

ТЕОРИЈА УЗОРАКА - ЗАДАЦИ СА ВЕЖБИ

1. Показати да је $MSE(\hat{\theta}) = D(\hat{\theta}) + b^2(\hat{\theta})$.
2. Нека су X_1, X_2, \dots, X_n независне случајне величине из исте расподеле са коначним очекивањем μ и коначном дисперзијом $\sigma^2, \sigma > 0$.
 - а) Одредити пристрасност оцене (за σ^2) $S_n^2 = \sum_{k=1}^n (X_k - \bar{X}_n)^2$.
 - б) Одредити непристрасну оцену за σ^2 .
3. Колико има простих случајних узорака без понављања, а колико са понављањем, ако се вади узорак обима 7 из популације која садржи 20 јединици?
4. Марко има 7 јабука, Петар 2, Јован 2 и Саша 6. Да ли је оцена укупног броја јабука непристрасна, ако се узорак обима 2 бира тако да је вероватноћа да су Марко и Петар у узорку $\frac{1}{3}$, вероватноћа да су Марко и Јован у узорку $\frac{1}{2}$, вероватноћа да су Петар и Јован $\frac{1}{6}$, док је вероватноћа избора свих осталих узорака једнака 0. (Као оцена укупног броја јабука користи се статистика $\hat{t} = N\bar{X}_n$).
5. Испитати да ли је боља (у средње квадратном смислу) оцена укупне суме обележја ($\hat{t} = N\bar{X}_n$) на основу узорака обима 2, или на основу узорака обима 3, из популације $\{1, 2, 3, 4\}$ ако се користи прост случајан узорак без понављања.
6. Из скупа $\{1, 2, 3, 4, 5, 6, 7, 8\}$ ваде се узорци без понављања обима 2, тако да сваки који садржи непаран број има вероватноћу 0. Они узорци који садрже 2 имају вероватноћу обрнуто пропорционалну другом елементу узорака, а они који садрже 4, а не садрже 2, имају вероватноћу $\frac{1}{8}$. Испитати непристрасност оцене средње вредности и одредити њену средње квадратну грешку.
7. Из популације $\{1, 2, \dots, 100\}$ извадити 15 простих случајних узорака без понављања обима 20. На сваком од њих наћи оцену укупне суме обележја и испитати који од узорака је најрепрезентативнији, тј. где је реализована вредност статистике $\hat{t} = N\bar{X}_n$ најближа стварној вредности.
8. Нека је $Z = \sum_{i=1}^N a_i I_i$, где су a_1, a_2, \dots, a_N константе, а I_i индикатори укључења i -те јединке у узорак, $i \in \{1, 2, \dots, N\}$. Одредити очекивање и дисперзију случајне величине Z (у зависности од вероватноћа укључења првог и другог реда).
9. Дата је популација $\{1, 2, 3, 4, 5, 6, 7, 8\}$ и размотрен је следећи план узорковања:

S	$\{1, 3, 5, 6\}$	$\{2, 3, 7, 8\}$	$\{1, 4, 6, 8\}$	$\{2, 4, 6, 8\}$	$\{4, 5, 7, 8\}$
$P(S)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

- а) Наћи вероватноће укључења π_i за сваки елемент i .
- б) Испитати непристрасност оцене $\hat{t} = N\bar{X}_n$.

1 ПРОСТ СЛУЧАЈАН УЗОРАК

1. У ресторану се служе четири врсте колача - баклаве, тулумбе, еклери и шампите. Десеторо људи је купило баклаву, двадесеторо тулумбу, двадесет и петоро еклер, а петнаесторо шампиту. На случајан начин се одаберу 2 различите врсте колача и бележи се број људи који су купили ту врсту.
 - а) Оценити укупан број особа које су купиле колач, а затим испитати непристрасност оцене која се користи.
 - б) Испитати да ли су узорачка дисперзија и узорачка стандардна грешка непристрасне оцене дисперзије и стандардне грешке на читавој популацији.
2. Узет је прост случајан узорак без понављања од 10 ученика од 100 ученика трећег разреда и бележене су њихове оцене из математике. Забележени су резултати: (4, 5, 5, 2, 3, 1, 3, 4, 4, 5). Оценити просечну оцену из математике, а затим израчунати оцену дисперзије те оцене.
3. Показати да је код простог случајног узорака без понављања s_n^2 непристрасна оцена за s^2 .

4. Одредити расподелу случајне величине μ_i која представља број појављивања i -те јединке популације, ($i \in \{1, 2, \dots, N\}$), у простом случајном узорку са понављањем обима n .
5. У кутији се налази 100 папирића на којима се налазе каро, пик, херц и треф. Извлачи се 10 папирића, тако што се папирић извуче, забележи се знак, а затим се врати у кутију. Добијен је узорак (К,П,П,Х,К,Т,Т,П,Х,Х).
- Ако каро има вредност 1 динар, пик 2, херц 3, а треф 4, оценити укупну суму новца која се налази у кутији.
 - Одредити оцену броја пикова.
 - Одредити непристрасне оцене дисперзија ових оцена.
6. Показати да је код простог случајног узорка са понављањем $\frac{N}{N-1}s_n^2$ непристрасна оцена за s^2 .
7. Одредити обим простог случајног узорка без понављања потребан за оцењивање пропорције неког својства обележја популације.
8. Истраживач жели да оцени пропорцију деце са плавим очима у вртићу од 4000 деце. Прихватљиво одступање му је 5% са ризиком 8%. Одредити обим простог случајног узорка без понављања који је потребан истраживачу. Ако је истраживач већ радио истраживање у другом вртићу и добио пропорцију $\frac{1}{3}$, који му је онда обим узорка потребан?
9. Узет је прост случајан узорак од 10 различитих кућа од 100 кућа које се налазе у једном насељу. Број становника у кућама из узорка је 2,5,1,4,4,3,2,5,2,3.
- Оценити укупан број становника у том насељу и оценити дисперзију те оцене.
 - Оценити просечан број становника по кући и оценити дисперзију те оцене.
 - Наћи приближни 90%-ни интервал поверења за укупан број становника, као и за просечан број становника по кући.
10. Оцењује се број дивљих животиња у неком региону који је подељен на 286 области. Изабран је прост случајан узорак без понављања од 15 области и дат је број животиња у њима:
- $$1, 50, 21, 98, 2, 36, 4, 29, 7, 15, 86, 10, 21, 5, 4$$
- Наћи узорачку дисперзију, оцену просечног броја животиња по области, оцену дисперзије те оцене, као и оцену стандардне грешке.
 - Наћи 90%-ни интервал поверења за укупан број животиња.
 - Коју величину узорка треба узети за оцењивање укупног броја животиња тако да одступање не буде веће од 2000 животиња са вероватноћом 0.9.
11. 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 45. Одредити величину простог случајног узорка без понављања, потребну да са вероватноћом 0.95 одступање не буде веће од 500 стабала.
12. Посматрамо популацију обима 5, чији су елементи означени бројевима 1, 2, 3, 4 и 5, и вредности обележја су редом 3, 1, 0, 1 и 5. Размотримо принцип простог случајног узорковања без понављања за узорак обима 3.
- Одредити све могуће узорке и вероватноће да сваки од тих узорка буде изабран.
 - За сваки узорак израчунати узорачку средњу вредност, узорачку дисперзију, оцену дисперзије узорачке средине и медијану.
 - Показати да је средња вредност обележја на узорку непристрасна оцена средње вредности обележја популације и да је узорачка дисперзија непристрасна оцена дисперзије популације.
13. У датотеци *deca.txt* дати су подаци о броју деце у свакој од 512 улица у неком граду. Наћи 95%-ни интервал поверења за укупан број деце у том граду користећи прост случајан узорак без понављања обима 200.
14. За који од следећих планова простог случајног узорковања без понављања ће бити добијена најпрецизнија оцена средње вредности обележја на популацији? Претпоставимо да обележје на свакој од популација има дисперзију 100.
- Узорак обима 400, добијен из популације обима 4000.
 - Узорак обима 30, добијен из популације обима 300.
 - Узорак обима 3000, добијен из популације обима 300000000.

2 УЗОРКОВАЊЕ СА НЕЈЕДНАКИМ ВЕРОВАТНОЋАМА

- Одредити *Hansen–Hurwitz*-ову оцену укупне суме обележја популације ако се користи прост случајан узорак са понављањем, а затим и дисперзију и *Hansen–Hurwitz*-ову оцену дисперзије те оцене.
- Из популације се вади узорак са понављањем обима n и претпоставља се да је p_i , вероватноћа избора i -те јединке у узорак у једном покушају, позната за свако $i \in \{1, 2, \dots, N\}$. Одредити π_i - вероватноћу укључења i -те јединке у узорак и π_{ij} - заједничку вероватноћу укључења јединки i и j ($i, j \in \{1, 2, \dots, N\}$).
- Популација се састоји од 15 јединки чије су величине 23, 30, 41, 26, 53, 60, 28, 52, 113, 72, 80, 35, 42, 38 и 52. Изабрати узорак са понављањем и узорак без понављања обима 5 са вероватноћама пропорционалним величини користећи Лахиријев метод.
- Изабран је узорак са вероватноћама пропорционалним величини, са понављањем, обима 3 из популације величине 10. Дате су вредности обележја изабраних елемената и вероватноће избора:

i	1	2	3
x_i	3	10	7
p_i	0.06	0.2	0.1

- Наћи оцену укупне суме обележја популације користећи *Hansen–Hurwitz*-ову оцену.
 - Оценити дисперзију те оцене.
 - Наћи оцену укупне суме обележја популације користећи *Horvitz–Thompson*-ову оцену.
 - Оценити дисперзију те оцене.
- Ако се бира узорак обима n без понављања, показати да важи:
 - $\sum_{i=1}^N \pi_i = n$.
 - $\sum_{j=1, j \neq i}^N \pi_{ij} = (n-1)\pi_i$, за свако фиксирано i , $i = \overline{1, n}$.
 - За испитивање загађености 320 језера укупне површине $80km^2$ изабран је узорак са понављањем обима 4. Прво језеро из узорка бирано је два пута, а остала два по једном. Концентрација загађености у та три језера у узорку је редом 2, 5 и 10, а површине тих језера (у km^2) су редом 1.2, 0.2 и 0.5. Наћи *Hansen–Hurwitz*-ову оцену средњег загађења по језеру у посматраној популацији, као и оцену дисперзије добијене оцене.
 - У датотеци *populacija.txt* налазе се величине 120 јединки из популације. Извадити узорак са понављањем и узорак без понављања обима 90 користећи метод кумуланти.
 - Из популације коју чине три поља на којима се узгаја пшеница бира се узорак обима 2 са вероватноћама пропорционалним величинама, са понављањем. У следећој табели су дати подаци о количини произведене пшенице на сваком пољу и вероватноће избора сваког поља.

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

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

- У датотеци *radnici.txt* дати су подаци о броју радника и производњи у 10 фабрика у индустријској зони. Изабрати узорак обима 3 са понављањем са вероватноћама избора пропорционалним броју радника у фабрици. Користећи добијени узорак одредити *Hansen–Hurwitz*-ову оцену укупне производње.
- Дата је популација од четири прашуме, њихове површине и бројеви тигрова који живе у њима:

Прашума	1	2	3	4
Површина (у km^2)	100	200	300	500
Број тигрова	11	20	23	54

Вади се узорак од 2 прашуме. Оценити укупан број тигрова, одредити дисперзију те оцене и наћи непристрасну оцену те дисперзије ако је узорковање вршено:

- а) Са понављањем (Користити *Hansen – Hurwitz*-ову оцену),
- б) Без понављања (Користити *Horvitz – Thompson*-ову оцену),

и извучен је узорак (1,2).

11. Популацију чини база *trees* која садржи податке о 31 дрвету.

- а) Изабрати прост случајан узорак без понављања обима 10 и оценити средњу вредност обележја *Volume* користећи добијени узорак. Затим поновити поступак 1000 пута и за тако добијене вредности нацртати хистограм. Упоредити оцену средње вредности добијену на основу једног узорка и ону добијену симулацијом.
- б) Изабрати узорак са понављањем обима 10 са вероватноћама пропорционалним обиму стабла (променљива *Girth*) и оценити средњу вредност обележја *Volume* користећи *Hansen – Hurwitz*-ову оцену.