

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 дупли 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 латентност, реакција скакаваца на звучне и визуелне стимулације. Подаци представљају време у милсекундама 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 дупли стабло-лист дијаграм за оба скупа података.
 - Да ли се може тврдити да расподела латентности прати облик звона у оба случаја?
 - У којем случају је латентност шире распрострањена?
- Подаци представљају вредности од 14.1 до 17.6 са размаком 0.1 и подељени су у 6 класа.
 - Одредити распон података.
 - Одредити најмању ширину класа потребну да се прекрије опсег.
 - Одредити стварну ширину класа која се користи да би се категорисали подаци.
 - Одредити границе класа.
 - Одредити средишње тачке класа.
 - 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?

7. 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 |

- а) 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?
8. 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?
9. 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 | | | | | |
| мушкараци | 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.
10. 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 алхола 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.
 в) Одредити $\sum_{i=1}^{15} x_i$ и $\sum_{i=1}^{15} x_i^2$.

г) Одредити дисперзију и стандардну девијацију.

11. 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 аутлејере.

12. 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, која 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 A.
в) Ако претпоставило да је податак 12.4 погрешно записан и избришемо децималну тачку, како се мењају претходни резултати?
г) Да ли постоје аутлејери 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. Спроведена 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. Подаци представљају вредности од 35.49 до 59.37 са размаком 0.01 који су подељени у 10 класа.
- а) Одредити распон података.
 б) Одредити најмању ширину класе потребну за покривање овог распона.
 в) Одредити стварну ширину класе која се користи у категоризацији података.
 г) Одредити границе класа.
 д) Одредити средишње тачке класе.

4. 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.
5. 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?
6. Дате су вредности променљивих X и Y :

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

а) Одредити $\sum_{i=1}^6 x$ и $\sum_{i=1}^6 y$.

б) Одредити $\sum_{i=1}^6 x^2$ и $\sum_{i=1}^6 y^2$.

в) Одредити \bar{x} и \bar{y} .

г) Одредити узорачку медијану за сваки скуп података.

д) Да ли неки од скупова има јединствену моду? Уколико има, која је вредност?

7. 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.
8. 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?
9. Компанија спроводи истраживање о својим клијентима. За два дана су прикупљени подаци о потрошеној количини новца:

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

- а) Одредити узорачку средину за сваки од скуп података.
 б) 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?
 в) 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?
 г) Пронаћи свеукупну средину за све куповине комбиновањем сва четири скупа података у један и одредити узорачку средину за нови скуп података.
 д) Пронаћи тежинску средину за сва четири скупа података и упоредити добијене резултате са резултатима добијеним под г).
 ђ) 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.
 е) Пронаћи свеукупну медијану за све куповине комбиновањем четири скупа података.
10. Подаци представљају дневну продају у хиљадама евра две франшизе истог ланца хамбургера:

| | | | | | | | | |
|------------|-----|-----|-----|-----|-----|-----|-----|-----|
| франшиза A | 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 B | 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 |

- а) Одредити узорачку средину за сваку франшизу. Да ли су, на основу ових резултата, просечне дневне продаје приближно једнаке?
- б) Одредити узорачку медијану за сваку франшизу. Да ли су, на основу ових резултата, медијане дневне продаје приближно једнаке?
- в) На основу дијаграма стабло-лист, да ли се може речи да дневне породаје имају исту расподелу. Који подаци имају већу распршеност?
- г) Доказати тврђење под в) одређивањем дисперзије.
- д) Да ли би било необично да у неком од следећих дана дневна продаја буде преко 4000 евра, а следећег дана 1000 евра за франшизу А? А за франшизу В?
11. Спроведено 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.
- а) Одредити узорачку средину и узорачку медијану.
- б) Одредити узорачки распон.
- в) Одредити узорачку дисперзију и стандардну девијацију.
12. 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 | | | | | | | | | |

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

Вероватноћа

1. Болница поседује 50 јединица крви означених са $A+$. Међутим, четири јединице су у ствари $A-$. Случајно се бира једна јединица крви. Ако свака јединица има исту вероватноћу да се изабере, одредити вероватноћу да се изабере погрешно означена јединица.
2. Према статистици кошаркаш погађа 82% слободних бацања. Ако га фаулирају и добије слободно бацање, одредити вероватноћу да погоди.
3. Код биљке грашка алел за висину (V) је доминантан у односу на алел за ниску стабљику (v), алел за жуто зрно (Z) је доминантан у односу на зелено (z), а алел за округло облику (O) је доминантан у односу на алел за наборано зрно (o). Претпоставимо да се укрштају две биљке, једна са генима $VVZZOo$ и друга са генима $VvZzOo$.
 - а) Описати родитељске биљке на основу наведених карактеристика.
 - б) Нацртати стабло могућих начина укрштања и описати биљке које могу да настану.
 - в) Одредити вероватноћу настанка високе биљке.
 - г) Одредити вероватноћу настанка високе биљке са округлим и жутиим зрном.
 - д) Одредити вероватноћу настанка биљке са зеленим зрном.
4. Код заморчића кратко крзно је доминантно над дугим, а црно над албино. Женка црне боје са кратким крзном је упарена са албино мужјаком дугог крзна.
 - а) Који су могући генотипи родитеља?
 - б) Нацртати стабло могућих начина укрштања за све генотипе родитеља.
 - в) Одредити вероватноћу настанка албино заморчета кратког крзна (у сваком од случајева).
5. Одредити узорачки простор за експеримент који представља крвне групе са Rh фактором. Дефинисани су догађаји: A - крв садржи A антиген, B - крв садржи B антиген, P - Rh фактор је позитиван. Одредити догађаје: $\bar{A}, A \cap \bar{P}, \bar{A} \cap \bar{B}, A \cup B, (A \cap B) \setminus P$.
6. Проучавање показује да 12% од свих људи који посете лекара буде задржано у болници. Од свих људи који дођу на преглед, 1% има лошу реакцију на одређени лек, а 12.4% се задржи у болници или има лошу реакцију на лек.
 - а) Одредити вероватноћу да случајно изабрана особа буде задржана у болници и има лошу реакцију на лек.
 - б) Одредити вероватноћу да случајно изабрана особа буде задржана у болници, али нема лошу реакцију на лек.
 - в) Одредити вероватноћу да случајно изабрана особа има лошу реакцију на лек, а није задржана у болници.
7. Експеримент се састоји од извлачења једне карте из добро промешаног шпила. Нека је H - извучен је штих (десетка, жандар, дама, краљ или кец), B - карта је црне боје. Одредити
 - а) $P(H)$
 - б) $P(B)$
 - в) $P(H \cap B)$
 - г) $P(H|B)$
 - д) $P(B|H)$
 - ђ) Да ли је $P(H|B) = P(B|H)$?
8. Здравствена организација процењује да 15% одрасле популације има хипертензију. Истраживање показује да 75% свих одраслих мисли да они лично немају хипертензију. Процењено је да 6% свих одраслих има хипертензију али не мисли да је има.
 - а) Ако одрасли пацијент сматра да нема хипертензију, одредити вероватноћу да је болест заправо присутна.

- б) Ако је болест присутна, одредити вероватноћу да пацијент сумња у њено присуство.
9. Лажно позитивна стопа медицинског теста је вероватноћа да тест покаже присуство болести када заправо болест не постоји.
- а) Да ли би лажно позитивна стопа требала да буде мала или велика?
- б) Проучавана је нова метода за детектовање болести бубрега. Према њој 49% испитаника је било позитивно. Коришћењем другог метода показано је да 39% испитаника у ствари има болест. Такође, показано је да 17% испитаника нема болест, али нови тест показује да има. Одредити лажно позитивну стоцу нове методе.
- в) Ако се узме нови тест и резултат је позитиван, да ли је то јасни доказ да је болест присутна?
- г) Како би се дефинисала лажно негативна стопа?
- д) Шта је повољније за пацијента: мала лажно позитивна стопа или мала лажно негативна стопа?
10. У некој популацији 9% људи има крв типа B , а 61% има позитиван Rh фактор. Одредити вероватноћу да случајно изабрана особа има $B+$ крв.
11. Ако брачни пар, где и мушкарац и жена имају један рецесивни (плави) и један доминантан (браон) ген за боју очију, има дете оно са вероватноћом $\frac{3}{4}$ има браон очи. Ако тај пар има два детета, одредити вероватноћу да
- а) оба имају браон очи;
- б) један има плаве, а други браон очи.
12. Са вероватноћом 0.2 особа која је изложена рубиоли је и добије. Ако је женска особа изложена ризику током трудноће, вероватноћа је 0.1 да ће дете имати дефект, иначе вероватноћа да има дефект је 0.01.
- а) Ако је дете рођено са дефектом, одредити вероватноћу да је мајка била изложена током трудноће.
- б) Ако је дете рођено без дефекта, одредити вероватноћу да је мајка била изложена током трудноће.
13. Проучавано је колико мушкарци и жене пију и добијени су следећи подаци:

| | мушкарци | жене |
|------------|----------|------|
| константно | 23% | 40% |
| често | 21% | 5% |
| умерено | 46% | 37% |
| ретко | 10% | 18% |

Претпоставља се да половину популације чине мушкарци, а половину жене.

- а) Случајно је изабрана особа и утврђено је да константно пије. Израчунати вероватноћу да је особа мушкарац.
- б) Случајно је изабрана особа и утврђено је да ретко пије. Израчунати вероватноћу да је особа жена.
14. Четири рибонуклеотиде: аденин, урацил, гуанин и цитозин, представљају основу RNA тако што формирају "речи", односно низове од три рибонуклеотиде, не обавезно различите.
- а) Колико речи се може формирати на овај начин?
- б) Колико од тих речи има све различите нуклеотиде?
- в) Колико од тих речи садржи најмање две исте нуклеотиде?
- г) Одредити вероватноћу да случајно изабрана реч почиње аденином и нема идентичних нуклеотида.

15. У медицинском проучавању врши се експеримент на 9 људи. Два лека и плацебо се тестирају. Сваки се користи на тачно 3 особе са случајним редоследом. Колико је таквих редоследа могуће?
16. У лабораторији је рођено 500 мишева. Четворо од њих има дефект при рођењу. Истраживачи узимају 20 мишева за експеримент.
 - а) Одредити вероватноћу да није изабран ниједан дефектни миш.
 - б) Одредити вероватноћу да је изабран тачно један дефектни миш.

Задаци за вежбу

1. Оштећен новчић се баца 100 пута. Ако је глава пала 75 пута, одредити вероватноћу да при следећем бацању падне писмо.
2. Стандардан шпил карата садржи 52 карте. Карте се промешају и вади се једна карта случајно. Израчунати вероватноћу да је извучена
 - а) црна карта;
 - б) црвена карта;
 - в) каро;
 - г) кец;
 - д) штих (краљ, краљица, жандар, десет, кец).
3. Биоактивни тетрапид има следеће аминокиселине у саставу: аланин, глутаминску киселину, лизин и хистидин.
 - а) Нацртати дрво од 24 могућа начина на које ове четири аминокиселине могу формирати ланац.
 - б) Ако је сваки ланац једнако вероватан, одредити вероватноћу догађаја K - глутаминска киселина се налази на једном од крајева ланца.
 - в) Одредити вероватноћу догађаја B - лизин се не налази ни на једном крају ланца.
4. Планински лавови који живе на јавним пашњацима могу представљати претњу за говеда и овце. Због тога је пожељно одредити број ових лавова који живе на одређеном простору. Њих 10 је ухваћено, обележено и пуш
 - а) Нацртати дрво од 16 могућа резултата експеримента.
 - б) Одредити догађај A - прва и последња ухваћена животиња су означене.
 - в) Одредити догађај B - тачно три животиње су означене.
 - г) Одредити догађај да се истовремено десе и A и B .
5. Нектарина је глатка, док бресква није. Алел за глаткост је рецесивни. Сваки тип воћа може бити жуто или бело, при чему је жути алел доминантан. Дрво беле брескве је укрштено са жутом нектарином.
 - а) Који су могући генотипи за дрво брескве?
 - б) Који су могући генотипи за дрво нектарине?
 - в) Нацртати дрво могућих генотипа од таквих родитељских биљака.
 - г) Одредити вероватноћу добијања дрвета беле брескве (у сваком од случајева).
6. Преучавају се следећа три догађаја: P - дете је превремено рођено, M - дететова мајка је пушач, B - дете има дефект при рођењу.
 - а) Описати сваки од следећих догађаја: $\bar{P}, (P \cap M) \cap \bar{B}, (\bar{P} \cap \bar{M}) \cup B, (P \cup B) \cap \bar{M}$.
 - б) Одредити догађај: дететова мајка није пушач.
 - в) Одредити догађај: дететова мајка није пушач, али је дете прерано рођено.
 - г) Одредити догађај: дететова мајка је пушач, дете је прерано рођено и има дефект.

7. Према подацима центар за трансфузију крви 0.1% свих донора су позитивни на *HIV*, а 1% позитивни на херпес. Ако је 1.05% позитивних на једну од болести, одредити вероватноћу да случајно изабрани донор није заражен. Одредити вероватноћу да случајно изабрани донор има обе болести.
8. Породица има троје деце.
 - а) Описати скуп свих могућности за полове деце.
 - б) Ако је свака комбинација деце једнако вероватна, одредити вероватноћу да су у породици тачно два дечака.
 - в) Ако се зна да су прва два детета дечаци, одредити вероватноћу да су тачно два дечака у породици.
9. Вероватноћа да ће негативац бити стављен на *FBI*-ову листу најтраженијих људи је 0.02, вероватноћа да ће негативац бити ухваћен је 0.80, а вероватноћа да ће негативац бити стављен на листу или ухваћен је 0.81. Одредити вероватноће да негативац
 - а) буде стављен на листу и ухваћен;
 - б) не буде ухваћен;
 - в) буде ухваћен, ако је стављен на листу најтраженијих;
 - г) не буде ухваћен, ако је стављен на листу најтраженијих.
10. Жена која је носилац класичне хемофилије са вероватноћом 0.5 је преноси на синове. Одредити вероватноћу да њено прво дете је мушко и има хемофилију. Ако има тачно три сина, одредити вероватноћу да сва три имају хемофилију, да ниједан нема болест и да је тачно један заражен.
11. Оцењено је да 50% становништва Америке има прекомерну тежину и да 20% има висок притисак. Такође, се мисли да 40% свих људи са високим притиском има прекомерну тежину. Ако се случајно изабере особа, одредити вероватноћу да особа има и висок притисак и претерану тежину. Одредити вероватноћу да особа има висок притисак али нема прекомерну тежину.
12. Скрининг тест за рак има ниску лажно позитивну стопу и високу лажно негативну стопу, односно болест се открива у 95% случајева. Претпоставимо да 4% популације има болест. Одредити вероватноћу да особа за коју је тест био позитиван, заиста и има болест.
13. Фирме изнајмљују аутомобиле од агенције *A* у 26% случајева, од агенције *B* у 38% случаја/-је/-ва, а иначе од агенције *C*. Познато је да 10% аутомобила агенције *A*, 20% аутомобила агенције *B* и 5% аутомобила агенције *C* има лоше гуме. Ако је аутомобил случајно одабран из складишта ове три агенције, одредити вероватноћу:
 - а) да аутомобил има лоше гуме;
 - б) да је аутомобил из агенције *A*, ако има лоше гуме;
 - в) да је аутомобил из агенције *C*, ако има лоше гуме.
14. На колико начина се могу распоредити слова *A, B, Q, Z, T, S, X* тако да чине шифру од
 - а) седам слова без понављања;
 - б) пет слова без понављања.
15. Десет медведа је ухваћено, обележено и враћено у дивљину. Касније је поново ухваћено 8 медведа и пребројано је колико њих је означено. Претпоставља се да је иста вероватноћа да се ухвати било који од 100 медведа из популације.
 - а) Колико група од 8 медведа се може направити?
 - б) Одредити вероватноћу да ниједан медвед ухваћен у другом хватању није обележен.
 - в) Одредити вероватноћу да су сви медведи ухваћени удругом хватању обележени.

Случајне величине и дискретне расподеле

- Одредити да ли је случајна величина дискретна или непрекидна и ако је дискретна одредити вредности које узима:
 - случајна величина Y представља број који је пао приликом бацања коцкице за игру;
 - случајна величина K представља количину преписаног течног лека пацијенту;
 - случајна величина S представља брзину којом се аутомобил креће поред контролне тачке;
 - случајна величина N представља број људи који су ухапшени током месец дана.
- Нека је H случајна величина која представља највећи број који је пао приликом бацања две коцкице.
 - Одредити закон расподеле ове случајне величине.
 - Објединити једном формулом претходни резултат.
 - Одредити: $P\{H > 4\}, P\{2 \leq H \leq 4\}, P\{H = 7\}$.
- Послодавац треба да запосли четири особе. Пријавило се осам кандидата са истим квалификацијама, четири мушкарца и четири жене. Њихова имена су записана на папир и убачена у кутију, одакле су случајно извучена четири имена. Нека је X број жена које су том приликом извучене.

- Одредити вероватноћу да буду запослена тачно два мушкарца и тачно две жене.
- Одредити вероватноћу да су већином изабрани мушкарци.
- Да ли било необично да се не запосли ниједна жена? А највише једна?

- Дат је закон расподеле за R , број забележених беснила код паса:

$$R: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0.22 & 0.24 & 0.30 & 0.15 & 0.08 & 0.01 \end{pmatrix}$$

- Одредити очекивање од R .
 - Одредити $E(R^2)$.
 - Одредити дисперзију и стандардно одступање.
- Нека је X случајна величина са очекивањем 5 и дисперзијом 9, а Y случајна величина са очекивањем 7 и дисперзијом 16.
 - Одредити $E(\frac{X-5}{3}), E(\frac{Y-7}{4})$.
 - Одредити $D(\frac{X-5}{3}), D(\frac{Y-7}{4})$.
 - Показати да ако је Z случајна величина са очекивањем μ и стандардним одступањем σ , онда је $E(\frac{Z-\mu}{\sigma}) = 0, D(\frac{Z-\mu}{\sigma}) = 1$.
 - Дејство лека је тестирано на албино пацовима. Десет животиња је третирано леком који спречава синтезу протеина. Вероватноћа да ће пацов угинути током експеримента је 0.2. Ако је X број пацова који су угинули током експеримента, одредити:
 - очекивану вредност од X ;
 - дисперзију и стандардно одступање од X ;
 - израз за расподелу вероватноће;
 - вероватноћу да ниједан пацов не угине;
 - вероватноћу да најмање један пацов угине;
 - вероватноћу да бар 7 пацова буде живо на крају експеримента.
 - Душевни болесник у просеку има 2 напада озбиљне депресије током периода од шест месеци.
 - Одредити вероватноћу да ће проћи шест месеци без депресије.

- б) Одредити вероватноћу да ће проћи годину дана без депресије.
8. У просеку, копир машина се заглави два пута дневно. Нека X означава број заглављивања машине током 7 дана.
- Одредити очекивање од X .
 - Одредити дисперзију од X .
 - Одредити вероватноћу да се машина заглави мање од 8 пута током тог периода.
 - Одредити вероватноћу да се машина заглави више од 12 пута током тог периода.
 - Одредити вероватноћу да се машина заглави између 14 и 18 пута током тог периода.
 - Ако се машина заглави више од 23 пута, да ли се може сумњати у њену исправност?
9. Познато је да у неком граду становник има бицикл са вероватноћом 0.03. Израчунати вероватноћу да од 100 случајно изабраних становника број оних који поседују бицикл буде између 2 и 6.

Задаци за вежбу

- Одредити да ли је случајна величина дискретна или непрекидна и ако је дискретна одредити вредности које узима:
 - случајна величина T представља време потребно спасилачкој служби да реагује на позив;
 - случајна величина X представља изгубљену телесну масу у једној недељи;
 - случајна величина L представља дужину трајања позитивног ефеката у лечењу леукемије након што пацијенти успешно приме прву терапију;
 - случајна величина P представља број газираних пића који се на дневном нивоу продају у продавници.
- Одредити закон расподеле случајне променљиве D која представља разлику између добијених вредности при бацању коцкица, једне црвене и једне беле (црвена минус бела).
- Број деце у породицама које примају социјалну помоћ је мањи него у прошлости. Расподела вероватноће за ову променљиву је приказана у табели:

$$X: \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0.10 & 0.15 & 0.27 & 0.30 & ? \end{pmatrix}$$

- Одредити $P\{X = 4\}$.
 - Одредити $P\{X \leq 2\}$ и $P\{X < 2\}$.
 - Одредити вероватноћу да случајно изабрана породица која прима социјалну помоћ има једно, двоје или троје деце.
4. Расподела вероватноће случајне величине X представља број курсева које студент слуша у последњем семестру последње године студија.

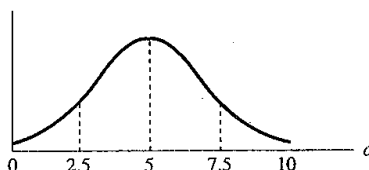
$$R: \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0.05 & 0.1 & 0.3 & 0.4 & 0.1 & 0.05 \end{pmatrix}$$

- Одредити средњу вредност за X
 - Одредити $E(X)$.
 - Одредити $E(X^2)$.
 - Одредити дисперзију и стандардно одступање за X .
5. Жена купује осигурање стана по цени од 125 евра годишње. Њен намештај је осигуран на максималну своту до 10000 евра. Осигуравајућа кућа сматра, на основу прошлих искустава, да су вероватноће за одштетним захтевом у висини од 500, 1000, 5000, 10000 евра 0.1, 0.02, 0.01, 0,001, редом. Да ли ће компанија остварити профит од 10 евра по полиси?

6. Претпоставља се да ће се код 40% ХИВ - позитвних особа вирус развити у сиду. Петнаест особа заражених ХИВ вирусом је случајно изабрано и тестирано на сиду. Нека је X број позитвних особа на ХИВ.
- Одредити $E(X), D(X), \sigma$.
 - Одредити израз за расподелу вероватноће.
 - Одредити вероватноћу да ниједна од тестираних особа није позитивна на сиду.
 - Одредити вероватноћу да је барем једна особа позитивна на сиду.
7. Од укупног броја штампача који се користе у кућним условима, 80% у почетку ради исправно. Осталим је потребне минимално подешавање. Продавац продаје 10 комада у посматраном месецу.
- Одредити вероватноћу да барем 9 штампача ради исправно након инсталације.
 - Претпоставимо да се у току 5 месеци прода 10 комада месечно. Одредити вероватноћу да ће барем 9 штампача радити исправно у сваком од 5 месеци?
8. Претпоставља се да код здраве особе број белих крвних зрнаца износи 6000 по кубном милиметру (mm^3) крви. Узето је на посматрање $0.001 mm^3$ крви. Нека је X број пронађених белих крвних зрнаца. Код здраве особе, која је очекивана вредност од X ? Ако је највише једна бела крвна ћелија пронађена, да ли то доказује недостатак белих крвних зрнаца у организму?
9. Државу погоди приближно 500 земљотреса годишње који су довољно јаки да се осете. Међутим, они са деструктивним магнитудама се јављају једном годишње. Одредити вероватноћу да ће у периоду од 6 месеци државу погодити земљотрес са деструктивним магнитудама. На основу вероватноће дешавања, одредити могућност 3 или више земљотреса са деструктивним магнитудама у посматраном периоду.

Непрекидне расподеле

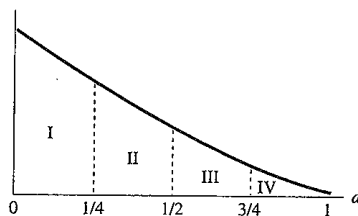
- Размотримо случајну величину L , дужина времена изношења завршне речи у судском процесу. Претпоставимо да има случајну вредност између 20 и 100 минута. Одредити:
 - расподелу од L ;
 - $P\{L \leq 30\}$ и $P\{L < 30\}$;
 - $P\{40 < L < 55\}$;
 - вероватноћу да траје дуже од сат времена;
 - вероватноћу да траје тачно сат времена.
- Нека је O прековремени рад запослених у некој фирми током недељу дана. Нека је функција расподеле од O дата на слици и нека је симетрична око 5 часова.



- Осечити област која одговара вероватноћи $P\{O < 2.5\}$.
 - Ако је вероватноћа из дела а) $\frac{1}{10}$, одредити вероватноћу да сакупљено време буде веће од 7.5 часова.
 - Одредити вероватноћу да сакупљено време буде тачно 5 часова.
 - Одредити вероватноћу да сакупљено време буде између 2.5 и 7.5 часова.
 - Ако се разматра период од 50 недеља, за колико недеља се очекује да прековремени сати буду између 2.5 и 7.5 часова.
- Код дијабетичара може се претпоставити да ниво глукозе X има нормалну расподелу са очекивањем $106 \text{ mg}/100\text{ml}$.
 - Одредити $P\{X \leq 120\}$.
 - Који проценат дијабетичара има ниво глукозе између 90 и 120?
 - Одредити вероватноћу да ниво глукозе буде између 106 и 110.
 - Одредити вероватноћу да ниво глукозе буде најмање 121
 - Одредити тачку x такву да 25% дијабетичара има ниво глукозе мањи од x .
 - Нека X има биномну расподелу са $n = 20$ и $p = 0.5$.
 - Одредити $P\{X \leq 5\}$.
 - Уз помоћ апроксимације одредити $P\{X \leq 5\}$.
 - Одредити $P\{X = 5\}$.
 - Уз помоћ апроксимације одредити $P\{X = 5\}$.
 - Истраживање показује да мање од $\frac{1}{3}$ свих пацијената који су имали први епилептични напад ће имати други у току наредне 3 године. Да би се то испитало посматрано је 150 пацијената од њиховог првог напада.
 - Ако је вероватноћа да се догоди други напад у наредних 3 године $p = \frac{1}{3}$, колики очекивани број пацијената из узорка који су имали други напад у току три године од првог?
 - Претпоставимо да је број пацијената који су имали други напад јнајвише 40, одредити вероватноћу да је направљена грешка у истраживању, односно да је вероватноћа да се други напад деси баш $\frac{1}{3}$.
 - Претпоставимо да је број пацијената који су имали други напад вечи од 40, одредити вероватноћу да је грешка неверовати истраживању, ако је вероватноћа да се други напад деси $\frac{1}{4}$.

Задаци за вежбу

1. Нека је D количина течног лека у кубним центиметрима (cc) које лекар треба да препише пацијенту да би лек постигао жељено дејство. Претпоставимо да је график вероватноће функције као на слици



Одредити које области на слици одговарају свакој од следећих вероватноћа:

- $P\{D \leq \frac{1}{4}\}$;
 - $P\{D < \frac{1}{4}\}$;
 - $P\{\frac{1}{4} < D < \frac{3}{4}\}$;
 - вероватноћи да најмање $\frac{1}{2}cc$ треба бити преписано;
 - вероватноћи да највише $\frac{3}{4}cc$ треба бити преписано;
 - Одредити нумеричку вредност области која одговара регијама I, II, III и IV заједно?
2. У нормалним условим одређени лек има Ph чија расподела има средњу вредност 9 и дисперзију $\frac{1}{4}$. Ако је Ph превисока или прениска, лек се одбацује као неприхватљив. Ако се узорак налази у горњих или доњих 15% расподеле, сматра се неприхватљивим. У ком Ph распону се узорак мора налазити да би лек био прихватљив?
3. Просечан износ новца преузет са банкомата је 6600 динара са стандардним одступањем 1000 динара. Случајна величина X представља количину новца која се подигне са банкомата и претпоставља се да је нормално расподељена.
- Одредити $P\{X > 7500\}$.
 - Одредити проценат људи који подиже више од 10000 динара.
 - Одредити проценат људи који подиже мање од 2500 динара.
 - Мање од колико новца подиже 25% људи?
4. Лек против главобоље је ефикасан у 80% случајева главобоља узрокованих нервозом. Ако тај лек узме 100 људи који пате од главобоље изазване нервозом, одредити вероватноћу да ће између 75 и 90 људи осетити олакшање.
5. Истраживање показује да је вероватноћа да је потребно извршити царски рез приликом порођаја једнака 0.05. Од следећих 100 порођаја, колико царских резова се очекује? Да ли је изненађујуће да бар 7 беба мора бити рођено царским резом? А најмање 10?

Закључивање о средњој вредности и дисперзији

1. Психолошко проучавање меморије код људи користи случајну листу речи. Сваком испитанику је дато 5 минута да проучи листу од 15 речи и након тога да понови што више њих. Добијени су следећи подаци: 10, 11, 14, 8, 11, 9, 8, 9, 10, 6. Оценити средњу вредност броја речи које су поновљене.
2. Ако је X случајна величина са очекивањем 5 и дисперзијом 20, за случајан узорак обима 8 одредити $E\bar{X}$ и $D\bar{X}$.
3. Размотримо експеримент бацања коцкице. Нека X представља број који је пао.
 - а) Одредити $D(X)$.
 - б) Нека X представља узорак од $n = 25$ бацања. Према централној граничној теореми, која је приближна расподела од \bar{X} ?
 - в) Предпоставимо да је у 25 бацања добијено $\bar{x} = 3.4$. Конструисати 95% интервал поверења за праву средину μ . Ако је $\mu = 3.5$, да ли интервал обухвата μ као што се очекује?
 - г) Конструисати 95% интервал поверења за μ на основу 25 бацања ако је $\bar{x} = 2.5$. Да ли овај интервал садржи μ ?
 - д) Да ли је 90% интервал дужи или краћи?
4. Случајна величина X представља број сати прековременог рада запослених на недељном нивоу. Узет је узорак и добијени су резултати: 1.6, 3.0, 4.1, 5.0, 2.0, 3.5, 4.0, 5.5, 2.1, 4.0, 4.5, 6.0, 2.2, 4.2, 5.0, 7.5.
 - а) Одредити тачкасте оцене μ, σ^2, σ .
 - б) Одредити 95% интервал поверења за μ .
 - в) Ако се сваки сат или део сата плаћа 200 динара за прековремени рад и ако је 10 запослених, колико новца је потребно да би се покрили трошкови са сигурношћу 95%?
5. Социолог проучава разводе у граду. У проучавању је случајно изабрано 500 разведених људи, а X је број година брака пре развода. Утврђено је из узорка да је $\bar{x} = 11.2$ година и $\beta = 4.2$ године.
 - а) Одредити 90% интервал поверења просека година брака пре развода.
 - б) Да ли би било изненађујуће да је просечна дужина брака 15 година?
6. Средњи ниво радијације у Америци је 0.3 рема годишње. Постоји страх да ће, као резултат повећања коришћења радиоактивних материјала доћи до пораста нивоа радијације.
 - а) Поставити нулту и алтернативну хипотезу
 - б) Објаснити последице прављења грешке I и II врсте.
7. Ранија проучавања показују да биоцид DDT се може акумулирати у организму. Средња концентрација DDT у организму појединца је била 9 јединица по милиону. Постоји претпоставка да као резултат контроле, ова концентрација опада. Поставити хипотезе за ово тврђење.
8. Један од ефеката DDT на птице је спречавање производње ензима кардонатне дехидратазе. Резултат тога је формирање јаја са знатно тањом љуском од нормалних. Да би се тестирала ова теорија, проучавани су врапци који су храњени мешавином 3 јединице по милиону (ppm) диелдрина и 15 ppm DDT . Тврдоћа љуске је поређена са средњом вредношћу љуске птица на које није утицао DDT . Процент опада ако се примети мекша љуска. Утврђено је на узорку величине 16 да је просечан проценат имао пад од 8% са стандардном девијацијом 5%. Тестирати хипотезу $H_0(\mu \leq 0)$ против $H_1(\mu > 0)$. Која је приближна p -вредност теста? Да ли је теорија потврђена статистички?

9. Просечно време за које кошаркашки тим даје погодак је 20 секунди. Постоји тврдња да то није тачно. Да би се тестирала та хипотеза узет је случајан узорак утакмица и мерено је време. Просечно време постизања поготка на основу 400 шанси је 18.8 са $s = 8$. Одредити p -вредност теста. Да ли би H_0 требало одбацити?
10. Власник музичке продавнице жели да процени старост својих купаца. Случајан узорак од 31-ог купца има средњу вредност 23 године са узорачком стандардном девојацијом 5 година.
- Конструисати 95% интервал поверења за средњи број година.
 - Конструисати 95% интервал поверења за дисперзију броја година.
 - Конструисати 95% интервал поверења за стандардну девијацију броја година.
11. Проучавана је јачина тона реклама и програма на телевизији. Променљива која нас интересује X је јачина звука у децибелима и њена дисперзија не сме бити сувише велика. Прихватљиво је да стандардно одступање буде 2 децибела или мање.
- Поставити нулту и алтернативну хипотезу којима се испитује да ли је дисперзија звука превелика.
 - Случајни узорак од 25 програма има узорачко стандардно одступање 2.2 децибела. Тестирати хипотезе из (а) са нивоом значајности $\alpha = 0.05$.
12. Чини се да је промена цена наочара постала превелика. Стандардно одступање од 15 евра је допустиво за трошкове продавца. За случајан узорак од 15 елемената добијено је $s^2 = 214.46$. Да ли се на основу ових података може закључити да је цена постала превелика?

Задаци за вежбу

1. Власник киоска са брзом храном је заинтересован за просечну количину новца коју посетиоци потроше. Добијена су три узорка у три различита дана:

| | | | | | | | | | | |
|-----------|------|------|------|------|------|------|------|------|------|------|
| узорак 1. | 1.50 | 2.50 | 1.75 | 3.25 | 1.35 | 2.16 | 2.25 | 1.21 | 1.53 | 1.16 |
| узорак 2. | 1.85 | 4.25 | 1.16 | 2.31 | 1.76 | 1.29 | 1.89 | | | |
| узорак 3. | 2.18 | 1.28 | 1.59 | 1.98 | 1.47 | 2.35 | | | | |

- Одредити $\bar{x}_1, \bar{x}_2, \bar{x}_3$.
 - Одредити тежинску средину вредности из узорка.
2. Нека X представља случајну цифру из табеле случајних бројева. Закон расподеле те случајне величине је

$$f(x) = \frac{1}{10}, x = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.$$

- Одредити $E(X)$, $E(X^2)$ и $D(X)$.
 - Нека је узет узорак од $n = 30$ случајних цифара и добијено је да је $\bar{x} = 4.2$. Конструисати 98% интервал поверења за праву средину μ . Ако је $\mu = 4.5$, да ли интервал обухвата μ као што се очекује?
3. Добијени су следећи подаци из узорка од 30 елемената:

37.5 38.9 38.9 38.1 39.5 39.2 39.1 40.0 40.1 40.1 40.8 40.5 40.6 40.4 40.3
41.7 41.2 41.5 41.6 41.4 41.1 42.0 42.2 43.3 43.1 43.2 44.0 44.9 45.0 45.7

- Одредити тачкасте оцене μ, σ^2, σ .
 - Одредити 90% интервал поверења за μ .
4. Да би имало смисла да радња буде отворена након 17:30 потребно је да се у том периоду заради у просеку 5000 динара. Случајан узорак од 36 дана показује да је $\bar{x} = 4500$ а дисперзија тог узорка је 121. Конструисати 95% интервал поверења за μ , просечну продају. На основу тог узорка, да ли радња треба да буде отворена увече?

5. У тесту за утврђивање присуства вируса сиде, спроводи се следеће тестирање H_0 (вирус је присутан) против H_1 (вирус није присутан). Објаснити шта се дешава ако се направи грешка прве врсте, а шта ако се направи грешка друге врсте.
6. Прописано је да ако је просечни ниво бактерија у води преко 70 онда вода није за пиће.
 - а) Поставити хипотезе којима се утврђује исправност воде.
 - б) Обајснити практични смисао грешака прве и друге врсте.
7. Средњи ниво угљен диоксида у ваздуху је 0.035%. Сматра се да је ниво одмах изнад земље виши од тога.
 - а) Поставити нулту и алтернативну хипотезу којима се то испитује.
 - б) Узето је 144 узорака у којима је тестиран ниво угљен диоксида. Узорачка средња вредност је 0.09% и стандардно одступање је 0.025%. Одредити p -вредност теста? Да ли треба тврђење прихватити статистички?
8. Претходна проучавања показују да је потребно 5 минута у просеку да се запамти листа од 15 речи. Психолог тврди да је потребно мање.
 - а) Поставити нулту и алтернативну хипотезу којима се то испитује.
 - б) Узет је узорак од 20 људи којима је било потребно 4 минута са стандардним одступањем 2.3 минута. Одредити p -вредност теста? Да ли треба одбацити нулту хипотезу?
9. Да би се поставио стандард за то шта треба сматрати нормалним нивоом холестерола за људе између 20 и 29 година узет је узорак од 30 особа. Урађен је тест крви и добијено је да је просечан ниво холестерола 180 mg/dl са стандардним одступањем 30 mg/dl .
 - а) Конструисати 95% интервал поверења за средњи број нивоа холестерола.
 - б) Конструисати 95% интервал поверења за дисперзију нивоа холестерола.
 - в) Конструисати 95% интервал поверења за стандардну девијацију нивоа холестерола.
10. Дужина времена трајања батеријске лампе након пуњења има нормалну расподелу са очекивањем 12 сати. Мисли се да стандардна девијација σ превазилази $\frac{1}{2}$ сата.
 - а) Поставити нулту и алтернативну хипотезу да је стандардна девијација времена више од $\frac{1}{2}$ сата.
 - б) Узет је узорак од 25 лампи и добијена је узорачка стандардна девијација 0.6. Да ли то потврђује да је стандардна девијација већа од $\frac{1}{2}$ сата?

Закључивање о проценту

- Ако ниво олова у организму детета прелази 30 милиграма по децилитру онда то може бити опасно. Случајан узорак од 1000 деце је узет и утврђено је да њих 200 има превелик ниво олова у организму.
 - Одредити проценат деце са превисоким нивоом олова.
 - Ако се узме узорак од 20000 деце, одредити број деце за које се очекује превисок ниво олова.
 - Конструисати 90% интервал поверења пропорције деце са превисоким нивоом олова.
- Постоји жеља да се утврди проценат p студената који желе да играју фудбал за факултет. Колики узорак је потребан да би се са сигурношћу 95% узорачки проценат налазило унутар 0.03 од p ?
- Претпоставља се да више од 85% деце која осете бол у грудима заправо има нормалан ехокардиограм. У узорку од 139 деце са болом у грудима 123 је имало нормалан ехокардиограм.
 - Поставити хипотезе којима би се испитала претпоставка.
 - Одредити тачкасту оцену процента деце са болом у грудима и нормалним ехокардиограмом.
 - Тестирати хипотезе из (а). Да ли треба одбацити H_0 са нивоом $\alpha = 0.10$?
- Постоји иницијатива да се сви студентски домови споје на једном месту. Због тога се жели утврдити који проценат студената и који проценат студенткиња жели то. У узорку од 50 студената њих 35 њих је за, а у узорку од 75 студенткиња њих 45 је за.
 - Одредити тачкасту оцену процента за сваку групу посебно.
 - Одредити тачкасту оцену разлике процената студената и студенткиња који су за.
 - Конструисати 95% интервал поверења за разлику из (б).
 - Да ли се може рећи да мушка популација фаворизује предлог?
- У медицини, прелазак са једног третмана на други је често врло скуп. Због тога пре промене је потребно утврдити да успешност новог третмана већа од старог и ако је већа за више од 0.10 третман се мања. Да ли има смисла промена ако су подаци следећи:

| | нови | стари |
|-----|------|-------|
| x | 15 | 60 |
| n | 20 | 100 |

Задаци за вежбу

- Процент дефектних производа у некој фабрици је p . Ако је узет узорак од 70 производа и међу њима је њих 10 дефектних, одредити тачкасту оцену од p и 95% интервал поверења за p .
- Одредити колико велики узорак је потребан да би се оценило p унутар 0.2 са 90% поверења.
- Озбиљна суша утиче и на сушење саднице и на стопу раста дрвета. Мисли се да већина дрвећа у подручју са сушом има величину стабла дупло мању него иначе. Узет је узорак од 250 стабала и међу њима је 150 са наведеном карактеристиком. Да ли ови подаци показују да је тврђење тачно?
- Узет је узорак од 400 студената и на случајан начин подељен у две једнаке групе. Свака група је позвана на вакцинацију против вируса грипа. Првој групи су показани слајдови и дат опис ефеката грипа, висок степен заразе и опасност од болести, на веома застрашујући начин. Другој групи је само дата брошура са описом болести и ништа није урађено што би изазвало страх. Од 200 студената у првој групи 44 је одбило вакцинацију, док је у другој групи одбило њих 38.

- а) Одредити тачкасте оцене процента у обе групе појединачно (p_1, p_2) .
 б) Одредити оцену од $p_1 - p_2$.
 в) Одредити 94% интервал поверења за $p_1 - p_2$.
 г) Да ли постоји велика разлика у групама?
5. Истраживач мисли да конзумирање цитрусног воћа током зимских месеци утиче на смањења стопе прехладе. Да би се то тестирало, 200 људи је узето случајно и подељено у две једнаке групе, контролну и експерименталну. Сваки члан експерименталне групе је јео три поморандже дневно од новембра до марта. Контролна група је јела нормалну количину. Добијени су резултати:

| | контролна | експериментална |
|-------------------------------------------|-----------|-----------------|
| укупно људи | 100 | 100 |
| број људи који је имао бар једну прехладу | 48 | 43 |

- а) Конструисати одговарајуће хипотезе.
 б) Тестирати хипотезе из (а) са нивоом $\alpha = 0.1$.

Поређење дисперзија и очекивања две нормалне популације

1. Социолог жели да врши експеримент на две групе људи и да расподела година у њима, у смислу средње вредности и дисперзије, буде приближно иста као пре дељења људи у групе.
 - а) Поставити хипотезе којима би могла да се детектује различитост у дисперзијама међу групама.
 - б) Узети су узорци величине 121 из сваке од популација и добијене су узорачке дисперзије $s_1^2 = 289$ и $s_2^2 = 225$. Да ли ови подаци указују на то да постоји разлика у дисперзијама међу групама?

2. Два хемичара тестирају мантиле који би се користили у лабораторији. Пожељнији је онај који има мању дисперзију јер је поузданији. Следећи подаци представљају дужине третирања мантила хемикалијама у сатима:

| | | | | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| I | 17.12 | 16.02 | 15.28 | 20.77 | 15.23 | 25.45 | 30.04 | 10.06 | 17.78 | 23.13 | | |
| II | 16.87 | 18.40 | 17.24 | 16.06 | 26.91 | 21.66 | 19.83 | 20.10 | 14.74 | 15.59 | 20.77 | 20.61 |

- а) Одредити тачкасту оцену од $\frac{\sigma_1^2}{\sigma_2^2}$.
 - б) Одредити 90% интервал поверења за $\frac{\sigma_1^2}{\sigma_2^2}$.
3. Постоји тврђење да је боља црна пластика од новина за спречавање корова код парадајза. Да би се то испитало, засађено је 9 биљака код којих је коришћена пластика и 10 биљака код којих су коришћене новине. Све биљке су имале исти третман и исти потенцијални раст. Након месец дана измерене су висине биљака:

| | | | | | | | | | | |
|---------------|------|------|------|------|------|------|------|------|------|------|
| црна пластика | 1.80 | 1.29 | 1.13 | 2.92 | 2.20 | 1.25 | 2.61 | 1.60 | 2.06 | |
| новине | 2.57 | 1.59 | 1.78 | 1.37 | 1.22 | 1.34 | 1.43 | 1.06 | 1.44 | 1.12 |

- а) Одредити тачкасту оцену од $\mu_1 - \mu_2$.
 - б) Одредити s_1^2 и s_2^2 . Тестирати једнакост дисперзија са нивоом $\alpha = 0.2$. Да ли треба одбацити нулту хипотезу?
 - в) Одредити s_p^2 и s_p . Конструисати 95% интервал поверења за $\mu_1 - \mu_2$. Да ли има значајне разлике у средњим вредностима?
4. Произвођач цигарета испитује који је бољи од два филтера. Филтер I је јефтинији од филтера II и користиће се изузев ако се не покаже да је филтер II ефективнији од филтера I. Да би се то испитало, узети су узорци од по 25 цигарета са сваки од филтера, механички су употребљене и измерени су нивои катрана и никотина. Добијени су резултати: $\bar{x}_1 = 1.10$, $s_1^2 = 0.05$, $\bar{x}_2 = 1.13$, $s_2^2 = 0.07$.
 - а) Поставити нулту и алтернативну хипотезу којом би се утврдило да ли филтер II даје већи ниво катрана и никотина.
 - б) Тестирати једнакост дисперзија са нивоом $\alpha = 0.20$.
 - в) Тестирати хипотезу из (а) са нивоом $\alpha = 0.05$. Шта се предлаже да компанија уради?
 5. У експерименту мерења слободног хлора у води, узети су узорци сувог праха и раствора који се мешају према јасним упутствима како би се добио узорак са познатом концентрацијом хлора и шаље се у лабораторије. Лабораторија је анализирано 17 узорака користећи једна метод, а 26 користећи други метод. Претпоставља се да први метод даје резултате који су у просеку виши од резултата другог метода. Добијени су резултати: $\bar{x}_1 = 72.1765$, $s_1^2 = 156.1544$, $\bar{x}_2 = 64.7308$, $s_2^2 = 4190.0725$.
 - а) Да ли треба употребити метод обједињене оцене за поређење средњих вредности?
 - б) Поставити хипотезе и тестирати тврђење да први метод тежи да покаже виши ниво хлора него други метод.

6. Проучава се количина опасног отпада на две локације. Отпад који интересује истраживаче је онај настао из домаћинства и малих предузећа. Добијени су подаци:

I локација: $n_1 = 96320, \bar{x}_1 = 16.59, s_1^2 = 25$
 II локација: $n_2 = 81609, \bar{x}_2 = 22.06, s_2^2 = 36$

Одредити 95% интервал поверења разлике просечне количине опасног отпада у узорку.

7. Микро метод је развијен за мерење количине угљен монооксида у ваздуху. Истраживачи желе да упореде тај метод са стандардним јод-пентоксид методом. Да би се то урадило, узет је узорак од 9 елемената и анализиран са оба метода и добијени су следећи резултати:

| | | | | | | | | | |
|------------|------|-------|------|-----|-------|------|------|-------|-------|
| микро | 95 | 184 | 40 | 261 | 215 | 26 | 56 | 128 | 155 |
| стандардни | 90.5 | 184.6 | 44.8 | 320 | 244.7 | 25.8 | 66.2 | 137.8 | 137.8 |

- а) Конструисати 95% интервал поверења за $\mu_1 - \mu_2$.
 б) Да ли се ова два метода разликују у просеку?
8. Психолози тврде да прворођено дете има виши IQ од другорођеног. Да би се тестирала та теорија узето је 200 парова деце и добијени су резултати: $D = X - Y, \sum d = 860, \sum d^2 = 4494$. Тестирати одговарајуће хипотезе и донети закључак.

Задаци за вежбу

1. Да би се упоредила цена поправке малих куних апарата проучаване су велике фирме које се баве поправком (прва популација) и мале радње за поправку (друга популација). Делује као да, иако је можда средња вредност цена поправки у ове две радње је иста, дисперзија је различита.
- а) Поставити хипотезе којима би могла да се детектује различитост у дисперзијама међу групама.
 б) Узети су узорци величине 30 из сваке од популација. Добијене су узорачке дисперзије $s_1^2 = 25$ и $s_2^2 = 16$. Одредити p -вредност теста? Да ли треба одбацити нулту хипотезу?

2. Две различите области тврде да је просечна температура 27°C . С обзиром да то тврђење може да буде погрешно, узети су узорци дневних температура:

| | | | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| област I | 28.47 | 29.81 | 27.43 | 25.67 | 23.00 | 27.32 | 25.49 | 26.96 | 28.99 | 23.90 |
| | 33.38 | 26.57 | 20.02 | 26.57 | 27.52 | | | | | |
| област II | 18.39 | 29.43 | 28.26 | 35.72 | 27.46 | 25.38 | 24.05 | 29.90 | 19.24 | 19.16 |
| | 20.86 | 29.34 | 29.88 | 28.81 | 27.33 | 29.43 | | | | |

Конструисати 90% интервал поверења за $\frac{\sigma_1^2}{\sigma_2^2}$. Да ли се може закључити која област има већу вариабилност у температури?

3. Проучавају се постигнућа ученика петог разреда у различитим периодима школске године. Узет је узорак од 121 резултата након првог полугодишта, узорачка средина је 3.25 а дисперзија 0.37. Узет је узорак исте величине на крају школске године и добијена је узорачка средња вредност 3.50 и дисперзија 0.44.
- а) Тестирати $H_0(\sigma_1^2 = \sigma_2^2)$.
 б) Ако се H_0 не одбацује, одредити s_p^2 и s_p .
 в) Одредити 95% интервал поверења за $\mu_1 - \mu_2$. Да ли делује као да има значајне разлике у средњим вредностима?
4. Мисли се да мушки адолесценти који пуше су почели да пуше раније него женски адолесценти. Ако су подаци

| | мушкарци | девојке |
|-------------------------|----------|---------|
| n | 33 | 14 |
| просечне године почетка | 11.3 | 12.6 |
| s^2 | 4 | 3.5 |

да ли је тврђење тачно?

5. Нови хемијски третман $NT1$ који не укључује притисак је уведен да би се постигло дуже трајање дрвене ограде. За случајан узорак од 122 слична елемента, на 61 је примењен $NT1$ а остали су третирани креозитним процесом са притиском, CPP . Након тога је ограда постављена у исти тип земљишта и добијени су подаци: $\bar{x}_1 = 25.9, s_1^2 = 31.36, \bar{x}_2 = 22.3, s_2^2 = 17.64$.

а) Да ли можемо закључити са нивоом $\alpha = 0.2$ да је $\sigma_1^2 \neq \sigma_2^2$?

б) Одредити 95% интервал поверења за $\mu_1 - \mu_2$.

6. Доктор сумња да вага за мерење телесне масе у његовој ординацији показује већу вредност тежине пацијената у односу на мале кућне ваге. Да би се ово тестирало узет је узорак од 10 пацијената са познатом вредношћу килаже коју показује кућна вага. У ординацији доктора су поново измерени и добијени су резултати:

| | | | | | | | | | | |
|-----|----|----|-----|----|----|----|----|----|----|-----|
| x | 75 | 74 | 101 | 85 | 65 | 55 | 90 | 66 | 69 | 105 |
| y | 77 | 73 | 104 | 87 | 64 | 56 | 91 | 66 | 65 | 107 |

где је x тежина на кућној ваги, а y тежина у ординацији. Тестирати одговарајуће хипотезе.

Регресија и корелација

1. Нека x представља број недеља у којима је особа била на дијети, а Y број килограма које је изгубила. Добијен је узорак: $(3,9)$, $(7,12)$, $(2,2)$, $(4,13)$, $(8,12)$, $(3,8)$.

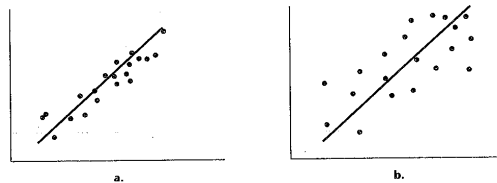
- Нацртати дијаграм распршености.
- Одредити једначину линеарне регресије Y у односу на x .
- Предвидети губитак након 5 година.

2. Дати су подаци:

| | | | | |
|-----|---|---|----|----|
| x | 1 | 2 | 3 | 4 |
| y | 6 | 7 | 10 | 13 |

- Одредити једначину линеарне регресије.
- Предвидети вредност Y за свако x .
- Одредити резидуале и SSE .
- Одредити $\hat{\sigma}^2$.
- Одредити суму резидуала.

3. За који од следеће две групе података се очекује да има већу оцењену дисперзију?



4. Проучава се однос између x , тежине аутомобила у тонама, и Y , потрошње бензина. Десет исправних аутомобила је вожено 1000 км и бележено је колико километара по литру је пређено. Резултати су: $n = 10$, $\sum x = 16.75$, $\sum x^2 = 28.64$, $\sum y = 170$, $\sum y^2 = 2900.46$, $\sum xy = 282.41$.

- Оценити коефицијент нагиба регресионе линије.
- Оценити σ^2 и σ .
- Одредити 95% интервал поверења просечне потрошње горива за све аутомобиле тежине 2 тоне.
- Одредити 95% интервал предвиђања просечне потрошње горива аутомобила тежине 2 тоне.

5. Модел линеарне регресије за величину оклопа рака (Y) и старости (x) рака добијен на основу узорка обима $n = 20$ је $\hat{y} = 17.1 + 3.2x$.

- Предвидети величину оклопа рака старости 4 године.
- Одредити вредност резидуала за $x = 4$, ако је вредност из узорка $y = 34$.
- Тестирати $H_0(\beta = 0)$ против $H_1(\beta \neq 0)$, ако је $SEE = 928.4$ и $S_{xx} = 321.24$.
- Ако је једначина логаритамске регресије $\hat{y} = 11.38 + 17.38 \ln x$, предвидети вредност величине оклопа за рака старости 4 године и одредити резидуале.

6. Власник фирме жели да проучи везу између наплате од стране купаца у првом месецу X и укупне годишње наплате од купца Y . Добијени су подаци:

| | | | | | |
|-----|-----|----|-----|-----|-----|
| x | 10 | 5 | 25 | 100 | 30 |
| y | 120 | 75 | 250 | 800 | 360 |

- Нацртати дијаграм распршености.
- Да ли изгледа да су корелисане X и Y , и ако јесту, како?

в) Одредити оцену корелације.

Задаци за вежбу

1. Нека x представља температуру у степенима целзијуса измађу 15 и 16 часова, а Y потрошњу електричне енергије изражене на скали од 1 до 10. Добијени су следећи подаци: (21,4), (31,7), (25,5), (34,8), (30,7), (24.4.5), (32,6), (30,6), (38,9), (38,9.5).

а) Нацртати дијаграм распршености.

б) Одредити линију регресије.

в) Предвидети вредност потребе за електричном енергијом ако је температура 35°C .

2. Нека x представља касе који раде у супермаркету, а Y време потребно да купац дође на ред. Добијени су подаци:

| | | | | | | | |
|-----|---|----|----|----|----|----|----|
| x | 3 | 4 | 5 | 10 | 5 | 7 | 8 |
| y | 9 | 11 | 14 | 26 | 15 | 17 | 17 |

а) Одредити једначину линеарне регресије.

б) Предвидети вредност Y за свако x .

в) Одредити резидуале и SSE .

г) Одредити $\hat{\sigma}^2$.

д) Одредити суму резидуала.

3. Проучава се веза између прихода домаћинства у хиљадама (x) и потрошње електричне енергије (Y). Добијени су подаци:

| | | | | | | | | |
|-----|------|------|------|------|------|------|------|------|
| x | 21.0 | 30.0 | 40.0 | 55.0 | 60.0 | 75.0 | 88.0 | 95.0 |
| y | 1.9 | 3.0 | 4.5 | 5.0 | 6.5 | 7.0 | 9.0 | 9.5 |

а) Оценити коефицијент нагиба регресионе линије.

б) Оценити σ^2 и σ .

в) Одредити 90% интервал поверења просечне потрошње електричне енергије породице са примањима 50000 годишње.

г) Одредити 90% интервал предвиђања потрошње електричне енергије породице са примањима 50000 годишње.

4. Конобар жели да одреди колико је просечно време које особа треба да чека да буде послужена у зависности од броја тренутно заузетих столова. Подаци су:

| | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|
| x | 5 | 5 | 10 | 12 | 15 | 20 | 20 | 25 | 27 | 30 |
| y | 10 | 12 | 14 | 13 | 18 | 19 | 21 | 23 | 30 | 35 |

а) Нацртати дијаграм распршености.

б) Тестирати $H_0(\beta = 0)$ против $H_1(\beta \neq 0)$, са нивоом $\alpha = 0.05$. Ако се одбацује нулта хипотеза, оценити параметре линеарне регресије.

5. Истраживач је заинтересован за проучавање везе између броја сати у току дана проведених испред телевизора, X , и нивоу читања у трећем разреду, Y . Добијени су следећи подаци:

| | | | | | | | | | | |
|-----|---|---|---|---|---|---|---|---|---|---|
| x | 3 | 5 | 6 | 2 | 0 | 1 | 3 | 5 | 6 | 2 |
| y | 4 | 3 | 2 | 4 | 5 | 5 | 3 | 2 | 1 | 5 |

а) Нацртати дијаграм распршености.

б) Да ли изгледа да су корелисане X и Y , и ако јесу, како?

в) Одредити оцену корелације.

Категоријски подаци

1. Проучава се нова вакцина против грипа. Узет је случајан узорак од 818 особа и класификовани су према томе да ли су вакцинисани и какво им је стање здравља. Подаци су дати у табели:

| | заражен | незаражен |
|--------------|---------|-----------|
| вакцинисан | 276 | 3 |
| невакцинисан | 473 | 66 |

- а) Да ли су неке маргиналне вредности задате од стране истраживача?
 б) Поставити хипотезе и тестирати да ли постоји веза између променљивих.
2. Спроведено је истраживање ефикасности новог система наручивања. Узет је узорак од 100 купаца послужених на стари начин и 100 купаца послужених на нови начин. Сваки купац је контактиран да би се изјаснио да ли је задовољан услугом. Добијени су следећи резултати:

| | задовољан | незадовољан | |
|-------|-----------|-------------|-----|
| нови | 82 | | 100 |
| стари | | 30 | 100 |

Тестирати одговарајуће хипотезе.

3. Новија проучавања показују да постоји веза између боје очију и брзине трчања атлетичара. Из случајног узорка од 100 атлетичара добијени су следећи подаци:

| | спор | средње | брз |
|-------|------|--------|-----|
| плава | 24 | 26 | 5 |
| браон | 21 | 14 | 10 |

Да ли ови подаци потврђују проучавања?

4. Проучавају се фактори утицаја лекара на одлуку о трансфузији. Узет је узорак од 49 лекара и 71 специјализаната и постављено им је питање о учесталости непотребне трансфузије коју је предложио други лекар. Подаци су дати у табели:

| | веома често | често | понекад | ретко | никад | |
|---------------|-------------|-------|---------|-------|-------|----|
| лекар | 1 | 1 | 3 | 31 | 13 | 49 |
| специјализант | 2 | 13 | 28 | 23 | 5 | 71 |

- а) Тестирати хипотезу о неповезаности величина.
 б) Да ли је ово тест независности или хомогености?

Задаци за вежбу

1. Спроведено је истраживање да би се видело да ли постоји веза између година и спремности да се користи компјутеризовани банковни систем. Подаци су добијени узимањем 500 случајно изабраних корисника банке којима је понуђено да користе систем 5 година:

| | користи | не користи |
|--------------------|---------|------------|
| испод 40 година | 150 | 75 |
| 40 или више година | 150 | 125 |

- а) Поставити одговарајуће нулту и алтернативну хипотезу којима се детектује веза између година и коришћења система.
 б) Тестирати хипотезе.
2. Узети су узорци од 100 студената и 100 студенткиња и спроведена је анкета о њиховом мишљењу о обавезном служењу војске. Резултати су дати у табели:

| | | | |
|----------|----|----------------------|-----|
| | за | неутрални или против | |
| мушкарци | | 75 | 100 |
| девојке | 15 | | 100 |

- а) Поставити одговарајућу нулту хипотезу да не постоји веза између пола и мишљења.
 б) Тестирати хипотезу.
3. Продавац жели да утврди да ли постоји веза између дана у недељи који је изабран за снижење и величине продаје. Током 2 године, 40 распродаја је спроведено током уторка, среде и четвртка, а распородаје су подељене у слабе, средње и велике. Добити су резултати:

| | | | | |
|----------|-------|--------|--------|----|
| | слаба | средња | велика | |
| уторак | 12 | 20 | 8 | 40 |
| среда | 12 | 22 | 6 | 40 |
| четвртак | 0 | 18 | 22 | 40 |

- а) Тестирати хипотезу о неповезаности дана у недељи и величине продаје.
 б) Да ли је ово тест независности или хомогености?
 в) Који дан бисте препоручили за распродају?

Анализа варијансе

1. Угљен-диоксид има лош утицај на микробиолошки развој. Мала количина CO_2 стимулише развој већине микроорганизама, док велика концентрација спречава развој већине микроорганизама. Проучава се утицај CO_2 на организме који кваре храну. Угљен-диоксид је постављен на 5 различитих атмосферских притисака. Резултати представљају проценте промене масе ћелија након сат времена. Коришћено је 10 култура за сваки ниво. Добијени су следећи резултати:

| | 0.0 | 0.083 | 0.29 | 0.50 | 0.86 |
|--|------|-------|------|------|------|
| | 62.6 | 50.9 | 45.5 | 29.5 | 24.9 |
| | 59.6 | 44.3 | 41.1 | 22.8 | 17.2 |
| | 64.5 | 47.5 | 29.8 | 19.2 | 7.8 |
| | 59.3 | 49.5 | 38.3 | 20.6 | 10.5 |
| | 58.6 | 48.5 | 40.2 | 29.2 | 17.8 |
| | 64.6 | 50.4 | 38.5 | 24.1 | 22.1 |
| | 50.9 | 35.3 | 30.2 | 22.6 | 22.6 |
| | 56.2 | 49.9 | 27.0 | 32.7 | 16.8 |
| | 52.3 | 42.6 | 40.0 | 24.4 | 15.9 |
| | 62.8 | 41.6 | 33.9 | 29.6 | 8.8 |

- а) Поставити одговарајуће хипотезе.
 б) Формирати АНОВА табелу.
 в) Одредити p -вредност теста.
2. Познато је да је токсични материјал испуштен у реку што је довело до великог сапинитета воде у области где је дозвољено пецање. Хидролози проучавају начин преношења токсичног материјала мерењем количине материјала нађеног у остригама упецаним на три различите локације. Добијени су резултати:

| | | | | | | | | |
|----------------------------|----|----|----|----|----|----|----|----|
| место 1 (ушће) | 15 | 26 | 20 | 20 | 29 | 28 | 21 | 26 |
| место 2 (далеко од залива) | 19 | 15 | 10 | 26 | 11 | 20 | 13 | 15 |
| место 3 (близу залива) | 22 | 26 | 24 | 26 | 15 | 17 | 24 | 18 |

- а) Тестирати разлике у просеку токсичног материјала нађеног у остригама на ове три локације.
 б) Ако је $H_0(\mu_1 = \mu_2 = \mu_3)$ одбачена, користећи Бонферонијев T тест прецизирати разлику која постоји.
3. Истраживач је заинтересован за тестирање два лека за регулисање крвног притиска. Сваки лек је доступан у облику пилуле и у течном облику. Због тога су могућа 4 третмана. Резултат мерења је проценат опадања притиска 15 минута након узимања лека. Истраживач жели да нађе разлике између ових третмана.
- а) Одредити облик контраста за поређење лекова A и B .
 б) Одредити облик контраста за поређење течног лека и пилуле.
 в) Узет је узорак од 20 особа са високим крвним притиском и случајно су подељени у 4 групе од по 5 људи. На сваку групу је примењен по један третман и добијени су резултати: $T_1 = 35, T_2 = 30, T_3 = 51, T_4 = 32, MSE = 5.48$. Користећи анализу варијанси тестирати разлике у средњим вредностима међу третманима, са нивоом 0.10.
 г) Користећи Шефев метод тестирати а) и б).
 в) Одредити p -вредност теста.
4. Спроведено је истраживање навика јелена у сезони. Изабране су четири стазе за које се зна да их јелени користе. Просечан број отисака, који се налазе недељно на одређеном делу сваке путање, одређује се пре почетка сезоне јелена, током сезоне, и после завршетка сезоне јелена. Стазе се третирају као блокови, и добијени су следећи подаци.

| стаза | пре | током | после |
|-------|------|-------|-------|
| 1 | 62.5 | 57.0 | 49.0 |
| 2 | 46.5 | 53.3 | 50.0 |
| 3 | 45.0 | 59.3 | 37.0 |
| 4 | 24.0 | 35.7 | 50.0 |

- а) Тестирати $H_0 : \mu_1 = \mu_2 = \mu_3$.
 б) Ако су пронађене разлике у делу (а), користити Бонферонијев Т тест са укупним α нивоом од највише 0.15 да се укаже на разлике које постоје.
 в) Оценити релативну ефикасност и прокоментарисати ефикасност блоковања.

Задаци за вежбу

1. Истраживач је заинтересован за утицај мањка сна на спретност. Узет је узорак од 22 особе и случајно су подељени у 4 групе величине 8. Након мерења количине сна, сви испитаницима је дат низ задатака да ураде. За сваку особу добијен је резултат од 0 до 10:

| група 1 | груп 2 | група 3 | група 4 |
|---------|--------|---------|---------|
| 8.95 | 7.70 | 5.99 | 3.78 |
| 9.04 | 5.81 | 6.79 | 3.35 |
| 7.72 | 6.61 | 6.43 | 2.45 |
| 6.21 | 6.07 | 5.85 | 4.27 |
| 6.48 | 8.04 | 5.78 | 4.87 |
| 7.81 | 5.96 | 7.60 | 3.14 |
| 7.50 | 7.30 | 5.78 | 3.98 |
| 6.90 | 7.46 | 6.00 | 2.47 |

Са нивоом значајности $\alpha = 0.05$ тестирати хипотезу да мањак сна не утиче на спретност.

2. Истраживач је заинтересован за проучавање три начина пречишћавања отпадних вода од угљеника. Добијени су следећи резултати:

| | | | | | | | | | | |
|------------|------|------|------|------|------|------|------|------|------|------|
| <i>I</i> | 34.6 | 35.1 | 35.3 | 35.8 | 36.1 | 36.5 | 36.8 | 37.2 | 37.4 | 37.7 |
| <i>II</i> | 38.8 | 39.0 | 40.1 | 40.9 | 41.0 | 43.2 | 44.9 | 46.9 | 51.6 | 53.6 |
| <i>III</i> | 26.7 | 26.7 | 27.0 | 27.1 | 27.5 | 28.1 | 28.1 | 28.7 | 30.7 | 31.2 |

- а) Тестирати $H_0(\mu_1 = \mu_2 = \mu_3)$ са нивоом значајности $\alpha = 0.1$.
 б) Ако се H_0 одбадију, користећи Бонферонијеве тестове утврдити разлике у популационим срединама.
3. Произвођач тениских лоптица је заинтересован за то како се нове лоптице понашају на различитим подлогама. Покренут је низ тестирања са бројањем игара одиграних пре него што се лоптица уништи. Коришћено је пет подлога и добијени су следећи резултати, на основу узорака обима 25, $T_1 = 1550, T_2 = 1700, T_3 = 1600, T_4 = 1250, T_5 = 1100, MS_E = 734.0$.
- а) Тестирати $H_0(\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5)$ са нивоом значајности $\alpha = 0.05$.
 б) Користећи Шефијев метод тестирати $H_0(\mu_2 - \mu_5 = 0)$.
4. У поређењу три лека за употребу у контролисању високог крвног притиска добијени су следећи подаци(праћено је смањење крвног притиска током периода од једне недеље):

| степен високог крвног притиска | Лек | | |
|--------------------------------|-----|---|---|
| | 1 | 2 | 3 |
| врло висок | 5 | 4 | 2 |
| висок | 4 | 5 | 4 |
| умерено висок | 6 | 8 | 8 |
| низак(мало висок) | 3 | 6 | 6 |

- а) Тестирати $H_0 : \mu_1 = \mu_2 = \mu_3$.
- б) Ако су пронађене разлике у делу (а), користити Бонферијев Т тест са укупним α нивоом од највише 0.15 да се укаже на разлике које постоје.
- в) Оценити релативну ефикасност и прокоментарисати ефикасност блоковања.

Тестови слободни од расподеле

1. Социолог жели да тестира хипотезу да је просечан број година када ступају у брак код Ескимки 18. Он сматра да је то сувише ниско и жели да потврди своју сумњу. Узет је случајан узорак од 20 жена и добијени су резултати:

18.2 17.4 15 18.1 14 22 15.2 18.3 27 17.8
16 21.6 18.5 23 17.6 18.3 16.7 20 25.7 19.6

Да ли се тестом знака са нивоом $\alpha = 0.05$ може закључити да је социолог у праву?

2. Медијана очекиване дужине живота мушкараца беле расе рођених 1900. године је била 46.6 година. Узет је узорак од 15 мушкараца рођених те године и дата су времена њихове смрти:

29.5 51.1 45.1 60 жив 42 15 55.8 53.1 25 47.5 37 жив 49.5 57.6

Да ли постоји доказ са нивоом значајности $\alpha = 0.05$ да је медијана очекиване дужине живота већа од 46.6? Користити Вилксонов тест.

3. Развијен је нови тип средства за полирање аутомобила и компанија тврди да је потребно мање времена за полирање него са другим. Да би се то потврдило узето је 15 парова аутомобила и при истим условима полирања добијени су резултати:

| | | | | | | | | | | | | | | | |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Нови: | 2.1 | 1.0 | 3.6 | 2.5 | 4.0 | 1.7 | 2.9 | 3.0 | 4.5 | 3.1 | 3.1 | 1.5 | 1.6 | 1.0 | 0.8 |
| Други: | 2.7 | 1.3 | 2.0 | 2.3 | 3.9 | 1.8 | 3.0 | 3.0 | 4.6 | 3.0 | 3.4 | 1.7 | 1.8 | 1.3 | 1.0 |

Користећи тест знака испити да ли је тврђење тачно.

4. Осигуравајућа компанија жели да покаже да су њихове годишње премије мање од конкурентских. Да би се потврдило, узет је узорак од 15 великих метропола и одговарајуће премије су:

| | | | | | | | | |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|
| Компанија | 500 | 498 | 505 | 495 | 490 | 498 | 480 | 501 |
| | 496 | 478 | 520 | 513 | 506 | 497 | 482 | |
| Други | 515 | 495 | 500 | 510 | 500 | 502 | 483 | 500 |
| | 498 | 490 | 526 | 515 | 513 | 499 | 490 | |

Текстом Вилксонова испитати да ли је компанија у праву.

5. Да би се утврдило да ли нови серум лечи леукемију, узето је 9 мишева са унапредовалим стадијумом болести. Пет мишева је добило третман, а 4 не. Дата су времена (у месецима) од почетка третмана до смрти:

Третирани: 2.1 5.3 1.4 4.6 0.9
Не третирани: 1.9 0.5 2.8 3.1

Са нивоом 0.05 испитати да ли је третман ефикасан.

6. Тестира се да ли су нови методи учења голфа ефикаснији од стандардног. Узето је 20 почетника и подељени су у 4 групе од по 5 људи. На једну групу је примењен стандардан метод, а на остале нови методи, а затим су измерене способности:

| <i>I</i> | <i>II</i> | <i>III</i> | <i>IV</i> (стандардна) |
|----------|-----------|------------|------------------------|
| 63.0 | 85.0 | 90.2 | 65.0 |
| 47.0 | 80.1 | 70.7 | 45.2 |
| 51.0 | 79.0 | 86.0 | 50.9 |
| 74.0 | 67.0 | 62.3 | 75.0 |
| 60.0 | 82.3 | 72.3 | 58.8 |

Користећи Крускал-Валисов тест са нивоом значајности $\alpha = 0.1$ проверити да ли постоји значајна разлика у резултатима.

7. Џокеј жели да нађе коња који је бржи од његовог. Разматра 5 коња које ће јахати 4 различита џокеја и мери се брзина:

| Џокеј | коњ 1 | коњ 2 | коњ 3 | коњ 4 | коњ 5 | тренутни |
|-------|-------|-------|-------|-------|-------|----------|
| A | 13.0 | 13.9 | 17.0 | 19.0 | 18.3 | 19.1 |
| B | 12.1 | 13.7 | 13.7 | 18.3 | 17.9 | 13.8 |
| C | 17.0 | 17.0 | 17.1 | 18.7 | 18.3 | 17.1 |
| D | 15.3 | 15.5 | 18.2 | 18.2 | 18.1 | 15.7 |

Да ли са нивоом поверења $\alpha = 0.1$ може да се закључи да се његов коњ не понаша исто као остали коњи?

Задаци за вежбу

1. Претпоставља се да је медијана дужине телефонског позива 3 минута. Телекомуникациона компанија се не слаже и мисли да је та вредност сувише мала. Да би се проверило ово тврђење узет је узорак од 10 позива и измерене су њихове дужине у минутима:

2.7 10.5 3.8 15.2 5.7 3.5 2.1 4.0 3.7 3.2

Да ли се са нивоом значајности 0.05 може закључити на основу теста знака да је компанија у праву?

2. Просечна цена свинског меса 1975. године је била 1.87 евра по килограму. Узет је узорак од 10 продавница са овим производом и проучаване су током 1975. године. Добијени су подаци о просечним годишњим ценама:

1.86 1.79 1.89 1.83 1.90 1.85 1.91 1.88 1.92

Користећи Вилкоксон тест проверити да ли се медијана цена поклапа са просечном годишњом ценом.

3. Како би развили боју на бази уља за сликање која се брзо суши, запослени у компанији покушавају са две формуле. Претпоставља се да се при истим условима тип II брже суши од типа I. Да би се то проверило, узето је 10 различитих површина за сликање и на сваку од њих су примењена оба типа боја. Добијена су времена сушења:

| | | | | | | | | | | |
|--------|------|------|------|------|-----|-----|------|------|------|-----|
| Тип I | 1.0 | 1.5 | 0.75 | 0.90 | 2.1 | 2.7 | 0.50 | 0.78 | 1.50 | 2.2 |
| Тип II | 0.75 | 1.25 | 0.80 | 0.85 | 2.0 | 2.7 | 0.45 | 0.75 | 1.25 | 2.1 |

Да ли ови подаци потврђују претпоставку да се тип II брже суши? Користити тест знака.

4. Психолог мисли да су жене више критичне по питању нечијег изгледа него мушкарци. Да би се тестирала та хипотеза, 12 особа које интервјуишу особе у великим компанијама је имало задатак да погледа слике са мушкарцима и женама за које се сматра да су неатрактивни и требали су да оцене атрактивност особа на слици на скали од -5 до 5. Добијени су подаци:

| | | | | | | | | | | | | |
|----------|----|------|------|---|-------|-----|-------|------|-----|------|------|-------|
| мушкарци | -3 | 0 | -1 | 1 | 1.5 | 1.7 | 1.3 | -4 | 0.1 | -3.1 | 0.12 | -0.75 |
| жене | -5 | -3.5 | -2.1 | 0 | -0.25 | 1.5 | -0.26 | -1.6 | 0.3 | -3.7 | 0.15 | -1.3 |

Да ли се на основу ових података може закључити да је психолог у праву? Претпоставља се симетрија података.

5. Сматра се да људи који возе црвена кола возе брже него они који немају црвена кола. Да би се то проверило, узет је узорак 15 аутомобила и добијени су резултати:

| | | | | | | | | | |
|--------|------|------|------|------|------|------|------|------|------|
| црвени | 55.1 | 50.5 | 65.3 | 69.2 | 58.0 | 60.0 | | | |
| други | 62.0 | 55.0 | 57.3 | 63.0 | 75.0 | 48.0 | 51.8 | 60.0 | 45.0 |

Да ли ови подаци потврђују претпоставку са нивоом 0.05?

6. У психолошком тесту, од три типа пацова, тренирани, делимично тренирани и нетренирани, је захтевано да пређу лавиринт. За сваког пацова је измерено време које му је било потребно за то. Добијени су следећи подаци:

| | | | | | | | | |
|---------------------|-----|-----|-----|-----|-----|-----|-----|------|
| тренирани | 0.7 | 1.3 | 1.6 | 0.8 | 1.2 | 0.9 | 0.4 | 1.8 |
| делимично тренирани | 1.9 | 1.8 | 1.3 | 2.7 | 1.4 | 2.7 | 2.1 | 1.6 |
| нетренирани | 1.3 | 2.4 | 2.8 | 1.9 | 1.8 | 1.9 | 1.3 | 1.6. |

Користећи Крускал-Валисов тест проверити да ли има разлике у временима, са нивоом 0.05.

7. Проучава се брзина којом 4 различита курира испоручује пакете. Идентични пакети су достављани на 8 различитих локација. Добијена су времена доставе:

| пакет | курир I | курир II | курир III | курир IV |
|-------|---------|----------|-----------|----------|
| 1 | 8 | 10 | 9 | 11 |
| 2 | 12 | 11 | 9 | 14 |
| 3 | 3 | 5 | 4 | 6 |
| 4 | 24 | 26 | 23 | 25 |
| 5 | 9 | 8 | 10 | 7 |
| 6 | 6 | 7 | 5 | 9 |
| 7 | 16 | 17 | 16 | 18 |
| 8 | 18 | 19 | 18 | 20 |

Тестирати хипотезу да су курири једнако брзи у достављању пакета.