Nika Novak

Študijsko gradivo

STATISTIKA

Naloge iz statistike za študente praktične matematike

Ljubljana, 2023

Fakulteta za matematiko in fiziko
Univerza v Ljubljani

1. Mečemo pošteno kocko. Slučajna spremenljivka X naj predstavlja število pik, ki padejo pri enem metu.

   a) Zapiši porazdelitev za slučajno spremenljivko X.
   b) Zapiši emprirično porazdelitev za X, če v desetih metih pade $5, 1, 6, 1, 3, 1, 3, 3, 1, 3$.
   c) Zapiši emprirično porazdelitev za X, če v dvajsetih metih pade $5, 1, 6, 1, 3, 1, 3, 3, 1, 3, 2, 5, 3, 5, 3, 1, 6, 3, 1, 6$.

   Rešitev:

   a) $X \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$

   b) $\text{Emp}(5, 1, 6, 1, 3, 1, 3, 3, 1, 3) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{4}{10} & 0 & \frac{4}{10} & 0 & \frac{1}{10} & \frac{1}{10} \end{pmatrix}$

   c) $\text{Emp}(5, 1, 6, 1, 3, 1, 3, 3, 1, 3, 2, 5, 3, 5, 3, 1, 6, 3, 1, 6) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{6}{20} & \frac{1}{20} & \frac{7}{20} & 0 & \frac{3}{20} & \frac{3}{20} \end{pmatrix}$

2. Ocene na izpitu

ocena x_i	frekvenca f_i
6	13
7	12
8	7
9	3
10	4

   Izračunaj povprečno oceno na izpitu in standardni odklon.

   Rešitev:

   $\overline{x} = \frac{6\cdot 13+7\cdot 12+8\cdot 7+9\cdot 3+10\cdot 4}{13+12+7+3+4} = 7.308$

   $\sigma = \sqrt{\frac{13(6-\overline{x})^2+12(7-\overline{x})^2+7(8-\overline{x})^2+3(9-\overline{x})^2+4(10-\overline{x})^2}{39}} = 1.284$

3. Izpelji formulo za povprečje in standardni odklon, če vsem vrednostim odštaješ isto konstanto.

   Rešitev:

   $\overline{x} = \frac{x_1+x_2+\cdots+x_n}{n} = \frac{(x_1-a)+\cdots(x_n-a)}{n} + a$

   Ker velja $(x_i - \overline{x})^2 = (x_i - a + a - \overline{x})^2 = (x_i - a)^2 + 2(x_i - a)(a - \overline{x}) + (a - \overline{x})^2$, lahko za vsoto zapišemo

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})^2 &= \sum_{i=1}^{n}(x_i - a)^2 + 2(a - \overline{x})\sum_{i=1}^{n}(x_i - a) + \sum_{i=1}^{n}(a - \overline{x})^2 = \\
&= \sum_{i=1}^{n}(x_i - a)^2 + 2(a - \overline{x})(n\overline{x} - na) + n(a - \overline{x})^2 \\
&= \sum_{i=1}^{n}(x_i - a)^2 - n(a - \overline{x})^2
\end{aligned}$$

Zato za standardni odklon velja

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 - (a - \overline{x})^2}$$

4. Za podatke iz naloge 2 zapiši empirično porazdelitev. Nariši histogram in porazdelitveno funkcijo.

Rešitev:

$\mathrm{Emp}(6,\ldots,10) = \begin{pmatrix} 6 & 7 & 8 & 9 & 10 \\ \frac{13}{39} & \frac{12}{39} & \frac{7}{39} & \frac{3}{39} & \frac{4}{39} \end{pmatrix}$ Verjetnosti pri posamezni oceni so relativne frekvence $\frac{f_i}{n}$. Izračunajmo še komulativne frekvence $F_i = \#(x \leq x_i)$ in relativne komulativne frekvence $\frac{F_i}{n}$:

x_i	f_i	$\frac{f_i}{n}$	F_i	$\frac{F_i}{n}$
6	13	$\frac{13}{39}$	13	$\frac{13}{39}$
7	12	$\frac{12}{39}$	25	$\frac{25}{39}$
8	7	$\frac{7}{39}$	32	$\frac{32}{39}$
9	3	$\frac{3}{39}$	35	$\frac{35}{39}$
10	4	$\frac{4}{39}$	39	1

5. Izračunaj povprečno izobrazbo zaposlenih v podjetju:

nedokončana osnovna šola	70
osnovna šola	5
poklicna srednja šola	2
gimnazija	1
fakulteta	22

Kaj pa, če so podatki malo bolj natančni?

nedokončana osnovna šola	70
osnovna šola	5
poklicna srednja šola	2
gimnazija	1
VSŠ	2
univerzitetni bolonjski program	0
univerzitetni program	0
2. stopnja univerzitetni bolonjski program	0
magisterij	0
doktorat	20

Rešitev:

X je urejenostna spremenljivka. Če ji želimo izračunati povprečje, ji dodelimo številske vrednosti $1-5$. Tedaj je povprečna izobrazba enaka $\overline{x}=2$, kar pomeni, da je povprečna izobrazba v podjetju zaključena osnovna šola.

Če pa bolj natančno opredelimo univerzitetno izobrazbo, je $\overline{x}=3$. Tako je povprečna izobrazba zaključena poklicna srednja šola.

6. Na banki nam ponujajo naložbeno shemo varčevanja, kjer bo glavnica 1000 EUR obrestovana prvo leto z obrestno mero 4%, drugo leto z obrestno mero 6% in tretje leto z obrestno mero 2%. Kakšna je povprečna obrestna mera, če je obrestovanje letno?

   Rešitev:

   Označimo glavnico s P. Tedaj ima investitor po prvem letu $P_1=P(1+R_1)$, po drugem letu $P_2=P_1(1+R_2)$ in po tretjem letu $P_3=P_2(1+R_3)=P(1+R_1)(1+R_2)(1+R_3)$. Če bi bila obretna mera konstantna v obdobju treh let, bi investitor dobil $P_3=P(1+R)^3$. Tako za povprečno obresno mero velja

$$R=\sqrt[3]{(1+R_1)(1+R_2)(1+R_3)}-1.$$

   Povprečna obresna mera v naložbeni shemi je $R=0.0398=3.98\%$.

7. Kot v nalogi 6 imamo za večletno varčevanje podane različne obrestne mere

obrestna mera R_i	frekvenca f_i
1.5%	2
2%	5
2.5%	3
3%	4
3.5%	1

   Izračunaj povprečno obrestno mero.

   Rešitev:

   $R=\sqrt[15]{1.015^2\cdot 1.02^5\cdot 1.025^3\cdot 1.03^4\cdot 1.035}-1=2.4\%$

8. Za podatke

   80, 80, 90, 110, 110, 130, 140, 140, 140, 370

   izračunaj mere centralne tendence (povprečje, mediana, modus) in mere razpršenosti (standardni odklon, variacijski razmik, interkvartilni razmik in povprečni absolutni odklon od povprečja in mediane).

   Rešitev:

   Mere centralne tendence:

   - povprečje $\overline{x}=139$
   - mediana m(Me)$\in[110,130]$
   - modus Mo $=140$

Mere razpršenosti:

- standardni odklon $\sigma = \sqrt{6449} = 80.3$
- variacijski razmik $\mathrm{VR} = \max - \min = 290$
- interkvartilni razmik $\mathrm{IQR} = q^+_{3/4} - q^-_{1/4} = 140 - 90 = 50$
- povprečni absolutni odklon $\mathrm{MAD}_u = \frac{1}{n}\sum_{i=1}^{n} |x_i - u|$
  $\mathrm{MAD}_{\overline{x}} = 46.8$
  $\mathrm{MAD}_{m^+} = 45$
  $\mathrm{MAD}_{m^-} = 45$

9. Vsem podatkom iz naloge 8 odštej 100 in ponovi izračune.

   Rešitev:

   Za popravljene podatke

   $$-20,\ -20,\ -10,\ 10,\ 10,\ 30,\ 40,\ 40,\ 40,\ 270$$

   izračunamo

   - mere centralne tendence: $\overline{x} = 39$, $\mathrm{m} \in [10, 30]$, $\mathrm{Mo} = 30$
   - mere razpršenosti: $\sigma = 80.3$, VR= 290, $\mathrm{IQR} = 50$, $\mathrm{MAD}_m = 45$

   Opazimo, da se mere centralne tendence tudi zmanjšajo za 100, mere razpršenosti pa ostanejo nespremenjene.

10. Med podatki iz naloge 8 poišči osamelce. Za preostale podatke ponovno poišči mere centralne tendence in mere razpršenosti.

    Rešitev:

    Osamelci so podatki, ki so izven intervala $[q^-_{1/4} - 1.5\mathrm{IQR}, q^+_{3/4} + 1.5\mathrm{IQR}]$.

    Ker je $q_{1/4} = 90$, $q_{3/4} = 140$ in $\mathrm{IQR} = 50$, so osamelci izven intervala $[15, 270]$. Vidimo, da je podatek 370 osamelec. Za podatke

    $$80,\ 80,\ 90,\ 110,\ 110,\ 130,\ 140,\ 140,\ 140$$

    ponovimo izračune.

    - mere centralne tendence: $\overline{x} = 113.3$, $\mathrm{m} = 110$, $\mathrm{Mo} = 140$
    - mere razpršenosti: $\sigma \doteq 24.04$, VR= 60, $\mathrm{IQR} = 50$, $\mathrm{MAD}_m = 21.1$

11. Za ocene izpita iz naloge 2 poišči mediano in prvi in tertji tercil.

Rešitev:

Kvantil $q_\alpha, \alpha \in (0,1)$ je tisti podatek, za katerega velja

$$P(X < q_\alpha) \leq \alpha \quad \text{in} \quad P(X \leq q_\alpha) \geq \alpha.$$

Verjetnosti lahko pretvorimo na število podatkov, ki so manjši oziroma manjši ali enaki določeni vrednosti

$$\#(X < q_\alpha) \leq \alpha n \quad \text{in} \quad \#(X \leq q_\alpha) \geq \alpha n.$$

Pomagamo si s komulativnimi frekvencami F_i

ocene x_i	f_i	F_i
6	13	13
7	12	25
8	7	32
9	3	35
10	4	39

Za mediano je $\alpha = \frac{1}{2}$. Zato mora veljati $\#(X < m) \leq \frac{39}{2} = 19.5$. Ker je $\#(X < 7) = 13$ in $\#(X < 8) = 25$, je $m \leq 7$.

Veljati mora še $\#(X \leq m) \geq 19.5$. Ker je $\#(X \leq 6) = 13$ in $\#(X \leq 7) = 25$, je $m \geq 7$. Mediana je tako enaka 7.

Za prvi tercil je $\alpha = \frac{1}{3}$. Iščemo $q_{1/3}$, za katerega velja $\#(X < q_{1/3}) \leq 13$ in $\#(X \leq q_{1/3}) \geq 13$. Ugotovimo, da mora veljati $q_{1/3} \leq 7$ in $q_{1/3} \geq 6$, zato je prvi tercil katerakoli vrednost na intervalu $[6,7]$.

Podobno izračunamo drugi tercil $q_{2/3} = 8$.

Za računanje kvantilov lahko podatke razvrstimo v ranžirno vrsto. Tedaj za kvantil q_α velja

$$\begin{aligned} n\alpha \notin \mathscr{Z} \Rightarrow &\quad q_\alpha = x_{(\lceil n\alpha \rceil)} \\ n\alpha \in \mathscr{Z} \Rightarrow &\quad q_\alpha \in [x_{(n\alpha)}, x_{(n\alpha+1)}] \end{aligned}$$

12. Gostota zvezne slučajne spremenljivke je

$$f(x) = \begin{cases} \frac{1}{(1+x)^2} & ; x > 0 \\ 0 & ; \text{sicer} \end{cases}$$

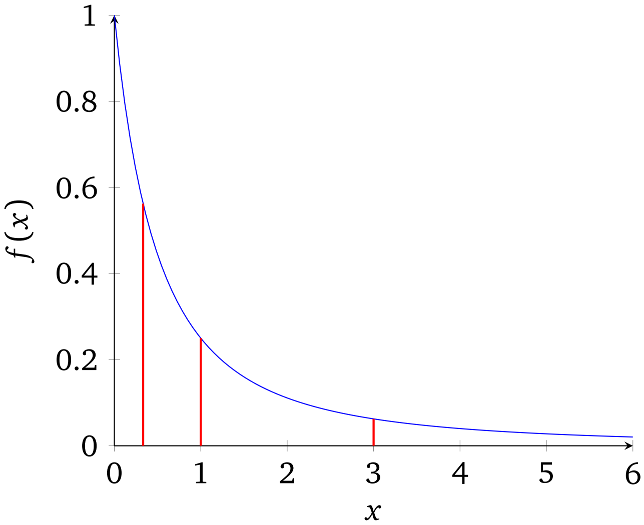
Izračunaj vse kvartile.

Rešitev:

Iz definicije kvantila $P(X < q_\alpha) \leq \alpha \leq P(X \leq q_\alpha)$ za zvezne slučajne spremenljivke sledi $F_X(q_\alpha) = P(X \leq q_\alpha) = \alpha$, kjer je $F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$ komulativna porazdelitvena funkcija slučajne spremenljivke X.

$$F_X(x) = \int_0^x \frac{1}{(1+t)^2}\,dt = \frac{-1}{1+t}\Big|_0^x = \frac{x}{x+1} \quad \text{za } x > 0$$

Iz enakosti $F_X(q_\alpha) = \frac{q_\alpha}{q_\alpha+1} = \alpha$ izrazimo $q_\alpha = \frac{\alpha}{1-\alpha}$. Tako je $q_{1/4} = \frac{1}{3}$, m$= q_{1/2} = 1$ in $q_{3/4} = 3$.



13. Za dihotomno spremnljivko podano s frekvenčno tabelo

x_i	frekvence f_i
1	6
2	14

izračunaj mere centralnih tendenc in mere razpršenosti. Doblene rezultate posploši na primer

x_i	relativna frekvenca
a	p
b	q

Rešitev:

Verjetnosti v empirični porazdelitvi $\mathrm{Emp}(X) = \begin{pmatrix} 1 & 2 \\ 0.3 & 0.7 \end{pmatrix}$ so enake relativnim frekvencam.

- mere centralne tendence: $\overline{x} = 1.7$, m $= 2$, Mo $= 2$
- mere razpršenosti: $\sigma = \sqrt{0.21}$, VR$= 1$, IQR $= 1$, $\mathrm{MAD}_m = 0.3$, $\mathrm{MAD}_{\overline{x}} = 0.42$

V splošnem za dihotomne spremenljivke velja

- povprečna vrednost $\overline{x} = ap + bq$

- mediana
$$\mathrm{m} = \begin{cases} a & ; p > q \\ b & ; p < q \end{cases},$$
za $p = q$ je m$\in [a, b]$
- standardni odklon $\sigma = \sqrt{\frac{(x_1-\overline{x})^2 f_1 + (x_2-\overline{x})^2 f_2}{n}} = \sqrt{p(a-\overline{x})^2 + q(b-\overline{x})^2}$
Ker je $a - \overline{x} = a - (ap + bq) = a(1-p) - bq = q(a-b)$ in $b - \overline{x} = p(b-a)$, je
$$\sigma = \sqrt{pq(a-b)^2(q+p)} = \sqrt{pq}\,|a-b|$$
- povprečni absolutni odklon
$$\mathrm{MAD}_{\overline{x}} = p|q(a-b)| + q|p(b-a)| = 2pq|a-b|$$
$$\mathrm{MAD}_{\mathrm{m}} = p|a-m| + q|b-m| = |a-b|\min\{p, q\}$$

14. Podatke $x_1 = 4$, $x_2 = 2$, $x_3 = 75$, $x_4 = 42$, $x_5 = 15$, $x_6 = 63$ razvrsti v ranžirno vrsto in določi rang števil 2 in 63.

Rešitev:

Ranžirna vrsta: $x_{(1)} = 2$, $x_{(2)} = 4$, $x_{(3)} = 15$, $x_{(4)} = 42$, $x_{(5)} = 63$, $x_{(6)} = 75$

Rang nam pove mesto v ranžirni vrsti, zato je $R(2) = 1$ in $R(63) = 5$.

15. Dana so števila 21, 27, 27, 27, 27, 28, 29, 29, 29, 32.

a) Določi vezani rang števil $27, 29, 20, 5, 40, 30$.
b) Določi relativni rang za števila $30, 40, 20, 29, 28, 27$.

Rešitev:

Za vsako število lahko določimo minimalni rang $R^-(x) = \#(X < x) + 1$ in maksimalni rang $R^+(x) = \#(X \leq x)$. Vezani rang je definiran kot
$$R(x) = \frac{R^+(x) + R^-(x)}{2}.$$
Relativni rang (kvantilni rang) pa je definiran kot
$$r(x) = \frac{R(x) - \frac{1}{2}}{n}.$$

a) $R(27) = \frac{5+2}{2} = 3.5$, $R(29) = 8$, $R(20) = \frac{1+0}{2} = 0.5$, $R(5) = 0.5$, $R(40) = \frac{11+10}{2} = 10.5$, $R(30) = 9.5$
Opazimo, da je $R(x) \in \left\{\frac{1}{2}, 1, \frac{3}{2}, 2, \ldots, n, n + \frac{1}{2}\right\}$. Če x ni v ranžirni vrsti, njegov rang ni celo število.
b) $r(30) = \frac{9.5-0.5}{10} = 0.9$, $r(40) = 1$, $r(20) = 0$, $r(29) = 0.75 (q_{3/4} = 29)$, $r(28) = 0.55$, $r(27) = 0.3$ ($q_{1/2} \in [27, 28]$)

16. Študentje so pisali dva različna kolokvija.

1. kolokvij		2. kolokvij	
Ambrož	83	Florian	84
Blaž	22	Gal	86
Cvetka	61	Helena	71
Darja	45	Iva	67
Emil	49	Jana	67
		Karmen	88
		Leon	89
		Mojca	64

Kdo je med svojimi pisal bolje, Cvetka ali Gal?

Rešitev:

Rezultate obeh kolokvijev razporedimo v ranžirni vrsti.

1. kolokvij: 22, 45, 49, 61, 83 $\Rightarrow$ $R(61) = 4,\ r(61) = 0.7$

2. kolokvij: 64, 67, 67, 71, 84, 86, 88, 89 $\Rightarrow$ $R(86) = 6,\ r(61) = 0.6875$

Če primerjamo relativna ranga, med njima ni bistvene razlike.

Primerjajmo še standardizirano vrednost $\dfrac{x - \overline{x}}{\sigma}$.

1. kolokvij: $\overline{x} = 52, \sigma_x = 20 \quad \Rightarrow \quad \frac{61-52}{20} = 0.45$

2. kolokvij: $\overline{y} = 77, \sigma_y = 10 \quad \Rightarrow \quad \frac{86-77}{10} = 0.9$

Glede na svoj kolokvij je Gal pisal bolje kot Cvetka.

17. Podatke

32, 32, 45, 55.5, 56, 56, 59, 68, 70, 72, 77, 78, 79, 80, 81, 84, 84.5, 90, 90, 99

razvrsti v razrede po

a) Riceovem pravilu;
b) Scottovem pravilu;
c) Modificiranem Freedman-Diaconisovem pravilu.

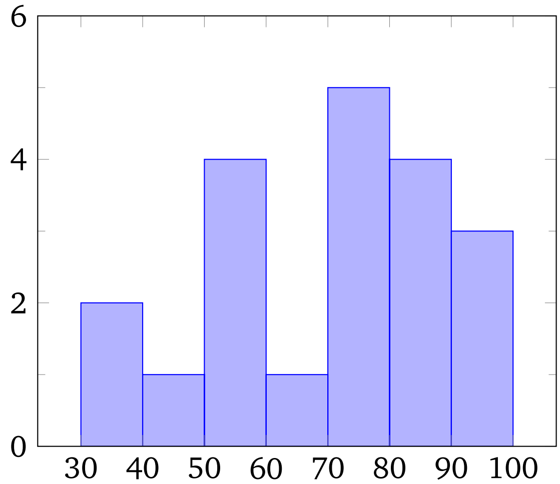
Za vsako od razporeditev nariši histogram in izračunaj povprečje iz frekvenčne tabele.

Rešitev:

a) Riceovo pravilo: število rezredov $\approx\ 2\sqrt[3]{n}$

Število razredov naj bo približno $2\sqrt[3]{20} = 5.4$. Izračunamo lahko širino razreda kot $\dfrac{\text{VR}}{2\sqrt[3]{n}} = \dfrac{67}{5.4} = 12.4 \approx 10$

razredi	f_i	$\overline{x_i}$
$30 \leq x < 40$	2	35
$40 \leq x < 50$	1	45
$50 \leq x < 60$	4	55
$60 \leq x < 70$	1	65
$70 \leq x < 80$	5	75
$80 \leq x < 90$	4	85
$90 \leq x < 100$	3	95

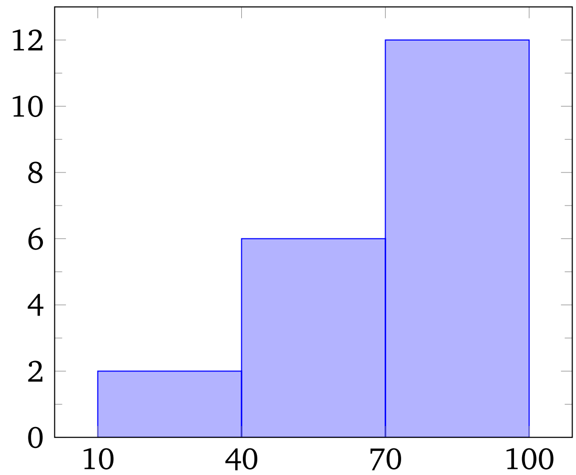


$\overline{x} \doteq \frac{2 \cdot 35 + 1 \cdot 45 + \cdots + 3 \cdot 95}{20} = 70$

b) Scottovo pravilo: širina razreda $\approx \dfrac{3.5\sigma}{\sqrt[3]{n}}$

$\overline{x} = 69.4,\ \sigma = 25.53 \quad \Rightarrow \quad$ širina razreda $\approx 32.9 \approx 30$

razredi	f_i	$\overline{x_i}$
$10 \leq x < 40$	2	25
$40 \leq x < 70$	6	55
$70 \leq x < 100$	12	85

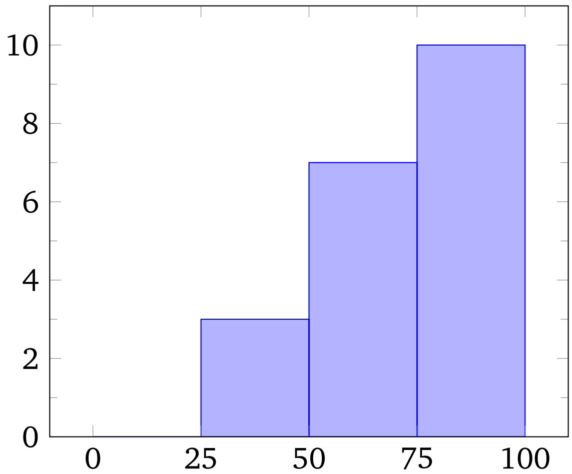


$\overline{x} \doteq \frac{2 \cdot 25 + 6 \cdot 55 + 12 \cdot 85}{20} = 70$

c) Modificirano Freedman-Diaconisovo pravilo (MFD): širina razreda $\approx \dfrac{2.6\mathrm{IQR}}{\sqrt[3]{n}}$

$\mathrm{IQR} = q^+_{3/4} - q^-_{1/4} = 84 - 56 = 28 \quad \Rightarrow \quad$ širina razreda $\approx 26.8 \approx 25$

razredi	f_i	$\overline{x_i}$
$0 \leq x < 25$	0	12.5
$25 \leq x < 50$	3	37.5
$50 \leq x < 75$	7	62.5
$75 \leq x < 100$	10	87.5

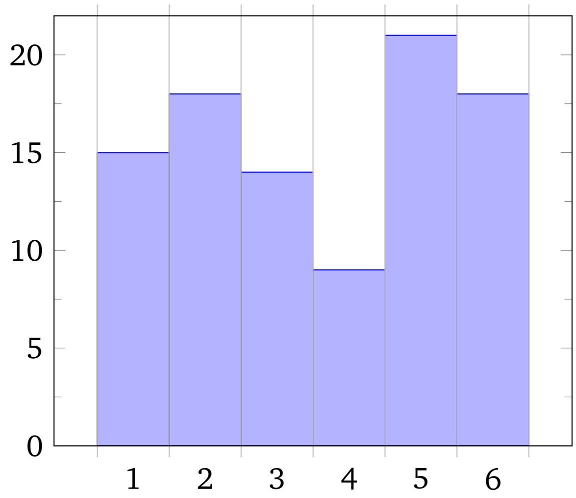


$\overline{x} \doteq \frac{3\cdot 37.5+7\cdot 62.5+10\cdot 87.5}{20} = 71.25$

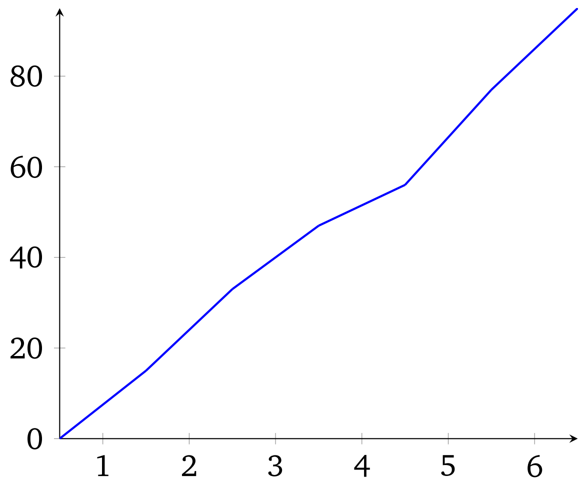
18. Pri metu kocke pade petnajstkrat 1, osemnajstkrat 2, štirinajstkrat 3, devetkrat 4, enaindvajsetkrat 5 in osemnajstkrat 6. Izračunaj vezane range, nariši histogram in ogivo. Izračunaj vrednosti ogive v x_i.

Rešitev:

x_i	f_i	F_i	$R^-(x_i)$	$R^+(x_i)$	$R(x_i)$
1	15	15	1	15	8
2	18	33	16	33	24.5
3	14	47	34	47	40.5
4	9	56	48	56	52
5	21	77	57	77	67
6	18	95	78	95	86.5



Ogiva/oživa



Poiščimo vrednost ogive pri $x = 1$. Na tem delu je ogiva premica $y = 15x - 7.5$, zato je $y(1) = 7.5$.

Pri $x = 2$ je ogiva premica $y = 18x - 12$, zato je $y(2) = 24$.

Dobimo $y(x_i) = R(x_i) - \frac{1}{2}$. Če bi ogivo risali z relativnimi frekvencami, je njena vrednost v posameznih točkah ravno relativni rang.

19. Za podatke

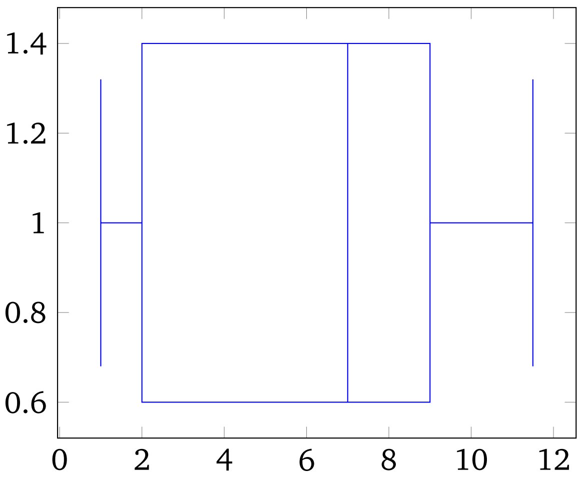
$$1,\ 1,\ 2,\ 2,\ 4,\ 6,\ 6.8,\ 7.2,\ 8,\ 8.3,\ 9,\ 10,\ 10,\ 11.5$$

nariši škatlo z brki.

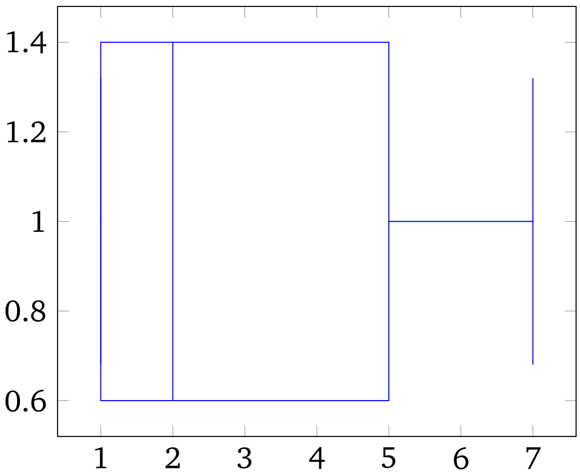
Rešitev:

$\min = 1, \max = 11.5, \mathrm{m} \in [6.8, 7.2], q_{1/4} = 2, q_{3/4} = 9$

Škatla z brki (box plot)



20. Iz škatle z brki razberi najmanjšo in največjo vrednost, mediano, prvi in tretji kvartil.



Rešitev:

$\min = 1, \max = 7, \mathrm{m} \ = 2, q^-_{1/4} = 1, q^+_{3/4} = 5$

21. Za podatke

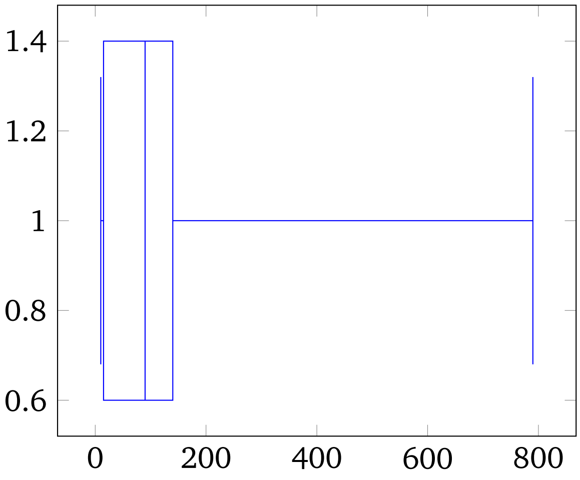
$$10,\ 10,\ 10,\ 15,\ 35,\ 75,\ 85,\ 90,\ 95,\ 100,\ 110,\ 140,\ 140,\ 175,\ 790$$

nariši škatlo z brki.

Nato preveri, če je med podatki osamelec in ga predstavi ločeno.

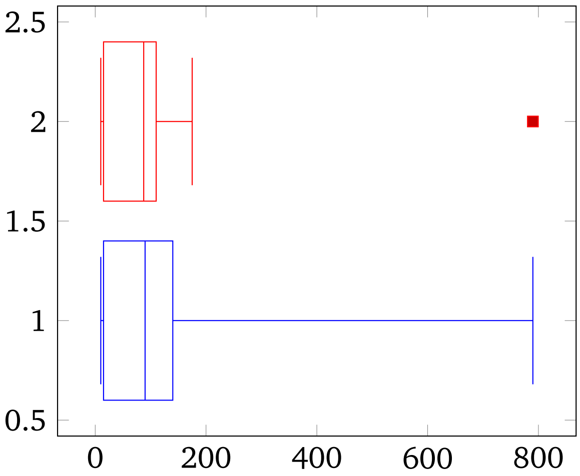
Rešitev:

$\min = 10, \max = 790, \mathrm{m} \ = 90, q_{1/4} = 15, q_{3/4} = 140$



Osamelci so elementi izven intervala $[q_{1/4}^{-} - \frac{3}{2}\ \mathrm{IQR}\ , q_{3/4}^{+} + \frac{3}{2}\ \mathrm{IQR}\ ] = [-170.5, 327.5]$. Torej je 790 osamelec.

$\min = 10, \max = 175, \mathrm{m}\ \in [85, 90], q_{1/4} = 15, q_{3/4} = 110$



22. Naj bo slučajna spremenljivka X porazdeljena enakomerno na $[a, b]$, $X \sim U[a, b]$. Tedaj je njena gostota

$$f(x) = \begin{cases} \frac{1}{b-a} & ; \ x \in [a, b] \\ 0 & ; \ \text{sicer} \end{cases}$$

in komulativna porazdelitvena funkcija

$$F(x) = \begin{cases} 0 & ; \ x < a \\ \frac{x-a}{b-a} & ; \ a \leq x \leq b \\ 1 & ; \ x > b \end{cases}$$

a) Za $X \sim U[0, 1]$ in $Y \sim U[0, 5]$ nariši kvantilni grafikon.
b) Za $X \sim U[1, 2]$ in $Y \sim U[2, 7]$ nariši kvantilni grafikon.

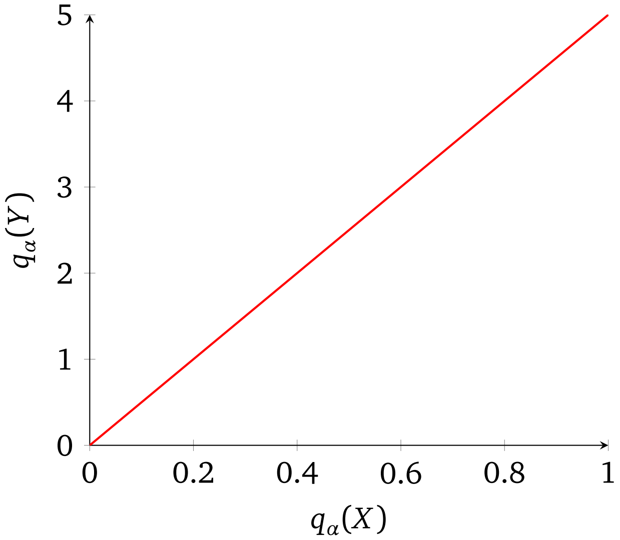
Rešitev:

Za kvantile zveznih slučajnih spremenljivk velja

$$\alpha = P(X \leq q_\alpha) = F_X(q_\alpha).$$

Tako lahko izračunamo $q_\alpha = F_X^{-1}(\alpha)$.

a) Ker je $F_X(x) = x$ za $0 \leq x \leq 1$, je $q_\alpha(X) = \alpha$. Po drugi strani je $F_Y(y) = \frac{y}{5}$ za $0 \leq y \leq 5$. Zato je $q_\alpha(Y) = 5\alpha = 5q_\alpha(X)$.



b) Ker je $F_X(x) = x-1$ za $1 \leq x \leq 2$, je $q_\alpha(X) = \alpha+1$. Iz $F_Y(y) = \frac{x-2}{5}$ za $2 \leq y \leq 7$, je $q_\alpha(Y) = 5\alpha + 2$.
Zato velja zveza $q_\alpha(Y) = 5q_\alpha(X) - 3$ za $q_\alpha(X) \in [1,2]$.

23. Naj bo X porazdeljena enakomerno na $[0,1]$ in naj bo Y eksponentna slučajna spremenljivka. Nariši kvantilni grafikon.

Rešitev:

Za $X \sim U(0,1)$ je $F_X(x) = x$ na intervalu $0 \leq x \leq 1$.

Gostota slučajne spremenljivke $Y \sim \text{Exp}\ (\lambda)$ je enaka

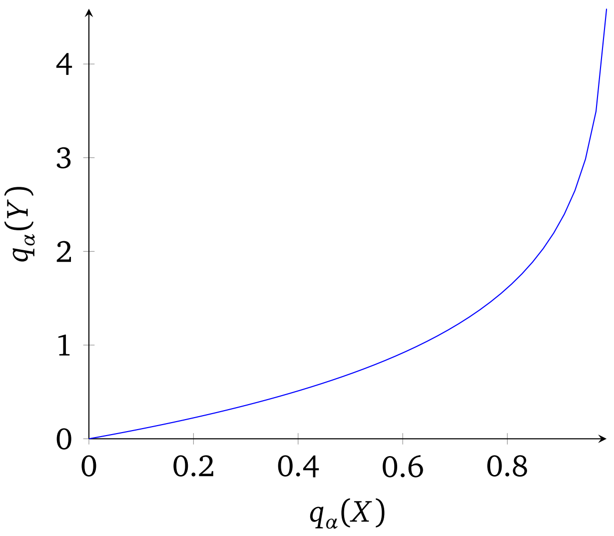
$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & ;\ y \geq 0 \\ 0 & ;\ \text{sicer} \end{cases},$$

komulativna porazdelitvena funkcija pa je enaka

$$F(x) = \begin{cases} 0 & ;\ y < 0 \\ 1 - e^{-\lambda y} & ;\ y \geq 0. \end{cases}$$

Za $p \in (0,1)$, je $q_p(X) = p$ in $1 - e^{-\lambda q_p(Y)} = p$. Tedaj iz enkosti $q_p(X) = 1 - e^{-\lambda q_p(Y)}$ dobimo

$$q_p(Y) = -\frac{1}{\lambda}\ \ln(1 - q_p(X)).$$



24. Nariši kvantilni grafikon za $X \sim U(0,5)$ in $Y \sim \begin{pmatrix} 1 & 2 & 3 & 5 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$.

Rešitev:

Slučajna spremenljivka X ima komulativno porazdelitveno funkcijo enako $F_X(x) = \frac{x}{5}$ za $0 \leq x \leq 5$. Zato je za $p \in (0,1)$ kvantil $q_p(X) = 5p$.

Oglejmo si še kvantile za Y.

$p < \frac{1}{4}: np < 1 \quad \Rightarrow \quad q_p(Y) = 1$

$p = \frac{1}{4}; np = 1 \quad \Rightarrow \quad q_p(Y) \in [1,2]$

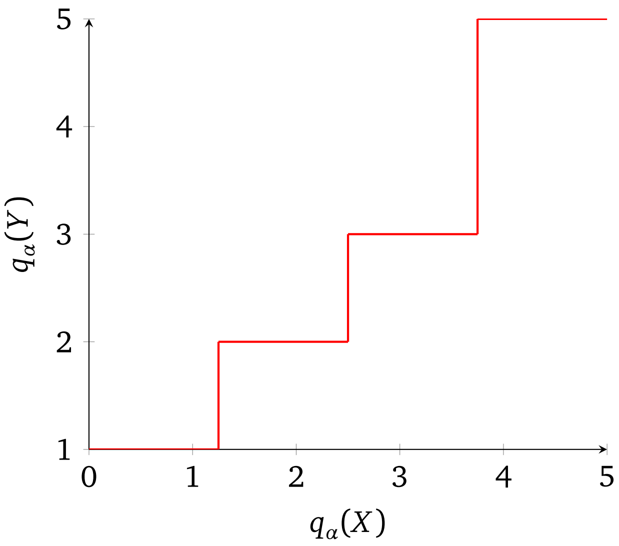
$\frac{1}{4} < p < \frac{2}{4}: 1 < np < 2 \quad \Rightarrow \quad q_p(Y) = 2$

$p = \frac{2}{4}; np = 2 \quad \Rightarrow \quad q_p(Y) \in [2,3]$

$\frac{2}{4} < p < \frac{3}{4}: 2 < np < 3 \quad \Rightarrow \quad q_p(Y) = 3$

$p = \frac{3}{4}; np = 3 \quad \Rightarrow \quad q_p(Y) \in [3,5]$

$p > \frac{3}{4}: 2 < np < 3 \quad \Rightarrow \quad q_p(Y) = 5$



Za empirićno podane slučajne spremenljivke računamo le kvatile za

$$p = \frac{k}{n+1}, \quad k = 1, 2, \ldots, n.$$

25. Poišči vse decile za standardno normalno slučajno spremenljivko.

Rešitev:

Gostota slučajne spremenljivke $X \sim N(0, 1)$ je

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}.$$

Komulativno porazdelitveno funkcijo

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{infty} e^{-\frac{t^2}{2}}\, dt$$

običajno označujemo s $\Phi(x)$ in je podana v tabeli.

Decili:

$$q_{0.1} = -1.28,\ q_{0.2} = -0.84,\ q_{0.3} = -0.52,\ q_{0.4} = -0.25,$$
$$q_{0.5} = 0.00,\ q_{0.6} = 0.25,\ q_{0.7} = 0.52,\ q_{0.8} = 0.84,\ q_{0.9} = 1.28$$

26. Dane podatke

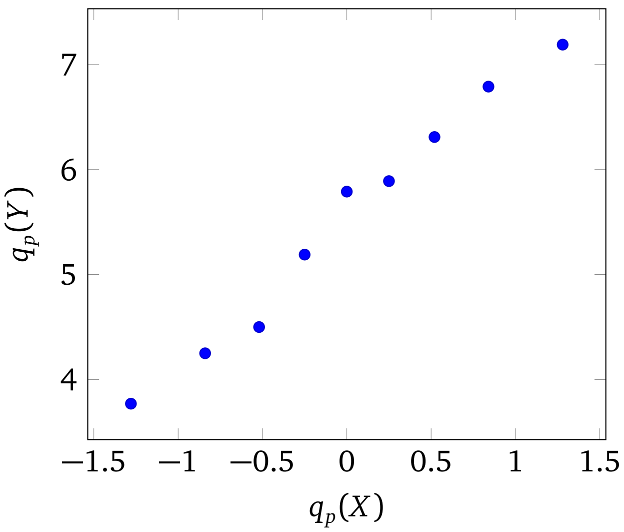
7.19, 6.31, 5.89, 4.5, 3.77, 4.25, 5.19, 5.79, 6.79

primerjaj z normalno porazdelitvijo, tako da narišeš kvantilni grafikon.

Rešitev:

Naj bo $X \sim N(0, 1)$ in $Y \sim \text{Emp}\,(7.19, \ldots, 6.79)$. Ker je Y empirično podana, rakunamo kvantile le za $p = \frac{k}{n+1} = \frac{k}{10}$.

p	$q_p(X)$	$q_p(Y)$
0.1	−1.28	3.77
0.2	−0.84	4.25
0.3	−0.52	4.5
0.4	−0.25	5.19
0.5	0.00	5.79
0.6	0.25	5.89
0.7	0.52	6.31
0.8	0.84	6.79
0.9	1.28	7.19



27. Podana je višina naključno izbranih očetov in njihovih sinov

oče	165	175	181	178	174	182	178
sin	168	177	190	178	173	186	185

Izračunaj Pearsonov korelacijski koeficient.

Rešitev:

Pearsonov korelacijski koeficient med dvema slučajnima spremnljivkama X in Y je

$$r_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \cdot \sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

očetje (X_i)	sinovi (Y_i)	$X_i - \overline{X}$	$Y_i - \overline{Y}$
165	168	-11	-12
175	177	-1	-3
181	190	5	10
178	178	2	-2
174	173	-2	-7
182	186	6	6
178	185	2	5

$\overline{X} = 176.14 \approx 176$, $\overline{Y} = 179.57 \approx 180$, $\sum(X_i - \overline{X})^2 = 195$, $\sum(Y_i - \overline{Y})^2 = 367$ $\sum(X_i - \overline{X})(Y_i - \overline{Y}) = 24.1$

$r_{X,Y} = 0.90$, kar pomeni, da je povezanost visoka.

28. V kavarni spremljajo, kako zunanja temperatura vpliva na dnevni zaslužek:

T v °C	23	25	27	32	30	29	30
dnevni zaslužek v EUR	190	210	260	290	350	290	300

Izračunaj Pearsonov in Spermanov korelacijski koeficient med temperaturo X in dnevnim zaslužkom Y.

Rešitev:

X_i	Y_i	$X_i - \overline{X}$	$Y_i - \overline{Y}$
23	190	-5	-80
25	210	-1	-60
27	260	-1	-10
32	290	4	20
30	350	2	80
29	290	1	20
30	300	2	30

$\overline{X} = 28$, $\overline{Y} = 270$

Pearsonov korelacijski koeficient:

$$r_{X,Y} = \frac{910}{\sqrt{60 \cdot 18200}} = 0.87$$

Spermanov korelacijski koeficient

$$\rho_{X,Y} = r_{R(X),R(Y)} = \frac{\sum_{i=1}^{n}(R(X_i) - \overline{R})(R(Y_i) - \overline{R})}{\sqrt{\sum_{n=1}^{n}(R(X_i) - \overline{R})^2 \cdot \sum_{n=1}^{n}(R(Y_i) - \overline{R})^2}},$$

kjer je $\overline{R} = \frac{n+1}{2}$.

X_i	Y_i	$R(X_i)$	$R(Y_i)$	$R(X_i) - \overline{R}$	$R(Y_i) - \overline{R}$
23	190	1	1	-3	-3
25	210	2	2	-2	-2
27	260	3	3	-1	-1
32	290	7	4.5	3	0.5
30	350	5.5	7	1.5	3
29	290	4	4.5	0	0.5
30	300	5.5	6	1.5	2

Spermanov korelacijski koeficient

$$\rho_{X,Y} = \frac{23}{\sqrt{27.5 \cdot 27.5}} = 0.84$$

29. Za dane podatke

X_i	1	2	3	4	5	6
Y_i	5	4	3	2	1	15

izračunaj Pearsonov in Spermanov korelacijski koeficient.

Rešitev:

X_i	Y_i	$X_i-\overline{X}$	$Y_i-\overline{Y}$	$R(X_i)$	$R(Y_i)$	$R(X_i)-\overline{R}$	$R(Y_i)-\overline{R}$
1	5	-2.5	0	1	5	-2.5	1.5
2	4	-1.5	-1	2	4	-1.5	0.5
3	3	-0.5	-2	3	3	-0.5	-0.5
4	2	0,5	-3	4	2	0.5	-1.5
5	1	1.5	-4	5	1	1.5	-2.5
6	15	2.5	10	6	6	2.5	2.5

$\overline{X}=3.5$, $\overline{Y}=5$, $\overline{R}=3.5$

Pearsonov korelacijski koeficient $r_{X,Y}=0.42$

Spermanov korelacijski koeficient $\rho_{X,Y}=-0.14$

30. Izračunaj Pearsonov in Spermanov korelacijski koeficient podatkov podanih s kontingenčno tabelo

X \ Y	1	2	3
2	8	1	0
3	2	1	0
5	2	4	2

Rešitev:

Izpišemo podatke za vsako spremenljivko posebaj.

X_i	f_i	$F(X_i)$	$X_i-\overline{X}$	$R(X_i)$	$R(X_i)-\overline{R}$
2	9	9	-1.5	5	-5.5
4	3	12	0.5	11	0.5
5	8	20	1.5	16.5	5.5

Y_i	f_i	$F(Y_i)$	$Y_i-\overline{Y}$	$R(Y_i)$	$R(Y_i)-\overline{R}$
1	12	12	-0.5	6.5	-4
2	6	18	0.5	15.5	5
3	2	20	1.5	19.5	9

$\overline{X}=3.5, \overline{Y}=1.5, \sum(X_i-\overline{X})^2=39, \sum(Y_i-\overline{Y})^2=9,$

$\sum(X_i-\overline{X})(Y_i-\overline{Y})=8\cdot(-1.5)(-0.5)+1\cdot(-1.5)(0.5)+0+\cdots+2(1.5)(1.5)=11$

Pearsonov korelacijski koeficient $r_{X,Y}=0.59$.

Spremanov korelacijski koeficient $\rho_{X,Y}=\dfrac{327}{\sqrt{561\cdot 504}}=0.61$

31. Podane imamo povprečne dnevne temperature delavnih dni v dveh zaporenih tednih.

    X : 5, 7, 8, 3, 2

    Y : 2, 2, 3, 11

    Kateri od tednov je bil toplejši?

    Rešitev:

    Točkovni biserialni koeficient

    Kadar želimo primerjati dve slučajni spremenljivki, ki sta podani na različnih prostorih, izračunamo točkovni biserialni koeficient r_{pb}. To je Pearsonov ko0relacijski koeficient ,ed združeno intevalsko spremenljivko in numerično vrednostjo skupine. Izbrani skupini priredimo višjo vrednost.

    Spremenljivki $x_1, \dots x_m$ in $y_1, \dots, y_n$ zdražimo v spremenljivko $u_1, u_2, \dots, u_{m+n}$ in definiramo $g_1, \dots, g_{m+n}$, kjer je $g_i = a$ za $i = 1, \dots, m$ in $g_i = b$ za $i = m+1, \dots, m+n$. Točkovni biserialni koeficient v korist prve skupine je

$$r_{pb} = r_{U,G} = \frac{K_{U,G}}{\sigma_U \sigma_G}$$

    za $a > b$.

    Če vzamemo $a = n$ in $b = -m$, dobimo

$$\begin{aligned}
E(G) &= n\frac{m}{n+m} - m\frac{n}{m+n} = 0 \\
\sigma_G^2 &= E(G^2) - E(G)^2 = E(G^2) = \frac{mn^2 + nm^2}{m+n} = mn \\
K_{U,G} &= E\left[(U - E(U))(G - E(G))\right] = E(UG) - E(U)E(G) = E(UG) = \\
&= \frac{n(x_1 + \cdots + x_m) - m(y_1 + \cdots + y_n)}{m+n} = \frac{mn}{m+n}(\overline{X} - \overline{Y})
\end{aligned}$$

$$r_{pb} = \frac{mn}{m+n}(\overline{X} - \overline{Y})\frac{1}{\sigma_U \sqrt{mn}} = \frac{\overline{X} - \overline{Y}}{\sigma_U} \cdot \frac{\sqrt{mn}}{m+n}$$

    Vemo še

$$\overline{U} = \frac{m}{m+n}\overline{X} + \frac{n}{m+n}\overline{Y}$$

    in

$$\sigma_U^2 = \underbrace{\frac{m}{m+n}\sigma_X^2 + \frac{n}{m+n}\sigma_Y^2}_{\sigma_W^2} + \underbrace{\frac{mn}{(m+n)^2}(\overline{X} - \overline{Y})^2}_{\sigma_B^2},$$

    kjer je σ_W^2 nepojasnjena varianca ozr. residualna varianca (uteženo povprečje varianc znotraj skupin) in σ_B^2 pojasnjena varianca (varianca mes skupinami. Opazimo, da je $\sigma_B^2 = r_{pb}^2 \sigma_U^2$.

    V našem vremenskem primeru je $\overline{X} = 5$ in $\overline{Y} = 4.5$, $\sigma_X^2 = \frac{26}{5}$ in $\sigma_Y^2 = \frac{57}{4}$. Za spremenljivko U velja

$\overline{U} = \frac{4}{9} \cdot 5 + \frac{5}{9} \cdot 92 = \frac{85}{18}$

$\sigma_U^2 = \frac{5}{9} \cdot \frac{26}{5} + \frac{4}{9} \cdot \frac{57}{4} + \frac{4 \cdot 5}{9^2}(5 - 4.5)^2 = 9.284$

$r_{pb} = 0.0815.$

32. Primerjamo rezultate dveh kolokvijev:

a : 22, 52, 39, 52, 43

b : 39, 73, 34, 34, 38, 53, 59

S točkovnim biserialnim koeficientom oceni, kateri kolokvij so piali boljše.

Rešitev:

$\overline{a} = 41.6, \overline{b} = 47.14$ Ker je povprečje drugega kolokvija višje, je slučajna spremenljivka G na drugem kolokviju pozitivna.

Definiramo novi spremenljivki

U:	39,	73,	34,	34,	38,	53,	59,	22,	52,	39,	52,	43
G:	5,	5,	5,	5,	5,	5,	5,	-7,	-7,	-7,	-7,	-7

Tedaj je $\overline{U} = \frac{5\overline{a}+7\overline{b}}{5+7} = 44.83$ in $\sigma_U = \sqrt{\frac{\sum(U_i - \overline{U})^2}{12}} = 13.03$. Ker je

$$r_{pb} = \frac{\sqrt{35}(47.14 - 41.6}{12 \cdot 13.03} = 0.2037,$$

so rezultati drugega kolokvija rahlo boljši od prvega kolokvija.

33. Gledalce so povpraševali, katero zvrst filma imajo najraje in kako pogosto so obiskali kino v zadnjem mesecu. Rezultati so podani v tabeli

	zvrst	0	1	2
X	komedija	4	2	2
Y	akcija	0	1	0
Z	romantični	0	3	1
W	drama	4	1	2
T	grozljivka	0	0	0

Preveri vpliv najljubše zvrsti filma na pogostost zahajanja v kino.

Rešitev:

Če primerjamo več spremenljivk

$$X_1, \ldots, X_n$$
$$Y_1, \ldots, Y_m$$
$$Z_1, \ldots, Z_k$$

jih združimo v eno spremenljivko

$U: X_1, \ldots, X_n, Y_1, \ldots, Y_m, Z_1, \ldots, Z_k$.

Tedaj je

$$\overline{U} = \frac{n}{n+m+k}\overline{X} + \frac{m}{n+m+k}\overline{Y} + \frac{k}{n+m+k}\overline{Z}$$

in

$$\sigma_U^2 = \underbrace{\frac{n}{n+m+k}\sigma_X^2 + \frac{m}{n+m+k}\sigma_Y^2 + \frac{k}{n+m+k}\sigma_Z^2}_{\sigma_W^2} +$$

$$+ \underbrace{\frac{n}{n+m+k}(\overline{X}-\overline{U})^2 + \frac{m}{n+m+k}(\overline{Y}-\overline{U})^2 + \frac{k}{n+m+k}(\overline{Z}-\overline{U})^2}_{\sigma_B^2}.$$

Opazujemo kvocient $\frac{\sigma_B^2}{\sigma_U^2}$.

Najprej izračunamo posamezna povprečja $\overline{X} = \frac{3}{4}$, $\overline{Y} = 1$, $\overline{Z} = \frac{5}{4}$, $\overline{W} = \frac{5}{7}$, $\overline{T} = 0$.

Tedaj je

$$\overline{U} = \frac{8 \cdot 3/4 + 1 \cdot 1 + 4 \cdot 5/4 + 7 \cdot 5/7 + 0 \cdot 0}{20} = \frac{17}{20}$$

$$\sigma_B^2 = \frac{8}{20}\left(\frac{3}{4} - \frac{17}{20}\right)^2 + \frac{1}{20}\left(1 - \frac{17}{20}\right)^2 + \frac{4}{20}\left(\frac{5}{4} - \frac{17}{20}\right)^2 + \frac{7}{20}\left(\frac{5}{7} - \frac{17}{20}\right)^2 + 0 = 0.0436$$

$$\sigma_U^2 = \frac{8 \cdot (0 - \frac{17}{20})^2 + 7 \cdot (1 - \frac{17}{20})^2 + 5 \cdot (2 - \frac{17}{20})^2}{20} = 0.6275$$

Ker je $\frac{\sigma_B^2}{\sigma_U^2} = 0.069$, je povezanost nizka. Torej ima najljubša zvrst malo vpliva na obisk kina.

34. Dan je vzorec:

$$30,\ 75,\ 50,\ 60,\ 20,\ 15,\ 90,\ 45.$$

a) Izračunaj nepristranski cenilki za pričakovano vrednost in varianco.
b) Izračunaj cenilki za prvi kvartil in $q_{5/9}$.

Rešitev:

a) Cenilka neke količine je nepristranska, če je enaka pričakovani vrednosti količine.
$\widehat{\mu} = \overline{x} = 49.4$
$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = 621.5$

b) Cenilke za kvantile: $h = (n-1)\alpha + 1, k = \lfloor h \rfloor \quad \Rightarrow \quad \widehat{q}_\alpha = x_{(k)} + (h-k)(x_{(k+1)} - x_{(k)})$.

Podatke najprejprej razporedimo v ranžirno vrsto:

$$15,\ 20,\ 30,\ 45,\ 50,\ 60,\ 75,\ 90.$$

$q_{1/4} = x_{(3)} = 30$, $\widehat{q}_{1/4} = x_{(3)} + 0 = 30$
$q_{5/9} \in [50, 60]$, $\widehat{q}_{5/9} = x_{(5)} + 0.\overline{4} \cdot (x_{(6)} - x_{(5)}) = 54.\overline{4}$

35. Naj bo X slučajna spremenljivka porazdeljena enakomerno na intervalu $[0, a]$. Po metodi momentov poišči cenilko parametra a, če imamo podan vzorec slučajne spremenljivke 1, 2, 1, 3, 10.

Rešitev:

$X \sim U[0, a]$:

gostota enakomerno porazdeljene slučajne spremenljivke $f_X(x) = \begin{cases} \frac{1}{a} & ; 0 \le x \le a \\ 0 & ; \text{sicer} \end{cases}$

	Momenti	cenilke
1. moment:	$m_1 = E(X)$	$\widehat{m_1} = \overline{X}$
2. moment:	$m_2 = E(X^2)$	$\widehat{m_2} = \frac{x_1^2 + \cdots + x_n^2}{n}$
$\vdots$		$\vdots$
k-ti moment:	$m_k = E(X^k)$	$\widehat{m_k} = \frac{x_1^k + \cdots + x_n^k}{n}$

Ker je $E(X) = \frac{a}{2}$, lahko izrazimo a s prvim momentom $a = 2E(X) = 2m_1$. Tedaj je cenilka $\widehat{a} = 2\overline{X} = 6.8$. Pri takem a ne bi mogli dobiti v vzorcu število 10.

Zato preverimo še z višjimi moment.

$E(X^2) = \frac{a^2}{3} \quad \Rightarrow \quad \widehat{a} = \sqrt{2\widehat{m_2}} = 8.31$

$E(X^3) = \frac{a^4}{5} \quad \Rightarrow \quad \widehat{a} = \sqrt[4]{5\widehat{m_4}} = 10.025$

36. Naj bo porazdelitev slučajne spremenljivke X enaka $X \sim \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1-2a-b & a & b & a \end{pmatrix}$. Po metodi momentov poišči cenilki za parametra a in b in ju oceni s pomočjo vzorca: -1, 0, 1, 1, 0, 0, -1, 2, 1, 2.

Rešitev:

Izračunajmo prva dva momenta: $m_1 = -1 + 4a + 2b$ in $m_2 = 1 + 2a$.

Tedaj sta cenilki za parametra enaki $\widehat{a} = \dfrac{\widehat{m_2} - 1}{2}$ in $\widehat{b} = \dfrac{\widehat{m_1} - 2\widehat{m_2} + 3}{2}$.

Iz vzorca izračunamo $\widehat{m_1} = \frac{1}{2}$ in $\widehat{m_2} = \frac{13}{10}$ in tako dobimo oceni $\widehat{a} = \frac{3}{20}$ in $\widehat{b} = \frac{9}{20}$.

37. Za slučajno spremnljivko iz naloge 36 poišči cenilki parametrov a in b še po metodi največjega verjetja.

    Rešitev:

    Metoda največjega verjetja

    Za vzorec $X_1, X_2, \ldots, X_n$ slučajne spremenljivke X definiramo funkcijo verjetja

    $$\ell(a) = f_X(X_1|a) \cdot f_X(X_2|a) \cdots f_X(X_n|a)$$

    in poiščemo njen maksimum. Zaradi lažjega računanja lahko iščemo maksimum funkcije

    $$\ln(\ell) = \ln f_X(X_1|a) + \ln f_X(X_2|a) + \cdots + \ln f_X(X_n|a).$$

    $\ell(a,b) = (1-2a-b)^2 a^5 b^3$, kjer je $a \geq 0$, $b \geq 0$ in $2a+b \leq 1$. Funkcija je nenegativna in na robu ničelna, zato bo v stacionarni točki dosegla maksimum. Izračunamo prva parcialna odvoda

    $$\frac{\partial \ell}{\partial a} = a^4 b^3 (1-2a-b)(5-14a-5b)$$
    $$\frac{\partial \ell}{\partial b} = a^5 b^2 (1-2a-b)(3-6a-5b)$$

    in poiščemo rešitve sistema $\frac{\partial \ell}{\partial a} = 0$ in $\frac{\partial \ell}{\partial b} = 0$. Tako dobimo oceni $\widehat{a} = \frac{1}{4}$ in $\widehat{b} = \frac{3}{10}$.

38. Za normalno porazdeljeno slučajno spremenljivko $X \sim N(\mu, \sigma)$ poišči cenilki za parametra μ in σ

    a) po metodi momentov.
    b) po metodi največjega verjetja.

    Rešitev:

    a) Metoda momentov:

       Vemo, da je $m_1 = E(X) = \mu$ in $m_2 = E(X^2) = \sigma^2 + \mu^2$. Zato je $\widehat{\mu} = \widehat{m}_1$ in $\widehat{\sigma}^2 = \widehat{m}_2 - \widehat{m}_1^2$.

    b) Metoda največjega verjetja

       Funkcija največjega verjetja je

       $$\ell(\mu, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right) e^{-\frac{1}{2\sigma^2}\left((x_1-\mu)^2 + \cdots + (x_n-\mu)^2\right)},$$

       kjer je $\mu \in \mathbb{R}$ in $\sigma \in (0, \infty)$. Ko se vrednosti približujejo robovom območja, gre funkcija verjetja proti 0. Funkcijo ℓ logaritmiramo

       $$\ln \ell(\mu, \sigma) = -n \ln \sqrt{2\pi} - n \ln \sigma - \frac{1}{2\sigma^2}\left((x_1-\mu)^2 + \cdots + (x_n-\mu)^2\right)$$

in izračunamo prva parcialna odvoda:

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2}\left((x_1-\mu)+\cdots+(x_n-\mu)\right)$$
$$\frac{\partial \ell}{\partial \sigma} = -\frac{n}{\sigma}+\frac{1}{\sigma^3}\left((x_1-\mu)^2+\cdots+(x_n-\mu)^2\right).$$

Za stacionarno točko mora veljati $\mu = \frac{x_1+\cdots+x_n}{n} = m_1$ in $\sigma^2 = \frac{x_1^2+\cdots+x_n^2}{n} - \mu^2$. Vidimo, da sta tako dobljeni cenilki $\widehat{\mu} = \widehat{m}_1$ in $\widehat{\sigma}^2 = \widehat{m}_2 - \widehat{m}_1^2$ enaki kot cenilki dobljeni po metodi momentov.

39. Za log-normalno porazdeljeno slučajno spremenljivko $X \sim \log N(\mu,\sigma)$ poišči cenilki za parametra μ in σ

    a) po metodi momentov.
    b) po metodi največjega verjetja.

Rešitev:

Slučajna spremenljivka X je porazdeljena log- normalno, če je $Y := \ln X \sim N(\mu,\sigma)$. Tedaj je $Z := \frac{Y-\mu}{\sigma} \sim N(0,1)$. Torej je $X = e^Y = e^{\sigma Z+\mu}$.

a) Metoda momentov

$$m_1 = E(X) = E(e^{\sigma Z+\mu}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sigma z+\mu}\cdot e^{-\frac{z^2}{2}}\,dz =$$
$$= \frac{e^{\mu+\frac{\sigma^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma)^2}\,dz = e^{\mu+\frac{\sigma^2}{2}}$$
$$m_2 = E(X^2) = E(e^{2\sigma Z+2\mu}) = e^{2\mu+2\sigma^2}$$

Če logaritmiramo obe enačbi, dobimo $\ln m_1 = \mu + \frac{\sigma^2}{2}$ in $\ln m_2 = 2\mu + 2\sigma^2$. Od tod lahko izračunamo

$$\mu = 2\ln m_1 - \frac{1}{2}\ln m_2 \quad \text{in} \quad \sigma^2 = \ln m_2 - 2\ln m_1$$

b) Metoda največjega verjetja
Najprej poiščimo gostoto log-normalno porazdeljene slučajne spremenljivke.

$$F_X(x) = P(X \le x) = P(e^{\sigma Z+\mu} \le x) = P\left(Z \le \frac{\ln x-\mu}{\sigma}\right) = \Phi\left(\frac{\ln x-\mu}{\sigma}\right)$$
$$f_X(x) = F_X'(x) = \Phi'\left(\frac{\ln x-\mu}{\sigma}\right)\cdot\left(\frac{\ln x-\mu}{\sigma}\right)' = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}\cdot\frac{1}{x\sigma}$$

Tako je funkcija verjetja enaka

$$\ell = \frac{1}{\sqrt{2\pi}^n\sigma^n x_1 x_2\cdots x_n}\,e^{-\frac{1}{2\sigma^2}\left((\ln x_1-\mu)^2+\cdots+(\ln x_n-\mu)^2\right)}$$
$$\ln \ell = -\frac{1}{2\sigma^2}\left((\ln x_1-\mu)^2+\cdots+(\ln x_n-\mu)^2\right) - \frac{n}{2}\ln 2\pi - n\ln\sigma - \ln x_1\cdots x_n$$

Izračunamo oba parcialna odvoda

$$\frac{\partial \ln \ell}{\partial \mu} = \frac{1}{\sigma^2}\left((\ln x_1 - \mu) + \cdots (\ln x_n - \mu)\right)$$
$$\frac{\partial \ln \ell}{\partial \sigma} = \frac{1}{\sigma^3}\left((\ln x_1 - \mu)^2 + \cdots + (\ln x_n - \mu)^2\right) - \frac{n}{\sigma}$$

Parcialna odvoda sta enaka nič, če velja

$$\mu = \frac{\ln x_1 + \cdots + \ln x_n}{n} \quad \text{in} \quad \sigma^2 = \frac{(\ln x_1 - \mu)^2 + \cdots + (\ln x_n - \mu)^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i)^2 - \mu^2$$

40. Statistična spremenljivka je porazdeljena normalno. Vrednosti na vzorcu so:

$$101, 91, 93, 103, 91, 101, 103, 95, 95.$$

Določi 95% interval zaupanja za μ, če je

a) $\sigma = 5$.
b) σ neznan.

Rešitev:

a) Interval zaupanja za μ pri normalni porazdelitvi z znanim standardnim odklonom σ
$$\overline{X} - Z_{1-\frac{\alpha}{2}} SE < \mu < \overline{X} - Z_{1-\frac{\alpha}{2}} SE,$$
kjer je $SE = \frac{\sigma}{\sqrt{n}}$ in $Z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1 - \frac{\alpha}{2})$.
$\overline{X} = 97, SE = \frac{5}{3}, Z_{0.975} = 1.96 \quad \Rightarrow \quad 93.73 < \mu < 100.27$

b) Interval zaupanja za μ pri normalni porazdelitvi z neznanim standardnim odklonom
$$\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\widehat{SE} < \mu < \overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\widehat{SE},$$
kjer je $\widehat{SE} = \frac{\widehat{\sigma}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ in $t_{1-\frac{\alpha}{2}}(n-1)$ kvantil Studentove porazdelitve z $n-1$ prostostnimi stopnjami.
$\overline{X} = 97, \widehat{SE} = \frac{5}{3}, t_{0.975} = 2.31 \quad \Rightarrow \quad 93.15 < \mu < 100.85$

41. Anica vsako pomlad kupi sadike in semena za svoj vrt. Za to je v zadnjih letih porabila

115, 95, 150, 89, 73, 93, 110 in 124 EUR.

Poišči interval zaupanja za povprečje denarja, ki ga potrebuje za zasaditev vrta, pri stopnji tveganja $\alpha = 0.05$, če predpostaviš, da je strošek porazdeljen normalno. Izračunaj še interval zaupanja, če je $\alpha = 0.1$.

Rešitev:

$\overline{X} = 106.125, \widehat{SE} = 8.48, t_{0.975}(7) = 2.36, t_{0.95}(7) = 1.89$

$\alpha = 0.05 \quad \Rightarrow \quad 86.12 < \mu < 126.13$

$\alpha = 0.1 \quad \Rightarrow \quad 90.10 < \mu < 122.15$

42. V drevesnici spremljajo rast dreves v enem letu. Sadike, ki so jih spremljali so zrastle za

20, 35, 44, 18, 27, 32 in 27 centimetrov.

Predpostavimo, da je letni prirast porazdeljen normalno. Poišči interval zaupanja za standardni odklon letnega prirastka pri stopnji tveganja $\alpha = 0.05$.

Rešitev:

Inteval zaupanja za standardni odklon normalne porazdelitve je

$$\sigma \in \left(\frac{\widehat{\sigma}\sqrt{n-1}}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}, \frac{\widehat{\sigma}\sqrt{n-1}}{\chi^2_{\frac{\alpha}{2}}(n-1)} \right),$$

kjer je $\widehat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ in $\chi_\alpha(n)$ kvantil porazdelitve hi-kvadrat z n prostosnimi stopnjami.

$\overline{X} = 29, \widehat{\sigma} = \sqrt{80}, \chi^2_{0.025}(6) = 1.2373442, \chi^2_{0.975}(6) = 14.449375 \quad \Rightarrow \quad \sigma \in (5.76, 19.7)$

43. Dvajsetkrat ponovimo isti poskus. Določi Waldov interval zaupanja za delež uspešnih poskusov pri $\alpha = 0.05$, če

a) so 4 poskusi uspeli;

b) ni uspel noben poskus;

c) so bili uspešni vsi poskusi.

Rešitev:

Waldov interval zaupanja za θ, kjer je $X \sim$ Bernoulli (θ), je za n ponovitev poskusa in S uspešnih poskusov enak

$$\widehat{\theta} - Z_{1-\frac{\alpha}{2}}\widehat{SE} < \theta < \widehat{\theta} + Z_{1-\frac{\alpha}{2}}\widehat{SE},$$

kjer je $\widehat{\theta} = \frac{S}{n}$, $\widehat{SE} = \sqrt{\frac{\widehat{\theta}(1-\widehat{\theta})}{n}}$ in Z_α kvantil standardne normalne porazdelitve.

a) $\widehat{\theta} = \frac{4}{20} = 0.2, \widehat{SE} = \sqrt{0.008}, Z_{0.975} = 1.96 \quad \Rightarrow \quad \theta \in (0.0528, 0.3472)$

b) $\theta \in \{0\}$

c) $\theta \in \{1\}$

44. Izpelji Wilsonov interval zaupanja za delež uspešnih poskusov.

Rešitev:

Uporabimo oceno

$$\widehat{\theta} - Z\sqrt{\frac{\theta(1-\theta)}{n}} < \theta < \widehat{\theta} + Z\sqrt{\frac{\theta(1-\theta)}{n}},$$

kjer pišemo Z namesto $Z_{1-\frac{\alpha}{2}}$. Tedaj velja neenakost

$$|\theta-\widehat{\theta}|<Z\sqrt{\frac{\theta(1-\theta)}{n}}.$$

Ko kvadriramo, dobimo kvadratno neenačbo za θ:

$$\theta^2\left(1+\frac{Z^2}{n}\right)+\theta\left(-2\widehat{\theta}-\frac{Z^2}{n}\right)+\widehat{\theta}^2<0.$$

Ker je diskriminanta $D=\left(-2\widehat{\theta}-\frac{Z^2}{n}\right)^2-4\theta^2(1+\frac{Z^2}{n})=4\frac{Z^2}{n}\widehat{\theta}(1-\widehat{\theta})+\frac{Z^4}{n^2}$ vedno pozitivna, ima kvadratna funkcija dve ničli

$$\theta_{1,2}=\frac{2n\widehat{\theta}+Z^2\pm\sqrt{4nZ^2\widehat{\theta}(1-\widehat{\theta})+Z^4}}{2n+2Z^2},$$

interval med njima pa je Wilsonov interval zaupanja.

45. Poišči Waldov in Wilsonov interval zaupanja za

 a) $n=10, S=2, \alpha=0.05$.
 b) $n=10000, S=5048, \alpha=10\%$

Rešitev:

 a) Waldov interval zaupanja: $-0.05<\theta<0.45$
 Wilsonov interval zaupanja: $0.06<\theta<0.51$
 b) Waldov interavl zaupanja: $0.4956761<\theta<0.5130239$
 Wilsonov interval zaupanja: $0.4965759<\theta<0.5130215$

46. Naj bo X normalno porazdeljena slučajna spremnljivka s standardnim odklonom $\sigma=5$ S pomočjo vzorca

101, 91, 93, 103, 91, 101, 103, 95, 95

preveri domnevo, da je $\mu=100$, če je stopnja tveganja $\alpha=0.05$ in alternativna domneva $H_1^{\pm}=\mu\neq 100, H_1^{+}:\mu>100$ ali $H_1^{-}:\mu<100$.

Rešitev:

Naj bo $X\sim N(\mu,\sigma)$, kjer je σ podan. Preizkušamo domnevo $H_0:\mu=\mu_0$. Tedaj je testna funkcija $Z=\frac{\overline{X}-\mu_0}{SE}$ porazdeljena standardno normalno, kjer je $SE=\frac{\sigma}{\sqrt{n}}$.

$H_1^{\pm}:\ \mu\neq\mu_0 \quad\Rightarrow\quad H_0$ zavrnemo, če je $|Z|>Z_{1-\frac{\alpha}{2}}$

$H_1^{+}:\ \mu>\mu_0 \quad\Rightarrow\quad H_0$ zavrnemo, če je $Z>Z_{1-\alpha}$

$H_1^{-}:\ \mu<\mu_0 \quad\Rightarrow\quad H_0$ zavrnemo, če je $Z<-Z_{1-\alpha}$

Iz podatkov izračunamo testno funkcijo $Z=\frac{97-100}{5/\sqrt{9}}=-1.8$.

$H_1^{\pm}: \mu \neq 100 \quad \Rightarrow \quad$ ker $|Z| = 1.8 < Z_{0.975} = 1.96, H_0$ ne zavrnemo.

$H_1^{+}: \mu > 100 \quad \Rightarrow \quad$ ker je $-1.8 < Z_{0.95} = 1.645, H_0$ ne zavrnemo.

$H_1^{+}: \mu < 100 \quad \Rightarrow \quad$ ker je $-1.8 < -Z_{0.95} = -1.645, H_0$ zavrnemo.

47. Dani podatki

$$105,\ 107,\ 95,\ 121,\ 107,\ 97,\ 111,\ 121,\ 95$$

so porazdeljeni normalno.

a) Preveri ničelno domnevo $H_0 : \mu = 100$ proti nasprotni domnevi $H_1 : \mu \neq 100$ pri stopnji tveganja $\alpha = 0.05$.

b) Ponovno preveri ničelno domnevo $H_0 : \mu = 100$ proti nasprotni domnevi $H_1 : \mu \neq 100$, če veš, da je standardni odklon $\sigma = 10$.

Rešitev:

Naj bo $X \sim N(\mu, \sigma)$, kjer σ ni znan. Preizkušamo domnevo $H_0 : \mu = \mu_0$. Tedaj je testna funkcija

$$T = \frac{\overline{X} - \mu_0}{\widehat{SE}} \sim \text{Student } (n-1),$$

kjer je $\widehat{SE} = \frac{\widehat{\sigma}}{\sqrt{n}}$ in $\widehat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$.

$H_1^{\pm}: \mu \neq \mu_0 \quad \Rightarrow \quad H_0$ zavrnemo, če je $|T| > t_{1-\frac{\alpha}{2}}(n-1)$

$H_1^{+}: \mu > \mu_0 \quad \Rightarrow \quad H_0$ zavrnemo, če je $T > t_{1-\alpha}(n-1)$

$H_1^{-}: \mu < \mu_0 \quad \Rightarrow \quad H_0$ zavrnemo, če je $T < -t_{1-\alpha}(n-1)$

a) Izračunamo $\overline{x} = 107, \widehat{\sigma} = 10, \widehat{SE} = \frac{10}{\sqrt{9}}4$ in testno funkcijo $T = 2.1$. Ker je $t_{0.975}(8) = 2.3060041 > |T|$, H_0 ne zavrnemo.

b) Če je σ znan, je testna funkcija $Z = 2.1$. Ker je $Z_{0.975} = 1.96 < |Z|$, H_0 zavrnemo.

48. Tovarna jamči, da je delež izdelkov z napako enak 20%.

a) V vzorcu 100 izdelkov je 24 izdelkov z napako. Pri stopnji tveganja $\alpha = 0.05$ preizkusi domnevo, da je verjetnost, da ima izdelek napako, enaka 0.2, proti alternativni domnevi, da je večja od 0.2

b) Kaj pa če je med 500 izdelki 120 okvarjenih?

c) Izračunaj mejno stopnjo tveganja, oziroma p-vrednost testa.

Rešitev:

Naj med n ponovljenimi poskusi S poskusov uspe. Preizkušamo domnevo $H_0 : \theta = \theta_0$. Tedaj je testna funkcija

$$Z = \frac{\widehat{\theta} - \theta_0}{SE} \sim N(0,1),$$

kjer je $SE = \sqrt{\frac{\theta_0(1-\theta_0)}{n}}$ in $\widehat{\theta} = \frac{S}{n}$.

$$
\begin{aligned}
H_1^{\pm}: \ \theta \neq \theta_0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } |Z| > Z_{1-\frac{\alpha}{2}} \\
H_1^{+}: \ \theta > \theta_0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } Z > Z_{1-\alpha} \\
H_1^{-}: \ \theta < \theta_0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } Z < -Z_{1-\alpha}
\end{aligned}
$$

a) Izračunamo $\widehat{\theta} = 0.24, SE = 0.04$ in testno funkcijo $Z = 1 < Z_{0.95} = 1.645$. Zato H_0 ne zavrnemo.
b) Kot prej je $\widehat{\theta} = 0.24$, a ker je n večji, je $Z = 2.24 > 1.645 = Z_{0.95}$ in H_0 zavrnemo.
c) Iščemo $\alpha(p)$, da je $Z = Z_{1-\alpha}$.
Ker je $Z = 2.24$, je $1-\alpha = 0.9875$ in $\alpha = 0.0125$.

49. Lastnik kavarne spremlja prodajo kave in slasdoleda

kava	63	71	65	67	62	75
sladoled	35	34	51	72	73	64

Predpostavi, da je prodaja kave in sladoleda porazdeljena normalno. Ali lahko pri stopnji značilnosti $\alpha = 0.05$ pokaže, da je povprečno število prodanih kav večje od povprečnega števila prodanih sladoledov? Kaj pa, če je stopnja značilnosti $\alpha = 0.01$?

Rešitev:

Preizkus enakosti sredin za parne vzorce

Preizkušamo domnevo $H_0 : \mu_X = \mu_Y$ nasproti alternativni domnevi $H_1^{\pm} : \mu_X \neq \mu_Y, H_1^{+} : \mu_X > \mu_Y$ ali $H_1^{-} : \mu_X < \mu_Y$. Razlika slučajnih spremenljivk $X - Y$ je tudi porazdeljena normalno. zato nalogo pretvorimo na preizkušanje domneve $H_0 : \mu_X - \mu_Y = 0$ nasproti alternativni domnevi $H_1^{\pm} : \mu_X - \mu_Y \neq 0$, $H_1^{+} : \mu_X - \mu_Y > 0$ ali $H_1^{-} : \mu_X - \mu_Y < 0$.

Za testno funkvijo vzamemo

$$T = \frac{\overline{X} - \overline{Y}}{\widehat{SE}} \sim \text{Student } (n-1),$$

kjer je $\widehat{SE} = \frac{\widehat{\sigma_{X_Y}}}{\sqrt{n}}$ in $\widehat{\sigma_{X-Y}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - y_i - \overline{x} - \overline{y})^2}$.

$$
\begin{aligned}
H_1^{\pm}: \ \mu_X - \mu_Y \neq 0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } |T| > t_{1-\frac{\alpha}{2}}(n-1) \\
H_1^{+}: \ \mu_X - \mu_Y > 0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } T > t_{1-\alpha}(n-1) \\
H_1^{-}: \ \mu_X - \mu_Y < 0 \quad &\Rightarrow \quad H_0 \text{ zavrnemo, če je } T < -t_{1-\alpha}(n-1)
\end{aligned}
$$

Opazujemo razliko podatkov 28, 37, 14, −5, −11, 11. Preizkušamo domnevo $H_0 : \mu_X - \mu_Y = 0$ nasproti domnevi $H_1^{+} : \mu_X - \mu_Y > 0$.

$\overline{X} - \overline{Y} = 12.3, \widehat{\sigma} = 18.5, T = 1.63$

a) Za $\alpha = 0.05$ je $t_{0.95}(5) = 2.0150 > T$, zato $H_=$ ne zavrnemo.
b) Za $\alpha = 0.01$ je $t_{0.99}(5) = 3.3649 > T$, zato $H_=$ ne zavrnemo.