји x особа не старијих од 50 година.

9. Расподела фреквенција (апсолутних, релативних и кумулативних) и њихово графичко приказивање

➲ **За ово питање пишете све у потпуности идентично као за питање 8!**

10. Дескриптивне мере

СУШТИНА ДЕСКРИПТИВНИХ МЕРА

Дескриптивне мере најчешће користимо да опишемо (сумирамо) одређену серију података. На пример, рачунамо аритметичку средину или дисперзију серије података. Дескриптивне мере још називамо и *карактеристичним вредностима*.

ПОДЕЛА ДЕСКРИПТИВНИХ МЕРА

Дескриптивне мере можемо да поделимо у три основне групе, које даље можемо класификовати.
Преглед дескриптивних мера је дат у наставку:

1. МЕРЕ ЦЕНТРАЛНЕ ТЕНДЕНЦИЈЕ (тзв. средње вредности)

a. ИЗРАЧУНАТЕ
  - аритметичка средина
  - геометријска средина
  - хармонијска средина

b. ПОЗИЦИОНЕ
  - модус
  - медијана

2. МЕРЕ ДИСПЕРЗИЈЕ (РАСПРШЕНОСТИ)

a. АПСОЛУТНЕ ИЗРАЧУНАТЕ
  - варијанса
  - стандардна девијација

b. АПСОЛУТНЕ ПОЗИЦИОНЕ
  - интервал варијације
  - интерквартилна разлика
  - квартили
  - перцентили

c. РЕЛАТИВНЕ
  - коефицијент варијације
  - стандардизовано одступање

3. МЕРЕ ОБЛИКА РАСПОДЕЛЕ

  - коефицијент асиметрије
  - коефицијент спљоштености

Битна напомена! У оквиру овог испитног питања напишите и градиво из питања 11 (мере централне тенденције), као и питања 13 (мере дисперзије). С обзиром да ово обухвата заиста пуно градива, препоручујемо да уколико добијете ово питање на испиту прво испишете све за друго испитно питање које сте добили, а потом све што стигнете за ово испитно питање.

11. Мере централне тенденције

ДЕФИНИЦИЈА МЕРА ЦЕНТРАЛНЕ ТЕНДЕНЦИЈЕ

Мере централне тенденције означавају тенденцију груписања или концентрације вредности обележја око средње вредности. Оне нам омогућавају да сумирамо тенденцију груписања вредности обележја око одређене средње вредности код неке серије података. Као једна група дескриптивних мера (поред мера дисперзије и мера облика расподеле), помажу нам да боље разумемо карактеристике дате серије података.

КОЈЕ СУ МЕРЕ ЦЕНТРАЛНЕ ТЕНДЕНЦИЈЕ?

Мере централне тенденције делимо на:

a. ИЗРАЧУНАТЕ
   - аритметичка средина
   - геометријска средина
   - хармонијска средина

b. ПОЗИЦИОНЕ
   - модус
   - медијана

Геометријском средином и хармонијском средином се не бавимо за испит Основи статистичке анализе. Стога, у наставку детаљно обрађујемо аритметичку средину, модус и медијану.

1. АРИТМЕТИЧКА СРЕДИНА

Аритметичка средина је једна од израчунатих мера централне тенденције, која представља „аритметички просек“ обележја одређене серије података. Аритметичка средина је:

- аритметички показатељ – јер се добија израчунавањем на основу података;
- синтетички показатељ – јер на њу утичу све вредности обележја.

Израчунавање аритметичке средине

Аритметичку средину рачунамо на различите начине, у зависности да ли имамо негруписане податке, груписане неинтервалне податке или груписане интервалне податке.

<u>Случај 1: Негруписани подаци</u>

а) Аритметичка средина основног скупа (μ)

Када имамо негруписане податке, аритметичка средина *основног скупа* се рачуна по формули:

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\sum_{i=1}^{N} x_i}{N}$$

где је μ аритметичка средина основног скупа, N број елемената основног скупа, а $x_1, x_2, \dots, x_N$ представља појединачне вредности обележја за елементе основног скупа. На пример, уколико су дати следећи негруписани подаци који се односе на основни скуп:

8 2 4 5 4 3 9 4 5

Аритметичку средину основног скупа добијамо као:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{8+2+4+5+4+3+9+4+5}{9} = \frac{44}{9} = 4{,}889$$

<u>Интерпретација:</u> Просек свих вредности обележја у датој серији података једнак је 4,889.

б) Аритметичка средина узорка ($\bar{x}$)

Када имамо негруписане податке, аритметичка средина *узорка* се рачуна по формули:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

где је $\bar{x}$ аритметичка средина узорка, n број елемената узорка, а x_1, $x_2, \dots, x_n$ представља појединачне вредности обележја за елементе изабраног узорка. На пример, уколико је из датог основног скупа изабран узорак:

5 4 3 9

Аритметичку средину узорка добијамо као:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5+4+3+9}{4} = 5{,}25$$

<u>Интерпретација:</u> Просек вредности обележја у изабраном узорку једнак је 5,25.

Случај 2: Груписани неинтервални подаци

а) Аритметичка средина основног скупа (μ)

Када имамо груписане неинтервалне податке, аритметичка средина *основног скупа* се рачуна по формули:

$$\mu = \frac{\sum xf}{N}$$

где је μ аритметичка средина основног скупа, N број елемената основног скупа, x_1, $x_2, \dots, x_N$ представља појединачне вредности обележја за елементе основног скупа, а f апсолутне фреквенције сваке вредности обележја. Пример се налази у скрипти за колоквијум/писмени испит на стр.37.

б) Аритметичка средина узорка ($\bar{x}$)

Када имамо груписане неинтервалне податке, аритметичка средина *узорка* се рачуна по формули:

$$\bar{x} = \frac{\sum xf}{n}$$

где је $\bar{x}$ аритметичка средина узорка, n број елемената узорка, x_1, $x_2, \dots, x_n$ представља појединачне вредности обележја за елементе изабраног узорка, а f апсолутне фреквенције сваке вредности обележја.

Случај 3: Груписани интервални подаци

а) Аритметичка средина основног скупа (μ)

Када имамо груписане интервалне податке, аритметичка средина *основног скупа* се рачуна по формули:

$$\mu = \frac{\sum x'f}{N}$$

где је μ аритметичка средина основног скупа, N број елемената основног скупа, $x_1,\ x_2, \dots, x_N$ представља појединачне вредности обележја за елементе основног скупа, f апсолутне фреквенције сваке вредности обележја, а x' средине интервала. Пример се налази у скрипти за колоквијум/писмени испит на стр.40.

б) Аритметичка средина узорка ($\bar{x}$)

Када имамо груписане интервалне податке, аритметичка средина *узорка* се рачуна по формули:

$$\bar{x} = \frac{\sum x'f}{n}$$

где је $\bar{x}$ аритметичка средина узорка, n број елемената узорка, $x_1, x_2, \dots, x_n$ представља појединачне вредности обележја за елементе изабраног узорка, f апсолутне фреквенције сваке вредности обележја, а x' средине интервала.

Карактеристике аритметичке средине

Најзначајније карактеристике аритметичке средине као сумарног показатеља су:

- **може бити и децимални број** – не мора да буде цео број, иако су нпр. све вредности обележја у серији података цели бројеви
- **не мора се поклапати ни са једном вредношћу обележја** – нпр. ако имамо негрупиcане податке 3 5 6 7, аритметичка средина је 5,25, што се не поклапа ни са једном вредношћу
- **ако су све вредности обележја једнаке, онда је и аритметичка средина исте вредности** – нпр. ако имамо негруписане податке 4 4 4 4, аритметичка средина је 4
- **налази се обавезно између најмање и највеће вредности обележја** – нпр. ако имамо негруписане податке 3 5 6 7, аритметичка средина је 5,25 – ово обавезно мора бити између 3 и 7, тј. између најмање и највеће вредности обележја, што заиста и јесте истина
- **код основног скупа је константа, а код узорка вредност зависи од изабраног узорка** – ово знамо јер за основни скуп узимамо све вредности обележја, а код узорка зависи које елементе из основног скупа смо изабрали
- **негативна и позитивна одступања вредности обележја од аритметичке средине се међусобно потиру** – ово математички можемо изразити и као:

  $$\sum_{i=1}^{N}(x_i - \mu) = 0$$

  На пример, ако имамо негруписане податке 3 6 9 12, аритметичка средина је 7,5. Одступања вредности обележја од аритметичке средине су:

  $x_1 - \mu = 3 - 7{,}5 = -4{,}5$
  $x_2 - \mu = 6 - 7{,}5 = -1{,}5$
  $x_3 - \mu = 9 - 7{,}5 = 1{,}5$
  $x_4 - \mu = 12 - 7{,}5 = 4{,}5$

  Уколико саберемо ово, добијамо нулу, тј. позитивна и негативна одступања се међусобно потиру.

Примена аритметичке средине

Када је добро применити аритметичку средину као сумарни показатељ, а када не?

- Аритметичка средина је добар показатељ када су вредности обележја које су најближе аритметичкој средини најзаступљеније (имају највећу фреквенцију), а удаљавањем вредности обележја од аритметичке средине број података равномерно се смањује (фреквенције опадају).
- Аритметичка средина *није* добар показатељ уколико постоје екстремне вредности у серији података, јер ће оне знатно смањити интерпретативну моћ аритметичке средине. На пример, уколико имамо серију података 4, 5, 6, 88, аритметичка средина је 25,75, али је очигледно да је погрешно закључити да је ово просечна вредност обележја, јер је интерпретативна моћ смањена због присуства екстремне вредности.

2. МОДУС

Модус је једна од позиционих мера централне тенденције, који је **типична најзаступљенија вредност обележја у серији података (са највећом фреквенцијом)**. Назив потиче од француске речи „mode“, која означава најчешћи, најпопуларнији предмет. На пример, уколико имамо серију података 4 5 6 7 7, модус је 7, јер та вредност обележја има највећу фреквенцију (јавља се 2 пута, док се остале вредности јављају једном).

Какве серије података могу бити?

У погледу броја модуса, разликујемо следеће серије података:

- **унимодалне** – то су серије података које имају један модус (нпр. серија података 4 5 6 7 7 има један модус – модус је 7)
- **бимодалне серије** – то су серије података које имају два модуса (нпр. серија података 4 4 5 6 7 7 има два модуса, 4 и 7)
- **мултимодалне серије** – то су серије података које имају више од два модуса (нпр. серија података 4 4 5 5 6 7 7 има три модуса – 4, 5 и 7)
- **безмодалне серије** – то су серије података које немају модус (нпр. серија података 1 3 5 6 нема модус јер се све вредности јављају исти број пута, тј. имају исту фреквенцију)

Утврђивање модуса

Модус утврђујемо на сличан начин независно од тога да ли имамо негруписане податке, груписане неинтервалне податке или груписане интервалне податке (исти је поступак, као и ознака и за основни скуп и за узорак).

<u>Случај 1: Негруписани подаци</u>

Модус утврђујемо тако што одредимо апсолутне фреквенције сваке вредности обележја и изаберемо вредност обележја где је највећа фреквенција. На пример, за негруписане податке 4 4 5 6 7, видимо да су апсолутне фреквенције за 5, 6 и 7 један, док је апсолутна фреквенција вредности обележја 4 два. Стога, модус је:

$$Mo = 4$$

Пазите! За модус узимамо вредност *обележја*, не вредност *фреквенције*!

Случај 2: Груписани неинтервални подаци

Модус утврђујемо на исти начин – вредност обележја која има највећу апсолутну фреквенцију. Овде је то још лакше, јер су нам подаци већ груписани. На пример:

Број чланова домаћинства (x)	Број радника (f)
1	17
2	34
3	126
4	57
5	37
6	12

Видимо да је највећа фреквенција за трочлана домаћинства, те је модус ове серије података:

$$Mo = 3$$

Случај 3: Груписани интервални подаци

Модус утврђујемо на исти начин – вредност обележја која има највећу апсолутну фреквенцију. На пример:

Време кашњења авиона у мин. (x)	Број авиона (f)
0 до 20	14
20 до 40	18
40 до 60	9
60 до 80	5
80 до 100	4

Видимо да је највећа фреквенција за време кашњења авиона 20 до 40, те је модус ове серије података:

$$Mo = 20 \text{ до } 40 \text{ минута}$$

Карактеристике модуса

Најзначајније карактеристике модуса као сумарног показатеља су:

- **неосетљив је на присуство екстремних вредности у серији података** – у том смислу је погоднији сумарни показатељ од аритметичке средине када постоје екстремне вредности у серији података
- **може се одредити и код квалитативних променљивих, за све мерне скале** (номиналну и ординалну) – што није случај за аритметичку средину која се може одредити само за нумеричке податке.

3. МЕДИЈАНА

Медијана је једна од позиционих мера централне тенденције, која представља **вредност средишњег члана серије, када су вредности рангиране по растућем или опадајућем поретку.**

Утврђивање медијане

Медијану утврђујемо на нешто различит начин, у зависности да ли имамо негрупусане податке, груписане неинтервалне податке или груписане интервалне податке.

Случај 1: Негруписани подаци

Да бисмо утврдили медијану, прво треба да рангирамо податке, а потом нађемо средишњу вредност. На пример, за негруписане податке 41 33 28 21 29 19 14 31 39 36, када их рангирамо добијамо:

14 19 21 28 29 31 33 36 39 41

С обзиром да имамо **паран број елемената** (укупно десет елемената), средишњу вредност ове серије података налазимо као аритметичку средину два средишња елемента.

14 19 21 28 **29** **31** 33 36 39 41

Медијана дате серије података износи:

$$Me = \frac{29 + 31}{2} = 30$$

С друге стране, уколико имамо **непаран број елемената** узимамо само средишњи члан. На пример, за серију података 8 2 4 5 4 3 9 4 5, када је рангирамо добијамо:

2 3 4 4 4 5 5 8 9

С обзиром да имамо непаран број елемената (укупно девет елемената), лако је наћи средишњу вредност ове серије података.

2 3 4 4 **4** 5 5 8 9

Медијана дате серије података је:

$$Me = 4$$

Интерпретација: 50% вредности у серији података је мање или једнако од 4, док је преосталих 50% вредности веће од 4.

Случај 2: Груписани неинтервални подаци

Када имамо груписане неинтервалне податке, поступак је другачији. Морамо прво да формирамо колону за кумулативне фреквенције и израчунамо позицију медијане (што је $\frac{N+1}{2}$ за основни скуп, а $\frac{n+1}{2}$ за узорак), а потом нађемо на коју се вредност обележја ова позиција односи. На пример, имамо основни скуп:

Успех (x)	Број ученика (f)
довољан	30
добар	34
врло добар	50
одличан	10

Одређујемо позицију медијане. С обзиром да имамо 124 елемента скупа (30+34+50+10), знамо да је $N = 124$. Значи, позиција медијане износи:

$$\text{позиција медијане} = \frac{N+1}{2} = \frac{124+1}{2} = 62{,}5$$

Кумулативну фреквенцију додајемо као колону:

Успех	Број ученика	Кумулативна фреквенција
довољан	30	30
добар	34	64
врло добар	50	114
одличан	10	124

Из кумулативне фреквенције видимо да успех „довољан“ покрива првих 30 елемената скупа. Успех „добар“ покрива следећих 34 елемената скупа, што чини укупно 64 елемената. Позиција 62,5 се налази управо унутар успеха „добар“. Стога, закључујемо да је:

$$Me = \text{добар}$$

<u>Случај 3: Груписани интервални подаци</u>

Када имамо груписане интервалне податке, поступак је идентичан као и за груписане неинтервалне податке. Пример можете видети на стр.41 у скрипти за колоквијум/писмени испит.

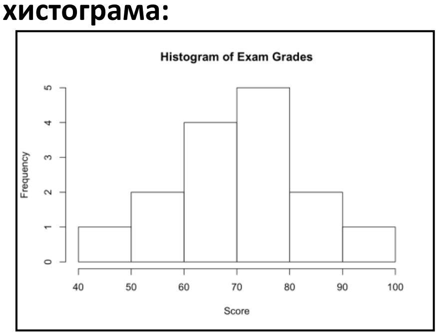
Карактеристике медијане

Најзначајније карактеристике медијане као сумарног показатеља су:

- **неосетљива је на присуство екстремних вредности у серији података** – те је погоднији сумарни показатељ од аритметичке средине када постоје екстремне вредности у серији
- **може се одредити и код квалитативних променљивих, за ординалну скалу** (не може за номиналну) – што није случај за аритметичку средину која се може одредити само за нумеричке податке, премда је модус бољи јер се може одредити и за номиналну скалу
- **медијана дели суму фреквенције на два једнака дела** – те је можемо интерпретирати као: 50% вредности обележја у серији података је мање или једнако од Me, док је преосталих 50% вредности обележја у серији података веће од Me.

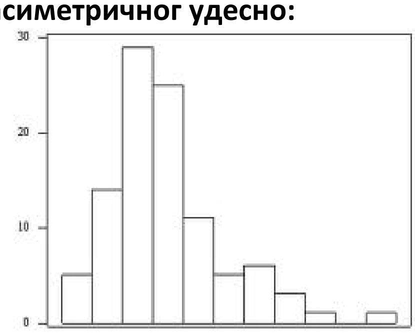
ОДНОСИ МЕРА ЦЕНТРАЛНЕ ТЕНДЕНЦИЈЕ ПРЕКО ХИСТОГРАМА

1) Код симетричног хистограма:



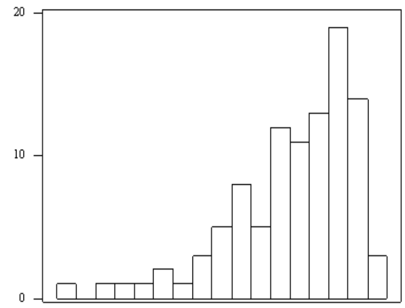
модус = медијана = аритметичка средина

2) Код хистограма асиметричног удесно:



модус < медијана < аритметичка средина

3) Код хистограма асиметричног улево:



модус > медијана > аритметичка средина