

Дескриптивне статистике

- Пре него што се нови производ стави у промет, обично се у одабраним радњама тестира реакција потрошача, тј. имају могућност да пробају производ и изразе мишљење. Да ли на овакав начин добијени подаци представљају популацију или узорак? Ако се већина потрошача који су пробали производ позитивно изјасни о куповини истог, да ли то гарантује да би већина свих потрошача купила тај производ уколико би био доступан на тржишту?
- Пре комплетирања пореске пријаве неке фирме за дату годину, потребно је утврдити вредност залиха у фирми тако што ће се пописати сви производи и њихове цене. Да ли ови подаци представљају популацију или узорак? Да ли је неопходна употреба статистичких метода за утврђивање вредности залиха?
- Испитавано је време реаговања возача у секундама након попијене две лименке пива. Дати подаци представљају разлику између времена реаговања након и пре попијеног пива:

0.1 0.4 0.6 2.7 1.1 0.6 1.6 0.3 2.4 0.2 0.1 2.5 0.8 2.1 2.0
0.7 3.2 1.5 1.3 3.5 0.5 0.9 0.5 1.5 4.0 1.1 1.7 2.3 1.0 4.6

Конструисати дијаграм стабло-лист. Да ли ови подаци имају расподелу у облику звона? Да ли се може тврдити да је време реаговања већине возача након попијеног пива 2 секунде спорије од нормалног?

- Анализирана је латенција, реакција скакаваца на звучне и визуелне стимулације. Подаци представљају време у милисекундама између пријема звучне или визуелне стимулације и померања главе скакавца што резултује променом правца кретања:

звучни	86	102	103	99	108	100	115	106	109	113	
	114	107	107	117	120	101	126	109	106		
визуелни	72	99	102	75	100	103	77	95	71	97	80
	104	101	78	73	90	71	70	81	103	89	

- Конструисати дијаграм стабло-лист за оба скупа података.
 - Да ли се може тврдити да расподела латенције прати облик звона у оба случаја?
 - Нацртати дупли дијаграм стабло-лист.
 - У којем случају је латенција шире распрострањена?
 - Која група има већу просечну вредност?
- Случајна величина X представља радни век у сатима литијумске батерије ручног дигитрона. Подаци су добијени на узорку од 50 батерија:

4285 602 2300 582 701 1137 1379 564 1379 478 1429 497 520
1278 2584 726 852 3367 261 205 14 510 1000 1402 2778 1850
318 2662 786 373 2066 3570 737 308 1100 414 604 99 3032
35 396 209 1009 2000 1560 99 83 454 349 981

- Конструисати хистограм релативних фреквенција
 - На основу хистограма, закључити да ли случајна величина X има искривљену расподелу и, уколико има, у коју страну?
 - Да ли се може тврдити да батерије трају најмање 1793 сата?
- Да би одлучио да ли да ради након 17 часова, власник радње је узео узорак купаца који су услужени након 17 часова на 36 случајно изабраних дана:

1 3 6 2 4 6 2 1 4 6 6 1 4 9 2 3 1 5
6 2 6 5 7 11 5 4 10 12 12 15 7 5 8 7 14 4

- Конструисати табелу фреквенција

- б) Конструисати хистограм фреквенција
 в) Да ли је расподела искривљена и, ако јесте, у ком смеру?
 г) На основу хистограма закључити, да ли је необично да има више од 4 купца увече?
 д) Да ли су сати након 17 часова популарни?

7. Акутна изложеност кадмијуму изазива респираторне поремећаје, оштећења бубрега и јетре, и може довести до смрти. Због тога је посматран ниво ваздушне кадмијумске прашине и кадмијум оксида у ваздуху. Ниво се мери милиграмима кадмијума по кубном метру ваздуха. Добијено је 35 података:

0.044	0.020	0.040	0.057	0.055	0.061	0.047	0.030	0.066	0.045	0.050	0.037
0.061	0.051	0.052	0.052	0.039	0.056	0.062	0.058	0.054	0.044	0.049	0.039
0.061	0.062	0.053	0.042	0.046	0.030	0.039	0.042	0.070	0.060	0.051	

- а) Конструисати дијаграм стабло-лист.
 б) Да ли се може рећи да расподела нивоа кадмијума у ваздуху има облик звона?
 в) Конструисати табелу фреквенција.
 г) Конструисати хистограм релативних фреквенција. Да ли он има облик звона?
8. Градска општина је заинтересована за отварање вечерње школе. Да би утврдили да ли је то потребно, узет је узорак од 20 жена и 15 мушкараца и записане су њихове године школовања:

жене	8	10	16	12	7	8	18	18	12	12					
	12	12	14	16	16	14	7	8	12	20					
мушкарци	8	12	12	14	14	16	16	16	16	12	7	9	20	20	12

- а) Одредити узорачке средине ових узорака.
 б) Одредити просечну вредност узорачких средина.
 в) Одредити тежинску средину.
 г) При којим условима би резултати под б) и в) били једнаки?
 д) Да ли можемо тврдити да је узорачка средина за жене једнака просечном броју година образовања у целој општини?
 ђ) Да ли је тачно тврђење да је просечан број година образовања у општини већи од 18?
 е) Одредити медијану узорка.
 ж) Одредити моду узорка.
9. Тврди се да је са концентрацијом од 0.1% алкохола у крви већина моторичких и сензорних способности поремећено. У проучавању конзумирања алкохола, испитаницима су дате по три флаше пива, а затим је одређена концентрација алкохола у крви. Добијени су следећи резултати:

0.10	0.08	0.11	0.09	0.09	0.08	0.12	0.09	0.07	0.08	0.10	0.12	0.09	0.06	0.13
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

- а) Одредити узорачку средину и медијану узорка.
 б) Одредити распон узорка.
 г) Одредити дисперзију и стандардну девијацију.

10. Висока концентрација озона у ваздуху може изазвати слепило и штетна је како за флору, тако и за фауну. Следећи подаци представљају ниво озона у ваздуху у пошумљеном делу земље:

160	165	170	172	161	176	163	196	162	160	162	185	167
180	168	163	161	167	173	162	169	164	179	163	178	

- а) Конструисати дијаграм стабло-лист. Да ли је расподела података асиметрична и, ако јесте, у коју страну?

б) Конструисати боксплот и идентификовати аутлајере.

11. Проучавана су два лека, амантадин и римантадин, која се користе у борби против вируса грипа. Доза од $100mg$ је дата здравим особама и измерено је време, у минутима, које је потребно да се достигне максимална концентрација плазме. Добијени су подаци:

A	105	126	120	119	133	145	200	123	108	112
	132	136	156	12.4	134	130	130	142	170	
R	221	261	250	230	253	256	227	264	236	246
	273	271	280	238	240	283	516			

а) Конструисати боксплот за сваку групу података.

б) Одредити $\bar{x}$ и s^2 за податке А.

в) Ако претпоставимо да је податак 12.4 погрешно записан и избришемо децималну тачку, како се мењају претходни резултати?

г) Да ли постоје аутлајери у скупу R? Ако их има, да ли је разумно избрисати их?

Задаци за вежбу

1. За експеримент се користи 40 пацова који су случајним избором подељени у две групе од по 20. Сваки пацов пролази кроз лавиринт и бележи се време проласка. Затим се једна група пацова учи како да прође кроз лавиринт. Након тога, свих 40 пацова пролазе кроз лавиринт и време проласка се поново бележи. Подаци представљају разлику у времену проласка између првог и другог мерења.

а) За коју групу се очекује да има већу просечну разлику?

б) На основу датих резултата у секундама, конструисати дупли дијаграм стабло-лист за сваку групу:

тренирани	4.0	3.2	4.1	4.9	4.2	3.7	4.3	4.2	4.4	3.6
	3.5	4.9	5.1	4.5	4.7	5.0	5.6	4.6	5.2	5.5
нетренирани	-2.1	-2.2	-1.1	-2.5	-1.2	2.0	-2.4	-0.6	1.3	-1.3
	-0.2	-2.7	1.4	0.9	2.2	2.1	1.8	2.1	1.1	2.6

в) Шта значи када се добије негативна вредност?

г) Да ли би било изнеђујуће рећи да подаци имају расподеле у облику звона?

2. Трагови цинка и бакра су присутни у храни под нормалним околностима. Међутим, ови хемијски елементи могу бити токсични и могу изазвати проблеме у организму ако се налазе у великој количини. Сprovedена је студија о нивоу сваког елемента у храни за бебе. Подаци представља просечну вредност у милиграмима по литру једнаких узорака изабраних из најпродаванијих 16 брендова:

цинк	3.0	5.8	5.6	4.8	5.1	3.6	5.5	4.7
	5.7	5.0	5.9	5.7	4.4	5.4	4.2	5.3
бакар	0.40	0.51	0.47	0.55	0.56	0.41	0.60	0.45
	0.60	0.51	0.48	0.63	0.50	0.45	0.62	0.57

а) Конструисати дупли дијаграм стабло-лист за сваки скуп података.

б) Да ли неки од скупова података има расподелу у облику звона?

в) Произведен је нови производ чији је просечан ниво бакра 0.53. Да ли је то значајно висок ниво у односу на већ постојеће стање на тржишту?

г) Да ли би било необично да је процењени ниво цинка испод 4.0 за нов производ?

3. Лекар испитује 60 пацијената у вези времена чекања (у минутима) од заказаног термина. Добијени су следећи подаци:

60	29	34	25	31	30	-1	17	6	50	10	18	38	25	35	36	31	23	12	52
8	27	27	30	42	9	47	31	27	6	45	23	25	37	3	50	53	28	16	19
32	36	9	33	36	33	58	26	18	32	12	32	8	16	48	36	46	-5	31	59

- а) Конструисати табелу фреквенције расподеле.
 б) Конструисати хистограм фреквенције.
 в) Шта значи када се јави негативна вредност?
 г) На основу облика хистограма, закључити да ли променљива има расподелу у облику звона.
 д) На основу хистограма, закључити да ли је необично чекати дуже од 5 минута на преглед.
4. Радна група проучава понашање возача на одређеној деоници пута. Случајним избором је посматрано 50 возила и забележене су њихове брзине у километрима на час:

49.1	60.0	72.3	77.2	64.1	59.7	52.0	68.2	71.0	90.0	65.1	45.0	59.5
84.1	75.0	60.0	55.0	55.0	67.2	68.0	62.1	62.0	58.0	53.0	85.0	74.9
62.3	54.7	61.0	76.0	65.1	64.2	66.0	57.5	58.0	75.0	69.0	63.5	57.8
77.2	65.0	65.8	65.0	61.1	50.0	63.0	76.0	68.3	59.9	78.3		

- а) Конструисати табелу фреквенције расподеле.
 б) Конструисати хистограм релативне фреквенције.
 в) Да ли се на основу хистограма може закључити да је брзина кретања већине возила приближно 65 километара на час?
5. Дате су вредности променљивих X и Y :

X	2	1	7	3	5	9
Y	2	0	3	8	1	3

- а) Одредити $\bar{x}$ и $\bar{y}$.
 б) Одредити узорачку медијану за сваки скуп података.
 в) Да ли неки од скупова има јединствену моду? Уколико има, која је вредност?
6. У два различита друштва социолог проучава у којим година се жене први пут удају. Добио је резултате:

A	14	20	25	21	26	22	14	24	13	19	15	15
	24	14	30	14	16	18	35	13	14	17	14	16
B	14	21	25	27	20	15	25	18	24	22	26	30
	18	19	27	26	31	26	32	35	20			

- а) Одредити узорачке средине. Да ли је просечан број година једнак у оба узорка?
 б) Конструисати дијаграм стабло-лист. Да ли су неки од података искривљени и, ако јесу, у коју страну?
 в) Одредити медијане узорака.
7. Анестетичар је проучавао промене, у процентима, CO_2 у артеријској крви две групе пацова након убризгавања идентичне дозе два лека. Добио је резултате:

I	27.2	30.1	30.5	28.4	30.7	31.3	30.5	30.1	29.6	30.2			
	31.7	32.0	28.6	29.2	33.0	31.7	32.6	28.2	29.1	30.7			
II	55.1	56.3	60.0	63.5	64.9	62.7	60.5	59.2	53.7	64.1	65.8	58.3	57.1
	55.4	56.5	55.1	57.0	59.3	60.7	62.1	63.6	64.0	65.3	62.8	59.5	

- а) Конструисати дијаграм стабло-лист. Који подаци су распршенији?
 б) Одредити узорачку средину и медијану узорака.
 в) Одредити узорачку дисперзију, узорачку стандардну девијацију и узорачки распон.
 г) Да ли је тачна тврдња да нема разлике у начину реаговања на лекове?

8. Компанија спроводи истраживање о својим клијентима. За два дана су прикупљени подаци о потрошеној количини новца:

<i>I</i> жене старије од 25 година	10.98	29.80	12.03	53.00	27.00	52.00	26.50
<i>II</i> жене млађе од 25 година	5.98	35.00	2.03	36.20	8.02	17.25	
<i>III</i> мушкарци старији од 25 година	90.00	18.21	75.00	37.50	5.98	110.00	16.25
<i>IV</i> мушкарци млађи од 25 година	10.02	18.35	17.98	6.95	22.50	12.03	7.32

- Одредити узорачку средину за сваки од скуп података.
 - Одредити тежинску средину куповине које су обављене од стране женских клијената. Да ли је тежинска средина једнака аритметичкој средини узорачких средина за скупове података *I* и *II*?
 - Одредити тежинску средину куповине које су обављене од стране мушких клијената. Да ли је тежинска средина једнака аритметичкој средини узорачких средина за скупове података *III* и *IV*? Ако јесте, због чега је тако?
 - Пронаћи свеукупну средину за све куповине комбиновањем сва четири скупа података у један и одредити узорачку средину за нови скуп података.
 - Пронаћи тежинску средину за сва четири скупа података и упоредити добијене резултате са резултатима добијеним под г).
 - Компанија сматра да просечна продаја мора прећи 25 евра да би пословали профитабилно. Да ли је то могуће на основу добијених резултата? Да ли резултати гарантују успех компаније? Објаснити.
 - Пронаћи свеукупну медијану за све куповине комбиновањем четири скупа података.
9. Подаци представљају дневну продају у хиљадама евра две франшизе истог ланца хамбургера:

франшиза <i>A</i>	0.9	1.7	2.9	3.5	2.5	3.2	3.8	0.7
	1.4	3.6	2.7	4.8	2.6	1.3	5.9	4.7
франшиза <i>B</i>	0.9	0.9	1.4	2.5	0.7	4.6	3.1	5.1
	4.5	0.3	5.2	1.7	0.6	5.3	4.2	5.0

- Одредити узорачку средину за сваку франшизу. Да ли су, на основу ових резултата, просечне дневне продаје приближно једнаке?
 - Одредити узорачку медијану за сваку франшизу. Да ли су, на основу ових резултата, медијане дневне продаје приближно једнаке?
 - На основу дијаграмо стабло-лист, да ли се може речи да дневне продаје имају исту расподелу. Који подаци имају већу распршеност?
 - Доказати тврђење под в) одређивањем дисперзије.
 - Да ли би било необично да у неком од следећих дана дневна продаја буде преко 4000 евра, а следећег дана 1000 евра за франшизу *A*? А за франшизу *B*?
10. Сprovedено је истраживање међу становницима једне области старијих од 65 година о томе колико различитих лекова користе током последњих 6 месеци. Добијен је узорак величине 10: 0, 3, 8, 1, 4, 6, 9, 1, 0, 8.
- Одредити узорачку средину и узорачку медијану.
 - Одредити узорачки распон.
 - Одредити узорачку дисперзију и стандардну девијацију.
11. Проучаван је утицај пушења на спавање. Подаци представљају време у минутима које је било потребно да испитаници заспе:

пушачи	69.3	56.0	22.1	47.6	53.2	48.1	52.7	34.4	60.2	43.8	23.2	13.8
непушачи	28.6	25.1	26.4	34.9	29.8	28.4	38.5	30.2	30.6	31.8	41.6	21.1
	36.0	37.9	13.9									

- Одредити узорачке медијане оба узорка.
- Конструисати боксплот за сваку групу.