

1 Прост случајан узорак

1. Узет је прост случајан узорак од 10 кућа из популације од 100 кућа. Број становника у кућама из узорка је: 2, 5, 1, 4, 4, 3, 2, 5, 2, 3.
 - а) Оценити укупан број становника у популацији и наћи дисперзију те оцене.
 - б) Одредити просечан број становника по кући и оценити дисперзију те оцене.
 - в) Наћи приближни 90% интервал поверења за тотал обележја популације, као и за средину обележја популације.
2. За популацију са елементима $[1, 2, 3, 4, 5, 6, 7, 8]$, размотрен је план узорковања:

s	$P(s)$
1, 3, 5, 6	1/8
2, 3, 7, 8	1/4
1, 4, 6, 8	1/8
2, 4, 6, 8	3/8
4, 5, 7, 8	1/8

- а) Наћи вероватноће укључења π_i за сваки елемент i .
 - б) Одредити узорачку расподелу за $\hat{t} = 8\bar{x}$.
3. Одређује се број дивљих животиња у неком региону који је подељен на 286 области. Изабран је прост случајан узорак од 15 области и дат је број животиња у њима:
$$1, 50, 21, 98, 2, 36, 4, 29, 7, 15, 86, 10, 21, 5, 4.$$
 - а) Наћи оцену средине и дисперзију узорка, оцену дисперзије и стандарде грешке.
 - б) Наћи 90% интервал поверења за укупан број животиња.
 - в) Коју величину узорка треба узети за оцену тотала популације тако да 90% интервал поверења не буде већи од $d = 2000$ животиња?
 4. Ботаничар жели да оцени број стабала брезе у некој области. Област је подељена на 1000 делова. Познато је из претходних испитивања да је дисперзија броја стабала по области приближно $s^2 \approx 45$. Код простог случајног узорка, одредити величину узорка тако да 95% интервал поверења за оцену укупног броја стабала у области која се посматра не буде већи од 500 стабала.
 5. У индустријском граду у јужној Индији већина породица се бави текстилном индустријом. У фајлу `baza1.txt` дати су подаци о броју машина за ткање у свакој од 51 улице у том месту. Одредити укупан број машина за ткање у граду користећи прост случајан узорак величине $n = 8$.

6. Популацију чини база података `trees`, која садржи податке о 31 дрвету. Изабрати прост случајан узорак обима $n = 10$ помоћу функције `sample()` и оценити средњу вредност променљиве `Volume` за популацију користећи добијени узорак. Затим поступак поновити 1000 пута и за тако добијене вредности одредити сумарне статистике и нацртати хистограм. Упоредити оцену средње вредности добијене на основу једног узорка и оцену добијену симулацијом.
7. Из популације која садржи $N = 250$ јединица изабрати узорак обима $n = 12$ помоћу таблица случајних бројева.
8. Посматрајмо популацију обима $N = 5$, чији су елементи означени бројевима 1, 2, 3, 4 и 5 и узимају вредности 3, 1, 0, 1, 5. Размотримо принцип простог случајног узорковања за узорак обима $n = 3$.
 - а) Одредити све могуће узорке обима 3, као и вероватноће да сваки од тих узорака буде изабран.
 - б) За сваки узорак, израчунати узорачку средњу вредност, дисперзију узорка, оцену дисперзије средине узорка и медијану.
 - в) Показати да је средина узорка непристрасна оцена средине обележја популације, да је дисперзија узорка непристрасна оцена дисперзије коначне популације.
9. За које од следећих планова простог случајног узорковања ће бити дата најпрецизнија оцена средње вредности популације? Претпоставимо да свака популација има дисперзију једнаку S^2 .
 - 1) Прост случајан узорак обима 400, добијен из популације обима 4000.
 - 2) Прост случајан узорак обима 30, добијен из популације обима 300.
 - 3) Прост случајан узорак обима 3000, добијен из популације обима 300000000.
10. Претпоставимо да желимо да оценимо пропорцију $\hat{p}$ популације о одређеној болести. Нека је y_i индикатор болести за особу i , тада је $\hat{p} = \bar{y}$. Показати да је оцена коефицијента коваријације

$$CV(\hat{p}) = \sqrt{\frac{N-n}{N-1} \frac{1-p}{np}}.$$

2 Узорак са неједнаким вероватноћама

1. Претпоставимо да имамо узорак са понављањем и да је p_i вероватноћа избора i -те јединице у једном покушају позната. Наћи вероватноћу π_i укључења јединице i у узорак и заједничку вероватноћу укључења π_{ij} јединица i и j .
2. Популација се састоји од 15 јединица чије су величине M_i једнаке редом: 23, 30, 41, 26, 53, 60, 28, 52, 113, 72, 80, 35, 42, 38 и 52. Изабрати узорак с неједнаким вероватноћама обима 3 користећи Лахиријев метод.

3. Изабран је узорак с вероватноћама пропорционалним величини с понављањем обима 3 из популације величине 10. Дате су вредности изабраних елемената и вероватноће избора: $x_1 = 3, p_1 = 0.06, x_2 = 10, p_2 = 0.2, x_3 = 7, p_3 = 0.1$.
- а) Наћи оцену тотала популације користећи *Hansen – Hurvitz*-ову оцену.
- б) Оценити дисперзију те оцене.
- в) Оценити тотал обележја популације користећи *Horvitz – Thompson*-ову оцену.
- г) Оценити дисперзију те оцене.
4. За испитивање загађености језера од 320 језера на површини од 80 km^2 изабран је узорак од 4 језера. Прво језеро из узорка бирано је два пута, а остала два по једном. Концентрација загађености за три језера у узорку су 2, 5 и 10, а величине тих језера (у km^2) су редом 1.2, 0.2 и 0.5. Наћи *Hansen – Hurvitz*-ову оцену средњег загађења по језеру у посматрној популацији, као и оцену дисперзије добијене оцене.
5. Из популације коју чине 3 поља на којима се узгаја пшеница бира узорак обима 2 са вероватноћама пропорционалним величини са понављањем. У следећој табели су дати подаци о количини произведене пшенице на сваком пољу и вероватноће избора сваког поља.

i	1	2	3
p_i	0.3	0.2	0.5
x_i	11	6	25

Налажењем свих могућих узорака одредити вероватноће укључења сваког елемента $\pi_i, i = 1, 2, 3$, *Hansen – Hurvitz*-ову и *Horvitz – Thompson*-ову оцену за укупну производњу пшенице.

6. Популацију чини база података **trees**, која садржи податке о 31 дрвету. Изабрати узорак ($n = 10$) са понављањем са вероватноћама избора пропорционалним обиму дрвећа (променљива **Girth**) и оценити средњу вредност обележја **Volume** за популацију користећи добијени узорак.
7. У фајлу **baza2.txt** дати су подаци о броју радника и производњи у 10 фабрика у индустриској зони. Из ове популације изабрати узорак обима 3 без понављања са вероватноћама избора пропорционалним броју радника у фабрици. Користећи добијени узорак одредити *Hansen – Hurvitz*-ову оцену укупне производње.

3 Стратификован узорак

1. У популацији од 6 елемената вредности 0, 1, 2 су у првом стратуму, а 4, 6, 11 у другом стратуму. Израчунати средње вредности стратума и популације, као и узоричке дисперзије стратума и дисперзију популације.

2. За популацију из претходног задатка са стратумима 0, 1, 2 и 4, 6, 11 у узорак су изабрани елементи 0 и 2 из првог стратума, а 6 и 11 из другог. Одредити оцену средине обележја популације, њену дисперзију и оцену те дисперзије. Упоредити добијену дисперзију са дисперзијом оцене средине добијене из простог случајног узорка. Оценити затим и тотал обележја популације, његову дисперзију и оцену те дисперзије.
3. Популација од 8 елемената подељена је на два стратума са елементима 3, 7, 4, 2 и 13, 10, 8, 5. У узорак су изабране јединице 3 и 7 из првог стратума, а 10 и 8 из другог стратума. Израчунати оцену средине обележја популације и оцену њене дисперзије. Оценити затим укупну суму обележја популације, израчунати дисперзију оцене и оцену те дисперзије.
4. Популација је подељена на три стратума чије су величине $N_1 = 150$, $N_2 = 90$ и $N_3 = 120$ и одговарајућа стандардна одступања су $s_1 = 100$, $s_2 = 200$ и $s_3 = 300$. Ако се бира стратификован узорак обима 12, одредити величину узорка из сваког стратума користећи:
 - а) пропорционални избор узорка;
 - б) Нојманов избор узорка.
5. Нека је популација од 200 елемената подељена у три стратума са $N_1 = 60$, $N_2 = 80$, $N_3 = 60$ и $s_1 = 2s_2 = 4s_3$. Ако се бира стратификован узорак обима 30, одредити величину узорка из сваког стратума користећи:
 - а) пропорционални избор узорка;
 - б) Нојманов избор узорка.
6. За испитивање просечне недељне потрошње бензина град је подељен на 4 дела, који се посматрају као стратуми. Забележена је потрошња бензина за протеклу недељу код сваког возача из узорка и добијени су следећи подаци:
 Стратум 1: $N_1 = 3750$, $n_1 = 50$, $\bar{x}_1 = 12.6$, $s_1^2 = 2.8$
 Стратум 2: $N_2 = 3272$, $n_2 = 45$, $\bar{x}_2 = 14.5$, $s_2^2 = 2.9$
 Стратум 3: $N_3 = 1387$, $n_3 = 30$, $\bar{x}_3 = 18.6$, $s_3^2 = 4.8$
 Стратум 4: $N_4 = 2475$, $n_4 = 30$, $\bar{x}_4 = 13.8$, $s_4^2 = 3.2$
 а) Оценити средњу недељну потрошњу за цео град.
 б) Наћи 95% интервал поверења добијене оцене.
 в) Ако треба изабрати стратификовани случајни узорак обима 1000, одредити величину узорка у сваком стратуму по пропорционалном распореду.
7. У фајлу `baza3.txt` дати су подаци о броју посета позоришту за ученике неке школе, као и стратум ком сваки ученик припада. Ученици су подељени у 3 стратума према успеху у школи. Изабрати стратификован узорак без понављања обима 30, користећи пропорционални распоред, а затим оценити просечан број посета позоришту за ученике те школе, дисперзију оцене и оцену те дисперзије.

4 Количничко и регресионо оцењивање

1. Дати су подаци за прост случајан узорак без понављања обима 10 из популације обима 53 у коме је x величина која се испитује, а y помоћна променљива:

x_i	8.75	2.55	9.00	1.10	7.50	5.00	1.15	3.40	2.00	1.25
y_i	8.35	1.50	10.00	0.60	7.50	7.95	0.95	4.40	1.00	0.50

Познато је да је популацијска средина обележја y једнака $\bar{y} = 3.33$.

- Наћи количничку оцену средине за обележје популације x .
 - Оценити дисперзију добијене количничке оцене.
 - Наћи 95% интервал поверења за средину обележја популације.
2. Нека је узет прост случајан узорак без понављања обима 64 из велике популације и нека су дати подаци: $\bar{y}_n = 5.00$, $\bar{y} = 5.50$, $\bar{x}_n = 60.5$. Наћи оцену количника $\hat{R}$ и количничку оцену средине обележја популације.
3. У граду од 72500 становника изабран је прост случајан узорак без понављања од четири домаћинства из популације од 25000 домаћинстава за испитивање просечних трошкова за исхрану за недељу дана. Прво домаћинство из узорка има 4 члана и троши укупно 1500 динара за храну недељно, друго домаћинство има 2 члана и троши 1000 динара, треће са 4 члана троши 2000 динара и четврто са 3 члана троши 1400 динара. Наћи количничку оцену за средње трошкове по домаћинству и оценити дисперзију те оцене.
4. У популацији обима 15 познате су вредности помоћне променљиве y_i : 2.3, 3.0, 4.1, 2.6, 5.3, 6.0, 2.8, 5.2, 11.3, 7.2, 8.0, 3.5, 4.2, 3.8, 5.2. Из популације је изабран прост случајан узорак и за те јединице су дате вредности обележја x_i :

x_i	40	40	70	82	52
y_i	4.1	5.2	7.2	8.0	5.2

Наћи количничку оцену средине обележја x .

5. Дати су подаци за два стратума исте величине из популације величине 12.

h	x_{hi}	y_{hi}
1	2, 2, 3, 4, 6, 7	6, 6, 7, 7, 8, 8
2	10, 11, 12, 14, 15, 16	10, 11, 16, 17, 18, 18

Из сваког стратума бирају се у узорак по 3 јединице. Вредности јединица укључених у узорак дате су у табели:

h	x_{hi}	y_{hi}
1	2, 3, 7	6, 7, 8
2	10, 14, 15	10, 17, 18

Наћи комбиновану и посебну количничку оцену средине обележја популације

6. За испитивање укупног приноса пшенице у области са 100 парцела узабран је прост случајан узорак од 4 парцеле и са сваке је измерен принос пшенице x_i . Принос пшенице зависи од количине ђубрива y_i , која је примењена и која је позната за сваку парцелу у популацији, тако да је $\bar{y} = 100$. Вредности парова (x_i, y_i) које су изабране у узорак су $(1410, 50)$, $(1690, 100)$, $(1680, 150)$, $(1850, 200)$. Регресионом методом оценити укупан принос пшенице и наћи оцену дисперзије добијене оцене. Упоредити добијене резултате са оценом тотала из простог случајног узорка.
7. Из популације од 1000 елемената изабран је прост случајан узорак обима 20. Дате су следеће сумарне статистике $\bar{x}_n = 10$, $\bar{y}_n = 6$, $\sum_{i=1}^{20} (x_i - \bar{x}_n)(y_i - \bar{y}_n) = -60$, $\sum_{i=1}^{20} (y_i - \bar{y}_n)^2 = 30$, $\sum_{i=1}^{20} (x_i - \bar{x}_n - b(y_i - \bar{y}_n))^2 = 80$ и $\bar{y} = 5$. Наћи регресиону оцену средину обележја x популације и оценити дисперзију ове оцене.
8. У фајлу `baza4.txt` дати су подаци о броју становника (у хиљадама) у два различита периода претходном y и каснијем x . Оценити просечан број становника по граду и укупан број становника у 75 градова чији је укупан број у претходном периоду био $\bar{y} = 5475$.
9. Вредности за x и y у популацији од 6 јединица су:

x_i	3, 5, 7, 6, 8, 13
y_i	1, 2, 2, 3, 3, 3

Рачунањем свих могућих простих случајних узорака обима 2 и свих узорака обима 3 показати да је $\hat{R}$ непристрасна оцена у оба случаја.

5 Кластер (групни) узорак

1. Популација од 12 елемената подељена је на три групе и подаци су дати у следећој табели:

	x_{ij}
1	2, 4, 6
2	3, 3, 4, 5, 5
3	1, 1, 2, 4

Изабран је узорак од прве и треће групе. Наћи оцену укупне суме обележја популације, дисперзију те оцене и количничку оцену тотала обележја популације.

2. Нека је популација, која садржи 100 елемената подељена на 10 група. Изабран је прост случајан узорак од 3 групе чије су величине и тотали обележја популације: $M_1 = 5$, $t_1 = 4$, $M_2 = 20$, $t_2 = 12$, $M_3 = 10$, $t_3 = 7$.
- а) Наћи непристрасну оцену тотала и средине обележја популације.
- б) Оценити дисперзију добијене оцене тотала.

- в) Наћи количничку оцену тотала обележја популације и одговарајућу оцену средине популације.
 - г) Оценити дисперзију добијене оцене тотала.
3. Нека је у претходном задатку изабран узорак од 3 скупине с вероватноћом која је пропорционална вечилини скупина.
- а) Наћи непристрасну оцену укупне суме обележја популације и оцену популацијске средине.
 - б) Оценити дисперзију добијене оцене укупне суме.

6 Системски (периодичан) узорак

1. Популација се састоји од 12 елемената чија је вредност обележја популације једнака њиховом редном броју. Формирати систематске узорке обима 4 и израчунати дисперзију средине систематског узорка.
2. Популацију чини 900 студената. Како изабрати узорак обима 120 користити систематски узорак?
3. Популација се састоји од 6 домаћинстава са одговарајућим величинама 2, 4, 3, 9, 1 и 2, где величина домаћинства представља број чланова домаћинства. Изабрана су 3 домаћинства без понављања, са вероватноћом пропорционалном његовој величини.
 - а) Наћи вероватноће укључења ових 6 домаћинстава.
 - б) Изабрати систематски узорак обима 2.
 - в) Користећи узорак добијен у делу б), одредити оцену просечног броја чланова домаћинства.