VAJE IZ OSNOV VERJETNOSTI IN STATISTIKE

Martin Raič

Datum zadnje spremembe: 6. maj 2019

Kazalo

1. Osnove kombinatorike

Pravilo vsote, pravilo produkta. Variacije, kombinacije in permutacije.

1. Manca ima 12 kap in 3 klobuke. Na koliko načinov se lahko pokrije?

2. Nace ima 4 puloverje, 2 srajci in 3 hlač. Na koliko načinov se lahko obleče?

3. Olga je kupila novo stanovanje. Na koliko načinov lahko opremi dnevno sobo, če ima na voljo 4 vrste parketa, 3 vrste nelesnih talnih oblog in 5 garnitur pohištva?

4. Peter je sprevodnik na progi Ljubljana – Litija, ki ima 7 postaj.

   a) Koliko je možnih enosmernih vozovnic?

   b) Največ koliko cen je lahko na ceniku, če privzamemo, da vožnja na nasprotni relaciji stane enako?

5. Jernej se uči poštevanke. Koliko računov se mora naučiti, če ve:

   - da se število ohrani, če ga pomnožimo z 1;
   - da se naravnemu številu zgolj pripiše ničla, če ga pomnožimo z deset;
   - da je množenje komutativno.

6. Koliko možnih besed (smiselnih ali nesmiselnih) lahko REZKA sestavi s premetavanjem črk svojega imena? Kaj pa TATJANA?

7. Urban je razrednik. 15 učencev v njegovem razredu obiskuje modelarski krožek, 21 likovni krožek, 3 pa oba krožka. Najmanj koliko učencev je v razredu?

8. 10 športnikov se pomeri na tekmovanju. Na koliko načinov lahko dobijo medalje? Delitve mest so izključene.

9. Na koliko načinov lahko 10 učencev med seboj izbere tričlansko delegacijo?

2. Elementarna verjetnost

Diskretni pogled na verjetnost v povezavi s kombinatoriko.

Če so vsi izidi enako verjetni, velja:

$$P = \frac{\text{število ugodnih izidov}}{\text{število vseh izidov}}$$

1. Kolikšna je verjetnost, da na poštenem kovancu pade grb?

2. Vržemo dva kovanca, vsi izidi, kjer kovanca ločimo, so enako verjetni. Kolikšna je verjetnost, da pade ena cifra in en grb?

3. Vržemo deset kovancev, vsi izidi, kjer kovance ločimo, so enako verjetni. Kolikšna je verjetnost, da pade en grb in devet cifer?

4. Vržemo pošteno kocko. Kolikšna je verjetnost, da pade:
   a) 6 pik?
   b) 1 pika?
   c) sodo mnogo pik?
   d) vsaj tri pike?

5. Vržemo dve kocki, vsi izidi, kjer kocki ločimo, so enako verjetni.
   a) Kolikšna je verjetnost, da na obeh pade enako mnogo pik?
   b) Kolikšna je verjetnost, da pade skupaj vsaj 10 pik?

Če sta dogodka A in B **nezdružljiva**, *t. j. $A \cap B = \emptyset$, velja:*

$$P(A \cup B) = P(A) + P(B)\,.$$

Splošneje, če so dogodki $A_1, A_2, \ldots, A_n$ paroma nezdružljivi, t. j. če sta poljubna dva izmed njih nezdružljiva, velja:

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)\,.$$

6. Petkrat vržemo kocko, vsi izidi, kjer mete ločimo, so enako verjetni. Kolikšna je verjetnost, da padejo tri, ne pa tudi štiri zaporedne šestice?

7. Iz dobro premešanega kupa standardnih 32 kart izvlečemo eno karto. Kolikšna je verjetnost, da bo karta:

a) pik?

b) as?

c) pik ali as?

Dogodka A in B sta si **nasprotna***, če sta nezdružljiva, njuna unija pa je gotovi dogodek. V tem primeru velja:*

$$P(B) = 1 - P(A)\,.$$

8. Iz dobro premešanega kupa standardnih 32 kart brez vračanja izvlečemo pet kart. Kolikšna je verjetnost, da je med njimi vsaj en as?

9. Pri igri Loto na kombinacijskem listku prekrižamo 7 številk izmed 39. Izžreba se 7 rednih številk in še ena dodatna. Možni so naslednji dobitki:

   - sedmica: vse prekrižane številke so redno izžrebane;
   - šest in dodatna: med prekrižanimi številkami je šest redno izžrebanih in ena dodatna;
   - šestica: natanko šest prekrižanih številk je redno izžrebanih, dodatna ni prekrižana;
   - petica: natanko pet prekrižanih številk je redno izžrebanih (dodatna pa je lahko prekrižana ali ne);
   - štirica: natanko štiri prekrižane številke so redno izžrebane (dodatna pa je lahko prekrižana ali ne);
   - tri in dodatna: natanko tri prekrižane številke so redno izžrebane, prekrižana pa je tudi dodatna številka.

   Izračunajte verjetnosti posameznih dobitkov.

10. Igra Loto pozna tudi sistemske listke, pri katerih prekrižamo več kot sedem številk. Sistem najmanj koliko številk bi morali vplačati, če naj bi bila verjetnost sedmice večja od polovice?

11. Študenti, ki bodo pisali izpit, se posedejo v tri vrste in tri kolone, tako kot je prikazano spodaj:

Aljaž	Brigita	Cveto
Dragica	Edo	Fani
Gregor	Hana	Iztok

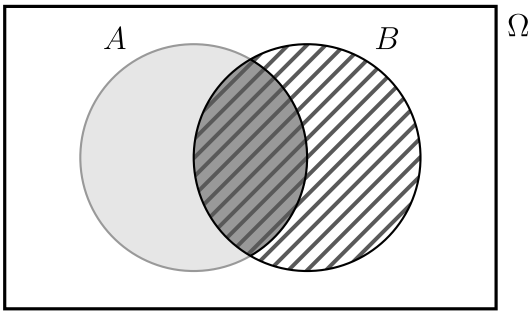
Asistent na slepo izbere tri študente in jih zamenja: prvega premesti na mesto drugega, drugega na mesto tretjega in tretjega na mesto prvega. Kolikšna je verjetnost, da Aljaž in Brigita po premestitvi še vedno sedita skupaj v isti vrsti?

3. Pogojna verjetnost

Računanje pogojne verjetnosti po definiciji. Izrek o popolni verjetnosti. Neodvisnost. Bayesova formula.

Definicija pogojne verjetnosti

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$



Če je dogodek B sestavljen iz samih enako verjetnih izidov, pa je tudi:

$$P(A \mid B) = \frac{|A \cap B|}{|B|} \,.$$

1. Desetkrat vržemo kovanec, vsi izidi so enako verjetni. Označimo naslednje dogodke:

$$\begin{aligned} A &:= \{\text{v vseh desetih metih pade cifra}\} \\ B &:= \{\text{v devetih metih pade cifra, v enem pa grb}\} \\ C &:= \{\text{v prvih devetih metih padejo same cifre}\} \end{aligned}$$

   Izračunajte $P(A)$ in $P(B)$, nato pa še $P(A \mid C)$ in $P(B \mid C)$.

2. Iz posode, v kateri je 5 rdečih in 2 zeleni kroglici, na slepo in brez vračanja vlečemo kroglice. Kolikšna je verjetnost:

   a) da bo prva izvlečena kroglica rdeča?
   b) da bosta prvi dve izvlečeni kroglici rdeči?
   c) da bo prva izvlečena kroglica rdeča, druga pa zelena?
   d) da bo prva izvlečena kroglica zelena, druga pa rdeča?
   e) da bodo prve tri izvlečene kroglice rdeče?
   f) da bodo prve tri izvlečene kroglice zelene?

Izrek o polni verjetnosti

Če $H_1, H_2, \ldots, H_n$ tvorijo popoln sistem dogodkov (t. j. vedno se zgodi natanko eden izmed njih), velja:

$$P(A) = P(H_1)\ P(A \mid H_1) + P(H_2)\ P(A \mid H_2) + \cdots + P(H_n)\ P(A \mid H_n)$$

3. Iz posode, v kateri je 5 rdečih in 2 zeleni kroglici, ponovno na slepo in brez vračanja vlečemo kroglice. Kolikšna je:

   a) verjetnost, da bo druga izvlečena kroglica rdeča?

   b) pogojna verjetnost, da je bila prva izvlečena kroglica rdeča, če vemo, da je druga izvlečena kroglica rdeča?

   c) pogojna verjetnost, da je bila prva izvlečena kroglica rdeča, če vemo, da sta prvi dve izvlečeni kroglici rdeči?

   d) pogojna verjetnost, da je bila prva izvlečena kroglica rdeča, če vemo, da je vsaj ena izmed prvih dveh izvlečenih kroglic rdeča?

   e) pogojna verjetnost, da je bila prva izvlečena kroglica rdeča, če vemo, da je izmed prvih dveh izvlečenih kroglic natanko ena rdeča?

4. Dan je standardni kup 32 kart. Njihove vrednosti od največje do najmanjše so as, kralj, dama, fant, 10, 9, 8 in 7. Iz kupa na slepo in brez vračanja izvlečemo dve karti. Kolikšna je:

   a) verjetnost, da bo prva karta as?

   b) verjetnost, da ima prva karta večjo vrednost od druge?

   c) pogojna verjetnost, da ima prva karta večjo vrednost od druge, če vemo, da je prva karta as?

   d) pogojna verjetnost, da je prva karta as, če vemo, da ima večjo vrednost od druge?

Dogodka A in B sta **neodvisna**, *če velja:*

$$P(A \cap B) = P(A)\ P(B)$$

(t. j. $P(A \mid B) = P(A)$). Splošneje, dogodki $A_1, A_2, A_3, \ldots$ so neodvisni, če za poljubne različne indekse $i_1, i_2, \ldots, i_k$ velja:

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})\ P(A_{i_2}) \ldots P(A_{i_k})\,.$$

5. Vržemo dva kovanca, vsi izidi so enako verjetni. Definirajmo naslednje dogodke:

$$\begin{aligned} A &:= \{\text{prvič pade cifra}\} \\ B &:= \{\text{drugič pade cifra}\} \\ C &:= \{\text{obakrat pade cifra}\} \\ D &:= \{\text{obakrat pade različno}\} \end{aligned}$$

Sta A in B neodvisna? Kaj pa A in C? Pa A in D? Pa B in D? Kaj pa dogodki A, B in D, ali so neodvisni?

Neodvisnost izpeljanih dogodkov

Naj bodo $A_1, A_2, \dots, A_m, B_1, B_2, \dots, B_n$ neodvisni dogodki. Nadalje naj bo A dogodek, izpeljan iz dogodkov $A_1, \dots, A_m$ z osnovnimi operacijami (unija, presek, komplement), B pa naj bo dogodek, izpeljan iz dogodkov $B_1, \dots, B_n$. Tedaj so naslednji dogodki neodvisni:

- *$A_1, A_2, \dots, A_m$ in B;*
- *A in $B_1, B_2, \dots, B_n$;*
- *A in B.*

6. Miha se odpravi na obisk k vinogradnikoma Janezu in Lojzu. Vsak mu ponudi kozarec vina. Janez mu ponudi šmarnico z verjetnostjo 60%, Lojz pa z verjetnostjo 30%. Verjetnost, da Miho boli glava, ne da bi pil šmarnico, je 40%, verjetnost, da ga boli po kozarcu šmarnice, 80%, po dveh kozarcih pa 100%. Kolikšna je verjetnost, da Miho po obisku obeh vinogradnikov boli glava? Privzamemo, da vinogradnika izbereta vrsto vina neodvisno drug od drugega.

Bayesova formula

Če $H_1, H_2, \dots, H_n$ tvorijo popoln sistem dogodkov, velja:

$$P(H_i \mid A) = \frac{P(H_i)\ P(A \mid H_i)}{P(H_1)\ P(A \mid H_1) + \cdots + P(H_n)\ P(A \mid H_n)}$$

7. Žena pošlja moža na trg po solato, ki jo prodajata dve branjevki, Francka in Micka. Verjetnost, da mož kupi solato pri Francki, je 40%, verjetnost, da kupi pri Micki, pa 60%. Francka ima 10%, Micka pa 20% nagnitih glav solate. Mož prinese domov nagnito glavo solate in žena ga nahruli: “Drugič raje glej solato, ne pa Micke!” Kolikšna je pogojna verjetnost, da je mož res kupil solato pri Micki? Privzamemo, da branjevki solato izbirata na slepo.

8. Miha se odpravi na obisk k vinogradnikoma Janezu in Lojzu. Vsak mu ponudi kozarec vina. Janez mu ponudi šmarnico z verjetnostjo 60%, Lojz pa z verjetnostjo 30%. Verjetnost, da Miho boli glava, ne da bi pil šmarnico, je 40%, verjetnost, da ga boli po kozarcu šmarnice (ne glede na to, čigave), 80%, po dveh kozarcih pa 100%.

   Naslednji dan Miho boli glava in prijatelj Tone mu pravi: "Janez ti je gotovo dal šmarnico!" Kolikšna je pogojna verjetnost, da ima prav? Privzamemo, da vinogradnika izbereta vrsto vina neodvisno drug od drugega.

9. Janez, Francelj in Tone gredo streljat zajce. Janez zadene z verjetnostjo $0\dot{}1$, Francelj z verjetnostjo $0\dot{}2$, Tone pa z verjetnostjo $0\dot{}3$, neodvisno drug od drugega.

   a) Vsi pomerijo, ustrelijo in zajec je zadet. Kolikšna je pogojna verjetnost, da ga je Janez zadel?

   b) Ko pridejo do zajca, se izkaže, da ga je zadel natanko eden. Kolikšna je pogojna verjetnost, da je bil to Janez?

4. Diskretne slučajne spremenljivke

Pojem porazdelitve. Matematično upanje, disperzija, kovarianca, korelacijski koeficient.

1. Vržemo dva neodvisna kovanca. Če pade cifra, kovanec dobimo. Slučajna spremenljivka S naj predstavlja skupni znesek v evrih, ki ga dobimo. Določite verjetnostni prostor G, S kot preslikavo iz G in porazdelitev slučajne spremenljivke S, če vržemo:

   a) en kovanec po 1€ in en kovanec po 2€;

   b) dva kovanca po 2€.

2. Petkrat vržemo pošteno kocko. Označimo z S število šestic, ki padejo. Določite:

   a) porazdelitev slučajne spremenljivke S;

   b) $P(S \text{ sodo})$;

   c) $P(S \geq 3)$;

   d) $P(S \text{ sodo} \mid S \geq 3)$.

3. Pošten kovanec mečemo, dokler ne pade cifra, a največ štirikrat. Označimo z M število metov. Zapišite porazdelitev te slučajne spremenljivke.

4. Pošten kovanec mečemo, dokler ne padeta dve cifri zapored, vendar največ sedemkrat. Označimo z M število metov. Zapišite porazdelitev te slučajne spremenljivke.

Matematično upanje: *če je* $X \sim \begin{pmatrix} a_1 & a_2 & \cdots a_n \\ p_1 & p_2 & \cdots p_n \end{pmatrix}$, *velja:*

$$E(X) = a_1p_1 + a_2p_2 + \cdots + a_np_n \,,$$
$$E\big(f(X)\big) = f(a_1)p_1 + f(a_2)p_2 + \cdots + f(a_n)p_n \,.$$

Disperzija (varianca):

$$D(X) = E\big((X - E(X))^2\big) = E(X^2) - \big(E(X)\big)^2$$

Standardni odklon: $\sigma(X) = \sqrt{D(X)}$.
U-metoda:

$$E(X) = u + E(X - u)$$
$$D(X) = D(X - u) = E\big((X - u)^2\big) - \big(E(X - u)\big)^2$$

5. Naj bo:

$$X \sim \begin{pmatrix} -1 & 0 & 1 & 4 \\ 0\text{'}1 & ? & 0\text{'}1 & 0\text{'}5 \end{pmatrix} .$$

   Izračunajte $P(X = 0)$, $E(X)$, $D(X)$ in $\sigma(X)$.

6. Naj bo:
$$X \sim \begin{pmatrix} 1001 & 1002 & 1003 & 1004 \\ 0{\cdot}2 & a & b & 0{\cdot}1 \end{pmatrix}.$$
Izračunajte a in b, če veste, da je $E(X) = 1002$.

7. Naj bo:
$$X \sim \begin{pmatrix} 1001 & 1002 & 1003 & 1004 \\ 0{\cdot}2 & a & b & 0{\cdot}1 \end{pmatrix}.$$
   a) Kolikšno je minimalno in kolikšno maksimalno matematično upanje slučajne spremenljivke X?
   b) Kolikšna je minimalna in maksimalna disperzija?

8. V posodi so 3 kovanci po 1 tolar, 6 kovancev po 2 tolarja in 1 kovanec po 5 tolarjev. Iz posode izvlečemo dva kovanca. Označimo z X_1 vrednost prvega v tolarjih, z X_2 pa vrednost drugega.
   a) Izračunajte $E(X_1)$ in $D(X_1)$.
   b) Naj bo $S = X_1 + X_2$. Izračunajte $E(S)$ in $D(S)$ tako v primeru, če kovance vračamo, kot tudi v primeru, če jih ne vračamo.

$$\boxed{\begin{aligned} &E(aX+b) = a\,E(X) + b \\ &D(aX+b) = a^2\,D(X) \end{aligned}}$$

9. Vržemo pet standardnih in neodvisnih kock, za vsako šestico, ki pade, dobimo 2€. Posamezna igra stane 1,70€. Naj bo S bilanca igre. Izračunajte $P(S > 0)$, $E(S)$ in $\sigma(S)$.

Kovarianca:

$$K(X,Y) := E\Big[\big(X - E(X)\big)\big(Y - E(Y)\big)\Big] = \\ = E(XY) - E(X)\,E(Y)$$

Če je $K(X,Y) = 0$, pravimo, da sta slučajni spremenljivki X in Y nekorelirani.
X, Y neodvisni $\Longrightarrow E(XY) = E(X)\,E(Y) \Longleftrightarrow X, Y$ nekorelirani
Nekoreliranost še ne pomeni neodvisnosti.

Korelacijski koeficient:

$$\rho(X,Y) = \frac{K(X,Y)}{\sigma(X)\,\sigma(Y)}$$

Velja $-1 \le \rho(X,Y) \le 1$. Če so a, b, c in d konstante ter $a, c > 0$, velja $\rho(aX+b, cY+d) = \rho(X,Y)$.

10. Za slučajni spremenljivki X_1 in X_2 iz 8. naloge izračunajte $K(X_1, X_2)$ in $\rho(X_1, X_2)$.

11. Navzkrižna porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	$0\text{˙}2$	$0\text{˙}2$	$0\text{˙}2$
$X = 1$	$0\text{˙}2$	0	$0\text{˙}2$

Dokažite, da sta slučajni spremenljivki X in Y nekorelirani. Sta tudi neodvisni?

12. Navzkrižna porazdelitev slučajnega vektorja (X, Y) je podana z naslednjo tabelo:

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	$0\text{˙}1$	$0\text{˙}2$	$0\text{˙}1$
$X = 2$			

Dopolnite tabelo, tako da bosta slučajni spremenljivki X in Y neodvisni.

5. Zvezne slučajne spremenljivke

Porazdelitvena gostota, kumulativna porazdelitvena funkcija. Matematično upanje in disperzija. Normalna porazdelitev, centralni limitni izrek.

Diskretne porazdelitve:	*Zvezne porazdelitve:*
$P(a \le X \le b) = \sum_{x; a \le x \le b} P(X = x)$	$P(a \le X \le b) = \int_a^b g_X(x)\,\mathrm{d}x$
$P(X = x) \ge 0$	$g_X(x) \ge 0$
$\sum_x P(X = x) = 1$	$\int_{-\infty}^{\infty} g_X(x)\,\mathrm{d}x = 1$
$E\big[f(X)\big] = \sum_x f(x)\,P(X = x)$	$E\big[f(X)\big] = \int_{-\infty}^{\infty} f(x) g_X(x)\,\mathrm{d}x$

1. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:

$$g_X(x) = \begin{cases} cx^2 & ;\ 0 \le x \le 3 \\ 0 & ;\ \text{sicer} \end{cases} .$$

Izračunajte konstanto c ter še $P(1 < X < 2)$, $P(X \le 1)$, $E(X)$ in $D(X)$.

Kumulativna porazdelitvena funkcija

$$F_X(x) = P(X < x)$$
$$a \le b \Longrightarrow P(a \le X < b) = F_X(b) - F_X(a)$$

Kumulativna porazdelitvena funkcija je naraščajoča ter velja $\lim_{x\to-\infty} F_X(x) = 0$ *in* $\lim_{x\to\infty} F_X(x) = 1$.
F_X *je zvezna in odsekoma zvezno odvedljiva* $\Longrightarrow$
$\Longrightarrow$ X *je zvezno porazdeljena* $\Longrightarrow$ F_X *je zvezna.*
Če je X *zvezno porazdeljena, je* $F_X(x) = \int_{-\infty}^{x} g_X(t)\,\mathrm{d}t$ *in skoraj povsod tudi* $g_X(x) = F_X'(x)$.

2. Izračunajte kumulativno porazdelitveno funkcijo slučajne spremenljivke X iz 1. naloge.

3. *Zvezna enakomerna porazdelitev* $\mathrm{U}(a, b)$ na intervalu (a, b) ima gostoto, ki je na navedenem intervalu konstantna, drugje pa je enaka nič.

a) Zapišite natančno formulo za gostoto.
b) Izračunajte kumulativno porazdelitveno funkcijo.
c) Izračunajte matematično upanje in disperzijo.

4. Dana je funkcija:
$$g(x) = \begin{cases} ae^{bx} & ; x \geq 0 \\ 0 & ; \text{sicer} \end{cases} .$$
Določite konstanti a in b, tako da bo g gostota neke porazdelitve z matematičnim upanjem 4.

5. Slučajna spremenljivka X je porazdeljena zvezno z gostoto:
$$g_X(x) = \begin{cases} 2 - kx & ; 0 \leq x \leq a \\ 0 & ; \text{sicer} \end{cases} .$$
Pri katerih a in k je matematično upanje $E(X)$ maksimalno?

Naj bo $\mu \in \mathbb{R}$ in $\sigma > 0$. **Normalna (Gaussova)** *porazdelitev* $\mathrm{N}(\mu, \sigma)$ *je porazdelitev z gostoto:*

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

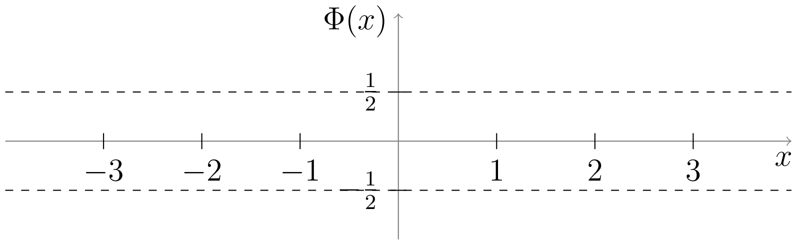
Brž ko je $a \le b$, velja:

$$P(a < X < b) = P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

Funkcija Φ je **Gaussov verjetnostni integral**:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} \, \mathrm{d}t$$

in je liha. Graf:



V literaturi so definicije funkcije Φ različne, zato je treba paziti!
Normalna porazdelitev $\mathrm{N}(\mu, 0)$ je porazdelitev, ki je skoncentrirana v μ ($X \sim \mathrm{N}(\mu, 0)$ pomeni $P(X = \mu) = 1$).
Parameter μ je tudi pričakovana vrednost (matematično upanje), parameter σ pa standardni odklon normalne porazdelitve; posledično je σ^2 njena disperzija.
Standardna normalna porazdelitev $\mathrm{N}(0, 1)$ *ima potemtakem gostoto:*

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

in kumulativno porazdelitveno funkcijo $F(x) = \frac{1}{2} + \Phi(x)$.

6. Slučajna spremenljivka X je porazdeljena normalno $\mathrm{N}(100, 10)$. Izračunajte $P(90 < X < 120)$ in $P(X < 0)$.

7. Avtobus spelje s postaje ob 7:00, sam pa pridem na postajo ob času, porazdeljenem normalno s sredino ob 6:59 in standardnim odklonom 3 minute. Kolikšna je verjetnost, da še ujamem avtobus?

Centralni limitni izrek

Če je S vsota veliko neodvisnih slučajnih spremenljivk z dovolj lepimi porazdelitvami, od katerih nobena posebej ne izstopa, je približno $S \sim \mathrm{N}(\mu, \sigma)$, kjer je $\mu = E(S)$ in $\sigma^2 = D(S)$.
Če je $S = X_1 + X_2 + \cdots + X_n$ in so $X_1, \ldots, X_n$ neodvisne, je seveda:

$$\mu = E(X_1) + E(X_2) + \cdots + E(X_n)$$
$$\sigma^2 = D(X_1) + D(X_2) + \cdots + D(X_n)\,.$$

8. Zavarovalnica pri neki polici z verjetnostjo 15% izplača 5 evrov, z verjetnostjo 10% 10 evrov, sicer pa ne izplača nič. Zavarovalna premija je 2 evra. Kolikšna je verjetnost, da ima zavarovalnica izgubo, če sklene:

   a) eno polico?
   b) tri police?
   c) pet polic?
   d) 500 polic (zgolj približno)?

   Pri eni, treh in petih policah točen rezultat primerjajte z verjetnostjo, dobljeno iz ustrezne normalne aproksimacije.

Naj bo $X \sim \mathrm{B}(n, p)$ ter $n \to \infty$ in še $p, 1 - p \gg 1/n$
(ali, ekvivalentno, $\sigma := \sqrt{np(1-p)} \to \infty$).
Če je $k \in \mathbb{Z}$ in $|k - np| \ll \sigma^{4/3}$, velja **Laplaceova lokalna formula:**

$$P(X = k) \sim \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(k-np)^2/(2\sigma^2)}\,.$$

Če je $a \le b$ ter še $|a - np| \ll \sigma^{4/3}$ **ali** *$|b - np| \ll \sigma^{4/3}$, velja*
Laplaceova integralska formula, *in sicer za $a, b \in \mathbb{Z}$ v obliki:*

$$P(a \le X \le b) \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sigma}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sigma}\right),$$

za $a, b \in \mathbb{Z} + \frac{1}{2}$ ali $b - a \gg 1$ pa v obliki:

$$P(a < X < b) \sim P(a \le X \le b) \sim \Phi\left(\frac{b - np}{\sigma}\right) - \Phi\left(\frac{a - np}{\sigma}\right)$$

Z drugimi besedami, $\mathrm{B}(n, p) \approx \mathrm{N}(np, \sigma)$.

9. V tovarni vsak dan proizvedejo 1600 izdelkov. Za vsakega je verjetnost, da bo okvarjen, enaka 10%.

a) Kolikšna je verjetnost, da bo okvarjenih natanko 160 izdelkov? Kolikšna pa je verjetnost, da bo okvarjenih natanko 175 izdelkov?

b) Kolikšna je verjetnost, da bo okvarjenih več kot 175 izdelkov? Kaj pa, da bo okvarjenih manj kot 150 izdelkov?

10. Zavarovalnica je proti nezgodi zavarovala 1000 oseb. Vsako od njih doleti nezgoda z verjetnostjo 0˙0015 in osebe so med seboj neodvisne. Kolikšna je verjetnost, da se noben zavarovanec ne ponesreči? Kolikšna pa je verjetnost, da se ponesrečijo več kot štirje? Točen rezultat primerjajte z rezultatom, dobljenem po Laplaceovi lokalni oziroma integralski formuli.

6. Kvantili

Pojem kvantila, mediana, kvartili, decili, percentili.

Kvantili

Število x_α *je* **kvantil** *slučajne spremenljivke* X *za verjetnost* α*, če velja:*

$$P(X < x_\alpha) \le \alpha \le P(X \le x_\alpha)\,.$$

Če je X *zvezno porazdeljena in ima v okolici točke* x_α *strogo pozitivno gostoto, pa velja:*

$$F_X(x_\alpha) = P(X < x_\alpha) = P(X \le x_\alpha) = \alpha\,.$$

Mediana: $m = x_{1/2}$
Tercila: $x_{1/3},\ x_{2/3}$
Kvartili: $x_{1/4},\ x_{1/2} = m,\ x_{3/4}$
Decili: $x_{0.1},\ x_{0.2}, \ldots, x_{0.9}$
Percentili: $x_{0.01},\ x_{0.02}, \ldots, x_{0.99}$

1. Slučajna spremenljivka X je porazdeljena zvezno z naslednjo gostoto:

$$g_X(x) = \begin{cases} \dfrac{1}{(1+x)^2} & ;\ x \ge 0 \\ 0 & ;\ \text{sicer} \end{cases}$$

   Izračunajte mediano, prvi kvartil, 9. decil in 99. percentil.

2. Izračunajte 99. percentil standardne normalne porazdelitve.

3. V tovarni vsak dan proizvedejo 1600 izdelkov. Za vsakega je verjetnost, da bo okvarjen, enaka 10%. Okvarjene izdelke spravijo v skladišče, kjer jih popravijo in ki se dnevno prazni. Najmanj kako veliko mora biti skladišče, če naj bo verjetnost, da bo premajhno, največ 0˙05?

4. Verjetnost, da je izdelek prvovrsten, je 60%. Najmanj koliko izdelkov moramo naročiti, če naj bo med naročenimi izdelki z verjetnostjo najmanj 0˙99 vsaj 59% izdelkov prvovrstnih?

7. Vzorčenje

1. Piščanci v populaciji imajo naslednje teže (v gramih):

$$900,\ 920,\ 930,\ 950,\ 950,\ 980,\ 1000,\ 1010,\ 1010,\ 1050\,.$$

   Vzamemo enostavni slučajni vzorec dveh piščancev. Kolikšna je verjetnost, da je vzorčno povprečje teže večje od enega kilograma, in kolikšen je 95. percentil, če gre za vzorec:

   a) brez ponavljanja?

   b) s ponavljanjem?

2. Statistična spremenljivka X na veliki populaciji ima naslednjo porazdelitev:

$$X \sim \begin{pmatrix} 1 & 2 & 4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}\,.$$

   Iz populacije vzamemo enostavni slučajni vzorec velikosti 100. Približno izračunajte:

   a) verjetnost, da vzorčno povprečje odstopa od populacijskega za manj kot $0{\cdot}1$;

   b) 95. percentil vzorčnega povprečja.

3. Na travniku rastejo bele in rdeče rože. Delež belih je 60%. Na slepo utrgamo 50 cvetov.

   a) Kolikšna je verjetnost, da delež belih odstopa od populacijskega za manj kot 5 absolutnih odstotkov?

   b) Izračunajte 95. percentil za delež belih cvetov v vzorcu.

8. Intervali zaupanja

Bernoullijevo zaporedje poskusov. Normalna porazdelitev z znanim σ. Normalna porazdelitev z obema neznanima parametroma, iščemo μ ali σ. Asimptotični intervali zaupanja za ne-Gaussove porazdelitve.

Interval zaupanja $(\zeta_{\min}, \zeta_{\max})$ *za karakteristiko* ζ *pri* **stopnji zaupanja** β *je določen z neenačbo:*

$$P\big(\zeta_{\min} < \zeta < \zeta_{\max}\big) \ge \beta$$

ki mora veljati za vse verjetnostne mere P *iz našega statističnega modela,* $\zeta_{\min}$ *in* $\zeta_{\max}$ *pa morata biti opazljivi. Če je res* $\zeta_{\min} < \zeta < \zeta_{\max}$*, pravimo, da pride do* **pokritosti**. *Če se da, interval izberemo tako, da je* β *natančna spodnja meja za verjetnost pokritosti. Tipično vzamemo* $\beta = 0\text{'}90$, $0\text{'}95$ *ali* $0\text{'}99$.
Če zaokrožujemo, spodnjo mejo vedno zaokrožimo navzdol, zgornjo pa navzgor.

Interval zaupanja za neznano verjetnost

Naj bo θ *verjetnost nekega dogodka. Izvedemo* n *poskusov in pri* S *se ta dogodek zgodi. Edina smiselna točkasta ocena za* θ *je seveda:*

$$\hat{\theta} = \frac{S}{n}\,.$$

Interval zaupanja:

$$\hat{\theta} - c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} - \frac{1}{2n} < \theta < \hat{\theta} + c\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} + \frac{1}{2n}\,,$$

kjer je $c = z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.
Ta interval zaupanja je le **približen**, *kar pomeni, da verjetnost, da je* θ *res notri, ni točno* β. *Minimalni pogoj za zadovoljivo natančnost je* $c^2 + 1 \le S \le n - c^2 - 1$ *(torej mora biti tudi* $n \ge 2(c^2+1)$*). Kaj storiti, če to ni izpolnjeno, glej kasneje.*

Opomba. Interval zapanja za neznano verjetnost ima veliko različic. Vsaka ima svoje prednosti in slabosti. Zgornja konstrukcija je *Waldov interval zaupanja s popravkom za zveznost.*

1. V 100 metih kocke je 20-krat padla šestica. Določite 95% in 99% interval zaupanja za verjetnost, da pade šestica.

2. Pri 10000 metih kovanca je padlo 5048 grbov. Določite 99% interval zaupanja za verjetnost, da pade grb.

Interval zaupanja za neznano verjetnost pri skrajnih frekvencah

Še naprej je θ verjetnost, da posamezen poskus uspe, n število izvedenih poskusov in S opaženo število uspelih. Interval zaupanja za θ:

$S = 0$:

$$\theta < \frac{1}{2n}\,\chi^2_{(1+\beta)/2}(2)$$

S *blizu* 0, $S \neq 0$:

$$\frac{1}{2n}\,\chi^2_{(1-\beta)/2}(2S) < \theta < \frac{1}{2n}\,\chi^2_{(1+\beta)/2}(2S+2)$$

S *blizu* n, $S \neq n$:

$$1 - \frac{1}{2n}\,\chi^2_{(1+\beta)/2}(2(n-S)+2) < \theta < 1 - \frac{1}{2n}\,\chi^2_{(1-\beta)/2}(2(n-S))$$

$S = n$:

$$\theta > 1 - \frac{1}{2n}\,\chi^2_{(1+\beta)/2}(2)$$

Tu je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami.

3. Janez je pri testu na vseh 20 vprašanj odgovoril pravilno. Določite 95% interval zaupanja za verjetnost dogodka, da Janez na posamezno vprašanje odgovori pravilno. Kaj pa, če bi bilo vprašanj 200?

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Ocenjujemo parameter μ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, kjer je σ znan, na voljo pa imamo vzorec neodvisnih opažanj $X_1, \ldots, X_n$. Tedaj ne glede na μ velja:

$$\frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)\,,$$

kjer je $\bar{X}$ vzorčno povprečje, definirano spodaj. Izračunamo torej:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$
$$c = z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$$
$$\Delta = \frac{c\sigma}{\sqrt{n}}$$

Interval zaupanja: $\bar{X} - \Delta < \mu < \bar{X} + \Delta$.

4. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, 5)$. Vrednosti na vzorcu so:

$$101,\ 91,\ 93,\ 103,\ 91,\ 101,\ 103,\ 95,\ 95$$

Določite 95% interval zaupanja za μ.

Sredina pri normalni porazdelitvi z neznanim standardnim odklonom

Ocenjujemo parameter μ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, kjer σ ni znan, na voljo pa imamo spet vzorec neodvisnih opažanj $X_1, \ldots, X_n$. Tedaj ne glede na μ in σ velja:

$$\frac{\bar{X} - \mu}{s} \sqrt{n} \sim \mathrm{S}(n-1)\,,$$

kjer je s popravljeni vzorčni standardni odklon, definiran spodaj. Izračunamo torej:

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c &= t_{(1+\beta)/2}(n-1) \\
s &= \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}} \\
\Delta &= \frac{cs}{\sqrt{n}}\,,
\end{aligned}$$

kjer je $t_p(df)$ kvantil Studentove porazdelitve z df prostostnimi stopnjami za verjetnost p.

Interval zaupanja: $\bar{X} - \Delta < \mu < \bar{X} + \Delta$.

Za velike vzorce konstrukcija asimptotično deluje tudi pri ne-Gaussovih spremenljivkah. V tem primeru lahko kvantil Studentove porazdelitve $t_{(1+\beta)/2}(n-1)$ zamenjamo s kvantilom normalne porazdelitve $z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.

5. Isto kot prejšnja naloga, le da σ zdaj ni znan.

6. Za 860 žensk poizvemo, koliko otrok imajo. Dobimo naslednjo frekvenčno porazdelitev:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

Poiščite 95% interval zaupanja za povprečno število otrok na žensko na celotni populaciji.

Opomba. Podatki so sicer izmišljeni, so pa ukrojeni po popisu Slovenije iz leta 2002 (števila žensk so deljena s 1000 in zaokrožena, izmišljen je tudi konec tabele).

Standardni odklon pri normalni porazdelitvi

Ocenjujemo parameter σ pri porazdelitvi $\mathrm{N}(\mu, \sigma)$, kjer μ ni znan. Tedaj je:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)\,.$$

Izračunamo torej:

$$\begin{aligned}
\bar{X} &= \frac{X_1 + X_2 + \cdots + X_n}{n} \\
c_1 &= \chi^2_{(1-\beta)/2}(n-1) \\
c_2 &= \chi^2_{(1+\beta)/2}(n-1) \\
s^2 &= \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}\,,
\end{aligned}$$

kjer je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost p.

Interval zaupanja:

$$s\sqrt{\frac{n-1}{c_2}} < \sigma < s\sqrt{\frac{n-1}{c_1}}$$

oziroma:

$$\sqrt{\frac{1}{c_2}\sum_{i=1}^{n}(X_i - \bar{X})^2} \le \sigma \le \sqrt{\frac{1}{c_1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

7. Isti podatki kot pri 4. nalogi, le da ocenjujemo σ.

8. Statistična spremenljivka je porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Vrednosti na vzorcu so:

$$124,\ 129,\ 126,\ 122,\ 124$$

Ocenite σ po občutku. Ali pride v 90% interval zaupanja?

9. Telesna teža v skupini 75 učencev ima naslednjo frekvenčno porazdelitev:

teža [kg]	39	40	41	42	43	44	45	46	47	48	49	50
št. učencev	1	3	5	8	8	7	9	8	6	6	4	3

teža [kg]	51	52	53	54	59
št. učencev	2	2	1	1	1

Privzemimo, da ta skupina predstavlja enostavni slučajni vzorec iz populacije, kjer je telesna teža porazdeljena normalno $\mathrm{N}(\mu, \sigma)$. Poiščite 99% interval zaupanja za μ in za σ (za vsakega posebej, pri čemer privzemite, da drugi parameter ni znan).

Asimptotični interval zaupanja za populacijski standardni odklon pri ne-Gaussovih spremenljivkah

$$\hat{\sigma}^2 - c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}} < \sigma^2 < \hat{\sigma}^2 + c\,\frac{\sqrt{\hat{\kappa}^4 - \hat{\sigma}^4}}{\sqrt{n}}\,,$$

kjer je:

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\,, \quad \hat{\kappa} = \sqrt[4]{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^4}\,.$$

in $c = z_{(1+\beta)/2} = \Phi^{-1}(\beta/2)$.

10. Poiščite 95% interval zaupanja za standardni odklon števila otrok na žensko za podatke iz 6. naloge:

število otrok	0	1	2	3	4	5	6	7	8	9	10
število žensk	227	168	320	96	29	11	4	2	2	0	1

9. Testi značilnosti

p-vrednosti. Test neznane verjetnosti. Z- in T-test sredine. T-test razlike sredin (za parne in neparne vzorce). Analiza variance z enojno klasifikacijo. Testiranje disperzije. Pearsonov test skladnosti. Test z znaki. Inverzijski test.

Želeli bi testirati, ali so opaženi podatki v vzorcu v skladu z ničelno hipotezo H_0 o porazdelitvi ali pa so morda bolj v skladu z alternativno hipotezo H_1. Pri testih značilnosti bodisi zavrnemo ničelno hipotezo bodisi pravimo, da odstopanja niso statistično dovolj **značilna**, *da bi jo zavrnili.*
Postopku, po katerem se odločimo, ali bomo ničelno hipotezo zavrnili ali ne, pravimo **test**. *Test ima* **stopnjo značilnosti** *α, če za vsako verjetnostno mero P, za katero velja H_0, velja:*

$$P(H_0 \text{ zavrnemo}) \leq \alpha\,.$$

Če se da, test načrtujemo tako, da je α natančna zgornja meja, z drugimi besedami, da je stopnja značilnosti **eksaktna**.
Če ničelno hipotezo zavrnemo pri $\alpha = 0\text{'}05$, pravimo, da so odstopanja statistično **značilna**. *Če se to zgodi pri pragu $0\text{'}01$, pa pravimo, da so statistično* **zelo značilna**.
Veliko testov poteka tako, da izračunamo testno statistiko T in H_0 zavrnemo, če T pade v kritično območje K_α. Tipična ničelna hipoteza je $\zeta = \zeta_0$, kjer je ζ kakšna karakteristika porazdelitve, tipične alternativne hipoteze pa so $\zeta \neq \zeta_0$ (dvostranski test) ali $\zeta < \zeta_0$ oz. $\zeta > \zeta_0$ (enostranski test).

Neznana verjetnost

Naj bo θ verjetnost nekega dogodka. Testiramo ničelno hipotezo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in pri S se ta dogodek zgodi. Ničelno hipotezo zavrnemo, če je p**-vrednost** *manjša ali enaka α, natančneje, $p(S) \leq \alpha$, kjer je funkcija p odvisna od alternativne hipoteze:*

- *pri $H_1\colon\ \theta > \theta_0$ je $p(k) = P(B \geq k)$;*
- *pri $H_1\colon\ \theta < \theta_0$ je $p(k) = P(B \leq k)$;*
- *pri $H_1\colon\ \theta < \theta_0$ je $p(k) = 2\min\{P(B \leq k), P(B \geq k)\}$.*

Tu je $B \sim \mathrm{B}(n, \theta_0)$.

1. Loterija za neko vrsto srečke trdi, da jih je vsaj pol dobitnih. Kupili smo osem srečk in le dve sta bili dobitni. Lahko pri stopnji značilnosti $\alpha = 0\text{'}05$ začnemo sumiti, da loterija laže? Seveda privzamemo, da so posamezne kupljene srečke med seboj neodvisne.

   Kaj pa, če bi kupili 16 srečk in bi 4 zadele?

2. Na neki fakulteti študira 70% žensk in 30% moških. Posebno priznanje za izjemne študijske dosežke je bilo podeljeno 5 ženskam in 7 moškim. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte ničelno hipotezo, da so izjemni študijski dosežki (po merilih komisije) neodvisni od spola, proti alternativni hipotezi, da so od spola odvisni.

Neznana verjetnost pri veliko poskusih

Naj bo θ verjetnost, da se zgodi določen dogodek. Testiramo ničelno hipotezo H_0, ki pravi, da je $\theta = \theta_0$. Izvedemo n poskusov in pri S se ta dogodek zgodi. Tokrat privzamemo, da je število poskusov dovolj veliko: za naše potrebe je to tedaj, ko je $n\theta \ge 5$ in $n(1-\theta) \ge 5$. Glede na alternativno hipotezo H_1 ničelno hipotezo zavrnemo:

- *pri H_1: $\theta \neq \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\left|\dfrac{S}{n} - \theta_0\right| - \dfrac{1}{2n}\right) \ge z_{(1-\alpha)/2}$;
- *pri H_1: $\theta > \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 - \dfrac{1}{2n}\right) \ge z_{1-\alpha}$;
- *pri H_1: $\theta < \theta_0$: če je* $\sqrt{\dfrac{n}{\theta_0(1-\theta_0)}}\left(\dfrac{S}{n} - \theta_0 + \dfrac{1}{2n}\right) \le -z_{1-\alpha}$.

Z $z_p = \Phi^{-1}\bigl(p - \frac{1}{2}\bigr)$ smo označili kvantil standardne normalne porazdelitve za verjetnost p.

3. Tovarna jamči, da je delež izdelkov z napako enak 20%. V vzorcu 100 izdelkov pa jih je 24 z napako. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je verjetnost, da ima izdelek napako, enaka $0{\cdot}2$, proti alternativni hipotezi, da je večja od $0{\cdot}2$. Kaj pa, če bi imelo napako 120 izdelkov izmed 500? In kaj, če bi v slednjem primeru vzeli stopnjo značilnosti $0{\cdot}01$?

4. 10000-krat vržemo kovanec in 5090-krat je padel grb. Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je kovanec pošten, proti alternativni hipotezi, da ni pošten. Kaj pa, če bi za alternativno hipotezo vzeli, da grb pade z večjo verjetnostjo kot cifra?

V vseh nadaljnjih nalogah tega razdelka imamo za vsako statistično spremenljivko X, definirano na populaciji, na voljo vzorec, na katerem ima ustrezna spremenljivka vrednosti $X_1, X_2, \ldots, X_n$, pri čemer so vrednosti neodvisne in porazdeljene enako kot X.

Sredina pri normalni porazdelitvi z znanim standardnim odklonom

Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. *Ker pri* H_0 *velja:*

$$Z := \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim \mathrm{N}(0,1)\,,$$

uporabimo Z**-test** *za testno statistiko* Z*, in sicer:*

- **dvostransko različico**, *če* H_1 *trdi, da je* $\mu \neq \mu_0$*;*
- **enostransko različico v desno**, *če* H_1 *trdi, da je* $\mu > \mu_0$*;*
- **dvostransko različico v levo**, *če* H_1 *trdi, da je* $\mu < \mu_0$*;*

Zgoraj omenjene različice Z*-testa za testno statistiko* Z *ničelno hipotezo zavrnejo:*

- **Dvostranska različica**: *če je* $|Z| \geq z_{1-\alpha/2}$*;*
- **Enostranska različica v desno**: *če je* $Z \geq z_{1-\alpha}$*;*
- **Enostranska različica v levo**: *če je* $Z \leq -z_{1-\alpha}$.

5. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, 5)$, dajo naslednje vrednosti:

$$101,\ 91,\ 93,\ 103,\ 91,\ 101,\ 103,\ 95,\ 95$$

Pri stopnji značilnosti $\alpha = 0{,}05$ testirajte ničelno hipotezo, da je $\mu = 100$, proti alternativni hipotezi, da je $\mu \neq 100$. Kaj pa, če bi za alternativno hipotezo vzeli $\mu < 100$ ali $\mu > 100$?

Sredina pri normalni porazdelitvi
z neznanim standardnim odklonom

Testiramo ničelno hipotezo $H_0\colon\ \mu = \mu_0$. *Ker pri* H_0 *velja:*

$$T := \frac{\bar{X} - \mu_0}{s}\sqrt{n} \sim \mathrm{S}(n-1)$$

kjer je s *definiran tako kot pri konstrukciji intervala zaupanja za ta primer, uporabimo* T**-test** *za testno statistiko* T *z* $n-1$ *prostostnimi stopnjami, in sicer:*

- **dvostransko različico**, *če* H_1 *trdi, da je* $\mu \neq \mu_0$*;*
- **enostransko različico v desno**, *če* H_1 *trdi, da je* $\mu > \mu_0$*;*
- **dvostransko različico v levo**, *če* H_1 *trdi, da je* $\mu < \mu_0$*;*

Zgoraj omenjene različice T*-testa za testno statistiko* T *z* df *prostostnimi stopnjami ničelno hipotezo zavrnejo:*

- **Dvostranska različica**: *če je* $|T| \geq t_{1-\alpha/2}(df)$*;*
- **Enostranska različica v desno**: *če je* $T \geq t_{1-\alpha}(df)$*;*
- **Enostranska različica v levo**: *če je* $T \leq -t_{1-\alpha}(df)$.

Tu je $t_p(df)$ *kvantil Studentove porazdelitve z* df *prostostnimi stopnjami za verjetnost* p.

6. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$109,\ 107,\ 95,\ 121,\ 107,\ 97,\ 111,\ 121,\ 95$$

Pri stopnji značilnosti $\alpha = 0{\dot{}}05$ testirajte ničelno hipotezo, da je $\mu = 100$, proti alternativni hipotezi, da je $\mu \neq 100$. Kaj pa, če bi vedeli, da je $\sigma = 10$?

7. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$46,\ 51,\ 48,\ 46,\ 52,\ 47,\ 51,\ 44,\ 47,\ 48$$

Pri stopnji značilnosti $\alpha = 0{\dot{}}05$ testirajte ničelno hipotezo, da je $\mu = 50$, proti alternativni hipotezi, da je $\mu \neq 50$. Kaj pa, če bi za alternativno hipotezo vzeli, da je $\mu < 50$?

Testiranje enakosti sredin za vzorce po parih

Večkrat hkrati izmerimo statistični spremenljivki X in Y. Privzamemo, da je $(X, Y) \sim \mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, pri čemer pa se μ_1 in μ_2 lahko spreminjata od meritve do meritve. Če meritve izhajajo iz enostavnega slučajnega vzorca iz velike populacije, to velja, če je porazdelitev vektorja (X, Y) na populaciji mešanica porazdelitev $\mathrm{N}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$, kjer se μ_1 in μ_2 lahko spreminjata. Testiramo ničelno hipotezo H_0, da je ves čas $\mu_1 = \mu_2$, alternativna hipoteza pa je lahko:

- *da je ves čas $\mu_X \geq \mu_Y$ in vsaj kdaj $\mu_X > \mu_Y$;*
- *da je ves čas $\mu_X \leq \mu_Y$ in vsaj kdaj $\mu_X < \mu_Y$;*
- *da velja ena izmed zgornjih dveh možnosti.*

Testiramo tako, da testiramo sredino razlike $X - Y$ glede na 0 (pri normalni porazdelitvi z neznanim standardnim odklonom: v prvih dveh primerih z enostranskim, v zadnjem primeru pa z dvostranskim testom).

POZOR! Če za alternativno hipotezo preprosto postavimo, da ni ves čas $\mu_X = \mu_Y$, ustrezni dvostranski test ni več **dosleden**, kar pomeni, da obstaja primer, ko velja alternativna hipoteza, a verjetnost, da hipotezo zavrnemo, z večanjem vzorca ne gre proti 1. To se zgodi, če je recimo na polovici populacije $\mu_X = \mu_Y + \delta$, na polovici pa $\mu_X = \mu_Y - \delta$.

8. Na desetih osebah so preizkušali učinek neke diete proti debelosti. Osebe so stehtali pred začetkom in po koncu diete. Podatki so naslednji:

Pred dieto	125	131	126	117	114
Po dieti	121	118	119	121	113

Pred dieto	134	123	135	100	117
Po dieti	118	111	130	97	118

Pri stopnji značilnosti $\alpha = 0{\dot{}}05$ testirajte ničelno hipotezo, da dieta nima učinka, proti alternativni hipotezi, da ima shujševalni učinek. Privzeti smete, da je vektor telesne teže pred in po dieti porazdeljen dvorazsežno normalno.

Testiranje enakosti sredin za neodvisne vzorce (homoskedastični model)

Naj bo $X \sim \mathrm{N}(\mu_1, \sigma)$ in $Y \sim \mathrm{N}(\mu_2, \sigma)$. Pri statistični spremenljivki X opazimo $X_1, \ldots, X_m$, pri Y pa $Y_1, \ldots, Y_n$. Privzamemo, da so vsa opažanja med seboj neodvisna. Če opažanja temeljijo na jemanju vzorcev, sta lahko X in Y definirani na različnih populacijah.

Testiramo hipotezo $H_0\colon\ \mu_1 = \mu_2$. Ker pri H_0 velja:

$$T := \frac{\bar{X} - \bar{Y}}{s} \sqrt{\frac{mn}{m+n}} \sim \mathrm{S}(m+n-2)\,,$$

kjer je:

$$s = \sqrt{\frac{(X_1 - \bar{X})^2 + \cdots + (X_m - \bar{X})^2 + (Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2}{m+n-2}}\,,$$

uporabimo T-test za testno statistiko T z $m+n-2$ prostostnimi stopnjami, in sicer:

- **dvostransko različico**, *če H_1 trdi, da je $\mu_1 \neq \mu_2$;*
- **enostransko različico v desno**, *če H_1 trdi, da je $\mu_1 > \mu_2$;*
- **dvostransko različico v levo**, *če H_1 trdi, da je $\mu_1 < \mu_2$;*

9. Vzorčne vrednosti statistične spremenljivke $X \sim \mathrm{N}(\mu_1, \sigma)$ iz prve populacije pridejo:

$$25,\ 16,\ 23,\ 17,\ 22,\ 18,\ 18,\ 21,\ 20\,,$$

vzorčne vrednosti statistične spremenljivke $Y \sim \mathrm{N}(\mu_2, \sigma)$ iz druge populacije pa pridejo:

$$19,\ 21,\ 23,\ 21,\ 25,\ 21,\ 24$$

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da je $\mu_1 = \mu_2$, proti alternativni hipotezi, da je $\mu_1 \neq \mu_2$.

10. Vzorčne vrednosti statistične spremenljivke $X \sim \mathrm{N}(\mu_1, \sigma)$ iz prve populacije pridejo:

$$102,\ 96,\ 103,\ 98,\ 105,\ 97,\ 103,\ 98,\ 100,\ 98,\ 99,\ 101$$

vzorčne vrednosti statistične spremenljivke $Y \sim \mathrm{N}(\mu_2, \sigma)$ iz druge populacije pa pridejo:

$$95,\ 97,\ 95,\ 99,\ 95,\ 95$$

Pri stopnji značilnosti $\alpha = 0{\cdot}01$ testirajte hipotezo, da je $\mu_1 = \mu_2$, proti alternativni hipotezi, da je $\mu_1 > \mu_2$.

Enakost sredin več normalnih statističnih spremenljivk: analiza variance (ANOVA) z enojno klasifikacijo

Danih je k populacij, na vsaki je definirana statistična spremenljivka, naj bodo to $X_1 \sim \mathrm{N}(\mu_1, \sigma)$, ..., $X_k \sim \mathrm{N}(\mu_k, \sigma)$. Iz vsake populacije vzamemo vzorec, pri čemer so vse enote vzorcev med seboj neodvisne. Vrednosti na vzorcu iz i-te populacije označimo z $X_{i1}, \ldots, X_{in_i}$. Testiramo hipotezo $H_0 \colon \mu_1 = \mu_2 = \ldots = \mu_k$, alternativna hipoteza H_1 pa je nasprotje H_0. Izračunajmo:

$$\bar{X}_i := \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}\,, \quad n := \sum_{i=1}^{k} n_i\,, \quad \bar{X} := \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} X_{ij} = \sum_{i=1}^{k} \frac{n_i}{n} \bar{X}_i\,,$$

$$S_{\mathrm{B}}^2 := \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2\,, \quad S_{\mathrm{W}}^2 := \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2\,.$$

Če velja H_0, sta S_{B}^2 in S_{W}^2 neodvisni ter $S_{\mathrm{B}}^2/\sigma^2 \sim \chi^2(k-1)$ in $S_{\mathrm{W}}^2/\sigma^2 \sim \chi^2(n-k)$, zato je:

$$F := \frac{S_{\mathrm{B}}^2/(k-1)}{S_{\mathrm{W}}^2/(n-k)} \sim \mathrm{F}(k-1, n-k)$$

kjer je $\mathrm{F}(k-1, n-k)$ Snedecorjeva porazdelitev. V skladu s tem uporabimo F**-test** *za testno statistiko F s $(k-1, n-k)$ prostostnimi stopnjami, in sicer enostransko različico v desno.*

F-test za testno statistiko F z (df_1, df_2) prostostnimi stopnjami ima sicer spet tri različice:

- **Dvostranska različica**: *H_0 zavrnemo, če je $F \le F_{\alpha/2}(df_1, df_2)$ ali $F \ge F_{1-\alpha/2}(df_1, df_2)$;*
- **Enostranska različica v desno**: *H_0 zavrnemo, če je $F \ge F_{1-\alpha}(df_1, df_2)$;*
- **Enostranska različica v levo**: *H_0 zavrnemo, če je $F \le F_{\alpha}(df_1, df_2)$.*

Tu je $F_p(df_1, df_2)$ kvantil Snedecorjeve porazdelitve z (df_1, df_2) prostostnimi stopnjami za verjetnost p.

11. Pacientom, ki so jim dajali določena zdravila, so merili neki parameter. Meritve so dale naslednje vrednosti:

Aspirin:	3, 5, 3, 5
Tilenol:	2, 2, 4, 4
Placebo:	2, 1, 2

Pri stopnji značilnosti $\alpha = 0{,}05$ testirajte hipotezo, da je vrednost parametra neodvisna od tega, ali pacient jemlje katero izmed obeh zdravil ali pa sploh nobenega.

Standardni odklon pri normalni porazdelitvi

Testiramo ničelno hipotezo $H_0\colon\ \sigma = \sigma_0$. Ker pri H_0 velja:

$$\chi^2 := (n-1)\,\frac{s^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2(n-1)$$

uporabimo **test hi hvadrat** *za testno statistiko χ^2 z $n-1$ prostostnimi stopnjami, in sicer:*

- **dvostransko različico**, *če H_1 trdi, da je $\sigma \neq \sigma_0$;*
- **enostransko različico v desno**, *če H_1 trdi, da je $\sigma > \sigma_0$;*
- **enostransko različico v levo**, *če H_1 trdi, da je $\sigma < \sigma_0$.*

Omenjene različice testa hi kvadrat za testno statistiko χ^2 z df prostostnimi stopnjami zavrnejo ničelno hipotezo:

- **Dvostranska različica**: *če je $\chi^2 \le \chi^2_{\alpha/2}(df)$ ali $\chi^2 \ge \chi^2_{1-\alpha/2}(df)$;*
- **Enostranska različica v desno**: *če je $\chi^2 \ge \chi^2_{1-\alpha}(df)$;*
- **Enostranska različica v levo**: *če je $\chi^2 \le \chi^2_{\alpha}(df)$.*

Tu je $\chi^2_p(df)$ kvantil porazdelitve hi kvadrat z df prostostnimi stopnjami za verjetnost p.

12. Meritve neke količine, porazdeljene normalno $\mathrm{N}(\mu, \sigma)$, dajo naslednje vrednosti:

$$99,\ 90,\ 108,\ 111,\ 97,\ 93,\ 90,\ 106,\ 104,\ 102$$

Pri $\alpha = 0{,}05$ testirajte:

a) ničelno hipotezo, da je $\sigma = 5$, proti alternativni hipotezi, da je $\sigma \neq 5$;

b) ničelno hipotezo, da je $\sigma = 10$, proti alternativni hipotezi, da je $\sigma < 10$.

Pearsonov test skladnosti s fiksno porazdelitvijo

Testiramo, ali je porazdelitev dane statistične spremenljivke enaka:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix} \qquad (r \geq 3)\,.$$

Pri tem so lahko $a_1, \ldots, a_r$ dejanske vrednosti spremenljivke ali pa le razredi, v katere pade. Vzorec velikosti n ima frekvenčno porazdelitev:

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_r \\ N_1 & N_2 & \cdots & N_r \end{pmatrix}$$

Izračunamo **pričakovane frekvence** $\tilde{N}_i := np_i$. *Ker tedaj pri veljavnosti ničelne hipoteze približno velja:*

$$\chi^2 := \sum_{i=1}^{r} \frac{(N_i - \tilde{N}_i)^2}{\tilde{N}_i} \sim \chi^2(r-1)$$

hipotezo o porazdelitvi testiramo s testom hi kvadrat za testno statistiko χ^2 z $r-1$ prostostnimi stopnjami, in sicer enostransko v desno.
Za naše potrebe je test dovolj natančen, če je $r \geq 3$ in $\tilde{N}_i \geq 5$ za vse i. Če dobimo $\tilde{N}_i < 5$, lahko razrede združimo. Za $r = 2$ pa lahko uporabimo kar dvostranski test uspeha poskusa.

13. V vzorcu so 2 osebka tipa RR, 5 tipa Rr in 4 tipa rr. Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte hipotezo, da je v populaciji 25% osebkov tipa RR, 50% tipa Rr in 25% tipa rr. Kaj pa, če bi bilo v vzorcu 20 osebkov tipa RR, 50 tipa Rr in 50 tipa rr?

14. Pri kvizu Lepo je biti milijonar od 22. novembra do 28. decembra 2003 je bil 21-krat pravilen odgovor A, 42-krat B, 77-krat C in 116-krat D. Pri stopnji značilnosti $\alpha = 0\text{'}01$ testirajte ničelno hipotezo, da so odgovori enakomerno porazdeljeni, proti alternativni hipotezi, da niso.

    Če izvzamemo prvih pet vprašanj, je bil A pravilen 21-krat, B 37-krat, C 53-krat in D 25-krat. Naredite isti test.

15. Podatki imajo naslednjo frekvenčno porazdelitev:

pod 10	10–20	20–30	30–40	nad 40
5	20	35	25	15

    Pri stopnji značilnosti $\alpha = 0\text{'}05$ testirajte hipotezo, da podatki izhajajo iz populacije, porazdeljene normalno $\mathrm{N}(25, 10)$.

Test z znaki

Večkrat hkrati izmerimo urejenostni statistični spremenljivki X in Y. Privzamemo, da sta X in Y ob vsaki meritvi neodvisni, sicer pa se lahko porazdelitvi od meritve do meritve spreminjata. Če meritve izhajajo iz enostavnega slučajnega vzorca iz velike populacije, to velja, če je porazdelitev vektorja (X, Y) na populaciji mešanica porazdelitev parov neodvisnih slučajnih spremenljivk.

Testiramo ničelno hipotezo H_0, da sta X in Y ves čas (ob vsaki meritvi) enako porazdeljeni. Za formulacijo alternativne hipoteze pa moramo razumeti **stohastično primerjavo** *porazdelitev slučajnih spremenljivk. Pravimo, da je slučajna spremenljivka X* **stohastično večja** *od Y, če je $F_X \le F_Y$, in X* **stohastično manjša** *od Y, če je $F_X \ge F_Y$. Alternativna hipoteza je lahko:*

- *H_1^+, da je X ves čas stohastično večja od Y in da X in Y nista ves čas enako porazdeljeni;*
- *H_1^-, da je X ves čas stohastično manjša od Y in da X in Y nista ves čas enako porazdeljeni;*
- *$H_1^\pm$, da velja ena od zgornjih dveh možnosti.*

Naj bo n število meritev. Z S^+ označimo število meritev, pri katerih je $X > Y$, z S^- pa število meritev, pri katerih je $X < Y$. Naj bo še $\tilde{n} = S^+ + S^-$ število meritev, pri katerih je $X \ne Y$ (primere, kjer pride enako, torej preprosto izločimo). Ničelno hipotezo zavrnemo, če je $p(S^+, S^-) \le \alpha$. Funkcija p (p-vrednost) je odvisna od ničelne hipoteze in ustreza testu uspeha poskusa pri ničelni hipotezi, da je le-ta enaka $1/2$:

- *pri H_1^+ postavimo $p(k^+, k^-) = P(S' \ge k^+) = P(S' \le k^-)$;*
- *pri H_1^- postavimo $p(k^+, k^-) = P(S' \ge k^-) = P(S' \le k^+)$;*
- *pri $H_1^\pm$ postavimo $p(k^+, k^-) = 2\min\{P(S' \ge k^+), P(S' \ge k^-)\} = 2\min\{P(S' \le k^+), P(S' \le k^-)\}$.*

Tu je $S' \sim \mathrm{B}(\tilde{n}, 1/2)$.

16. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	28	14	16	16	31	17	13	14	12	13
Y_i	26	29	31	18	37	10	19	33	23	45

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti alternativni, da je X ves čas stohastično manjša od Y in vsaj kdaj stohastično strogo manjša od Y. Nato naredite ustrezni enostranski test povprečij.

17. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	25	28	30	23	28	26	29	23	33	21	33	28
Y_i	35	27	29	21	18	25	28	27	31	19	32	26

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti alternativni, da je X ves čas stohastično večja od Y in vsaj kdaj stohastično strogo večja od Y. Nato naredite ustrezni enostranski test povprečij.

18. Meritve količin X in Y so zapisane v naslednji tabeli:

X_i	30	24	22	28	26	19	25	31	36	21	25	26	29	29	19	18
Y_i	28	121	21	25	25	17	122	129	34	20	22	23	126	26	18	17

.

Pri stopnji značilnosti $\alpha = 0{\dot{}}05$ testirajte hipotezo, da sta X in Y vsakič enako porazdeljeni, proti obema enostranskima alternativnima hipotezama. Nato naredite še ustrezna enostranska testa povprečij.

Test z znaki za veliko število meritev

Če je število meritev dovolj veliko (za naše potrebe vsaj 10), lahko test z znaki nadomestimo s približnim testom, pri katerem ničelno hipotezo zavrnemo:

- *proti* H_1^+, *če je* $\dfrac{S_+ - S_- - 1}{\sqrt{\tilde{n}}} \geq z_{1-\alpha}$;
- *proti* H_1^-, *če je* $\dfrac{S_+ - S_- + 1}{\sqrt{\tilde{n}}} \leq -z_{1-\alpha}$.
- *proti* $H_1^\pm$, *če je* $\dfrac{|S_+ - S_-| - 1}{\sqrt{\tilde{n}}} \geq z_{1-\alpha/2}$;

19. 50 ljudi so pred ogledom in po ogledu filma povpraševali, kako se počutijo: zelo slabo, slabo, srednje, dobro ali zelo dobro. Rezultati so naslednji:

pred	po
srednje	srednje
dobro	zelo dobro
srednje	zelo dobro
dobro	srednje
srednje	zelo dobro
dobro	dobro
srednje	dobro
dobro	dobro
dobro	zelo dobro
zelo dobro	zelo dobro
dobro	zelo dobro
zelo dobro	dobro
dobro	srednje
zelo dobro	srednje
srednje	dobro
srednje	dobro
dobro	zelo dobro
srednje	dobro
dobro	zelo dobro
zelo dobro	dobro
zelo dobro	zelo dobro
zelo dobro	dobro
slabo	dobro
dobro	srednje
srednje	zelo dobro

pred	po
dobro	zelo dobro
dobro	dobro
zelo dobro	zelo dobro
dobro	dobro
srednje	zelo slabo
srednje	zelo dobro
zelo dobro	srednje
dobro	dobro
dobro	dobro
srednje	slabo
slabo	srednje
srednje	srednje
zelo slabo	slabo
slabo	srednje
slabo	srednje
slabo	zelo dobro
zelo slabo	srednje
srednje	slabo
srednje	slabo
zelo slabo	srednje
srednje	dobro
slabo	zelo dobro
slabo	slabo
slabo	slabo
zelo slabo	srednje

Pri stopnji značilnosti $\alpha = 0{,}05$ testirajte ničelno hipotezo, da ogled filma ne spremeni počutja, proti alternativni hipotezi, da ga spremeni.

Inverzijski (Wilcoxon–Mann–Whitneyjev) test

Testiramo, ali sta **urejenostni** *statistični spremenljivki X in Y enako porazdeljeni. Pri statistični spremenljivki X opazimo $X_1, \ldots, X_m$, pri Y pa $Y_1, \ldots, Y_n$. Privzamemo, da so vsa opažanja med seboj neodvisna. Če opažanja temeljijo na jemanju vzorcev, sta lahko X in Y definirani na različnih populacijah.*

Opažene vrednosti združimo in jih uredimo po velikosti, recimo od namanjše do največje. Naj bodo $R_1, \ldots, R_m$ mesta (rangi), ki pripadajo opažanjem spremenljivke X. Privzemimo, da sta vzorca dovolj velika.

Tudi tu ločimo enostransko in dvostransko različico testa. Obravnavali bomo torej tri alternativne hipoteze:

- *H_1^+, da je X stohastično strogo večja od Y, t. j. $F_X \le F_Y$, za določen x pa tudi $F_X(x) < F_Y(x)$;*
- *H_1^-, da je X stohastično strogo manjša od Y, t. j. $F_X \ge F_Y$, za določen x pa tudi $F_X(x) > F_Y(x)$;*
- *$H_1^\pm$, da velja ena od zgornjih dveh možnosti.*

Če je število meritev dovolj veliko, ničelno hipotezo zavrnemo:

- *proti H_1^+, če je $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1) - 1\right) \ge z_{1-\alpha}$;*
- *proti H_1^-, če je $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1) + 1\right) \le -z_{1-\alpha}$;*
- *proti $H_1^\pm$, če je $\sqrt{\dfrac{3}{mn(m+n+1)}}\left(\left|2\sum_{i=1}^m R_i - m(m+n+1)\right| - 1\right) \ge z_{1-\alpha/2}$.*

Z drugimi besedami, izvedemo Z-test na testni statistiki

$$\sqrt{\frac{3}{mn(m+n+1)}}\left(2\sum_{i=1}^m R_i - m(m+n+1)\right) \text{ s popravkom } \sqrt{\frac{3}{mn(m+n+1)}}.$$

POZOR! Če bi dvostranska alternativna hipoteza preprosto trdila, da je spremenljivka na prvi skupini drugače porazdeljena kot na drugi, test ne bi bil več dosleden!

20. Tekmovalci dveh ekip, "zelenih" in "oranžnih", so se pomerili v teku. Vrstni red tekmovalcev je naslednji:

$$Z,\ Z,\ O,\ Z,\ Z,\ O,\ Z,\ Z,\ O,\ Z,\ O,\ O,\ O,\ Z,\ O,\ O,\ O,\ O,\ Z,\ O$$

(t. j. prvi, ki je prispel na cilj, je bil član "zelenih", drugi prav tako, tretji je bil član "oranžnih" itd.). Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte hipotezo, da so zeleni enako dobri kot oranžni, proti alternativni hipotezi, da je med njimi razlika.

21. Dijaki so se pomerili v teku na 60 metrov. Določeni so izjavili, da so se prej pripravljali, določeni pa, da ne. Rezultati tistih, ki so se pripravljali, so:

$$7{\cdot}6\,,\ 7{\cdot}6\,,\ 7{\cdot}7\,,\ 7{\cdot}8\,,\ 7{\cdot}8\,,\ 8{\cdot}0\,,\ 8{\cdot}1\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 9{\cdot}3\,,$$

rezultati tistih, ki se niso pripravljali, pa so:

$$7{\cdot}9\,,\ 8{\cdot}2\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}3\,,\ 8{\cdot}4\,,\ 8{\cdot}7\,,\ 8{\cdot}8\,.$$

Z inverzijskim testom testirajte ničelno hipotezo, da tisti, ki se pripravljajo, tečejo enako kot tisti, ki se ne pripravljajo, proti alternativni hipotezi, da tisti, ki se pripravljajo, tečejo bolje od tistih, ki se ne pripravljajo. Kaj pa pravi T-test? Stopnja značilnosti naj bo obakrat $\alpha = 0{\cdot}05$.

22. Med 17 študenti so izvedli anketo z naslednjima vprašanjema:

    1. Ocenite stopnjo stresa pri vas v zadnjih dveh tednih.
       (1 – zelo majhna, 2 – majhna, 3 – srednja, 4 – velika, 5 – zelo velika)
    2. Ali ste se v zadnjih dveh tednih posvečali študiju bolj kot ponavadi?
       (da/ne)

    Rezultati ankete so:

	1	2	3	4	5
da	0	2	1	5	2
ne	0	5	2	0	0

    Pri stopnji značilnosti $\alpha = 0{\cdot}01$ testirajte hipotezo, ali so študenti, ki se v zadnjih dveh tednih bolj posvečajo študiju, v tem času enako pod stresom kot tisti, ki se ne, proti alternativni hipotezi, da so študenti, ki se v zadnjih dveh tednih bolj posvečajo študiju, v tem času bolj pod stresom kot tisti, ki se ne.

10. Povezanost dveh številskih spremenljivk

Interval zaupanja za korelacijski koeficient. Testiranje nekoreliranosti. Kontingenčni test. Enostavna linearna regresija.

Interval zaupanja za korelacijski koeficient

Naj bo $\rho = \rho(X, Y)$. Najprej izračunamo **vzorčni korelacijski koeficient** *R:*

$$\begin{aligned} C_x^2 &= \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \\ C_y^2 &= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 \\ C_{xy} &= \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} \\ R &= \frac{C_{xy}}{C_x C_y} \end{aligned}$$

in ga normaliziramo:

$$Z := \operatorname{Arth} R = \frac{1}{2} \ln \frac{1+R}{1-R}$$

Približen interval zaupanja za ρ:

$$\operatorname{th}\left(Z - \frac{c}{\sqrt{n-3}}\right) \le \rho \le \operatorname{th}\left(Z + \frac{c}{\sqrt{n-3}}\right)$$

kjer je $c = z_{(1+\beta)/2}$ in $\operatorname{th}(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$.

1. Meritve krvnega pritiska so zbrane v naslednji tabeli:

sistolični	130	120	120	125	125	125	105	130	130	135
diastolični	80	80	85	80	75	75	75	80	85	70

sistolični	130	125	140	130	120
diastolični	75	80	90	80	85

Poiščite 95% interval zaupanja za korelacijski koeficient med sistoličnim in diastoličnim pritiskom.

Testiranje nekoreliranosti (Gaussov model)

Če velja ničelna hipoteza H_0, da sta X in Y nekorelirani, je približno:

$$T := \frac{R}{\sqrt{1-R^2}}\sqrt{n-2} \sim \mathrm{S}(n-2)$$

kjer je R vzorčni korelacijski koeficient. Tako nekoreliranost testiramo s T-testom za testno statistiko T z $n-2$ prostostnimi stopnjami, in sicer z:

- *dvostransko različico, če H_1 trdi, da sta X in Y korelirani;*
- *enostransko različico v desno, če H_1 trdi, da sta X in Y pozitivno korelirani;*
- *enostransko različico v levo, če H_1 trdi, da sta X in Y negativno korelirani.*

2. Za meritve krvnega pritiska iz 1. naloge pri stopnji značilnosti $\alpha = 0{\cdot}1$ testirajte ničelno hipotezo, da sta sistolični in diastolični pritisk nekorelirana, proti alternativni hipotezi, da sta korelirana.

Kontingenčni test neodvisnosti

Testiramo, ali sta spremenljivki X in Y, definirani na isti populaciji, neodvisni, pri čemer v vzorcu kombinacija $X = a_i, Y = b_j$ nastopa N_{ij}-krat ($i = 1, \ldots, r$, $j = 1, \ldots, s$). Ker pri neodvisnosti približno velja:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(N_{ij}-\tilde{N}_{ij})^2}{\tilde{N}_{ij}} \sim \chi^2\big((r-1)(s-1)\big)$$

kjer je:

$$N_{i.} = \sum_{j=1}^{s} N_{ij}\,, \quad N_{.j} = \sum_{i=1}^{r} N_{ij}\,, \quad n = \sum_{i=1}^{n}\sum_{j=1}^{n} N_{ij}\,, \quad \tilde{N}_{ij} = \frac{N_{i.}N_{.j}}{n}$$

neodvisnost testiramo s testom hi kvadrat z $(r-1)(s-1)$ prostostnimi stopnjami, in sicer z enostransko različico v desno. Za naše potrebe je test dovolj natančen, če je $\tilde{N}_{ij} \geq 5$ za vse i in j.

3. Na vzorcu 62 oseb dobimo naslednjo navzkrižno frekvenčno porazdelitev barve oči in las:

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	12	1	13
zelene	14	9	23
rjave	4	22	26
Skupaj	30	32	62

Pri stopnji značilnosti $\alpha = 0{\cdot}01$ testirajte hipotezo, da sta barva las in barva oči neodvisni.

Kontingenčni test neodvisnosti za dihotomni spremenljivki

Če sta spremenljivki X in Y dihotomni in so navzkrižne frekvence podane s tabelo:

A	B
C	D

velja:

$$\chi^2 = \frac{(A+B+C+D)(AD-BC)^2}{(A+B)(C+D)(A+C)(B+D)} .$$

4. Rezultati ankete z dvema vprašanjema 'Ali verjamete v horoskop?' in 'Ali verjamete v NLP-je?' so zbrani v naslednji tabeli:

Horoskop\NLP	vsaj malo	ne	Skupaj
vsaj malo	5	7	12
ne	6	9	15
Skupaj	11	16	27

Pri stopnji značilnosti $\alpha = 0{\cdot}05$ testirajte neodvisnost verovanja ljudi v horoskop in v NLP-je.

5. Na neki spletni strani je 1990 ljudi glasovalo, kateri film bo najverjetneje dobil oskarja. Anketirance so razdelili v tri razrede glede na starost. Rezultati ankete so naslednji:

	do 20	od 20 do 40	nad 40
Gospodar prstanov	350	250	180
Skrivnostna reka	80	100	100
Seabiscuit	70	90	130
Zgubljeno s prevodom	50	80	110
Gospodar in bojevnik	200	150	50

S kontingenčnim testom preizkusite domnevo, da je mnenje o oskarjih neodvisno od starosti. Stopnja značilnosti naj bo $0{\cdot}05$.

Enostavna linearna regresija

Spremenljivki X in Y zadoščata zvezi:

$$Y = a + bX + R$$

kjer je $R \sim \mathrm{N}(0, \sigma)$ neodvisna od X in kjer so a, b in σ neznani parametri. Želeli bi točkasto oceniti a in b (t. j. potegniti premico skozi podatke), poleg tega pa še točkasto in intervalsko oceniti vrednost spremenljivke Y za dano realizacijo, pri kateri poznamo X. Cenilki za a in b sta:

$$\hat{b} = \frac{C_{xy}}{C_x^2}\,, \quad \hat{a} = \bar{Y} - \hat{b}\bar{X}$$

Cenilka za Y pri danem X pa je:

$$\hat{Y} = \hat{a} + \hat{b}X$$

Interval zaupanja:

$$\hat{Y} - \Delta \le Y \le \hat{Y} + \Delta$$

kjer je:

$$\Delta = t_{(1+\beta)/2}(n-2)\, S \sqrt{1 + \frac{1}{n} + \frac{(X - \bar{X})^2}{C_x^2}}$$

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{a} - \hat{b}X_i)^2$$

6. Meritve dajo naslednje vrednosti:

X_i	1	2	3	4	5
Y_i	2	6	7	10	10

Določite regresijsko premico, napovejte Y pri $X = 10$ in poiščite 95% interval zaupanja.

REŠITVE

1. Osnove kombinatorike

1. $12 + 3 = 15$.

2. Če mora obleči natanko eno srajco, natanko ene hlače in natanko en pulover, na $4 \cdot 2 \cdot 3 = 24$ načinov. Če ni nujno, da nosi pulover, pa na $5 \cdot 2 \cdot 3 = 30$ načinov.

3. $(4 + 3) \cdot 5 = 35$.

4. Možnih vozovnic je $7 \cdot 6 = 42$, možnih cen pa $7 \cdot 6/2 = 21$.

5. Dejansko se mora naučiti produkte števil od 2 do 9, pri čemer lahko upošteva še komutativnost. Teh pa je:

$$\frac{8 \cdot 7}{2} + 8 = \frac{8 \cdot 9}{2} = 1 + 2 + \cdots + 8 = 36\,.$$

6. Rezka $5! = 120$, Tatjana pa $\dfrac{7!}{2!\,3!} = 420$.

7. $15 + 21 - 3 = 33$.

8. $10 \cdot 9 \cdot 8 = 720$.

9. $\dfrac{10 \cdot 9 \cdot 8}{3!} = \dbinom{10}{3} = 120$.

2. Elementarna verjetnost

1. $1/2$.

2. $1/2$.

3. $10/2^{10} \doteq 0\text{˙}00977$.

4. $P(6 \text{ pik}) = P(1 \text{ pika}) = 1/6$. $P(\text{sodo mnogo pik}) = 1/2$. $P(\text{vsaj tri pike}) = 2/3$.

5. $1/6$ (oboje).

6. $\dfrac{1 \cdot 1 \cdot 1 \cdot 5 \cdot 6}{6^5} + \dfrac{5 \cdot 1 \cdot 1 \cdot 1 \cdot 5}{6^5} + \dfrac{6 \cdot 5 \cdot 1 \cdot 1 \cdot 1}{6^5} = \dfrac{85}{7776} \doteq 0\text{˙}0109$.

7. $P(\text{pik}) = 1/4$, $P(\text{as}) = 1/8$, $P(\text{pik ali as}) = 11/32$.

8. $1 - \dfrac{\binom{27}{4}}{\binom{32}{4}} = 1 - \dfrac{\binom{28}{5}}{\binom{32}{5}} \doteq 0\text{˙}512$.

9. Verjetnosti dobitkov lahko računamo na dva načina. Pri prvem načinu si predstavljamo, da so prekrižane številke fiksne, nakar gledamo vsa možna žrebanja (*pogled igralca*). Lahko pa si predstavljamo tudi, da so fiksne izžrebane številke, nakar gledamo vsa možna križanja (pogled Loterije). Dobimo:

$$
\begin{aligned}
P(\text{sedmica}) &= \frac{1}{\binom{39}{7}} = \frac{1}{15.380.937} \doteq 6\text{˙}50 \cdot 10^{-8} \\
P(\text{šest in dodatna}) &= \frac{\binom{7}{6}\binom{32}{1}}{\binom{39}{7} \cdot 32} = \frac{\binom{7}{6}\binom{1}{1}\binom{31}{0}}{\binom{39}{7}} \approx \frac{1}{2.197.277} \doteq 4\text{˙}55 \cdot 10^{-7} \\
P(\text{šestica}) &= \frac{\binom{7}{6}\binom{32}{1} \cdot 31}{\binom{39}{7} \cdot 32} = \frac{\binom{7}{6}\binom{1}{0}\binom{31}{1}}{\binom{39}{7}} \approx \frac{1}{70.880} \doteq 1\text{˙}41 \cdot 10^{-5} \\
P(\text{petica}) &= \frac{\binom{7}{5}\binom{32}{2}}{\binom{39}{7}} \approx \frac{1}{1477} \doteq 6\text{˙}77 \cdot 10^{-4} \\
P(\text{štirica}) &= \frac{\binom{7}{4}\binom{32}{3}}{\binom{39}{7}} \approx \frac{1}{88\text{˙}6} \doteq 0\text{˙}0113 \\
P(\text{tri in dodatna}) &= \frac{\binom{7}{3}\binom{32}{4} \cdot 4}{\binom{39}{7} \cdot 32} = \frac{\binom{7}{3}\binom{1}{1}\binom{31}{4}}{\binom{39}{7}} \approx \frac{1}{97\text{˙}8} \doteq 0\text{˙}0102
\end{aligned}
$$

10. Če vplačamo sistem k številk, je verjetnost sedmice enaka $\binom{k}{7}/\binom{39}{7}$. Očitno je ta izraz naraščajoč v k. Tabela nekaj verjetnosti:

k	$P(\text{sedmica})$
39	1
38	0˙821
37	0˙669
36	0˙543
35	0˙437

Torej bi bilo potrebno vplačati sistem najmanj 36 številk.

11. Dogodek, čigar verjetnost iščemo, je disjunktna unija naslednjih treh dogodkov:

- Asistent ne premesti niti Aljaža niti Brigite. Naj bo to dogodek S_0.
- Asistent premesti Aljaža ali Brigito, Cveta in še enega od ostalih, pri čemer pa mora Cveto v ciklu premestitev slediti Aljažu oz. Brigiti. Naj bo to dogodek S_1.
- Asistent premesti Aljaža, Brigito in Cveta, pri čemer pa mora Cveto v ciklu premestitev slediti Brigiti. Naj bo to dogodek S_2.

Verjetnostni prostor lahko razdelimo na $9 \cdot 8 \cdot 7 = 504$ enako verjetne izide – glede na to, katere študente asistent izbere za premestitev in v katerem vrstnem redu. Vsak tak izid lahko ponazorimo z zaporedjem črk oblike xyz, pri čemer bo črka A označevala Aljaža, črka B Brigito, črka C Cveta, črke X, Y in Z katerega koli, ki ni Aljaž ali Brigita, črka W pa katerega koli, ki ni Aljaž, Brigita ali Cveto.

- Dogodek S_0 sestavljajo izidi oblike XYZ, ki jih je $7 \cdot 6 \cdot 5 = 210$.
- Dogodek S_1 sestavljajo vsi izidi oblike ACW, BCW, WAC, WBC, CWA in CWB. Teh je $6 \cdot 6 = 36$.
- Dogodek S_2 pa sestavljajo izidi oblike ABC, BCA in CAB, ki so trije.

Iskana verjetnost je tako enaka:

$$P(S_0) + P(S_1) + P(S_2) = \frac{210 + 36 + 3}{504} = \frac{249}{504} = \frac{83}{168} \doteq 0˙494 \,.$$

Opomba. Po trije in trije izidi, kot smo jih definirali, določajo enako premestitev. Tako bi lahko za izide vzeli tudi kar premestitve. Teh je $\dfrac{9 \cdot 8 \cdot 7}{3} = 2 \cdot \binom{9}{3} = 168$.

3. Pogojna verjetnost

1. $P(A) = \frac{1}{1024} \doteq 0{\cdot}000977$, $P(B) = \frac{10}{1024} \doteq 0{\cdot}00977$,
$P(A \mid C) = P(B \mid C) = \frac{1}{2} = 0{\cdot}5$.

2. a) $\frac{5}{7} \doteq 0{\cdot}714$, b) $\frac{5}{7} \cdot \frac{4}{6} = \frac{10}{21} \doteq 0{\cdot}476$, c) $\frac{5}{7} \cdot \frac{2}{6} = \frac{5}{21} \doteq 0{\cdot}238$, d) $\frac{2}{7} \cdot \frac{5}{6} = \frac{5}{21} \doteq 0{\cdot}238$,
e) $\frac{5}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} = \frac{2}{7} \doteq 0{\cdot}286$, f) 0.

3. a) $\frac{5}{7} \doteq 0{\cdot}714$, b) $\frac{2}{3} \doteq 0{\cdot}667$, c) 1, d) $\frac{5}{7}\Big/\left(1 - \frac{2}{7} \cdot \frac{1}{6}\right) = \frac{3}{4} = 0{\cdot}75$, e) $\frac{1}{2} = 0{\cdot}5$.

4. a) $\frac{1}{8} = 0{\cdot}125$, b) $\frac{1}{2}\left(1 - \frac{3}{31}\right) = \frac{14}{31} \doteq 0{\cdot}452$, c) $\frac{28}{31} \doteq 0{\cdot}903$, d) $\frac{1}{8} \cdot \frac{28}{31}\Big/\frac{14}{31} = \frac{1}{4} = 0{\cdot}25$.

5. A in B sta neodvisna, prav tako tudi A in D ter B in D.
A in C sta odvisna.
A, B in D so odvisni.

6. $0{\cdot}4 \cdot 0{\cdot}7 \cdot 0{\cdot}4 + \big(0{\cdot}6 \cdot 0{\cdot}7 + 0{\cdot}4 \cdot 0{\cdot}3\big) \cdot 0{\cdot}8 + 0{\cdot}6 \cdot 0{\cdot}3 \cdot 1 = 0{\cdot}724$.

7. $\dfrac{0{\cdot}6 \cdot 0{\cdot}2}{0{\cdot}6 \cdot 0{\cdot}2 + 0{\cdot}4 \cdot 0{\cdot}1} = 0{\cdot}75$.

8. $\dfrac{0{\cdot}6 \cdot 0{\cdot}7 \cdot 0{\cdot}8 + 0{\cdot}6 \cdot 0{\cdot}3 \cdot 1}{0{\cdot}4 \cdot 0{\cdot}7 \cdot 0{\cdot}4 + \big(0{\cdot}6 \cdot 0{\cdot}7 + 0{\cdot}4 \cdot 0{\cdot}3\big) \cdot 0{\cdot}8 + 0{\cdot}6 \cdot 0{\cdot}3 \cdot 1} \doteq 0{\cdot}713$.

9. a) $\dfrac{0{\cdot}1}{1 - 0{\cdot}9 \cdot 0{\cdot}8 \cdot 0{\cdot}7} \doteq 0{\cdot}202$,

b) $\dfrac{0{\cdot}1 \cdot 0{\cdot}8 \cdot 0{\cdot}7}{0{\cdot}1 \cdot 0{\cdot}8 \cdot 0{\cdot}7 + 0{\cdot}9 \cdot 0{\cdot}2 \cdot 0{\cdot}7 + 0{\cdot}9 \cdot 0{\cdot}8 \cdot 0{\cdot}3} \doteq 0{\cdot}141$.

4. Diskretne slučajne spremenljivke

1. $G = \{\mathrm{gg}, \mathrm{gc}, \mathrm{cg}, \mathrm{cc}\}$, S kot preslikava:

ω	gg	gc	cg	cc
$P(\omega)$	1/4	1/4	1/4	1/4
$S(\omega)$ – primer a)	0	2	1	3
$S(\omega)$ – primer b)	0	2	2	4

Porazdelitev slučajne spremenljivke S:

Primer a): $S \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix}$.

Primer b): $S \sim \begin{pmatrix} 0 & 2 & 4 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}$.

2. a) $S \sim \mathrm{B}(5, 1/6)$, torej za $k = 0, 1, 2, 3, 4, 5$ velja $P(S = k) = \binom{n}{k}\left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$. Približna porazdelitev:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 0.402 & 0.402 & 0.161 & 0.0322 & 0.00322 & 0.000129 \end{pmatrix}.$$

b) $P(S\,\mathrm{sodo}) = P(S = 0) + P(S = 2) + P(S = 4) \doteq 0{\cdot}566$,
c) $P(S \ge 3) \doteq P(S = 3) + P(S = 4) + P(S = 5) \doteq 0{\cdot}0355$,
d) $P(S\,\mathrm{sodo} \mid S \ge 3) = \dfrac{P(S = 4)}{P(S = 3) + P(S = 4) + P(S = 5)} \doteq 0{\cdot}0906$.

3. $M \sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{pmatrix}$.

4. $M \sim \begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 1/4 & 1/8 & 1/8 & 3/32 & 5/64 & 21/64 \end{pmatrix}$.

5. $P(X = 0) = 0{\cdot}3$, $E(X) = 2$, $D(X) = 4{\cdot}2$, $\sigma(X) \doteq 2{\cdot}05$.

6. $a = 0{\cdot}7$, $b = 0$.

7. a) Iz $E(X) = 1002 + b$ dobimo $\min E(X) = 1002$ in $\max E(X) = 1002{\cdot}7$.
b) Iz $D(X) = 0{\cdot}6 + b - b^2$ dobimo, da je $\min D(X) = 0{\cdot}6$ in $\max D(X) = 0{\cdot}85$.

8. $E(X_1) = 2$, $D(X_1) = 1{\cdot}2$.
$E(S) = 4$ v obeh primerih.
$D(S) = 2{\cdot}4$, če vračamo, in $32/15$, če ne vračamo.

9. $P(S > 0) \doteq 0{\cdot}598$, $E(S) \doteq -0{,}033$€, $\sigma(S) \doteq 1{,}67$€.

10. Če vračamo, je $K(X_1, X_2) = 0$ in seveda tudi $\rho(X_1, X_2) = 0$.
Če ne vračamo, je $K(X_1, X_2) = -2/15$ in $\rho(X_1, X_2) = -1/9$.

11. Slučajni spremenljivki X in Y sta odvisni, čeprav sta nekorelirani. To lahko vidimo iz poljubne navzkrižne verjetnosti.

12. Dopolnjena tabela:

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	0'1	0'2	0'1
$X = 2$	0'15	0'3	0'15

5. Zvezne slučajne spremenljivke

1. $c = 1/9$, $P(1 < X < 2) = 7/27$, $P(X \le 1) = 1/27$,
$E(X) = 9/4$, $D(X) = 27/80$.

2. $F_X(x) = \begin{cases} 0 & ; x \le 0 \\ \frac{1}{27}x^3 & ; 0 \le x \le 3 \\ 1 & ; x \ge 3 \end{cases}$

3. $g(x) = \begin{cases} \dfrac{1}{b-a} & ; a < x < b \\ 0 & ; \text{sicer} \end{cases}$, $F(x) = \begin{cases} 0 & ; x \le a \\ \dfrac{x-a}{b-a} & ; a \le x \le b \\ 1 & ; x \ge b \end{cases}$,

$E(X) = (b-a)/2$, $D(X) = (b-a)^2/12$.

4. $a = b = 1/4$.

5. Da je g_X sploh gostota, mora veljati $a > 0$, $ka \le 2$ in $2a - ka^2/2 = 1$, torej $k = 4/a - 2/a^2$ in $0 < a \le 1$. Nadalje velja $E(X) = a^2 - ka^3/3 = (2a - a^2)/3$, kar je maksimalno pri $a = 1$ in $k = 2$.

6. $P(90 < X < 120) = \Phi(2) - \Phi(-1) \doteq 0{\cdot}81859$,
$P(X < 0) = \Phi(-10) + \frac{1}{2} \doteq 7{\cdot}62 \cdot 10^{-24}$.

7. $1/2 + \Phi(1/3) \doteq 0{\cdot}63056$.

8. Za izračun verjetnosti iz normalne aproksimacije moramo izračunati matematično upanje in varianco. Če je X bilanca zavarovalnice pri posamezni polici, je $E(X) = 0{\cdot}25$ in $D(X) = 10.6875$. Normalna aproksimacija za verjetnost izgube pri n policah je tako $\dfrac{1}{2} - \Phi\left(\dfrac{0{\cdot}25\, n}{\sqrt{10{\cdot}6875\, n}}\right) = \dfrac{1}{2} - \Phi\left(\sqrt{\dfrac{n}{171}}\right)$.

$n = 1$: Točen rezultat: $0{\cdot}25$, normalna aproksimacija: $0{\cdot}4695$.

$n = 3$: Zavarovalnica nima izgube, če bodisi ne plača ničesar bodisi plača enkrat po 5 evrov. Verjetnost, da ima izgubo, je torej:

$$1 - 0{\cdot}75^3 - 3 \cdot 0{\cdot}75^2 \cdot 0{\cdot}15 = 0{\cdot}325\,.$$

Normalna aproksimacija: $0{\cdot}4473$.

$n = 5$: Zavarovalnica nima izgube, če bodisi plača največ eno škodo bodisi plača dve škodi po 5 evrov. Verjetnost, da ima izgubo, je torej:

$$1 - 0{\cdot}75^5 - 5 \cdot 0{\cdot}75^4 \cdot 0{\cdot}25 - 10 \cdot 0{\cdot}75^3 \cdot 0{\cdot}15^2 \doteq 0{\cdot}272\,.$$

Normalna aproksimacija: $0{\cdot}4321$.

$n = 500$: Točen rezultat je težko izračunati, znaša pa $0{\cdot}04265$.
Normalna aproksimacija: $0{\cdot}04364$.

9. Označimo z X število okvarjenih izdelkov.

a) Laplaceova aproksimacija za $P(X = 160)$: 0˙033245
Točen rezultat za $P(X = 160)$: 0˙033228.
Laplaceova aproksimacija za $P(X = 175)$: 0˙015221.
Točen rezultat za $P(X = 175)$: 0˙014929.

b) Laplaceova aproksimacija za $P(X > 175)$: 0˙09824
Točen rezultat za $P(X > 175)$: 0˙09944.
Laplaceova aproksimacija za $P(X < 150)$: 0˙19079.
Točen rezultat za $P(X < 150)$: 0˙19147.

10. Označimo z X število ponesrečenih.

Laplaceova lokalna formula za $P(X = 0)$: 0˙15381,
Točen rezultat za $P(X = 0)$: 0˙22288.
Laplaceova integralska formula za $P(X > 4)$: 0˙00712.
Točen rezultat za $P(X > 4)$: 0˙01849.

6. Kvantili

1. Za $0 < \alpha < 1$ velja $x_\alpha = \alpha/(1-\alpha)$. Tako je $x_{1/4} = 1/3$, $m = 1$, $x_{0.9} = 9$ in $x_{0.99} = 99$.

2. $2{\cdot}33$.

3. Označimo z x potrebno velikost skladišča. Po Laplaceovi integralski formuli je x približno najmanjše celo število, za katerega velja:

$$\Phi\left(\frac{x - 159{\cdot}5}{12}\right) \geq 0{\cdot}45\,.$$

Torej približno velja $x = \lceil y \rceil$, kjer je y rešitev enačbe:

$$\Phi\left(\frac{y - 159{\cdot}5}{12}\right) = 0{\cdot}45$$

in iz $y \doteq 179.24$ dobimo $x = 180$.

Dejansko je verjetnost, da bo pokvarjenih izdelkov (strogo) več kot 179, enaka $0{\cdot}0539$, da jih bo več kot 180, pa $0{\cdot}0457$.

4. Po Laplaceovi integralski formuli (brez uporabe polovičk) dobimo, da moramo naročiti najmanj 13030 izdelkov.
V resnici je najmanjše možno število naročenih izdelkov, ki ustrezajo zahtevi, že 12922. Ne ustreza pa *vsako* število izdelkov, ki je večje ali enako 12922: prvo število, od katerega naprej vsako število naročenih ustreza, je 13096. Nekaj točnih verjetnosti, kjer z n označimo število izdelkov, z S pa število prvovrstnih:

$$\begin{aligned}
n = 12921 &: \quad P(S \geq 7624) \doteq 0{\cdot}9897146436 \\
n = 12922 &: \quad P(S \geq 7624) \doteq 0{\cdot}9900021378 \\
n = 13095 &: \quad P(S \geq 7727) \doteq 0{\cdot}9899715692 \\
n = 13096 &: \quad P(S \geq 7727) \doteq 0{\cdot}9902509306
\end{aligned}$$

7. Vzorčenje

1. a) Možnih vzorcev, če elementov ne uredimo, je $\binom{10}{2} = 45$. Vzorci, pri katerih je povprečje teže večje od 1000 gramov, so:

$$(1010, 1000)\ (2\times)\,,\ (1010, 1010)\,,\ (1050, 980)\,,\ (1050, 1000)\,,\ (1050, 1010)\ (2\times)\,.$$

Verjetnost, da bo vzorčno povprečje večje od 1000 gramov, je torej $7/45 \doteq 0\text{˙}156$. Vzorčna povprečja tež uredimo po velikosti od največjega navzdol:

$$1030,\ 1030,\ 1025,\ 1015, \ldots$$

Torej je $P(\bar{X} \le 1025) = 43/45 > 0\text{˙}95$ in $P(\bar{X} < 1025) = 42/25 < 0\text{˙}95$. Torej je 1025 95. percentil. Če je $q > 1025$, je $P(\bar{X} < q) \ge 43/45 > 0\text{˙}95$, torej q ni 95. percentil. Če pa je $q < 1025$, je $P(\bar{X} \le q) \le 42/45 < 0\text{˙}95$ in q prav tako ni 95. percentil. Torej je 1025 edini 95. percentil.

b) Možnih vzorcev je 100. Vzorci, pri katerih je povprečje teže večje od 1000 gramov, so:

$$(1010, 1000)\ (4\times)\,,\ (1010, 1010)\ (4\times)\,,\ (1050, 980)\ (2\times)\,,\ (1050, 1000)\ (2\times)\,,$$
$$(1050, 1010)\ (4\times)\,,\ (1050, 1050)\,.$$

Verjetnost, da bo vzorčno povprečje večje od 1000 gramov, je torej $17/100 = 0\text{˙}17$. Vzorčna povprečja tež spet uredimo po velikosti od največjega navzdol:

$$1050,\ 1030,\ 1030,\ 1030,\ 1030,\ 1025,\ 1025,\ 1015, \ldots$$

Brž ko je $1025 \le q \le 1030$, je $P(\bar{X} \le q) \ge 0\text{˙}95$ in $P(\bar{X} < q) \le 0\text{˙}95$, torej je q 95. percentil. Če je $q > 1030$, je $P(\bar{X} < q) \ge 0\text{˙}99 > 0\text{˙}95$, torej q ni 95. percentil. Če pa je $q < 1025$, je $P(\bar{X} \le q) \le 0\text{˙}93 < 0\text{˙}95$ in q spet ni 95. pecentil. 95. percentili so torej zajeti v intervalu $[1025, 1030]$.

2. Za normalno aproksimacijo potrebujemo $E(X) = 2$ in $D(X) = 3/2$.

a) Iz normalne aproksimacije izraza $P(1\text{˙}9 < \bar{X} < 2\text{˙}1)$ dobimo $0\text{˙}58578$, iz normalne aproksimacije izraza $P(1\text{˙}905 < \bar{X} < 2\text{˙}095)$ pa $0\text{˙}56206$.
Točen rezultat: $0\text{˙}56133$.

b) Iz normalne aproksimacije dobimo $2\text{˙}201$.
V resnici je (edini) 95. percentil enak $2\text{˙}2$.

3. a) Iz Laplaceove integralske formule dobimo $0\text{˙}52951$.
Točen rezultat: $0\text{˙}52914$.

b) Iz Laplaceove integralske formule dobimo $0\text{˙}714$.
V resnici je (edini) 95. percentil enak $0\text{˙}72$.

8. Intervali zaupanja

1. 95% interval: $(0\text{˙}1166, 0\text{˙}2834)$.
99% interval: $(0\text{˙}0919, 0\text{˙}3081)$.

2. $(0\text{˙}4918, 0\text{˙}5178)$.

3. Za 20 vprašanj pride $(0\text{˙}631, 1]$.
Za 200 vprašanj pride $(0\text{˙}963, 1]$.

4. $\bar{X} = 97$, $\Delta \doteq 3\text{˙}27$ (zaokroženo navzgor),
$93\text{˙}73 < \mu < 100\text{˙}27$.

5. $s = 5$, $t_{0.975}(8) \doteq 2\text{˙}31$, $\Delta \doteq 3\text{˙}85$ (zaokroženo navzgor),
$93\text{˙}15 < \mu < 100\text{˙}85$.

6. $\bar{X} \doteq 1\text{˙}55$, $\sigma \doteq s \doteq 1\text{˙}30$. Interval zaupanja za μ: $1\text{˙}46 < \mu < 1\text{˙}64$.

7. $\chi^2_{0.025}(8) \doteq 2\text{˙}180$, $\chi^2_{0.975}(8) \doteq 17\text{˙}53$,
$3\text{˙}37 < \sigma < 9\text{˙}58$ (spodnja meja zaokrožena navzdol, zgornja pa navzgor).

8. $\bar{X} = 125$, $s \doteq 2\text{˙}646$, $\chi^2_{0.05}(4) \doteq 0\text{˙}711$, $\chi^2_{0.95}(4) \doteq 9\text{˙}49$, $1\text{˙}72 < \sigma < 6\text{˙}28$.

9. $\bar{X} \doteq 45\text{˙}51$, $s \doteq 3\text{˙}71$,
$t_{0.995}(74) \doteq 2\text{˙}64$, $\Delta \doteq 1\text{˙}14$ (zaokroženo navzgor), $44\text{˙}37 < \mu < 46\text{˙}65$,
$\chi^2_{0.005}(74) \doteq 46\text{˙}4$, $\chi^2_{0.995}(74) \doteq 109\text{˙}1$, $3\text{˙}05 < \sigma < 4\text{˙}69$.
Pri obeh intervalih zaupanja je spodnja meja zaokrožena navzdol, zgornja pa navzgor.

10. $\bar{X} \doteq 1\text{˙}55$, $\hat{\sigma} \doteq 1\text{˙}30$, $\hat{\kappa} \doteq 2\text{˙}09$.
Interval zaupanja za σ: $1\text{˙}19 < \sigma < 1\text{˙}41$.

9. Testi značilnosti

1. Za alternativno hipotezo je smiselno postaviti, da je dobitnih manj kot pol srečk. Če je torej $B \sim \mathrm{B}(8, 1/2)$, je:

$$p(2) = P(B \le 2) = \binom{8}{0}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^8 + \binom{8}{1}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^7 + \binom{8}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^6 =$$
$$= \frac{37}{256} \doteq 0\text{'}145\,,$$

torej ne moremo trditi, da loterija laže.

Če pa bi kupili 16 srečk in bi 4 zadele, bi vzeli $B \sim \mathrm{B}(8, 1/2)$ in bi dobili:

$$p(4) = P(B \le 4) \doteq 0\text{'}0384$$

in bi lahko (pri 5% tveganju) trdili, da loterija laže.

2. Izvedemo dvostranski test. Če je $S \sim \mathrm{B}(12, 0\text{'}7)$, je $P(S \le 5) \doteq 0\text{'}0386$ in $P(S \ge 5) \doteq 0\text{'}9905$, torej je $p(5) \doteq 0\text{'}772$. Hipoteze ne moremo zavrniti (zavrnili pa bi jo, če bi za alternativno hipotezo postavili, da so izjemni študijski dosežki pristranski v korist moških).

3. $n = 100, S = 24, \alpha = 0\text{'}05$: $0\text{'}875 < 1\text{'}645$, hipoteze ne moremo zavrniti.
$n = 500, S = 120, \alpha = 0\text{'}05$: $2\text{'}18 \ge 1\text{'}645$, hipotezo zavrnemo.
$n = 500, S = 120, \alpha = 0\text{'}01$: $2\text{'}18 < 2\text{'}33$, hipoteze ne moremo zavrniti.

4. $n = 10000, \quad S = 5090, \quad H_0\colon\ \theta = 0\text{'}5 : H_1\colon\ \theta \ne 0\text{'}5$: $1\text{'}79 < 1\text{'}96$, hipoteze ne moremo zavrniti.
$n = 10000, \quad S = 5090, \quad H_0\colon\ \theta = 0\text{'}5 : H_1\colon\ \theta > 0\text{'}5$: $1\text{'}79 > 1\text{'}645$, hipotezo zavrnemo.

5. $\bar{X} = 97, \quad Z = -1\text{'}8$.
Pri alternativni hipotezi $\mu \ne 100$ upoštevamo $z_{0.975} \doteq 1\text{'}96$ in hipoteze ne moremo zavrniti.
Pri alternativni hipotezi $\mu < 100$ upoštevamo $z_{0.95} \doteq 1\text{'}645$ in hipotezo zavrnemo.
Pri alternativni hipotezi $\mu > 100$ tudi upoštevamo $z_{0.95} \doteq 1\text{'}645$, a hipoteze ne zavrnemo.

6. $\bar{X} = 107, \quad s = 10, \quad T = 2\text{'}1, \quad t_{0.975}(8) \doteq 2\text{'}31$.
Hipoteze ne moremo zavrniti.
Če bi vedeli, da je $\sigma = 10$, pa bi upoštevali $z_{0.975} \doteq 1\text{'}96$ in hipotezo bi zavrnili.

7. $\bar{X} = 48, \quad s \doteq 2\text{'}58, \quad T = -2\text{'}45, \quad df = 9$.
Če za H_1 vzamemo, da je $\mu \ne 50$, upoštevamo $t_{0.975}(9) \doteq 2\text{'}26$ in hipotezo zavrnemo.
Če za H_1 vzamemo, da je $\mu < 50$, pa upoštevamo $t_{0.95}(9) \doteq 1\text{'}83$ in hipotezo prav tako zavrnemo.

8. Označimo z Δ razliko v teži posamezne osebe (teža po dieti minus teža pred dieto). Tedaj je $\Delta \sim \mathrm{N}(\mu, \sigma)$ in testiramo ničelno hipotezo, da je $\mu = 0$, proti alternativni hipotezi, da je $\mu < 0$.
Velja $\bar{\Delta} \doteq -5{}^{\cdot}6$, $\quad s \doteq 6{}^{\cdot}433$, $\quad T \doteq -2{}^{\cdot}75$, $\quad t_{0.95}(9) \doteq 1{}^{\cdot}83$.
Ničelno hipotezo zavrnemo, torej sprejmemo hipotezo, da dieta deluje.

9. $\bar{X} = 20, \quad \bar{Y} = 22, \quad s = 2{}^{\cdot}65, \quad T = -1{}^{\cdot}5, \quad t_{0.975}(14) \doteq 2{}^{\cdot}14.$
Hipoteze ne moremo zavrniti.

10. $\bar{X} = 100, \quad \bar{Y} = 96, \quad s = 2{}^{\cdot}5, \quad T = 3{}^{\cdot}2, \quad t_{0.99}(16) \doteq 2{}^{\cdot}58.$
Hipotezo zavrnemo.

11. $\bar{X}_1 = 4, \quad \bar{X}_2 = 3, \quad \bar{X}_3 \doteq 1{}^{\cdot}667,$
$S_{\mathrm{B}}^2 \doteq 9{}^{\cdot}333, \quad S_{\mathrm{W}}^2 \doteq 8{}^{\cdot}667, \quad F \doteq 4{}^{\cdot}31, \quad F_{0.95}(2, 8) \doteq 4{}^{\cdot}46.$
Hipoteze ne moremo zavrniti.

12. $s = 7{}^{\cdot}45.$

a) $\chi^2 = 20, \quad \chi^2_{0.025}(9) \doteq 2{}^{\cdot}70, \quad \chi^2_{0.975}(9) \doteq 19{}^{\cdot}0,$ hipotezo zavrnemo.

b) $\chi^2 = 5, \quad \chi^2_{0.05}(9) \doteq 3{}^{\cdot}33,$ hipoteze ne moremo zavrniti.

13. Če so frekvence 2, 5 in 4, je: $\chi^2 = 1, \quad \chi^2_{0.95}(2) \doteq 5{}^{\cdot}99.$
Hipoteze ne moremo zavrniti.

Pri frekvencah 20, 50 in 40 pa je $\chi^2 = 10$. Ker je kritično območje isto, hipotezo zdaj zavrnemo.

14. Pri frekvencah 21, 42, 77 in 116 je: $\chi^2 \doteq 81{}^{\cdot}3 \quad \chi^2_{0.99}(3) \doteq 11{}^{\cdot}3.$
Hipotezo zavrnemo.

Če so frekvence 21, 37, 53 in 25, pa je $\chi^2 \doteq 18{}^{\cdot}2$ in hipotezo še vedno zavrnemo.

15. Diskretizirana porazdelitev iz ničelne hipoteze:

pod 10	10–20	20–30	30–40	nad 40
0˙0668	0˙2417	0˙3829	0˙2417	0˙0668

$\chi^2 \doteq 11{}^{\cdot}8, \quad df = 4, \quad \chi^2_{0.95}(4) \doteq 9{}^{\cdot}49.$
Hipotezo zavrnemo.

16. Velja $S^+ = 2$ in $S^- = 8$. Ker za slučajno spremenljivko $B \sim \mathrm{B}(10, 1/2)$ velja $P(B \le 2) \doteq 0{}^{\cdot}0547$, ničelne hipoteze o enakosti porazdelitev s testom z znaki pri tej stopnji značilnosti ne moremo zavrniti.

Pač pa lahko zavrnemo hipotezo, da imata enaki povprečji:
$\bar{X} = 17{}^{\cdot}4, \quad \bar{Y} = 27{}^{\cdot}1, \quad s \doteq 11{}^{\cdot}27, \quad T \doteq -2{}^{\cdot}72, \quad t_{0.95}(9) \doteq 1{}^{\cdot}83.$

17. Velja $S^+ = 10$ in $S^- = 2$. Ker za slučajno spremenljivko $B \sim \mathrm{B}(12, 1/2)$ velja $P(B \le 2) \doteq 0{}^{\cdot}0193$, ničelno hipotezo o enakosti porazdelitev zavrnemo: sprejmemo hipotezo, da je X stohastično večja in vsaj kdaj tudi strogo večja od Y.

Pač pa ne moremo zavrniti hipoteze, da imata X in Y enaki povprečji:
$\bar{X} = 27{}^{\cdot}25, \quad \bar{Y} = 26{}^{\cdot}5, \quad s \doteq 4{}^{\cdot}575, \quad T \doteq 0{}^{\cdot}57, \quad t_{0.95}(11) \doteq 1{}^{\cdot}80.$

18. Velja $S^+ = 12$ in $S^- = 4$. Ker za slučajno spremenljivko $B \sim \mathrm{B}(16, 1/2)$ velja $P(B \le 4) \doteq 0{\cdot}0193$, ničelno hipotezo o enakosti porazdelitev zavrnemo, če jo testiramo proti hipotezi, da je X stohastično večja in vsaj kdaj strogo večja od Y.

Oglejmo pa si zdaj, kaj pravita ustrezna testa povprečij. Iz:

$$\bar{X} = 25{\cdot}5\,, \quad \bar{Y} = 48{\cdot}375\,, \quad s \doteq 44{\cdot}36\,, \quad T \doteq 2{\cdot}06\,, \quad t_{0.95}(15) \doteq 1{\cdot}75$$

dobimo, da lahko ničelno hipotezo zavrnemo v korist hipoteze, da ima Y ves čas večje povprečje od X. Tako smo pri testu povprečij zavrnili ničelno hipotezo ravno v nasprotno smer kot pri testu z znaki.

Opomba. Pri tej in tudi pri prejšnji nalogi smo pri obeh testih dobili kontradiktorne rezultate. To kaže na to, da je nekaj narobe z našim modelom, recimo s tem, da je X ves čas stohastično večja od Y.

19. Ko preštejemo, dobimo, da se je 12 ljudi pred ogledom počutilo boljše kot po ogledu, 25 ljudi pa po ogledu boljše kot pred ogledom; 13 ljudi se je pred in po ogledu počutilo enako. Ker je:

$$Z = \frac{|12 - 25| - 1}{\sqrt{37}} \doteq 1{\cdot}97 > 1{\cdot}96\,,$$

hipotezo zavrnemo.

20. $m = 9$, $n = 11$, $\sum_{i=1}^{9} R_i = 70$, $Z \doteq -1{\cdot}82$, $z_{0.975} \doteq 1{\cdot}96$.
Hipoteze ne moremo zavrniti.

21. Pripravljalo se jih je 12, 8 pa se jih ni pripravljalo. Vsota rangov tistih, ki se niso pripravljali, je 110. Testna statistika v korist teh rangov pride $Z \doteq 1{\cdot}97$, kar je več od kritične vrednosti $1{\cdot}645$. Ničelno hipotezo tako zavrnemo in sprejmemo hipotezo, da so tisti, ki so se pripravljali, tekli bolje od tistih, ki se niso.

Pri T-testu pa povprečje tistih, ki so se pripravljali, pride $8{\cdot}083$, povprečje tistih, ki se niso pripravljali, pa $8{\cdot}362$. Velja še $s \doteq 0{\cdot}4066$, testna statistika pa pride $T \doteq 1{\cdot}50$, kar primerjamo s kritično vrednostjo $t_{0.95}(18) \doteq 1{\cdot}73$. Ničelne hipoteze tokrat ne moremo zavrniti.

22. Iz tabele:

	1	2	3	4	5	Skupaj
da	0	2	1	5	2	10
ne	0	5	2	0	0	7
Skupaj	0	7	3	5	2	17
Kumulativno	0	7	10	15	17	
Vezani rang	–	4	9	13	16·5	

razberemo vsoto rangov tistih, ki so odgovorili z da:

$$2 \cdot 4 + 1 \cdot 9 + 5 \cdot 13 + 2 \cdot 16{\cdot}5 = 115\,,$$

Testna statistika pride 2˙39, kar presega kritično vrednost $z_{0.99} \doteq 2˙33$. Torej zavrnemo hipotezo, da so oboji študenti enako pod stresom, in sprejmemo hipotezo, da so tisti, ki so se bolj posvečali študiju, bolj pod stresom kot tisti, ki se niso. Enostranska odstopanja so prišla zelo značilna.

10. Povezanost dveh številskih spremenljivk

1. Označimo z X sistolični, z Y pa diastolični pritisk. Izračunajmo:
$\bar{X} = 126$, $\bar{Y} \doteq 79{\cdot}67$, $C_x \doteq 30{\cdot}17$, $C_y \doteq 19{\cdot}32$, $C_{xy} = 130$, $R \doteq 0{\cdot}223$, $Z \doteq 0{\cdot}227$, $c \doteq 1{\cdot}96$.
Interval zaupanja: $-0.327 < \rho < 0.656$.

2. $T \doteq 0{\cdot}825$, $df = 13$, $t_{0.95}(13) \doteq 1{\cdot}77$.
Hipoteze ne moremo zavrniti.

3. Teoretične frekence:

oči\lasje	rdeči, blond	rjavi, črni	Skupaj
modre	6˙29	6˙71	13
zelene	11˙12	11˙87	23
rjave	12˙58	13˙42	26
Skupaj	30	32	62

so vse višje od 5. Velja $\chi^2 \doteq 22{\cdot}8$, kar primerjamo z $\chi^2_{0.99}(2) \doteq 9{\cdot}21$.
Hipotezo o neodvisnosti zavrnemo: barva las in barva oči sta bili statistično zelo značilno povezani.

4. Velja $\chi^2 \doteq 0{\cdot}0077$, kar primerjemo s $\chi^2_{0.95}(1) \doteq 3{\cdot}84$.
Hipoteze o neodvisnosti ne moremo zavrniti.

5. Vse opažene in zato tudi vse pričakovane frekvence so enake vsaj 5.
$\chi^2 \doteq 169$, $\chi^2_{0.95}(8) \doteq 15{\cdot}5$. Hipotezo zavrnemo.

6. $\bar{X} = 3$, $\bar{Y} = 7$, $C_x^2 = 10$, $C_{xy} = 20$.
Regresijska premica: $y = 2x + 1$.
Pri $X = 10$ je $\hat{Y} = 21$.
$t_{0.975}(3) \doteq 3{\cdot}18$, $S \doteq 1{\cdot}155$, $\Delta \doteq 9{\cdot}08$.
Interval zaupanja: $11{\cdot}92 < Y < 30{\cdot}08$.