

1 Прост случајан узорак

1. Из основог скупа $y = (1, 2, 3, 4)'$ написати све уређене и неуређене узорке обима $n = 2$ без понављања и са понављањем.
2. Из основог скупа од $N = 250$ домаћинстава изабрати случајан узорак обима $n = 12$.
3. За оцену

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{m}_Y)^2$$

дисперзије у простом случајном узорку

$$S^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - m_Y)^2.$$

известити следеће облике који су погодни код израчунавања:

а)

$$\bar{S}^2 = \frac{1}{n-1} \left(\sum_{i=1}^n y_i^2 - n\hat{m}_Y^2 \right),$$

б)

$$\bar{S}^2 = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \right],$$

в)

$$\bar{S}^2 = \frac{1}{n(n-1)} \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right].$$

4. Потписи једне петиције сређени су са 676 листова. Сваки лист има довољно места за 42 потписа, али многи листови имају мањи број потписа. Потписи су избројани на узорку од 50 листова (узорак око 7%) и резултати су дати у табели:

y_i	42	41	36	32	29	27	23	19	16	15	14	11	10
n_i	23	4	1	1	1	2	1	1	2	2	1	1	1

9	7	6	5	4	3
1	1	3	2	1	1

где је y_i број потписа на i -том листу, а n_i број листова са y_i потписа. Оценити укупан број потписа $\hat{\tau}_Y$ и наћи дисперзију средине узорка \bar{S}^2 .

5. Треба оценити количину новца коју студенти потроше на одлазак у биоскоп за месец дана. Узет је прост случајан узорак од $n = 160$ студената из популације на универзитету од $N = 32400$ студената и нађене су вредности $\hat{m}_Y = 105.30$ динара и $\bar{S}^2 = 453.60$. Наћи 95% интервал поверења за оцену средине популације.

6. За оцену како ученици основне школе користе своје време узет је прост случајан узорак од 12 ученика од укупно 30 ученика неке школе. Ученици су одговорили колико времена проведу у учењу. Добијене су вредности

$$\hat{m}_Y = 9.1 \text{ сати и } \bar{S}^2 = 22.3.$$

Наћи 95% оцену грешке границе за оцену просечног броја часова проведених недељно у учењу.

7. 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.

- Одредити просечан број становника по кући и оценити дисперзију те оцене.
- Оценити укупан број становника у популацији. Наћи оцену дисперзије добијене оцене.
- Наћи приближни 90% интервал поверења за средину и за тотал обележја популације.

8. Ботаничар жели да оцени број стабала брезе у некој области. Област је подељена на 1000 делова. Познато је из претходних испитивања да је дисперзија у броју стабала по области приближно $\bar{S}^2 \approx 45$. Код простог случајног узорка, одредити величину узорка тако да 95% интервал поверења за оцену укупног броја стабала у области која се проучава не буде већи од: а) 500 стабала; б) 1000 стабала; в) 2000 стабала.

9. Одређује се број дивљих животиња у неком региону. Регион 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.

- Наћи средину и дисперзију узорка, оцену дисперзије и стандардне грешке средине узорка, оцену укупног броја дивљих животиња, дисперзију и стандардну грешку те оцене.
- Наћи 90% интервал поверења за укупан број животиња.
- Коју величину узорка треба узети за оцену тотала популације тако да 90% интервал поверења не буде већи од $\Delta = 2000$ животиња?

10. У индустријском граду у јужној Индији већина породица се бави текстилном индустријом. У фајлу `bazal.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 $n = 8$.

11. Популацију чини база `trees`, која садржи податке о 31 дрвету. Изабрати прост случајан узорак обима $n = 10$ 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.

12. За процену броја бирача који би гласали за одређени предлог на референдуму, узет је прост случајан узорак од 1200 бирача, од којих се 552 бирача изјаснило да би подржали предлог. Број бирача који имају право гласа је 1 800 000. Наћи оцену пропорције броја бирача који би гласали за предлог и одредити 95% интервал поверења за пропорцију.
13. У простом случајном узорку обима $n = 100$, из популације од $N = 500$ елемената, 37 елемената је из класе C . Наћи 90% интервал поверења за пропорцију и укупан број елемената који припадају класи C .

2 Узорак са неједнаким вероватноћама

14. Нека је дат узорак са понављањем и нека је ψ_i вероватноћа избора i -те јединице у једном покушају. Наћи вероватноћу π_i укључења јединице i у узорак и заједничку вероватноћу укључења π_{ij} јединица i и j .
15. Изабран је узорак са вероватноћама пропорционалним величини са понављањем обима $n = 3$ из популације величине $N = 10$. Дате су вредности изабраних елемената и вероватноће са којим су елементи изабрани: $y_1 = 3, \psi_1 = 0.06, y_2 = 10, \psi_2 = 0.20, y_3 = 7, \psi_3 = 0.10$.
- а) Оценити тотал и средину обележја популације користећи Horvitz-Thompson-ову оцену.
- б) Оценити дисперзије тих оцена.
16. За узорак из задатка 15:
- а) Наћи оцену тотала обележја популације користећи Hansen-Hurwitz-ову оцену.
- а) Оценити дисперзију те оцене.
17. 80 км² изучаване области је под језерима. За испитивање загађености језера од 320 језера 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 км²) су, редом, 1.2, 0.2 и 0.5.
- а) Наћи непристрасну оцену средњег загађења по језеру у посматраној популацији.
- б) Оценити дисперзију добијене оцене.
18. 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 за популацију користећи добијени узорак.
19. У фајлу baza2.txt дати су подаци о броју радника и производњи у 10 фабрика у индустријској зони. Из ове популације изабрати узорак обима 3 са понављањем са вероватноћама избора пропорционалним броју радника у фабрици. Користећи добијени узорак одредити Hansen - Hurvitz-ову оцену укупне производње.

i	1	7	10	11	14
y_i	25	30	70	82	40
x_i	2.3	2.8	7.2	8.0	3.8

Наћи линеарну регресиону оцену средине обележја популације y .

24. За испитивање укупног приноса пшенице у области са $N = 100$ парцела изабран је прост случајан узорак од $n = 4$ парцеле и са сваке је измерен принос пшенице y_i . Принос пшенице зависи од количине ђубрива x_i која је применјена и која је позната за сваку парцелу у популацији, тако да је $m_X = 100$. Вредности парова (x_i, y_i) су $(50, 1410)$, $(100, 1690)$, $(150, 1680)$, $(200, 1850)$. Регресионом методом оценити укупан принос пшенице и наћи оцену дисперзије добијене оцене. Упоредити добјене резултате са оценом тотала из простог случајног узорка.
25. Дати су подаци о броју становника (у хиљадама) у два различита периода претходном x и каснијем y :

Редни број	x	y	Редни број	x	y	Редни број	x	y
1	18	25	10	17	28	19	25	32
2	30	48	11	19	20	20	170	185
3	42	35	12	93	105	21	60	72
4	125	180	13	106	125	22	43	50
5	73	80	14	43	62	23	72	83
6	69	70	15	190	205	24	65	72
7	160	185	16	18	28	25	43	59
8	76	115	17	82	98			
9	12	16	18	60	58			

Оценити просечан број становника по граду и укупан број становника у 75 градова чији је укупан број у претходном периоду износио $\tau_X = 5475$.

4 Стратификован узорак

26. У земљи има 120 продавница хране. За оцену средње месечне продаје, прикупљени су подаци из 20 продавница. Познато је да неке фирме имају ланце продавница што је приказано у табели

Фирма	Број продавница
А	50
Б	30
В	20
Г	10
Д	10
Укупно	120

Како је продаја у продавницама исте фирме слична, а из различитих фирми различита, формирају се стратуми састављени од продавница исте фирме. Из сваког стратума бира се прост случајан узорак величине редом: $n_1 = 8, n_2 = 5, n_3 = 3, n_4 = 2, n_5 = 2$. У следећој табели дата је вредност месечне продаје (у 1000 динара) за сваку изабрану продавницу:

Стратум	Вредност продате робе							
1	250	330	210	280	380	190	220	305
2	420	380	470	520	440			
3	200	190	160					
4	100	78						
5	70	58						

Наћи оцену просечне вредности продате робе, оцену дисперзије добијене оцене и оцену укупне вредности продаје.

27. Ако је у претходном задатку узет прост случајан узорак од 20 продавница, упоредити прецизност оцена средина обележја популације за прост случајан и стратификован случајан узорак.
28. Резултати за стратификован узорак дати су у табели

Стратум	N_h	n_h	\hat{m}_h	\bar{S}_h^2
1	20	5	1.6	3.3
2	9	3	2.8	4.0
3	12	4	0.6	2.2

Оценити средину и тотал обележја целе популације и наћи оцене дисперзија тих оцена.

29. Добијени су следећи резултати за стратификован узорак:

Стратум 1: $N_1 = 100, n_1 = 50, \hat{m}_1 = 10, \bar{S}_1^2 = 2800$;

Стратум 2: $N_2 = 50, n_2 = 50, \hat{m}_2 = 20, \bar{S}_2^2 = 700$;

Стратум 3: $N_3 = 300, n_3 = 50, \hat{m}_3 = 30, \bar{S}_3^2 = 600$.

- а) Оценити средину обележја целе популације и наћи оцену дисперзије те оцене.
- б) Наћи 95% интервал поверења за средину.

30. Популација је подељена на три стратума чије су величине $N_1 = 150, N_2 = 90$, и $N_3 = 120$, а стандардна одступања $\sigma_1 = 100, \sigma_2 = 200$ и $\sigma_3 = 300$. Поделити узорак величине $n = 12$ на три одговарајућа стратума:

- а) користећи пропорционалан распоред;
- б) користећи оптималан распоред (за једнаке трошкове).

31. Нека је популација од 200 елемената подељена у три стратума са $N_1 = 60, N_2 = 90, N_3 = 50$ и $\sigma_1 = 2\sigma_2 = 4\sigma_3$. Бира се стратификован узорак обима $n = 30$, одредити величину узорка узетог из сваког стратума:

- а) по пропорционалном распореду;
- б) по оптималном распореду (за једнаке трошкове).

32. Популација од $N = 8$ елемената подељена је на два стратума са елементима 3, 7, 4, 2 и 13, 10, 8, 5 и у узорак су изабране јединице 3 и 7 из првог стратума, а 10 и 8 из другог стратума. Израчунати оцену средине обележја популације \hat{m}_Y^{str} , њену дисперзију $D(\hat{m}_Y^{str})$ и оцену дисперзије $\hat{D}(\hat{m}_Y^{str})$. Оценити затим тотал обележја популације, израчунати његову дисперзију и оцену дисперзије.

33. У фајлу rozoriste.txt дати су подаци о броју посета позоришту ученика неке школе, као и подела ученика по стратумима. Ученици су подељени у три стратума према успеху у школи. Изабрати стратификован узорак обима 30 користећи пропорционални распоред, а затим оценити просечан број посета позоришту ученика те школе, дисперзију оцене и оцену те дисперзије.

5 Групни узорак

34. Популација од $M = 12$ јединица подељена је на $N = 3$ групе и подаци су дати у следећој табели

Група	y_{ij}
1	1,2,4,5
2	3,3,4,6
3	1,2,2,3

Изабран је узорак од прве и треће групе. Наћи оцену тотала обележја популације, дисперзију те оцене и одговарајућу оцену обележја популацијске средине по елементу.

35. Популација од 500 домаћинстава подељена је на 20 група свака са по 25 домаћинстава, затим је изабран прост случајан узорак од 3 групе и у њима су избројани чланови:

Група l	5	13	17
Средина групе m_l	5.52	5.68	4.28

Наћи оцену просечне величине домаћинства и оцену дисперзије те оцене.

36. Нека је популација подељена на $N = 10$ група (примарних јединица) и нека има укупно $M = 100$ секундарних јединица у популацији. Изабран је прост случајан узорак од $n = 3$ групе чије су величине M_l и тотали обележја популације τ_l : $\tau_1 = 4, M_1 = 5, \tau_2 = 12, M_2 = 20, \tau_3 = 7, M_3 = 10$.

- Наћи непристрасну оцену тотала и средине по елементу обележја популације.
- Оценити дисперзију добијене оцене тотала.
- Наћи количничку оцену тотала обележја популације и одговарајућу оцену средине по елементу.
- Оценити дисперзију добијене оцене тотала.

37. Нека је у претходном задатку изабран узорак од $n = 3$ групе са вероватноћом која је пропорционална величини група.

- Наћи непристрасну оцену тотала обележја популације и одговарајућу оцену популацијске средине по елементу.
- Оценити дисперзију добијене оцене тотала.

6 Систематски узорак

38. Популација се састоји од $N = 12$ елемената чија је вредност обележја популације једнака њиховом редном броју. Формирати систематске узорке обима $n = 4$ и израчунати дисперзију средине систематског узорка.
39. За податке из табеле (где је $N = 40, n = 4, k = 10$)

Бројеви систематског узорка										
	1	2	3	4	5	6	7	8	9	10
	0	1	1	2	5	4	7	7	8	6
	6	8	9	10	13	12	15	16	16	17
	18	19	20	20	24	13	25	28	29	27
	26	30	31	31	33	32	35	37	38	38
$\sum_{i=1}^4 y_{ij}$	50	58	61	63	75	71	82	88	91	88

израчунати $D(\hat{m}_Y^{sys})$.

7 Вишеетапни узорак

40. Бира се прост случајан узорак од $n = 2$ примарне јединице из популације од $N = 3$ примарне јединице од којих прва има $M_1 = 3$, друга $M_2 = 4$ и трећа $M_3 = 5$ секундарних јединица. Из сваке изабране примарне јединице бирају се по две секундарне јединице. Нека су изабране прва и друга примарна јединица и из прве примарне јединице изабране су вредности 2 и 5, а из друге 5 и 7.

- а) Оценити тотал обележја популације.
 б) Израчунати оцену дисперзије добијене оцене.

41. Ученици једног разреда неке школе подељени су у 10 одељења односно група. Изабран је прост случајан узорак од одељења 3, 5 и 8. Треба оценити колико су пута у току школске године ученици били у биоскопу. Из трећег одељења изабрани ученици су одговорили на следећи начин: 2, 4, 5, 7, 10, 1, 6, из петог одељења 5, 6, 8, 10, 4, 3 и из осмог 3, 2, 6, 4, 5. Број ученика по одељењима је $M_1 = 25, M_2 = 35, M_3 = 33, M_4 = 30, M_5 = 31, M_6 = 28, M_7 = 37, M_8 = 29, M_9 = 28, M_{10} = 27$.