Verjetnost in statistika - rešene naloge iz Kvizkota

Jernej Vivod

September 22, 2018

1 Verjetnost in kombinatorika

1.1 00 - Dobrodošli

1. Moževa sestra iz ZDA je sporočila: We are expecting twins. Kateri od naslednjih odgovorov je bolj verjeten? (predpostavi, da ne pričakuje enojajčnih dvojčkov)

   rešitev:
   Verjetnost, da je novorojenček fantek oz. punčka je enaka.
   Zapišimo vse možnosti: FF FP PF PP
   Vidimo lahko, da je verjetnost za dva fantka in dve punčki enaka in sicer $P(FF) = P(PP) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.
   Verjetnost, da se bo rodil en fantek in ena punčka pa je $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. Torej je bolj verjetno, da se bo rodil en fantek in ena punčka.
   Slučajna spremenljivka X, ki predstavlja število rojenih deklic, kot tudi slučajna spremenljvika Y, ki predstavlja število rojenih dečkov, se porazdeljujeta binomsko. ($X \sim B(2, k)$, kjer k označuje število rojenih punčk). Enako velja za slučajno spremenljivko Y.

2. Iz kupa 52ih običajnih kart vzamemo 4 ase, 4 dvojke in 4 trojke, nato pa med temi 12imi kartami naključno izberemo 5 kart. Kakšna je verjetnost, da sta med njimi vsaj 2 črni?

   rešitev:
   V vsakem izboru štirih asov, dvoj in trojk sta po dve rdeči in po dve črni karti vsakega tipa. Iz tega sledi, da je med temi dvanajstimi kartami 6 črnih in 6 rdečih.
   Verjetnost, da sta med izbranimi petimi kartami manj kot dve črni je enaka $P(A < 2) = \frac{\binom{6}{5} + \binom{6}{4} \cdot \binom{6}{1}}{\binom{12}{5}} = 0.1212$. Iz tega sledi, da je verjetnost nasprotnega dogodka (da sta bili izbrani vsaj dve črni karti) enaka $1 - 0.1212 = 0.8788$.

1.2 01 - Kombinatorika

1. Študent napiše C++ program, ki izpiše vsa možna zaporedja dolžine 3, z uporabo naslednjih črk: A, B, E, T in O. Nato program popravi tako, da ima vsako od tričlenih zaporedij vsaj en samoglasnik in nobenih zaporedoma ponavljajočih se črk.

   Če sta a in b števili različnih tričlenih zaporedij, ki jih je potrebno izpisati z originalnim in s popravljenim programom zaporedoma, kakšen je par (a, b)?

   rešitev:
   Če želimo izpisati vsa možna zaporedja dolžine 3 z uporabo 5 črk, je takih zaporedij $5^3 = 125$, saj imamo za vsako mesto 5 možnosti. Če želimo izpisati vsa tričlena zaporedja, ki vsebujejo vsaj en samostalnik in nobenih zaporedoma ponavljajočih črk, potem je takih možnosti $5 \cdot 4 \cdot 4 - 2 \cdot 1 \cdot 1 = 78$. Leva stran odštevanja nam predstavlja vse možnosti za zaporedja, kjer nimamo dveh zaporednih enakih črk. V tem primeru je za prvo mesto 5 možnosti, za drugo 4 (ne upoštevamo možnosti črke, ki je enaka prvi) ter za zadnjo spet 4 možnost (spet ne upoštevamo možnosti, ki je enaka drugi črki). Od tega odštejemo vse možnosti zaporedij, kjer ni zaporednih črk in kjer nastopajo sami soglasniki.

2. Če imamo tabelo naključnih števk (tj. števil 0,1,..,9) in vsako med njimi podvojimo, ali smo potem dobili dobro tabelo za naključna števila od 0 do 18?

rešitev:
Če premislimo lahko vidimo, da pri podvajanju prve tabele ne moremo dobiti nekaterih vrednosti, kot sta npr. 3 in 5. Druga tabela tako zagotovo ne vsebuje nekaterih števil iz intervala $[0, 18]$ in zato ni dobra.

3. Študent ima 3 matematične in 5 računalniških učbenikov. Ob začetku šolskega leta mora kupiti 2 nova učbenika, zato se odloči dva stara prodati. Oba prodana učbenika morata biti iz istega področja, ker s tem pridobi dodatni popust. Na koliko načinov ju lahko izbere?

rešitev:
Število načinov je vsota števila načinov da izmed treh matematičnih učbenikov izberemo 2 ter števila načinov, da izmed računalniških učbenikov izberemo 2. Rezultat je torej $\binom{3}{2} + \binom{5}{2} = 13$.

4. Osem vojakov je potrebno poslati na stražo v štiri postojanke (po dva v vsako). Recimo, da želimo vsak dan izbrati drugačno razporeditev. Na koliko načinov lahko to storimo?

rešitev:
Za prvo postojanko moramo izmed osmih vojakov izbrati dva. To lahko naredimo na $\binom{8}{2}$ načinov. Za drugo postojanko izbiramo izmed preostalih šestih vnovič dva, torej imamo $\binom{6}{2}$ možnosti in podobno dalje za ostale postojanke. Če ta števila možnosti med sabo zmnožimo dobimo rezultat 2520.

5. Zdravnik preučuje povezavo med krvnim tlakom (visok, normalen, nizek) in srčnim utripom (enakomeren, neenakomeren) svojih bolnikov. Ugotovi, da ima

   (i) 14% visok krvni tlak,

   (ii) 22% nizek krvni tlak,

   (iii) 15% neenakomeren srčni utrip,

   (iv) od tistih z neenakomernim srčnim utripom ima ena tretjina visok krvni tlak,

   (v) od tistih z normalnim krvnim tlakom ima ena osmina neenakomeren srčni utrip.

   Kakšen je delež bolnikov, ki imajo enakomeren srčni utrip in nizek krvni tlak?

rešitev:
TODO

6. Učitelj na podružnic ima v razredu samo štiri učence. Zastavi jim tri vprašanja.

   Na koliko načinov lahko razporedi vprašanja med učence, če je pomembno samo to, kolikokrat je odgovarjal posamezen učenec?

rešitev: Za vsako vprašanje lahko izberemo enega učenca, lahko tudi večkrat. Možnosti imamo torej $\binom{4+3-1}{3} = 20$ (kombinacije s ponavljanjem).

7. Iz kupa igralnih kart (52) na slepo izberemo eno karto in jo nato vrnemo nazaj. Postopek ponovimo 5-krat.

   Kakšna je verjetnost, da bomo videli dvakrat srce, po enkrat pika, križa in karo?

rešitev: Verjetnost, da bomo najprej vlekli srce je $P = 13/52$. Verjetnost, da bomo spet vlekli srce je $P = 13/52$. Verjetnost, da vlečemo nato pika je zopet enaka in enako naprej za križ in karo. Za vlečenje teh kart obstaja več načinov. Teh je natanko $\frac{5!}{2!}$. Število načinov pomnožimo z verjetnostmi ugodnih vlečenj in dobimo rezultat $(\frac{13}{52})^5 \cdot \frac{5!}{2!} = 0.0586$.

8. Iz kraja A v kraj B vodi 5 različnih poti, iz kraja B v kraj C pa 6. Na koliko načinov lahko potujemo iz kraja A v kraj B, če imamo med tema krajema še 7 direktnih povezav?

   rešitev: Za vsako pot iz kraja A v kraj B imamo šest možnosti za nadaljno pot iz kraja B v kraj C. Temu prištejemo še 7 možnosti za potovanje z direktnimi povezavami. Vseh možnosti je torej $5 \cdot 6 + 7 = 37$.

9. Koliko je vseh trimestnih števil?

   rešitev: Za prvo števko iz leve imamo 9 možnosti, za ostali dve števki pa 10 možnosti. Vseh trimestnih števil je torej $9 \cdot 10 \cdot 10 = 900$.

10. Koliko je vseh sodih trimestnih števil?

    rešitev: Število je sodo, če ima na zadnjem mestu števko 0, 2, 4, 6 ali 8. Za tretjo števko imamo 9 možnosti, za drugo pa 10. Takih števil je torej $9 \cdot 10 \cdot 4 = 450$.

11. Koliko je vseh trimestnih števil s sodo prvo števko?

    rešitev: Za sodost prve števke (iz leve) imamo štiri možnosti - 2, 4, 6 ali 8. Za ostali dve števki imamo 10 možnosti. Takih števil je torej $3 \cdot 10 \cdot 10 = 400$.

12. Koliko je vseh trimestnih števil s samimi enakimi števkami?

    rešitev: Za prvo števko iz leve imamo 9 možnosti. Ker morajo biti ostali dve števki enaki, je takih števil torej 9.

13. Koliko je vseh trimestnih števil s samimi različnimi števkami?

    rešitev: Za prvo števko iz leve imamo 9 možnosti, za drugo imamo 9 možnosti, ki niso enake prvi števki in za tretjo 8 možnosti, ki niso enake ne prvi, ne drugi števki. Takih števk je torej $9 \cdot 9 \cdot 8 = 648$

14. Koliko je vseh trimestnih števil, ki so palindromi?

    rešitev: Za prvo števko iz leve imamo 9 možnosti. Zadnja števka se mora z njo ujemati. Za sredno števko pa imamo 10 možnosti. Takih števil je torej $9 \cdot 10 = 90$.

15. Iz škatle s petimi različnimi kroglicami vzamemo dve kroglici.

    - na koliko načinov lahko to storimo, če kroglice vračamo in je vrstni red jemanja pomemben?

      rešitev: Za vsako prvo krolico lahko izvlečemo 5 različnih drugih kroglic. Možnosti je torej $5 \cdot 5 = 25$

    - Na koliko načinov lahko to storimo, če kroglic ne vračamo in je vrstni red jemanja pomemben.

      rešitev: Za prvo kroglico imamo 5 možnosti, za drugo pa 4, saj smo eno izvlekli v prvem vlečenju. Načinov je torej $5 \cdot 4 = 20$.

    - Na koliko načinov lahko to storimo, če kroglic ne vračamo in vrstni red jemanja ni pomemben?

      rešitev: Izmed petih kroglic si izberemo dve. To lahko storimo na $\binom{5}{2} = 10$ načinov.

    - Na koliko načinov lahko to storimo, če kroglice vračamo in vrstni red jemanja ni pomemben?

      rešitev:

      Izmed petih kroglic si izberemo dve. Lahko tudi enaki. To lahko storimo na $\binom{5+2-1}{2} = 15$ načinov (kombinacije s ponavljanjem).

16. Na polici razporedimo 3 leposlovne knjige, 4 učbenike in 2 strokovni knjigi.

- Na koliko načinov lahko to storimo, če vse knjige med seboj razlikujemo?

rešitev: To je preprosti primer štetja permutacij. To lahko storimo na $9! = 362880$ načinov.

- Na koliko načinov lahko to storimo, če vse knjige med seboj razlikujemo, vendar morajo prve v vrsti stati leposlovne knjige, nato učbeniki in za njimi strokovni knjigi?

rešitev: Permutiramo vsako zvrst posebej in množimo števila permutacij. Rezultat je $3! \cdot 4! \cdot 2! = 288$

- Na kliko načinov lahko to storimo, če vse knjige med seboj razlikujemo in knjige iste vrste morajo stati skupaj?

rešitev: Rezultat prejšne podnaloge množimo s $3!$, kar predstavlja permutiranje postavitev skupkov knjig iste vrste. Rezultat je $288 \cdot 3! = 1728$.

- Na koliko načinov lahko to storimo, če knjig iste vrste ne razlikujemo?

rešitev: "Oddelimo" permutacije, kjer ločimo knjige iste vrste. Načinov je torej $\frac{9!}{3! \cdot 4! \cdot 2!} = 1260$.

17. V škatli se nahaja 6 belih in 4 modre kroglice. Kroglice med seboj ločimo, vrstni red jemanja pa ni pomemben.

- Na koliko načinov lahko ven vzamemo 3 bele in 3 modre kroglice?

rešitev: To lahko storimo na $\binom{6}{3} \cdot \binom{4}{3} = 80$ načinov.

- Na koliko načinov lahko ven vzamemo 4 kroglice, med katerimi je vsaj ena bela in ena modra?

rešitev: Najprej izračunamo število načinov, da izvlečemo 4 izključno modre kroglice ter 4 izključno bele kroglice. To lahko storimo na $\binom{6}{4} + \binom{4}{4} = 16$ načinov. To število odštejemo od vseh načinov, da vlečemo 4 kroglice in dobimo rezultat $\binom{10}{4} - 16 = 194$.

18. Na koliko načinov lahko postavimo v vrsto n belih in k modrih kroglic, če zahtevamo, da ni dveh zaporednih belih kroglic?

rešitev: Nalogo si lahko predstavljamo na sledeči način: v vrsto postavimo izmenoma škatlo, belo kroglico, škatlo, belo kroglico,...,škatlo, belo kroglico, škatlo (torej imamo v vrsti n + 1 škatel in n belih kroglic). V škatle med belimi kroglicami damo po eno modro kroglico (za kar porabimo $n-1$ modrih kroglic) s čimer ravno zadostimo pogojem naloge. Preostalih $k-(n-1)$ modrih kroglic lahko poljubno porazdelimo po $n+1$ škatlah. Teh porazdelitev je $\binom{(k-(n-1))+(n+1)-1}{(n+1)-1} = \binom{k+1}{n}$.

19. Na koliko načinov lahko m modrih, r rdečih in z zelenih kroglic razdelimo v s škatel?

rešitev: Ker lahko n kroglic porazdelimo v k škatel (če so te lahko tudi prazne) na $\binom{n-1}{k-1}$ načinov, je rešitev $\binom{m+s-1}{s-1} \cdot \binom{r+s-1}{s-1} \cdot \binom{z+s-1}{s-1}$.

20. Avtomobilske tablice na Havajih vsebujejo tri črke, ki jim sledijo še tri števke.

- Koliko je različnih možnih havajskih tablic?

rešitev: Za črke imamo 26^3 možnosti, za števke pa 10^3 možnosti. Torej je možnih $26^3 \cdot 10^3$ havajskih tablic.

- Ko obiščemo Honolulu, opazimo, da se vse tablice začenjajo z enim E, F, G ali H. Koliko tablic je možnih, če se lahko začnejo le s katero od teh štirih črk?

rešitev: Možnih je $4 \cdot 26 \cdot 26 \cdot 10^3 = 2704000$ tablic.

- Recimo, da bi država dovolila uporabo katerihkoli šestih črk ali števk v poljubnem vrstem redu. Koliko različnih tablic bi bilo možnih v tem primeru (dovolimo ponavljanja)?

  rešitev: Možnih bi bilo 36^6 različnih tablic.

1.3 02 - Verjetnost

1. Naj bo A poljuben dogodek. Potem dogodek A in njegov nasprotni dogodek $\bar{A}$ sestavljata popoln sistem dogodkov.

   rešitev: Nasprotni dogodek $\bar{A}$ je dogodek, ki se zgodi vedno, ko se dogodek A ne zgodi. Popoln sistem dogodkov pa predstavlja neka množica dogodkov, od katerih se pri vsaki ponovitvi zgodi vsaj eden. Torej je odgovor "vedno drži".

2. Pod določenim pogojem je možno, da je vsota verjetnosti vseh vzorcev v vzorčnem prostoru manjša kot 1.

   rešitev: V teoriji verjetnosti je vzorčni prostor množica vseh možnih rezultatov nekega poskusa. Torej se ob vsaki ponovitvi zagotovo zgodi vsaj eden. Odgovor je torej "ne drži".

3. Za poljubna dogodka velja $P(A + B) = P(A) + P(B) + P(AB)$.

   rešitev: Odgovor je "ne drži", saj namreč velja $P(A+B) = P(A)+P(B)-P(AB)$ (glej princip inkluzije in ekskluzije).

4. Denimo da so A, B in C paroma nezdružljivi dogodki, za katere velja $P(A\cup B) = 0.3$, $P(B\cup C) = 0.4$ in $P(A \cup C) = 0.5$. Koliko je P(A)?

   rešitev:
   $P(A) = \frac{P(A\cup B)+P(A\cup X)}{P(B\cup C)} = \frac{0.3+0.5}{0.4} = 0.2$

5. Zapisa $A \cap B \subset A$ in $A \cap B \subset C$ govorita o:

   rešitev: Govorita o produktu dogodkov, kar lahko jasno razberemo iz uporabe simbola $\cap$. Zapomnimo si, da je vsota dogodkov dogodek, ki se zgodi, če se zgodi vsaj eden izmed dogodkov v vsoti in produkt dogodek, ki se zgodi, če se zgodijo vsi dogodki, ki nastopajo v produktu.

6. Produkt nezdružljivih dogodkov A in B je:

   rešitev: Dogodka sta nezdružljiva, če se ne moreta zgoditi hkrati. Oziroma, če je verjetnost njunega produkta enaka 0. Torej je njun produkt nemogoč dogodek.

7. Poljuben dogodek in njegova negacija sta:

   rešitev: Negacija dodogka A je dogodek, ki se zgodi vedno, ko se ne zgodi dogodek A. Torej se ne moreta zgoditi hkrati (sta nezdružljiva). Dogodka sta nasprotna, če ko se eden izmed njiju zgodi, se drugi ne zgodi in obratno. Torej sta dogodek in njegova negacija nezdružljiva in nasprotna dogodka.

8. Klasična in statistična verjetnost se razlikujeta po tem, da je ocena za verjetnost dogodkov...

> **rešitev:** Pri statistični verjetnosti za definicijo verjetnosti vzamemo razmerje med ugodnimi in neugodnimi izidi, proti kateremu to razmerje konvergira v limiti, ko gre število poskusov proti neskončnosti. Torej je ocena za verjetnost dogodkov dobljena računsko.
>
> Pri klasični definiciji verjetnosti pa za verjetnost uporabimo razmerje med številom ugodnih izidov in številom neugodnih izidov, ki so vsi enako verjetni. Torej je ocena za verjetnost dobljena empirično.

9. Elementarni dogodki, ki jih obravnavamo s klasično definicijo verjetnosti...

> **rešitev:** Morajo sestavljati popoln simetrični sistem dogodkov. Simetrični sistem v tem primeru pomeni, da so vsi dogodki enako verjetni. Moramo seveda zajeti vse možne dogodke in iz njih izbrati ugodne. Torej mora biti sistem popoln.

10. Dva moška in pet žensk naključno razporedimo v vrsto. Kakšna je verjetnost, da bosta moška sedela skupaj na začetku in na koncu vrste?

> **rešitev:** Permutacij, kjer moška sedita na začetku vrste je $2! \cdot 5!$. Permutacij, kjer moška sedita skupaj na koncu vrste je isto število. Vseh permutacij je $7!$. Verjetnost za ugodni izid je torej $\frac{2 \cdot 2! \cdot 5!}{7!} = \frac{2}{21}$.

11. Hkrati vržemo 3 poštene igralne kocke. Kakšna je verjetnost, da pade ena šestica, hkrati pa vsaka kocka pokaže različno število pik?

> **rešitev:** Možnosti za ugodni izzid je $3 \cdot (1 \cdot 5 \cdot 4) = 60$, saj lahko šestica pade na kateri koli od treh kock.
> Za preostali dve pa imamo 20 možnosti.
> Vseh možnih izzidov pa je 6^3.
> Rešitev je torej $\frac{60}{6^3} = \frac{5}{18}$

12. V veslaškem klubu so 3 krmarji in 10 veslačev, od tega 5 članov in 5 mladincev. Trener mora v naglici izbrati posadko za četverec, zato jo sestavi kar na slepo.

    Kakšna je verjetnost, da bo izbral najboljšega krmarja, najboljšega člana in najboljšega mladinca?

> **rešitev:** Vseh rešitev je $\binom{3}{1} \cdot \binom{10}{4}$.
> Ugodnih izzidov pa je $1 \cdot \binom{8}{2}$ (Najboljši krmar ter najboljša veslača ter še dva veslača izmed preostalih osmih).
> Verjetnost je torej $\frac{2}{45} = 0.0444$.

13. Problem rešujeta neodvisno drug od drugega dva učenca. Učenec A rešuje probleme tako, da je verjetnost za posamezno rešitev 0.09, učenec B pa z verjetnostjo 0.6.

    Kakšna je verjetnost, da bo problem rešen?

rešitev: Velja

$$P(A+B) = P(A) + P(B) - P(AB)$$

Za $P(AB)$ računamo

$$P(AB) = P(A|B) \cdot P(B)$$

Ker sta dogodka A in B neodvisna velja

$$P(AB) = 0.9 \cdot 0.6.$$

Torej velja

$$P(A+B) = 0.9 + 0.6 - 0.9 \cdot 0.6 = 0.96.$$

14. Kakšna je verjetnost, da bo rešil nalogo le učenec A?

rešitev: Zanima nas vrednost

$$P(A\bar{B})$$

Računajmo:

$$P(A\bar{B}) = P(A|\bar{B}) \cdot P(\bar{B})$$
$$P(A\bar{B}) = 0.9 \cdot 0.4 = 0.36$$
$$P(\bar{A}B) = P(\bar{A}|B) \cdot P(B)$$
$$P(\bar{A}B) = 0.1 \cdot 0.6 = 0.06$$

Seštejemo dobljeni verjetnosti, da dobimo verjetnost, da se je zgodil natanko eden od dogodkov A in B ter dobimo 0.42.
Verjetnost $P(A\bar{B})$ normaliziramo s to vrednostjo, da dobimo rezultat

$$P(A\bar{B}) = \frac{0.36}{0.42} = 0.857$$

15. Pri poskusu sodelujejo štiri osebe. Vsaka si povsem naključno in neodvisno od drugih izbere neko naravno število, manjše od 100.

Kakšna je verjetnost, da bo vsaj eno od teh števil deljivo s 3?

rešitev: Število je deljivo s 3, ko je vsota števk deljiva s 3. Izračunajmo število ugodnih izidov. Število večkratnikov števila 3 manjših od 100 preštejemo tako, da računamo $\lfloor \frac{100}{3} \rfloor = 33$.
Izračunajmo verjetnost nasprotnega dogodka - nobena od štirih oseb si ne izbere števila, ki je deljivo s 3.
Število ugodnih izzidov je:

$$66^4$$

število vseh izzidov je:

$$99^4$$

Verjetnost ugodnega izzida (rešitev naloge) je torej

$$1 - \frac{66^4}{99^4} = \frac{65}{81} = 0.8025$$

16. Študent obvlada 8 izpitnih vprašanj od 12-tih. Na izpitu mora na slepo izbrati 4 vprašanja. Pozitivno oceno doseže, če pravilno odgovori vsaj na 2 vprašanji.

    Kakšna je verjetnost, da bo študent opravil izpit?

rešitev: Za študenta neugodnih izzidov izbiranj vprašanj je:

$$\binom{4}{4} + \binom{4}{3} \cdot \binom{8}{1} = 33$$

Vseh možnih izborov vprašanj je

$$\binom{12}{4} = 495$$

Verjetnost, da bo študent opravil izpit je torej

$$1 - \frac{33}{495} = \frac{14}{15}$$

17. V prvi vrečki so 3 bele kroglice, v drugi pa 2 rdeči. Na slepo izberemo eno od obeh vrečk, v njej izberemo kroglico in jo prenesemo v sosednjo vrečko. Naposled še enkrat sežemo na slepo v eno od obeh vrečk in v njej na slepo izberemo kroglico.

    Kakšna je verjetnost, da bo nazadnje izbrana kroglica bele barve?

rešitev:
Ker dvakrat izbiramo med dvema vrečkama (tudi veznik ali na to nakazuje) uporabimo operator + pri prvem in pri drugem izboru vrečke. Za vsak primer premestitve upoštevamo število belih kroglic v vsaki vreči.
Rezultat torej izračunamo na naslednji način:

$$\frac{1}{2} \cdot (\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{3}) + \\ + \frac{1}{2} \cdot (\frac{1}{2} \cdot \frac{3}{4}) = \frac{25}{48}$$

18. V športnem oddelku gimnazije, ki ga obiskuje 24 dijakov, se jih 15 ukvarja z nogometom, 15 jih kolesari in 10 šahira. Nogomet in šah igra 6 dijakov. Z nogometom in kolesarjenjem se jih

ukvarja 10, 7 pa jih kolesari in igra šah. Vse tri športe gojijo 3 dijaki. Izračunajte verjetnosti dogodkov:

- da se slučajno izbran dijak ne ukvarja z nobenim športom,

  rešitev:

- da slučajno izbran dijak kolesari in šahira, a ne igra nogometa

  rešitev:

- da se med dvema slučajno izbranima dijakoma eden ukvarja samo z nogometom, drugi pa samo kolesari

  rešitev:

19. continue at previous (q0-21)

1.4 03 - Pogojna verjetnost

1. Ali drži naslednja trditev? Če za dogodka A in B velja $P(A) = \frac{1}{3}$ in $P(B) = \frac{1}{3}$, potem je $P(AB) = \frac{1}{9}$.

   rešitev: Velja
   $$P(AB) = P(A|B) \cdot P(B)$$
   Torej bi v našem primeru takšen rezultat dobili samo v primeru, da sta oba dogodka neodvisna, kar pa ni nujno.

2. Ali je naslednja trditev pravilna? Dogodka A in B, ki sta nezdružljiva in za katera velja $A \neq N \neq B$, sta lahko tudi neodvisna.

   rešitev: Če sta nezdružljiva to pomeni, da če se eden zgodi, se drugi zagotovo ne. Torej iz informacije o tem, da se je eden od dogodkov zgodil lahko razberemo, da se drugi ni in posledično ne moreta biti neodvisna.

3. Igralno kocko vržemo petkrat zapored.

   - Kolikšna je verjetnost, da bo v prvem in zadnjem metu padlo enako število pik?

     rešitev: Verjetnost, da bo v zadnjem metu padlo isto število pik kot v prvem je 1/6

   - Kolikšna je verjetnost, da bo padlo pri tem sodo število pik vsaj enkrat?

     rešitev: Računajmo verjetnost nasprotnega dogodka. Verjetnost, da ne bo nikoli padlo sodo število pik je $(\frac{3}{6})^5 = 1/32$
     Iz tega sledi, da je verjetnost nasprotnega dogodka, da je vsaj enkrat padlo sodo število pik enaka $1 - \frac{1}{32} = \frac{31}{32}$.

   - Kolikšna je verjetnost, da bo šele v tretjem metu prvič padlo manj kakor tri pike?

     rešitev: Računajmo verjetnost, da bo v prvih dveh metih padlo več kot tri pike, v tretjem poskusu pa manj kot tri pike.
     $$(\frac{4}{6})^2 \cdot \frac{2}{6}$$

- Kolikšna je verjetnost, da bo padlo v vsakem naslednjem metu več pik kakor v prejšnem?

rešitev: Ugodni izzidi so:

1 2 3 4 5 1 2 3 4 6 1 2 3 5 6 1 2 4 5 6 1 3 4 5 6 2 3 4 5 6

Ugodnih izzidov je torej 5.

Verjetnost je torej:

$$\frac{6}{6^5} = \frac{1}{1296}$$

4. Ali je naslednja trditev resnična? Če sta dogodka A in B neodvisna, sta njuni pogojni verjetnosti $P(B|A)$ in $P(A|B)$ lahko različni.

rešitev: Če sta dogodka A in B neodvisna, je po definiciji nedovisnosti $P(A|B) = P(A)$ in $P(B|A) = P(B)$. Dogodka pa seveda nista nujno enako verjetna, torej je odgovor drži.

5. Na promociji v veleblagovnici vsak kupec dobi možnost, da iz bobna naključno izbere karto. V bobnu je 100 kart in na 20ih piše "Zadetek", na preostalih pa "Žal nam je. Poskusite znova prihodnjič." Predpostavimo, da dva kupca preizkusita svojo srečo in da izvlečeno karto prvega kupca NE nadomestimo pred drugim kupcem.

   Kolikšna je verjetost, da drugi kupec zadane, če vemo, da je zadel prvi kupec?

   rešitev: Verjetnost je enaka 19/99. V bobnu je še 19 ugodnih kart ter 99 vseh kart.

6. Ali velja naslednja trditev? Za poljubna nezdružljiva dogodka velja $P(AB) = P(A) \cdot P(B)$.

   rešitev: Trditev ne velja, saj je za poljubna nezdružljiva dogodka verjetnost $P(AB)$ vedno 0.

7. V škatli imamo 3 neizpravne in 17 izpravnih enot. Dve enoti si izberemo (brez vračanja).

   Kakšna je verjetnost, da sta obe enoti neizpravni, če je bila prva, ki so jo izbrali neizpravna?

   rešitev: Verjetnost je enaka $2/19 = 0.105$. V škatli so še 2 neizpravni ter 19 vseh enot.

8. Deset odstotkov zavarovancev so kadilci, ostali so nekadilni. Za vsakega nekadilca je verjetnost, da tekom leta umre, enaka 0.01, za vsakega kadilca pa je verjetnost, da v istem času umre, enaka 0.05. Recimo, da zavarovanec umre.

   Kakšna je pogojna verjetnost, da je bil zavarovanec kadilec?

   rešitev: Naj dogodek K pomeni, da je zavarovanec kadilec ter dogodek U, da umre. Računajmo z uporabo Bayesovega pravila.

$$P(K|U) = \frac{P(U|K) \cdot P(K)}{P(U|K) \cdot P(K) + P(U|\bar{K}) \cdot P(\bar{K})} = \frac{0.05 \cdot 0.1}{0.05 \cdot 0.1 + 0.01 \cdot 0.9} = 0.3571$$

9. Ali velja naslednja trditev? Za vsaka neodvisna dogodka A in B sta pogojni verjetnosti $P(A|B)$ in $P(B|A)$ enaki.

   rešitev: To ne drži. Intuitiven protiprimer je npr. verjetnost, da se počutimo slabo, če imamo steklino ni enaka verjetnosti, da imamo steklino, če se počutimo slabo.

10. Na letalo ustrelimo dvakrat. Pri prvem strelu je verjetnost za zadetek 0.3, pri drugem pa 0.6. Enkrat zadeto letalo se zruši z verjetnostjo 0.2, dvakrat zadeto pa z verjetnostjo 0.9.

Kolikšna je verjetnost, da s tema dvema streloma zrušimo letalo? Denimo, da smo letalo uspešno zrušili. Kolikšna je verjetnost, da se nam je prva sestrelitev posrečila že s prvim strelom?

rešitev: Seštejemo verjetnosti vseh možnosti.

$$(0.3 \cdot 0.6) \cdot 0.9 + (0.3 \cdot 0.4) \cdot 0.2 + (0.7 \cdot 0.6) \cdot 0.2 = 0.27$$

Verjetnost, da s tema dvema streloma zrušimo letalo je torej 0.27.

Pogojno verjetnost, da smo ga sestrelili že s prvim strelom izračunamo kot

$$P(P|S) = \frac{P(PS)}{P(S)} = \frac{0.3 \cdot 0.2}{0.27} = 0.222$$

Intuitivno je produkt verjetnosti, da smo ga sestrelil s prvim strelom ter da smo ga sestrelili enaka verjetnosti, da smo ga sestrelili s prvim strelom.

11. Spodnja tabela prikazuje starostno skupino in izbrano Colo za 40 oseb iz nekega bloka, ki sodelujejo pri testiranju okusa.

starost	Cola A	Cola B
$10-19$	3	x
$20-29$	8	6
$30-39$	y	5

Opazimo, da je tiskarski škrat nadomestil dve števili v tabeli z x in y. Njegova mama nam zaupa, da je njen otrok dolgo naključno anketiral osebe iz tega bloka, in tako ugotovil, da je verjetnost, da je oseba iz prve starostne skupine, pri tem, da si je izbrala Colo A, enaka $1/7$.

Koliko je par (x, y)?

rešitev: Vemo, da velja $P(prvastarostnaskupina|A) = 1/7$

Ker se pogojne verjetnosti seštejejo v 1 lahko računamo:

$$\frac{1}{7} \cdot s = 3$$
$$s = 21 \to y = 10$$

Iz tega sledi:

$$x = 40 - 21 - 5 - 6 = 8$$

Torej je rezultat $(8, 10)$.

12. V kupu 16 igralnih kart imamo po 4 karte vsake barve. Na slepo izberemo 3 karte. Kakšna je verjetnost, da bodo vse tri karte različnih barv, če sta prvi dve karti različnih barv?

rešitev: Če sta prvi dve karti že različnih barv, je verjetnost, da bo tretja različna od ostalih dveh enaka $\frac{8}{14}$. Saj je v kupu še 14 kart, od teh 8 ugodnih.

13. Vrgli smo dve igralni kocki in dobili vsoto pik na obeh kockah 6. Kolikšna je verjetnost, da smo na drugi kocki dobili liho število?

rešitev: Rezultat je kvocient med številom ugodnih izzidov ter številom vseh izzidov, kjer je vsota 6.
Meti, ki dajo vsoto šest:

15

24

33

42

51

Verjetnost je torej $\frac{3}{5}$.

14. Denimo, da je verjetnost, da študent reši pravilno prvo vprašanje štirikrat manjša od verjetnosti, da pravilno reši drugo vprašanje in za eno polovico manjša od verjetnosti, da pravilno reši vsaj eno od obeh vprašanj. , Na naši fakulteti so si študentje izbirne predmete izbrali takole: 55 študentov statistiko, 80 študentov kriptografijo, 75 študentov verjetnost, 25 študentov kriptografijo in verjetnost, 20 študentov statistiko in verjetnost, 10 statistiko in kriptografijo, 5 študentov vse tri predmete.

- Definiraj nezdružljiva dogodka ter popoln sistem dogodkov in pojasni, kako lahko slednjega uporabimo za klasično definicijo verjetnosti.

  rešitev: Nezdružljiva dogodka sta dogodka, ki se ne morate zgoditi hkrati. Torej je verjetnost njunega produkta vedno enaka 0. Popoln sistem dogodkov je množica dogodkov, ki so med sabo nezdružljivi in se pri vsaki ponovitvi poskusa zgodi natanko eden izmed njih.
  Naj je verjetnostni poskus, ki ima n med seboj enakovrednih izidov (enakovrednost izidov pomeni, da se vsi izidi pojavijo približno enako pogosto, če se poskus ponovi večkrat). Opazuje se dogodek A, za katerega je ugodnih m izidov. Po klasični definiciji je verjetnost dogodka A razmerje med številom ugodnih izidov in številom vseh možnih izidov. Med seboj enakovredni izidi, ki tvorijo dogodke, predstavljajo popoln sistem dogodkov.

- Definiraj vsoto in produkt dogodkov ter utemelji zvezo med verjetnostmi dveh dogodkov, njuno vsoto ter njunim produktom.

  rešitev: Vsota dogodkov je dogodek, ki se zgodi, če se zgodi vsaj eden od dogodkov v vsoti. Produkt dogodkov je dogodek, ki se zgodi, če se zgodijo vsi dogodki, ki predstavljajo produkt.
  Velja:

  $$P(A+B) = P(A) + P(B) - P(AB)$$

- Določi verjetnost, da je študent pravilno rešil obe vprašanji, če veš, da je ta verjetnost, za tretino manjša od razlike verjetnosti, da je pravilno rešil drugo vprašanje in da je pravilno rešil prvo vprašanje.

  rešitev: Velja $4P(A) = P(B)$ oziroma za $P(A) = x$ je $P(B) = 4x$ in $P(A) + \frac{1}{2} = P(A+B)$ oziroma $P(A+B) = x + \frac{1}{2}$ ter $P(AB) + \frac{1}{3} = P(B) - P(A)$ oziroma $P(AB) = 3x - \frac{1}{3}$. Končna zveza $P(A) + P(B) = P(A+B) + P(AB)$ preide v $x + 4x = x + \frac{1}{2} + (3x - \frac{1}{3})$ oziroma $x = \frac{1}{6}$

- Koliko študentov je v tem letniku?

  rešitev: TODO

- Koliko študentov je izbralo dva predmeta od vseh treh naštetij?

rešitev: TODO

- Koliko študentov je izbralo samo en predmet od vseh treh naštetih?

rešitev: TODO

- Koliko študentov je izbralo kriptografijo in ne verjetnosti?

rešitev: TODO

- Izračunaj verjetnost dogodka, da je slučajno izbran študent izbral statistiko.

rešitev: TODO

- Izračunaj verjetnost dogodka, da je slučajno izbran študent izbral verjetnost, pri pogoju, da je izbral statistiko.

rešitev: TODO

- Napiši pravilo o vključitvi in izključitvi ter na osnovi tega posploši zvezo iz zgornjega vprašanja na štiri dogodke.

rešitev: Glej zapiske za teoretični del.

15. pogojna verjetnost

- Definiraj oziroma pojasni, kaj pomeni $P(A|B)$ in kako jo izračunamo?

rešitev: $P(A|B)$ pomeni pogojno verjetnost, da se zgodi dogodek A, če vemo, da se je zgodil dogodek B.

- Verjetnost, da študent opravi izpit, če se je pripravljal, je 90%. Verjetnost, da se je naključno izbran študent pripravljal in opravil izpit, je 50%. Kolikšna je verjetnost, da se je naključno izbran tudent pripravljal na izpit?

rešitev:

$$P(O|P) = 0.9$$
$$P(OP) = 0.5$$

Iz tega sledi:

$$P(O) = \frac{P(OP)}{P(O|P)} = \frac{0.5}{0.9} = 0.5556$$

- Kako izračunamo pogojno verjetnost dveh neodvisnih dogodkov?

rešitev: Pogojna verjetnost dveh neodvisnih dogodkov je kar verjetnost dogodkov samih. Za neodvisna dogodka A in B velja $P(A|B) = P(A)$ in $P(B|A) = P(B)$.

- Utemelji zvezo iz točke (a) (bodisi s klasično ali statistično definicijo verjetnosti).

rešitev: TODO

- Napiši in utemelji Bayesovo formulo

rešitev: Glej zapiske za teoretični del.

16. Dvofazni poskus, formula za popolno verjetnost

rešitev: Glej zapiske za teoretični del.

- Definiraj pogojno verjetnost in podaj formulo za njen izračun.

rešitev: Glej nazaj.

- Definiraj popoln sistem dogodkov ter podaj formulo za popolno verjetnost.

rešitev: Glej zapiske za teoretični del.

- Kako lahko izračunamo verjetnost, da je dogodek A nastopil skupaj z določenim dogodkom (hipotezo) H_i iz prve faze?

rešitev:

$$P(AH_i) = P(A|H_i) \cdot P(H_i)$$

- Oglejmo si preprost primer poligrafskih testov. Študije so pokazale, da v primeru, ko testirani pri preiskusi laže, poligraf le-to (laž) zazna v 88% primerov; v primeru, ko testirani govori resnico, pa je pri odkrivanju resnice uspešen v 86% primerov. Denimo, da 99% vseh testirancev na testu govori resnico. Vzemimo neko osebo, ki ji je poligraf pripisal laž. Kolikšna je verjetnost, da dejansko laže?

rešitev: To je tipičen primer uporabe Bayesovega obrazca... Glej rešeno podobno nalogo o testu za bolezen.

- Izpelji Bayesov obrazec

rešitev: Glej izpiske za teoretični del.

17. Diskretne slučajne spremenljivke

rešitev:

- Definiraj slučajno spremenljivko in pojasni, kdaj je le-ta diskretna. Kako lahko podaš/predstaviš diskretno slučajno spremenljivko?

rešitev: Slučajna spremenljivka je spremenljivka, katere vrednosti so rezultati nekega naključnega pojava. Diskretna je, ko lahko zavzame le končno mnogo vrednosti. Podamo jo lahko z porazdelitveno shemo.

- Kaj je porazdelitvena funkcija in kako izgleda v primeru diskretne slučajne spremenljivke?

rešitev: Porazdelitvena funkcija je funkcija, ki zbira komulativno verjetnost, da je vrednost s.s. manjša od parametra funkcije. V primeru diskretne slučajne spremenljivke ima stopničasto obliko.

- Naštej dva primera diskretnih slučajnih spremenljivk, ter pri vsaki opiši tudi njeno verjetnostno tabelo in zgled uporabe.

rešitev: Glej zapiske za teoretični del ali zbirko informacij o pogostih porazdelitvah.

- Kaj je to slučajni vektor in kako opišemo v primeru diskretne dvorazsežne spremenljivke njegovo verjetnostno funkcijo? Ali se da iz nje ugotoviti kdaj sta njegovi komponenti neodvisni?

rešitev: Slučajni vektor je n-terica slučajnih spremenljivk. V primeru diskretne dvorazsežnostne porazdelitve ga opišemo s kontingenčno tabelo. Njegovi komponenti sta neodvisni, ko je verjetnost v vsaki celici enaka produktu komponente robnih porazdelitev, ki ustrezajo tisti celici.

- Razloži kaj si predstavljamo pod pogojno porazdelitvijo (lahko se omejiš na diskretni primer) in kaj je to pogojna verjetnostna funkcija.

rešitev: Pogojna porazdelitev je porazdelitev neke slučajne spremenljivke, če vemo, da je neka druga slučajna spremenljika zavzela neko vrednost. Označimo jo kot npr. $P(X|Y = y)$. Pogojna verjetnostna funkcija je funkcija, ki prireja verjetnosti vsaki vrednosti v domeni neke spremenljivke ob vedenju, kakšno vrednost je zavzela neka druga slučajna spremenljivka.

- Definiraj pogojno pričakovano vrednost in izračunaj pričakovano vrednost slučajne spremenljivke $E(X|Y)$.

rešitev: Pogojna pričakovana vrednost je dolgoročna povprečna vrednost, ki jo zavzame slučajna spremenljivka, če vemo, kakšno vrednost je zavzela neka druga slučajna spremenljivka (fiksiramo njeno vrednost).

$$E(X|Y = y_k) = \sum_{i=1}^{\infty} x_i p_{i|k} = \frac{1}{q_k} \sum_{i=1}^{\infty} x_i p_{ik}.$$

18. Vržemo standardno pošteno kocko. Naj bo A dogodek, da je padlo liho število pik.

- Naj bo B dogodek, da padejo več kot tri pike. Ali sta dogodka A in B neodvisna?

rešitev: Izračunajmo verjetnost, da je padlo liho število pik. $P(A) = \frac{1}{2}$. Verjetnost, da je padlo več kot tri pike pa je $\frac{1}{2}$. Izračunajmo še verjetnost, da je padlo hkrati liho število pik ter več kot tri pike: $P(AB) = \frac{1}{6}$. Ugoden je le rezultat, da je padlo 5 pik.
Opazimo, da velja $P(AB) = \frac{1}{6} \neq P(A)P(B) = \frac{1}{4}$. Dogodka torej nista neodvisna.

- Naj bo B dogodek, da padejo več kot štiri pike. Ali sta dogodka A in B neodvisna?

rešitev: Preveri, če velja $P(AB) = P(A) \cdot P(B)$.
Verjetnost, da pade liho število pik je $\frac{1}{2}$. Verjetnost, da pade več kot tri pike je $\frac{1}{2}$. Verjetnost, da pade liho število pik ter hkrati več kot tri pike je $\frac{1}{6}$.
Opazimo, da velja $\frac{1}{4} \neq \frac{1}{6}$. Dogodka torej nista neodvisna.

19. Vržemo dve pošteni kocki. Naj bo A dogodek, da je na prvi kocki padlo 6 pik in B dogodek, da je vsota pik na obeh kockah 8 pik.

- Izračunaj $P(A|B)$.

rešitev: Zapišimo vse možne mete, pri katerih je vsota pik enaka 8.

26

35

44

53

62

Verjetnost je torej $\frac{1}{5} = 0.2$.

- Izračunaj $P(B|A)$.

rešitev:

Zapišimo vse možnosti dveh metov, pri katerih je na prvi kocki padlo 6 pik.

616263646566

Vidimo, da je izzid, kjer je vsota pik enaka 8, samo eden. Torej je verjetnost $\frac{1}{6} = 0.1667$

20. Vržemo standardno kocko. Naj bo A dogodek, da padejo vsaj štiri pike, B dogodek, da pade šest pik, L pa dogodek, da pade liho mnogo pik.

- Izračunaj $P(A|L)$

rešitev: Če je se je zgodil dogodek L, je padlo bodisi 1, 3 ali 5 pik. Dogodek, ki zadošča A in L je samo ta, da je padlo 5 pik.

Pogojna verjetnost je torej $P(A|L) = \frac{1}{3}$.

- Izračunaj $P(B|L)$

rešitev: Verjetnost, da je padlo 6 pik, če vemo, da jih je padlo liho število je enaka 0 zaradi očitnih razlogov.

- Denimo sedaj, da kocka ni poštena, pač pa ena pika pade z verjetnostjo 0.3. Izidi z dvema, tremi, štirimi in petimi pikami imajo verjetnost 0.15, šest pik pa pade z verjetnostjo 0.1. Izračunaj $P(A|L)$.

rešitev: Na standardni kocki imamo 3 liha števila. Od tega je eno večje ali enako štiri.

Verjetnost je torej $\frac{1}{3} = 0.3333$

- Izračunaj $P(B|L)$ ob pogojih iz prejšnega vprašanja.

rešitev: Verjetnost je enaka 0, saj če vemo, da je padlo liho mnogo pik ne more pasti šest pik.

- Denimo sedaj, da kocka ni poštena, pač pa ena pika pade z verjetnostjo 0.3, izidi z dvema, tremi, štirimi in petimi pikami imajo verjetnost 0.15, šest pik pa pade z verjetnostjo 0.1. Izračunaj $P(A|L)$.

rešitev: Verjentost, da je padlo 5 pik, če vemo, da jih je padlo liho število je enako normalizirani verjetnosti, da je padlo 5 pik. Vsota verjetnosti, da je padlo liho število pik je 0.6.

Verjetnost je torej $\frac{0.15}{0.6} = 0.25$

- Izračunaj $P(B|L)$ ob pogojih iz prejšnega vprašanja.

rešitev: Še vedno ni mogoče, da bi padlo šest pik, če vemo, da je padlo liho število pik.

21. Dva igralca imata v žari tri bele in tri črne kroglice. Najprej prvi vrže kovanec. Če pade grb, odstrani eno črno kroglico, če pade cifra, pa doda eno črno kroglico. Nato drugi igralec na slepo izbere eno kroglico.

- Kakšna je verjetnost da bo drugi igralec iz žare izbral črno kroglico?

rešitev: Verjetnost je vsota verjetnosti, da bo prvi igralec odstranil eno ter da bo drugi vlekel črno ter da bo prvi igralec dodal eno ter da bo drugi vlekel črno.

$$P() = \frac{1}{2} \cdot (\frac{2}{5}) + \frac{1}{2} \cdot (\frac{4}{7}) = \frac{17}{35} =$$

- Kakšna je verjetnost, da je na kovancu padla cifra, če vemo, da je drugi igralec izbral črno kroglico?

rešitev:

$$P(C|) = \frac{P(C)}{P()}$$

$$P(C|) = \frac{\frac{1}{2} \cdot \frac{4}{7}}{\frac{17}{35}} = \frac{10}{17} = 0.588$$

22. V posodi imamo 3 bele, 4 črne in 5 modrih kroglic. Iz posode izberemo dve kroglici (kroglic ne vračamo, vsako kroglico izberemo z enako verjetnostjo).

- Kakšna je verjetnost, da bomo izbrali dve krogilci enakih barv?

rešitev: Seštejemo verjetnosti, da izberemo najprej belo in nato spet belo in tako dalje za vse ostale barve. Naj bo A dogodek, da smo izvlekli kroglice enakih barv. Računajmo.

$$P(A) = \frac{3}{12} \cdot \frac{2}{11} + \frac{4}{12} \cdot \frac{3}{11} + \frac{5}{12} \cdot \frac{4}{11} = \frac{19}{66} = 0.2879$$

- Kakšna je verjetnost, da bomo izbrali dve kroglici različnih barv?

rešitev: To je nasprotni dogodek od dogodka, da smo izbrali kroglici enakih barv. Računamo: $1 - \frac{19}{66} = \frac{47}{66} = 0.712$

- Kakšna je verjetnost, da bomo izbrali dve kroglici enakih barv, če je prva kroglica bela?

rešitev: Če vemo, da je prva kroglica bela, je pogojna verjetnost preprosto verjetnost, da bo tudi naslednja kroglica bela:

$$p = \frac{2}{11} = 0.182$$

- Kakšna je verjetnost, da smo izbrali dve beli kroglici, če vemo, da smo izbrali dve kroglici enakih barv?

rešitev:

$$p = \frac{P(BB)}{P(O)} = \frac{\frac{3}{12} \cdot \frac{2}{11}}{\frac{19}{66}} = \frac{3}{19} = 0.1579$$

23. V gledališču je 142 sedežev in na blagajni prodajajo 142 oštevičenih kart. 142 ljudi je kupilo karte za sedeže 1,2,...,142 in prihajajo v gledališče v tem vrsten redu. Neolikani gost pride prvi in se usede na naključen sedež. Ostali gosti se usedejo na svoj sedež, če je prost, sicer pa naključno izberejo med prostimi sedeži.

    Kolikšna je verjetnost, da bo sedež gosta, ki pride tretji v gledališče, ob njegovem prihodu prost?

    rešitev: Če je prvi gost zasedel mesto tretjea gosta, potem njegov sedež ni prost. Če je prvi gost zasedel mesto drugega gosta, potem ta sedi na mestu tretjega gosta z verjetnostjo $\frac{1}{141}$. V ostalih primerih bo sedež tretjega gosta prost.
    Verjetnost, da bo sedež tretjega gosta prost, je torej:

$$\frac{1}{142} \cdot \frac{140}{141} + \frac{140}{142} = \frac{140}{141}$$

24. Dva topa izstrelita vsak po eno zrno na sovražen cilj. Verjetnost, da zadane prvi top, je 0.2. Verjetnost, da zadane drugi top pa 0.6. Sovražni cilj bo gotovo uničen, če padeta nanj dve zrni, če pa ga zadane le eno, je verjetnost uničenja enaka 0.3.

    Kolikšna je verjetnost, da bo cilj po obeh strelih uničen?

    rešitev: Računajmo:

$$P(U) = 0.2 \cdot 0.6 + 0.2 \cdot 0.4 \cdot 0.3 + 0.8 \cdot 0.6 \cdot 0.3 = 0.288$$

25. Na predavanja hodi 60% študentov. Če študent hodi na predavanja, naredi izpit v 90% primerov. Če ne hodi, ga opravi samo v 10% primerov.

    Kolikšna je verjetnost, da je slučajno izbrani pozitivni študent hodil na predavanja?

rešitev:

$$P(H|P) = \frac{P(P|H)\cdot P(H)}{P(P|H)\cdot P(H) + P(P|\bar{H})\cdot \bar{H}} = \frac{0.9\cdot 0.6}{0.9\cdot 0.6 + 0.1\cdot 0.4} = 0.931$$

26. Dane so tri škatle. V eni sta dva zlata kovanca, v drugi en zlat in en srebrn, v tretji pa dva srebrna. Dovoljeno nam je, da na slepo izberemo en kovanec (tj. vseh šest z enako verjetnostjo). Če uganemo, kakšen je drugi kovanec v škatli, ki smo jo izbrali, dobimo kovanec.

    Kolikšna je verjetnost, da dobimo kovanec?

    Initially **GG**, **SS** and **GS** are equally likely $\left(\mathrm{i.e.}, \mathrm{P(GG)} = \mathrm{P(SS)} = \mathrm{P(GS)} = \frac{1}{3}\right)$. Therefore, by Bayes rule the conditional probability

    $$\mathrm{P(GG \mid see\ gold)} = \frac{\mathrm{P(see\ gold \mid GG)} \times \frac{1}{3}}{\mathrm{P(see\ gold \mid GG)} \times \frac{1}{3} + \mathrm{P(see\ gold \mid SS)} \times \frac{1}{3} + \mathrm{P(see\ gold \mid GS)} \times \frac{1}{3}} = \frac{1}{1+0+\frac{1}{2}} \times \frac{\frac{1}{3}}{\frac{1}{3}} = \frac{2}{3}$$

    rešitev:

    Razmislek: recimo, da je kovanec zlat. Potem vemo, da je prišel ali iz škatle z dvema zlatima kovancema ali pa iz škatle z enim zlatim in enim srebrnim kovancem. Ker sta obe škatli enako verjetni, je verjetnost, da bomo uganili, enaka 1/2, ne glede na to, kaj rečemo. Je s tem razmislekom kaj narobe?

    rešitev: Če smo izvlekli zlat kovanec je večja verjetnost, da je prišel iz škatle z dvema zlatima kovancema. Verjetnost, da je prišel iz škatle z dvema zlatima kovancema je $\frac{2}{3}$.

27. V posodi je 6 rdečih, 4 bele in 5 črnih kroglic. Na slepo izvlečemo kroglici in z X označimo število izvlečenih črnih kroglic.

    - Določite porazdelitveno shemo slučajne spremenljivke X.

      rešitev: Zaloga vrednosti je $0, 1, 2$, saj se lahko zgodi, da ne izvlečemo nobene črne kroglice, izvlečemo eno ali pa sta obe izvlečeni kroglici črni.
      Z nekaj računanja pridemo do porazdelitve:

      $$X \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{3}{7} & \frac{10}{21} & \frac{2}{21} \end{pmatrix}$$

    - Določite matematično upanje slučajne spremenljivke X.

      rešitev: Naredimo preprost skalarni produkt vrednosti z verjetnostmi in dobimo rezultat 0.6667.

28. Za diskretno slučajno spremenljivko X z zalogo vrednosti $1, 2, 3, 4$ velja $P(X = k) = C \cdot k^2$.

    Določite konstanto C.

    rešitev: Računamo:

    $$c\cdot 1 + c\cdot 4 + c\cdot 9 + c\cdot 16 = 1$$
    $$c\cdot(1+4+9+16) = 1$$
    $$c = \frac{1}{30}$$

29. Petkrat vržemo pošteno kocko. Označimo s S število šestic, ki padejo.

    - Določi porazdelitev slučajne spremenljivke S.

      rešitev: Slučajna spremenljivka S je porazdeljena binomsko $S \sim B(5, \frac{1}{6})$.

- Izračunaj $P(S \quad sodo)$

  rešitev: Zanima nas verjetnost, da pade sodo število šestic. Računamo $P(S=0)+P(S=2)+P(S=4)=0.565844$.

  Za računanje verjetnosti smo uporabili Bernoullijev obrazec.
- Izračunaj $P(S \geq 3)$

  rešitev: To je nasproten dogodek kot $P(S<3)$. Verjetnost je torej:

  $$P(S \geq 3)=1-P(S=0)-P(S=1)-P(S=2)=0.0355$$
- Izračunaj $P(S \quad sodo|S \geq 3)$.

  rešitev: Verjetnost, da je število štestic sodo, če vemo, da jih je več ali enako tri, je enako kvocientu verjetnosti, da jih je sodo ter da je $S \geq 3$ ter verjetnosti, da jih je sodo mnogo. Dobimo rezultat 0.0906.

30. Janez in Meta igrata modificirano verzijo igre človek ne jezi se, pri kateri namesto ene igralne kocke mečeta dve. Janez lahko začne igro, če je vsota pik na obeh kockah pri njegovem metu enaka 6, Metka pa če je vsota pik pri njenem metu enaka 7. Z X označimo število metov, ki jih za začetek potrebuje Janez in z Y število metov, ki jih potrebuje Metka.

- Izračunajte porazdelitvi teh dveh slučajnih spremenljivk.

  rešitev: X je porazdeljena geometrijsko, kjer za verjetnost vzamemo verjetnost, da je vsota pik enaka 6. Podobno je porazdeljena tudi s.s. Y.
- Koliko metov bo v povprečju za začetek igre potreboval Janez?

  rešitev: Matematično upanje geometrijsko porazdeljene s.s. je $E(X)=\frac{1}{p}=\frac{35}{6}=7.2$.
- Koliko metov bo v povprečju za začetek igre potrebovala Metka?

  rešitev:

  $$E(Y)=\frac{1}{p}=6$$

31. Trije lokostrelci streljajo na tarčo. Verjetnost, da jo v sredino zadane $i-ti$ lokostrelec je enaka $\frac{1}{2i}$. Z X označimo število zadetkov tarče po tem, ko vsi sprožijo svoj strel.

    Določite njeno porazdelitev ter z njeno pomočjo izračunajte verjetnost dogodka, da tarčo zadaneta vsaj 2 strela.

    rešitev: Računajmo verjetnosti. Verjetnost, da ga zadane en lokostrelec je $\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}+\frac{1}{2}\cdot\frac{1}{4}\frac{5}{6}+\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{1}{6}=0.4791$. Verjetnost, da jo zadaneta dva je $\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{5}{6}+\frac{1}{2}\cdot\frac{3}{4}\frac{1}{6}+\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{6}=0.1875$. Verjetnost, da jo zadanejo vsi trije je $\frac{1}{2}\cdot\frac{1}{4}\cdot\frac{1}{6}=0.0208$. Verjetnost, da jo nihče ne zadane je $\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}=0.3125$.
    Verjetnost, da zadaneta vsaj dva strela je torej $0.0208+0.1875=0.2083$.

32. Dvestokrat vržemo pošten kovanec in gledamo, kolikokrat pade grb.

- Kolikšna je verjetnost, da pade grb stokrat?

  rešitev: Reši z MATLAB ali R (funkciji *binopdf, binocdf*)
- Za vse možnosti k poišči verjetnsot, da pade grb k-krat.

  rešitev: Reši z MATLAB ali R
- Kolikšna je verjetnost, da bo padel grb vsaj stodesetkrat?

  rešitev: Reši z MATLAB ali R

- Kolikšna je verjetnost, da bo padel grb vsaj petindevetdesetkrat in kvečjemu stodesetkrat?

  rešitev: Reši z MATLAB ali R

- Kolikšna je verjetnost, da bo grab padel manjkrat kot cifra?

  rešitev: Reši z MATLAB ali R

- Kolikšno je pričakovano število grbov?

  rešitev: Reši z MATLAB ali R

33. V kupu 16 igralnih kart imamo po 4 karte vsake barve. Na slepo izberemo 3 karte. Kakšna je verjetnost, da bodo vse tri karte različnih barv, če sta prvi dve karti različnih barv?

    rešitev: Že rešeno. Glej nazaj.

34. Vrgli smo dve igralni kocki in dobili vsoto pik na dveh kockah 6. Kakšna je verjetnost, da smo na drugi kocki dobili liho število?

    rešitev: Že rešeno. Glej nazaj.

35. Vrgli smo dve igralni kocki. Kakšna je verjetnost, da na drugi kocki dobimo strogo več pik kot na prvi?

    rešitev: Ugodni izzidi so:

    $$12$$
    $$23$$
    $$34$$
    $$45$$
    $$56$$

    Ugodnih je torej 5 izzidov. Vseh izzidov je 36 . Torej je verjetnost $\frac{5}{36} = 0.1389$.

1.5 04 - Slučajne spremenljivke

1. Po definiciji je vrednost slučajne spremenljivke naključno število.

   rešitev: Ne drži vedno, saj je domena lahko omejena ter vrednost ni nujno število.

2. Naj bo $p(x) = (6 - |x - 7|)/36$ za $x = 2, 3, ..., 12$ verjetnostna funkcija. Koliko je P(6)?

   rešitev:

   $$p(6) = (6 - |6 - 7|)/36 = \frac{5}{36}$$

3. V povprečju pride na postajo 5 potnikov na minuto. Predpostavimo, da je število potnikov porazdeljeno po Poissonovi porazdelitvi. Kolikšna je verjetnost, da bodo prišli v naslednji minuti trije potniki?

   rešitev: Verjetnostna funkcija Poissonove porazdelitve je podana z enačbo:

   $$p_k(\lambda) = \lambda^k \frac{e^{-\lambda}}{k!}$$

   za λ vstavimo 5, za k pa 3 in izračunamo desno stran enačbe, da dobimo verjetnost.

4. V povprečju pride na postajo 5 potnikov na minuto. Predpostavimo, da je število potnikov porazdeljeno po Poissonovi porazdelitvi. Kako je porazdeljena časovna razlika dveh zaporednih potnikov?

rešitev: Porazdelitev, ki dobro opisuje časovno razliko med dvema dogodkoma v Poissonovem procesu je eksponentna porazdelitev.

5. Ali drži naslednja trditev? Če v kocki s stranico 1 enakomerno izbiramo slučajne točke, njihove koordinate pa spremlja slučajni vektor (X, Y, Z), potem slučajne spremenljivke (X, Y, Z) niso porazdeljene enakomerno, so pa neodvisne.

rešitev: zbiranje točk si lahko predstavljamo kot neodvisno izbiranje točk na treh daljicah dolžine 1. Ker ima vsaka točka na vsaki daljici enako možnost, da je izbrana, je porazdelitev vseh spremenljivk enakomerna. Spremenljivke so pa med seboj tudi neodvisne.

6. Naj bosta slučajni spremenljivki $XinY$ standardno normalno porazdeljeni in neodvisni

   Ali drži, da so zadnje tri trditve pravilne, ostale pa niso pravilne?

   - $X + Y$ je porazdeljena $N(0, 1)$.
   - $X + Y$ je neodvisna od X.
   - $X^2 + Y^2 = 1$.
   - $X - Y$ je normalno porazdeljena.
   - $P(X > Y) = P(Y > Y)$.
   - $P(X > 0, Y > 0) = \frac{1}{4}$.

rešitev: Za neodvisni, normalno porazdeljeni slučajni spremenljivki velja:

$$Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Torej je v našem primeru $X + Y$ porazdeljena kot $N(0, \sqrt{2})$.
Druga trditev iz očitnih razlogov ne velja.
Vsota kvadratov normalnih spremenljivk je porazdeljena po hi-kvadrat porazdelitvi z dvema stopnjama prostosti. Ta porezdelitev pa ni vedno enaka 1.
Če sta s.s. spremenljivki X in Y neodvisni, potem je njuna razlika porazdeljena kot

$$Z \sim N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Torej je porazdeljena normalno.
Verjetnost, da bo s.s. X zavzela večjo vrednost kot s.s. Y je enaka varjetnosti, da bo Y zavzela večjo vrednost kot X, saj imata obe isto porazdelitev.
Verjetnost, da bosta obe s.s. zavzeli vrednost večjo od 0 je enaka produktu verjetnosti, da posamezna s.s. zavzame vrednost večjo od 0. Torej trditev drži.

7. Koliko je enaka ploščina desno od $z = 2.1$ standardni normalni porazdelitvi.

rešitev:
Iz tabele odčitamo, koliko je ploščina levo od te točke in to število odštejemo od 1.
Rezultat je 0.01786 (napaka v rešitvah).

8. Vržemo pošteno kocko. Naj bo A dogodek, da na kocki pade 6 pik, in naj bo X število pik, ki je padko na kocki.

   Kateri od naštetih zapisov nima smisla?

rešitev:
Zapis $P(X) > 0$ ni smiselen, saj moramo za zapis verjetnosti zapisati, pri kateri prednosti s.s. X to velja.

9. Ali naslednja trditev drži? Če je slučajna spremenljivka X porazdeljena standardno normalno $N(0, 1)$, potem za vsako realno število x velja

$$P(X < x) = 1 - P(X > x)$$

rešitev:
Trditev drži, saj je ploščina pod krivuljo enaka 1. Torej predstavlja ploščina za navpično črto skozi x nek delež ploščine, ki se skupaj z deležem ploščine nad to črto sešteje v 1.

10. Ali je naslednja trditev vedno pravilna? Če je slučajna spremenljivka X porazdeljena Binomsko $B(n, p)$, potem za vsako celo število k velja:

$$P(X < k) = 1 - P(X > k)$$

rešitev:
Preberimo, kaj ta zapis pomeni. Verjetnost, da bo število ugodnih dogodkov manj kot k je enaka 1 - verjetnost, da bo ugodnih izzidov več kot k. Binomska porazdelitev je diskretna, torej bi morali upoštevati, da X zavzame lahko tudi vrednost k. Trditev torej ne drži vedno.

11. Naj bo slučajna spremenljivka X porazdeljena geometrijsko s parametrom p. Če je $P(X = 4) = \frac{8}{81}$, potem je p enak koliko?

   rešitev: Rešujmo enačbo:

$$(1 - p)^3 p = \frac{8}{81}$$

   Rešimo za p in dobimo rezultat $\frac{1}{3}$.

12. Slučajne spremenljivke

    - Kaj je to slučajna spremenljivka in kaj moramo vedeti o njej, da lahko rečemo, da jo poznamo?

      rešitev:

      Slučajna spremenljivka je spremenljivka, katere možne vrednosti so rezultati naključnega fenomena. Vedeti moramo, kakšne vrednosti lahko zavzame (torej kakšna je njena domena) ter kakšne so verjetnosti, da zavzame posamezno vrednost iz domene (torej kako je porazdeljena).

    - Navedi osnovne lastnosti porazdelitvenih funkcij.

      rešitev: Glej izpiske za teoretični del.

    - Kako delimo slučajne spremenljivke (naštej po tri najbolj pomembne predstavnike vsake vrste) in pojasni, zakaj je razlikovanje pomembno?

      rešitev: Glej izpiske za teoretični del oziroma učbenik str. 117.

- Naj bo X število dvojk, ki padejo v dvanajstih neodvisnih metih standardne kocke. Zapišite in poimenujte porazdelitev te slučajne spremenljivke.

rešitev: Slučajna spremenljivka je porazdeljena binomsko $X \sim B(12, \frac{1}{6})$.

- Študent dobi na izpitu pozitivno oceno z verjetnostjo $\frac{1}{4}$. Na izpit hodi, dokler ga prvič ne opravi. Naj X označuje število opravljanj izpita. Zapišite in poimenujte porazdelitev te slučajne spremenljivke.

rešitev: Iščemo porazdelitev, ki opisuje število potrebnih poskusov do prvega uspeha. Takšna je geometrijska porazdelitev.

- Koliko je verjetnost, da bo študent šel na vsaj 8 izpitov?

rešitev: Seštejemo verjetnosti, da ga bo opravil v 1., 2., ...,7. izpitu in to vsoto odštejemo od 1.

$$p = (\frac{3}{4})^7$$

Rezultat je 0.1335

- Kolikokrat bo v povprečju moral na izpit?

rešitev: Zanima nas pričakovana vrednost. Računamo: $E(X) = \frac{1}{p} = 4$.

- Definiraj slučajni vektor ter robno porazdelitveno funkcijo in pojasni, kdaj so slučajne spremenljivke med seboj neodvisne. Podaj kakšen primer.

rešitev: Slučajni vektor je n-terica slučajnih spremenljivk. Neodvisni sta, ko velja $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$ za $\forall x, y$ v domeni X in Y.
Robna porazdelitev slučajnega vektorja je porazdelitev poljubnega dela vektora. Robna porazdelitev slučajnega vektorja X, Y je porazdelitev ene izmed slučajnih spremenljivk, ki sestavljata slučajni vektor.
Robni verjetnostni funkciji slučajnega vektorja X, Y sta $p_X(x_i)$ in $p_Y(y_j)$, robni porazdelitveni funkciji pa sta $F_X(x)$ in $F_Y(y)$.
Pri slučajnih vektorjih pravimo funkciji $F_i(x_i) = F(\infty, ..., \infty, x_i, \infty, ..., \infty)$ robna porazdelitvena funkcija spremenljivke X_i.

13. Slučajne spremenljivke in neodvisnost

- Kaj moramo vedeti o slučajni spremenljivki, da lahko rečemo, da jo poznamo (bodite bolj konkretni v primeru diskretne in zvezne slučajne spremenljivke)?

rešitev: Moramo vedeti kakšna je njena domena in kakšna je verjetnost, da zavzame posamezno vrednost iz domene. Za zvezno slučajno spremenljivko potrebujemo neko funkcijo, ki nam opisuje porazdelitev verjetnosti. Porazdelitev poznamo, če imamo funkcijo $F(x) = P(X \leq x)$, ki opisuje verjetnost, da s.s. X zavzame vrednost manjšo od x.

- Definiraj slučajni vektor, njegovo porazdelitveno funkcijo in jo opiši v primeru diskretne dvorazsežne porazdelitve.

rešitev: Slučajni vektor je n-terica slučajnih spremenljivk. Porazdelitveno funkcijo zapišemo kot $F(x_1, ..., x_n) = P(X_1 \leq x_1, ..., X_n \leq x_n)$. V primeru diskretne slučajne spremenljivke ima porazdelitvena funkcija, predstavljena v kartezičnih koordinatah stopničasto obliko.

- Opiši, kdaj lahko rečemo, da je slučajni vektor zvezno porazdeljen. Kako v tem primeru z gostoto verjetnosti opišemo njegovo porazdelitveno funkcijo in obratno, kako iz porazdelitvene funkcije dobimo gostoto (lahko se omejite na dvorazsežni primer)?

rešitev: Slučajni vektor $X = (X_1, X_2, ..., X_n)$ je zvezno porazdeljen, če obstaja integrabilna funkcija (gostota verjetnosti) $p(x_1, x_2, ..., x_n) \geq 0$ z lastnostjo

$$F(x_1, ...x_n) = \int_{-\infty} x_1 \int_{-\infty} x_2 \cdots \int_{-\infty} x_n p(t_1, ..., t_n) dt_1 dt_2 ... dt_n$$

$$F(\infty, ..., \infty) = 1.$$

- V primeru zvezne porazdelitve definiraj robno porazdelitveno funkcijo in robno verjetnostno gostoto ter pojasni kdaj so slučajne spremenljivke med seboj neodvisne.

rešitev: V primeru zvezne dvorazsežne porazdelitve imamo robni verjetnostni gostoti, ki sta enaki:

$$p_X(x) = F_X'(x) = \int_{-\infty}^{\infty} p(x, y) dy$$

$$p_Y(y) = F_Y'(y) = \int_{-\infty}^{\infty} p(x, y) dx$$

- Ali se da iz verjetnostne funkcije diskretnega dvorazsežnega slučajnega vektorja ugotoviti neodvisnost njegovih komponent (pojasni kako oziroma zakaj ne)?

rešitev: Da. Če sta neodvisni, za vsak par vrednosti s.s. $X = x$ in $Y = y$ velja $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$.

- Pri diskretni dvorazsežni porazdelitvi definiraj $P(X = x_k | Y = y_k)$. Razloži s pomočjo verjetnostne tabele, kaj si (v tem primeru) predstavljamo pod pogojno porazdelitvijo (glede na pogoj $Y = y_h$.

rešitev: To je verjetnost, da s.s. X zavzame vrednost x_k, če vemo, da je y zavzela vrednost y_k. Če gledamo verjetnostno tabelo, kjer so vrednosti s.s. X na vrhu in vrednosti Y na levi, potem je ta pogojna verjetnost enaka vrednosti v celici, ki jo določata x_k in y_k. Pogojna porazdelitev spremenljivke X, ob pogoju, da je $Y = y_k$ pa je vrstica verjetnosti, kjer je $Y = y_k$.

14. Normalna porazdelitev

- Definiraj standardni odklon in pričakovano vrednost slučajne spremenljivke.

rešitev: Pričakovana vrednost je dolgoročna povprečna vrednost, k kateri konvergira povprečna vrednsot s.s., ko gre število poskusov proti neskončnosti. Standardni odklon pa je mera, ki kvantificira količino variacije oziroma disperzije neke množice podatkov oziroma vrednosti, ki jih zavzame slučajna spremenljivka.

- Opiši postopek za standardizacijo slučajne spremenljivke in izračunaj njeno pričakovano vrednost ter njen odklon.

rešitev: Slučajno spremenljivko standardiziramo, tako da od vsake vrednosti, ki jo zavzame odštejemo njeno pričakovano vrednost in delimo s standardnim odklonom. Tako dobimo standardizirano slučajno spremenljivko s pričakovano vrednostjo 0 in standardnim odklonom 1.

- Opiši normalno porazdelitev in razloži, zakaj je pomembna.

rešitev: Normalna porazdelitev je zelo pogosta zvezna porazdelitev, ki je pogosto uporablja, ko porazdelitve realnih slučajnih spremenljivk niso poznane. Zelo dobro jo opisujeta pričakovana vrednost oz. povprečje ter standardni odklon. itd. (Glej zapiske za teoretični del)

- Naj bo X normalna slučajna spremenljivka s parametroma $\mu = 19$ in $\sigma^2 = 9$. Izračunaj naslednje verjetnosti: $P(X > 21)$ in $P(X \leq 20)$.

rešitev:

$$z = \frac{21-19}{3} = 0.6667$$

Pogledamo v tabelo, kakšna je verjetnost, da se v normalni porazdelitvi s.s. nahaja 0.6667 standardne deviacije od povprečja in dobimo verjetnost $p = 0.7486$.

- Opiši večrazsežno normalno porazdelitev.

rešitev: Glej zapiske za teoretični del.

15. Neodvisno vržemo dva kovanca in standardno kocko. Za vsako piko dobimo 1 evro, za vsako cifro, ki pade, pa 2 evra. Slučajna spremenljivka S naj predstavlja skupni znesek v evrih, ki ga dobimo.

    Zapiši porazdelitev slučajne spremenljivke S in izračunaj $E(S)$.

    rešitev: Seštejemo vse verjetnosti za vse možne zneske in naredimo verjetnostno (kontigenčno) tabelo. Pričakovano vrednost dobimo tako, da vzamemo skalarni produkt vektorja vrednosti slučajne spremenjivke z vektorjem verjetnosti teh vrednosti.

16. Za neko stavo je treba plačati a evrov. Vržemo kovanec in če pade cifra, dobimo 4 evre.

    Koliko naj bo a, da bo pričakovan dobiček 1 EUR na igro.

    rešitev: Računajmo pričakovano vrednost.

    $$E(X) = -a + \frac{1}{2} \cdot 4 = 1$$

    Rešimo enačbo, da dobimo $a = 1$.

17. Naj bo X slučajna spremenljivka, ki nam pove, kolikokrat je v štirih neodvisnih metih standardne kocke padlo liho mnogo pik.

    Zapiši porazdelitev te slučajne spremenljivke in izračunaj $E(X)$.

    rešitev: Spremenljivka X je porazdeljena binomsko. Pričakovano vrednost izračunamo kot $E(X) = np$, kjer je $n = 4$ in $p = \frac{1}{2}$.

18. Med 12 kartami so štirje piki. Na slepo in brez vračanja izvlečemo tri karte. Naj bo X število pikov med njimi.

    Zapiši in poimenuj porazdelitev s.s. X in izračunaj E(X).

    rešitev:

    To je primer hipergeometrijske porazdelitve. Pričakovano vrednost $E(X)$ izračunamo kot $E(X) = n \cdot \frac{K}{N} = 3 \cdot \frac{4}{12} = 1$

19. Cesto pred vrtcem v povprečju prevozi 100 avtomobilov na uro.

    - Kolikšna je verjetnst, da bosta v treh minutah cesto prevozila manj kot 2 avtomobila?

      rešitev: Uporabimo Poissonovo porazdelitev. Če velja, da na 60 minut v povprečju cesto prevozi 100 avtomobilov, jih v treh minutah v povprečju 5.

      $$P(X < 2) = 5^0 \frac{e^{-5}}{0!} + 5^1 \frac{e^{-5}}{1!} = 0.04$$

- Kolikšna je verjetnost, da v treh minutah cesto prevozijo več kot trije avtomobili?

rešitev:

$$P(X > 3) = 1 - (5^0 \frac{e^{-5}}{0!} + 5^1 \frac{e^{-5}}{1!} + 5^2 \frac{e^{-5}}{2!} + 5^3 \frac{e^{-5}}{3!}) = 0.735$$

- Skupina iz vrtca potrebuje 1 minuto, da prečka cesto. Kolikšna je verjetnost, da v času, ko ta skupina prečka cesto, mimo ne pripelje noben avto?

rešitev:

$$P(X = 0) = 1.6667^0 \cdot \frac{e^{-1.6667}}{0!} = 0.1889$$

20. Na prvem tradicionalnem FRI teku sodeluje 10 žensk in 15 moških. Pred štartom študentka izbere 3 tekmovalce za intervju.

- Kolikšna je verjetnost, da bo izbranih več žensk kot moških?

Da je izbranih več žensk pomeni, da sta izbrani dve ženski in en moški. Verjetnost tega dogodka je $\frac{\binom{10}{2} \cdot \binom{15}{1}}{\binom{25}{3}} = \frac{27}{92}$

- Koliko žensk pričakujemo, da bo izbranih za intervju?

Pričakovana vrednost hipergeometrijske porazdelitve je $E(X) = n \cdot \frac{K}{N} = 3 \cdot \frac{10}{25} = 1.2$.

- Kolikšna je verjetnost, da se dejansko število žensk od pričakovanega razlikuje za kvečjemu 1?

Seštejemo verjetnost, da jih ne bo izbranih 0 ter verjetnost, verjetnost, da bo izbrana ena, ter verjetnost, da bosta izbrani 2.

21. Slučajna spremenljivka X ima zalogo vrednosti $1, 2, ..., 10$. Verjetnost, da X zavzame vrednost k, je enaka $c \cdot k$.

- Izračunaj konstanto c.

rešitev: Verjetnosti se morajo sešteti v 1. Računamo:

$$1 \cdot c + 2 \cdot c + ... + 10 \cdot c = 1$$

Rešimo enačbo, da dobimo $55 \cdot c = 1$ ter nato $c = \frac{1}{55}$.

- Izračunaj $P(X \leq 5)$.

rešitev:

$$P(X \leq 5) = 1 \cdot \frac{1}{55} + ... + 5 \cdot \frac{1}{55} = 0.2727$$

- Izračunaj matematično upanje slučajne spremenljivke X.

rešitev: Izvedemo skalarni produkt med verjetnostmi in vrednostmi. Dobimo rezultat 7.

- Izračunaj disperzijo D(X).

rešitev:

Disperzijo izračunamo po obrazcu:

$$D(X) = E(X^2) - E(X)^2 = 55 - 49 = 6$$

- Izračunaj standardni odklon $\sigma(X)$

rešitev:

Standardni odklon je koren variance oz. disperzije. Torej velja $\sigma = \sqrt{6}$

22. Dana je slučajna spremenljivka X:

$$X \sim \begin{pmatrix} 2 & 7 & 9 & 10 \\ 0.1 & 0.3 & 0.2 & 0.4 \end{pmatrix}$$

- Izračunaj E(X)

rešitev: S skalarnim produktom izračunamo in dobimo rezultat 8.1.

- Izračunaj D(X)

rešitev:

$$D(X) = E(X^{)} - E(X)^2 = 71.3 - 65.61 = 5.69$$

- Izračunaj E(4X + 2)

rešitev:

```
>> v = [2 7 9 10]

v =

     2     7     9    10

>> p = [0.1 0.3 0.2 0.4]

p =

    0.1000    0.3000    0.2000    0.4000

>> (4.*v + 2) * p'

ans =

   34.4000
```

- Izračunaj D(4X + 2)

```
>> E1 = ((4.*v + 2).^2) * p'

E1 =

   1.2744e+03

>> E2 = ((4.*v + 2) * p').^2

E2 =

   1.1834e+03

>> E1 - E2

ans =

   91.0400
```

rešitev:

23. V galaksiji kot je naša je v povprečju 5 supernov na stoletje. Zadnjo, za katero obstajajo zgodovinski dokazi, je leta 1604 opazoval Kepler.

    - Kolikšna je verjetnost, da od Keplerjevega opažanja dalje v Rimski cesti res ni bilo nobene supernove?

rešitev: Razmak med dvema pojavoma v Poissonovem procesu modeliramo s eksponentno porazdelitvijo. Računajmo verjetnost, da se v 414 letih ni zgodila nobena supernova.

$$P(X = 414) = 0.01 \cdot e^{-0.01 \cdot 414} = 0.00016$$

- Kolikšna je verjetnost, da bomo pojav supernove znotraj Rimske ceste dočakali v naslednjih 30 letih?

rešitev: Vemo, da je funkcija "brez spomina". Torej velja:

$$P(X = 30) = 0.01 \cdot e^{-0.01 \cdot 30} = 0.0074$$

24. V ribniku plava 10 rib. Ribič prvi dan ulovi 2 ribi, ju označi in vrne nazaj v ribnik. Drugi dan se ribič zopet poda na lov, vendar tokrat ulovi 5 rib. Naj bo X število označenih med njimi.

- Kako je porazdeljena slučajna spremenljivka X?

rešitev: Spremenljivka je porazdeljena hipergeometrijsko.

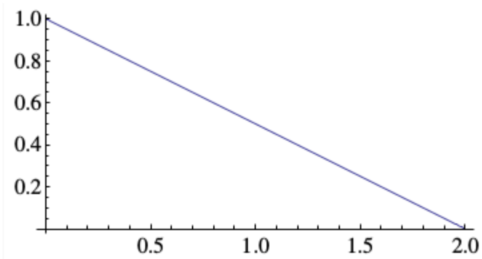
- Koliko označenih rib lahko pričakujemo v ulovu drugega dne?

rešitev: Pričakovana vrednost hipergeometrijske spremenljivke je $E(X) = n \cdot \frac{K}{N} = 5 \cdot \frac{2}{10} = 1$.

- Kolikšna je verjetnost, da jih bo več kot pričakovano?

rešitev: Izračunaj, da jih bo manj ali enako kot pričakovano (komulativna funkcija verjetnosti) ter od 1 odštej to vrednost.

25. Naj bo X zvezna slučajna spremenljivka, ki zavzame vrednosti med 0 in 2, njena gostota verjetnosti pa je spodnja funkcija.



Izračunaj mediano m in ugotovi, kje glede nanjo se nahaja pričakovana vrednost μ.

rešitev: Verjetnostna funkcija je $P(x) = 1 - \frac{1}{2}x$. (Z integriranjem lahko preverimo, da se verjetnosti res seštejejo v 1).
Pričakovano vrednost dobimo tako, da integriramo konvolucijo vrednosti s.s. in verjetnostne funkcije:

$$E(X) = \int_0^2 x \cdot (1 - \frac{1}{2}x)dx = 0.667$$

26. ???

27. Tarča je sestavljena iz črnega kroga v belem kvadratu. Strelec trikrat ustreli proti tarči. Naj bo X število zadetkov tarče in Y število zadetkov v črno. Ali sta spremenljivki X in Y neodvisni?

rešitev:

Nista neodvisni. Če poznamo vrednost s.s. spremenljivke X, lahko, če vemo, da ni zadel tarče, takoj sklepamo, da tudi ni zadel tarče. Torej se med spremenljivkama prenaša neka informacija.

28. Ali za $X \sim \chi^2(25)$ velja

$$P(X > 38.5) \in [0.05, 0.1]?$$

rešitev: Preverimo tabelo za hi-kvadrat porazdelitev z 25 stopnjami prostosti.

Opazimo, da se verjetnosti seštejejo v vrednost izven tega intervala.

29. Kovanec mečemo, dokler trikrat ne pade grb. Katera ocena za verjetnost, da bomo kovanec vrgli vsaj sedemkrat, je najboljša?

rešitev: Za izračun rezultata uporabimo negativno binomsko porazdelitev:

```
>> sum_prob = 0;
>> for i = 3:6
sum_prob = sum_prob + nchoosek(i - 1, 3 - 1)*(0.5^3)*(0.5^(i-3));
end
>> 1 - sum_prob

ans =

    0.3438
```

30. Pet vstopnic za rock koncert moramo razdeliti med 25 razgrajaških članov našega kluba. Označimo člane s številkami 1 do 25. Naključno izberi 5 srečnežev, ki bodo dobili vstopnice.

   - Izkaže se, da je med 25 člani kluba 10 žensk, ki jim pripadajo števila od 1 do 10. Dvajsetkrat naključno izberi vzorec s pet enotami po zgornjem postopku. Zabeleži število žensk v vsakem od vzorcev. Nariši histogram, ki predstavi dobljene rezultate. Izračunaj povprečno število žensk v teh 20 vzorcih.

     rešitev:

     TODO

   - Ali meniš, da bi morali člani kluba posumiti, da gre za diskriminacijo, če nobene od vstopnic ne bi dobila ženska? (Naj slučajna spremenljivka X označuje število žensk v vzorcu. Kako je X porazdeljena? Kolikšna je verjetnost, da v izbranem vzorcu ni žensk? Kolikšno je pričakovano število žensk v vzorcu?)

     rešitev:

     TODO

31. Naključna števila lahko uporabimo za simulacijo rezultatov slučajnega vzorčenja. Recimo, da izberemo enostavni slučajni vzorec velikosti 25 iz velikega števila srednješolcev in da si 20 ne poišče dela med poletnimi počitnicami. Da simuliramo ta enostavni slučajni vzorec, generiramo v programu R 25 naključnih števil med 0 in 9 z ukazom sample(0:9,25,TRUE). Števili 0 in 1 naj pomenita nezaposlene dijake, ostale števke pa tiste z zaposlitvijo. To je točna imitacija enostavnega slučajnega vzorca, ker 0 in 1 predstavljata ravno 20 izmed 10 enako verjetnih števk.

    Simuliraj rezultate petdesetih vzorcev tako, da preštejeh ničle in enice med 25 števkami v vsakem od 50 naključno generiranih seznamov. Nariši histogram, ki prikaže dobljene rezultate. Ali je resnično stanje populacije (20 nezaposlenih, tj. 5 v vzorcu 25) blizu sredine tega histograma? Kaj sta največje in najmanjše dobljeno število nezaposlenih dijakov med temi 50 vzorci? Kolikšen odstotek vzorcev je imel med 4 in 6 nezaposlenih? Rezultat vpiši v odstotkih, pri tem pa ga zaokroži na celo število.

rešitev:

TODO

32. Nek mladi par si želi tri otroke. Skupaj je 8 možnih razporeditev dečkov in deklic. Dogovorimo se, da oznaka ŽŽM pomeni, da sta prva dva otroka deklici in tretji deček, ter podobno za ostale možnosti. Vseh 8 razporeditev je (približno) enako verjetnih.

- Napravi seznam vseh razporeditev. Kolikšna je verjetnost vsake izmed njih?

  rešitev:

  FFF

  FFŽ

  FŽF

  FŽŽ

  ŽFF

  ŽFŽ

  ŽŽF

  ŽŽŽ

  Vsi izidi so enako verjetni (1/8).

- Na podlagi tega verjetnostnega modela napiši porazdelitev slučajne spremenljivke X, ki šteje število deklic med tremi otroki.

  rešitev:

  Slučajna spremenljivka je porazdeljena binomsko. $X \sim B(3, 0.5)$.

- Uporabi model iz prejšnega vpračanja in izračunaj verjetnost, da sta med otroki vsaj dve deklici.

  rešitev:

  $P(X \geq 2) = P(X = 2) + P(X = 3) = 0.5$

  Uporabi ukaz *binopdf*

- Vrnimo se k modelu iz prvega vprašanja. Uporabi ta model za izračun verjetnosti, da sta med tremi otroki vsaj dve deklici. Primerjaj rezultat s tistim iz prejšne točke.

  rešitev:

  Preštejemo število ugodnih nizov in to število delimo s številom vseh izzidov. Dobimo enak rezultat.

1.6 05 - Sredine

1. Za slučajno spremenljivko X, ki je binomsko porazdeljena z $n = 20$ in $p = 0.3$, izračunaj $\sum_{k=0}^{n} k^2 p_k$.

rešitev:

Rešimo nalogo z MATLAB.

k = [1:20];

p = 0.3;

n = 20;

probabilities = binopdf(k, n, p);

K_ sqr = k.^2;

*res = k_ sqr * probabilities';*

V spremenljivki se nahaja vrednost 40.2, ki je tudi rešitev naloge.

2. Vsaka srečka stane 2 EUR. Prodanih je bilo 400 srečk, med njimi ena dobi 500 EUR, dve drugi pa vsaka po 50 EUR.

   Kakšen je povprečen dobiček, ki ga pričakuješ z nakupomene srečke?

rešitev: Računajmo pričakovano vrednost spremenljivke X, ki beleži količino, ki smo jo zadeli.

$$E(X) = 500 \cdot \frac{1}{400} + 2 \cdot 50 \cdot \frac{2}{400} = 1.75$$

Povprečen dobiček je torej 1.75 EUR.

3. Momenti so vzorčne statistike, kar pomeni, da so simetrične funkcije in zato z njimi ne moremo izmeriti asimetrije.

rešitev: Ne drži.

4. ??

5. Mediana za slučajno spremenljivko, ki je zvezno porazdeljena, je točka, na realni osi, ki je enaka oddaljena od najmanjšega in od največjega števila iz zaloge vrednosti te slučajne spremenljivke.

rešitev: Ne drži. Pri zveznih porazdelitvah imamo lahko tudi opravka z neskončnimi domenami slučajne spremenljivke. V tem primeru niti ne moremo govoriti o najmanjši oziroma največji vrednosti oziroma o sredini med njima.

6. Ali je pričakovana vrednost poljubne simetrične porazdelitve vedno enaka mediani?

rešitev: Da. Mediana loči zgornjo polovico podatkov oz. domene slučajne spremenljivke od spodnje. Pri simetričnih porazdelitvah se bodo mediana, povprečna vrednost in pričakovana vrednost ujemale.

7. Pričakovana vrednost slučajne spremenljivke deluje linearno, tj. če jo uporabimo na neki linearni kombinaciji slučajnih spremenljivk, je enaka linearni kombinaciji pričakovanih vrednosti posameznih slučajnih spremenljivk z enakimi koeficienti.

rešitev: To drži. Pomembna lastnost pričakovane vrednosti je:

$$E(aX + bY) = aE(X) + bE(y).$$

8. Ali disperzija deluje linearno, kadar so slučajne spremenljivke med seboj neodvisne?

rešitev: To ne drži, saj za disperzijo velja že na primer $D(aX) = a^2 D(X)$. Torej ni homogen operator in posledično tudi ne linearen.

9. Vsak kolokvij pri nadaljevalnem predmetu iz statistike je vreden 100 točk. Po treh kolokvijih ima študent povprečje 88 točk in odklon 3. Koliko mora pisati četrti in peti kolokvij, da bo njegovo povprečje 90 točk, odklon pa bo ostal enak?

rešitev: Nastavimo enačbi.

$$9 = \frac{9 + (x_1 - 90)^2 + (x_2 - 90)^2}{5}$$

in

$$90 = \frac{88 * 3 + x_1 + x_2}{5}$$

Enačbi rešimo. Najlaže s pomočjo ustrezne programske opreme (R, MATLAB)

10. Za neko binomsko porazdeljeno slučajno spremenljivko B(n, p) je pričakovana vrednost $\mu = 4$ in standardni odklon $\sigma = \sqrt{3}$. Določi vrednost p.

rešitev: Nastavimo sistem enačb:

$$E(X) = np = 4$$

in

$$SE(X) = \sqrt{npq} = \sqrt{np(1-p)} = \sqrt{3}$$

Rešujmo

$$p = \frac{4}{n}$$

$$n(\frac{4}{n})(1 - \frac{4}{n}) = 3n = 16 \Rightarrow p = \frac{4}{16} = \frac{1}{4}$$

11. Slučajni vektor (X, Y) je porazdeljen diskretno z naslednjo verjetnostno tabelo.

$Y \setminus X$	0	1	2
-1	0	1/4	0
0	0	1/4	1/4
1	1/4	0	0

Izračunaj K(X, Y).

rešitev: Kovarianco iz verjetnostje tabele izračunamo z uporabo formule

$$E(X, Y) = E(XY) - E(X)E(Y).$$

Za $E(X)$ in $E(Y)$ dobimo vrednosti 1 in 0.
Za $E(XY)$ dobimo rezultat $-\frac{1}{4}$.
Rezultat je torej

$$E(X, Y) = -\frac{1}{4} - 1 \cdot 0 = -\frac{1}{4}.$$

12. Slučajni spremenljivki X in Y sta porazdeljeni enakomerno na intervalu $(0, 1)$. (Ni se težko prepričati, da je $D(X) = \frac{1}{12}$).

Če velja $K(X, Y) = \frac{1}{24}$, koliko je potem $D(X + 2Y)$?

rešitev:
TODO

13. Naj bosta X in Y enako porazdeljeni slučajni spremenljivki.

    Čemu je izraz K(X, Y) - D(X) vedno enak?

    rešitev: TODO

14. Naj bosta X in Y neodvisni slučajni spremenljivki.

    Čemu je izraz $K(X,Y) - D(X)$ vedno enak?

    rešitev: TODO

15. Katera porazdelitev ima mediano manjšo od pričakovane vrednosti (matematičnega upanja)?

    rešitev: Geometrijska porazdelitev $G(2/3)$ ima mediano manjšo od pričakovane vrednosti, saj je ta lastnost neujemanja mediane in pričakovane vrednosti značilna za nesimetrične porazdelitve, kot je geometrijska porazdelitev.
    (Naredi si skice verjetnostnih funkcij tipičnih porazdelitev)

16. Vsak test na tečaju statistike je vreden 100 točk. Po štirih testih ima študent povprečje 85.0. Njegov rezultat na petem testu pa je 95.0.

    Kakšno je povprečje in standardni odklon za njegovih prvih pet testov?

    rešitev: Povprečje po petih testih izračunamo kot

    $$\frac{85 \cdot 4 + 95}{5} = 87$$

    Za izračuna standardnega odklona pa nimamo dovolj podatkov. Potrebovali bi rezultate vseh testov.

17. Diskretne slučajne spremenljivke

    - Definiraj slučajno spremenljivko in pojasni, kdaj je le-ta diskretna. Kako lahko podaš/predstaviš diskretno slučajno spremenljivko?

      rešitev: Slučajna spremenljivka je spremenljika, katere možne vrednosti so izzidi slučajnih pojavov. Diskretna je, ko ima števno domeno. Podamo jo lahko z verjetnostno tabelo, kjer se v stolpcu nahajata vrednost spremenljivke in verjetnost, da zavzame to vrednost.

    - Kaj je porazdelitvena funkcija in kako izgleda v primeru diskretne slučajne spremenljivke?

      rešitev: Zbirna funkcija verjetnosti ali porazdelitvena funkcija (oznaka cdf iz cumulative distribution function) je v verjetnostnem računu funkcija, ki opisuje verjetnostno porazdelitev realne slučajne spremenljivke X. Označuje se jo z $F(X)$.
      Meri verjetnost, da slučajna spremenljivka zavzame vrednost manjšo ali enako x. V primeru diskretne s.s. ima stopničasto obliko.

    - Naštej dva primera diskretnih slučajnih spremenljivk, ter pri vsaki opiši tudi njeno verjetnostno tabelo in zgled uporabe.

      rešitev: Primera diskretne slučajne spremenljivke sta spremenljivka porazdeljena po binomski porazdelitvi in spremeljivka porazdeljena po enakomerni diskretni porazdelitvi. Binomsko porazdeljena slučajna spremenljivka meri, koliko poskusov, pri katerih se lahko zgodita dva medsebojno disjunktna dogodka, je bilo uspešnih izmed vseh. Enakomerno porazdeljena s. s. pa lahko z enako verjetnostjo zavzame eno od n vrednosti. Primer je slučajna spremenljivka, ki beleži koliko pik je padlo na kocki.

- Kaj je to slučajni vektor in kako opišemo v primeru diskretne dvorazsežne porazdelitve njegovo verjetnostno funkcijo? Ali se da iz nje ugotoviti kdaj sta njegovi komponenti neodvisni (pojasni kako oziroma zakaj ne)?

  rešitev: Verjetnostni vektor ali stohastični vektor je vektor z nenegativnimi vrednostmi, ki se seštejejo v 1. Predstavlja nam verjetnosti, da slučajna spremenljivka zavzame določeno vrednost iz domene. Slučajni vektor je lahko tudi n-terica slučajnih spremenljivk $X = (X_1, ..., X_n)$. Tudi za slučajni vektor opišemo porazdelitveni zakon s porazdelitveno funkcijo ($x_i \in R$).

- Razloži, kaj si predstavljamo pod pogojno porazdelitvijo (lahko se omejiš na diskretni primer) in kaj je to pogojna verjetnostna funkcija

  rešitev: Pogojna porazdelitev je porazdelitev slučajnih spremenljivk X in Y, kjer predstavimo verjetnosti vseh možnih kombinacij vrednosti teh dveh slučajnih spremenljivk. Pogojna verjetnostna funkcija s. s. X je funkcija, ki vrednostim iz domene s.s. X dodeli verjetnosti, če vemo, kakšno vrednost je zavzela slučajna spremenljivka Y.

- Definiraj pogojno pričakovano vrednost in izračunaj pričakovano vrednost slučajne spremenljivke E(X|Y).

  rešitev: Pogojna pričakovana vrednost je povprečna vrednost spremenljivke v limiti, ko gre število poskusov proti neskončnosti, če vemo, kakšno vrednost je zavzela neka druga spremenljivka Y.
  $E(X|Y)$ izračunamo kot:

  $$E(X|Y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(Y - \mu_y)$$

18. Zvezne slučajne spremenljivke

- Definiraj slučajno spremenljivko in pojasni kdaj je le-ta zvezna. Kako z gostoto porazdelitve predstaviš verjetnostno porazdelitev zvezne slučajne spremenljivke?

  rešitev: Slučajna spremenljivka je spremenljivka, katere možne vrednosti so rezultati nekega naključnega pojava. Zvezna je, ko lahko zavzame neštevno mnogo vrednosti. Njena verjetnostna funkcija je zvezna. Gostota verjetnosti je enaka seštevku verjetnosti vseh možnih vrednosti manjših od x, ki jih s.s. lahko zavzame. Torej je enaka določenemu integralu $\int_{-\infty}^{x} p(t)dt$.

- Napiši vsaj tri lastnosti porazdelitvene funkcije. Kako izgleda le-ta v primeru diskretne slučajne spremenljivke (lahko poveš tudi samo, kako izgleda graf le-te)?

  rešitev: Porazdelitvena funkcija v primeru diskretne slučajne spremenljivke ima stopničasto obliko. Nekaj lastnosti:

  – Je nenegativna.

  – Je desno zvezna.

  – V limiti, ko x narašča proti neskončnosti, se približuje 1.

- Naštej dva primera zveznih slučajnih spremenljivk, ter pri vsaki podaj tudi njeno gostoto verjetnosti.

rešitev: Primera sta slučajna spremenljivka, porazdeljena s Poissonovo porazdelitvijo ter slučajna spremenlivka, porazdeljena z enakomerno zvezno porazdelitvijo.
Poissonova porazdelitev : $p_k(n) = \lambda^k \frac{e^{-\lambda}}{k!}$ Enakomerna zvezna porazdelitev:

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & sicer \end{cases}$$

- Kaj je to slučajni vektor? Kako z gostoto verjetnosti opišemo v primeru zvezne dvorazsežne porazdelitve njegovo verjetnostno funkcijo? Kakšnemu pogoju mora ustrezati verjetnostna funkcija, da so komponente slučajnega vektorja neodvisne?

rešitev: Definicija slučajnega vektorja - glej nekaj vprašanj nazaj.
V primeru zmezne dvorazsežne porazdelitve je porazdelitvena funkcija dvojni integral, ki meri prostornino pod funkcijo. Ta prostornina se v limiti akumulira do 1. Da so komponente slučajnega vektorja neodvisne, mora za gostoto verjetnosti veljati:

$$P(X = x \cap Y = y) = P(X = x) \cdot P(Y = y)$$

- Pri dvorazsežni diskretni slučajni spremenljivki definiraj $P(X = x_k|Y = y_h)$. Razloži s pomočjo verjetnostne tabele, kaj si (v tem primeru) predstavljamo pod pogojno porazdelitvijo (glede na pogoj $Y = y_h$).

rešitev: Ta vrednost funkcije nam predstavlja relativno verjetnost, da s.s. X zavzame vrednost x, če je s.s. Y zavzela vrednost y. Pri verjetnostni tabeli, kjer vsak stoplec predstavlja vrednost, ki jo lahko zavzame s.s. X in vsaka vrstica vrednost, ki jo lahko zavzame Y, gledamo stolpec označen z x_k in vrstico označeno z y_k. Vrednost, ki se nahaja v celici pri teh indeksih, je pogojna verjetnost, ki nas zanima.

- Definiraj pogojno pričakovano vrednost in izračunaj pričakovano vrednost slučajne spremenljivke $E(X|Y)$, tj. $E(E(X|Y))$ izrazi z $E(X)$.

rešitev: Pogojna pričakovana vrednost je dolgoročno povprečje vrednosti, ki jih zavzame slučajna spremenljivka X, če vemo katero vrednost je zavzela slučajna spremenljivka Y.

$$E(E(X|Y)) = \sum_{k=1}^{\infty} q_k E(X|y_k) = \sum_{k=1}^{\infty}\sum_{i=1}^{\infty} x_i p_{ik} = \sum_{i=1}^{\infty} x_i \sum_{k=1}^{\infty} p_{ik} = \sum_{i=1}^{\infty} x_i p_i = E(X)$$

19. Pričakovana vrednost slučajne spremenljivke

- Definiraj pričakovano vrednost (tj. matematično upanje) slučajne spremenljvike.

rešitev: Pričakovana vrednost slučajne spremenljivke je dolgoročna povprečna vrednost slučajne spremenljivke v limiti, ko gre število poskusov proti neskončnosti.

- Vsaka srečka stane 2 EUR. Prodanih je bilo 400 srečk, med njimi pa ena dobi 250 EUR, dve drugi pa vsaka po 25 EUR. Izračunaj, kakšen povprečen izkupiček lahko pričakuješ z eno srečko!

rešitev: Najprej izračunajmo verjetnosti za dobitke.
Verjetnost, da smo zadeli 250 EUR je $\frac{1}{400}$. Verjetnost, da smo zadeli 25 EUR pa je $\frac{2}{400}$.
Pričakovana vrednost dobitka je torej $E(X) = \frac{1}{400} \cdot 250 + \frac{2}{400} \cdot 25 = 0.75$ EUR.
Ker srečka stane 2 EUR, je pričakovana vrednost izkupička enaka $-2 + 0.75 = -1.25$.

- Naj bosta X in Y slučajni spremenljivki. Izpelji koliko je pričakovana vrednost spremenljivke $aX + bY$, kjer sta a in b realni števili.

rešitev:

$$E(aX + bY) = E(aX) + E(bY) = aE(X) + bE(Y)$$

- Izračunaj pričakovano vrednost slučajne spremenljivke W, ki je porazdeljena normalno $N(3, 14)$.

rešitev: Normalna porazdelitev je simetrična. Torej se pričakovana vrednost (povprečna vrednost) ter mediana ujemata. Pričakovana vrednost (oziroma povprečje) je ena od lastnosti, ki opisujejo normalno porazdelitev, zato jo lahko iz notacije kar razberemo. Enaka je $E(X) = \mu = 3$.

- Izračunaj pričakovano vrednost slučajne spremenljivke W, ki je porazdeljena binomsko $B(100, 0.25)$.

rešitev: pričakovano vrednost binomsko porazdeljene slučajne spremenljivke dobimo s formulo $E(X) = np$.
Torej velja:

$$E(X) = 100 \cdot 0.25 = 25$$

- Naj bo spremenljivka X porazdeljena po shemi

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ \frac{1}{10} & \frac{2}{10} & \frac{3}{10} & ? \end{pmatrix}$$

določi pričakovano vrednost $E(X)$

rešitev: Pričakovano vrednost lahko dobimo kot skalarni produkt vektorja, ki vsebuje verjetnosti z vektorjem, ki na enakem mestu vsebuje vrednost slučajne spremenljivke, ki ima to verjetnost.
Neznan parameter mora biti seveda tak, da se verjetnosti seštejejo v 1. Torej je verjetnost, da s.s. zavzame vrednost 4 enaka 4/10.
Rešitev (MATLAB sintaksa)

$$x * p' = 3$$

- Spremenljivka Y ima gostoto verjetnosti $f(y) = cy(1 - y)$ za $0 < y < 1$. Določi pričakovano vrednost E(Y).

rešitev: Množimo vrednosti s.s. z verjetnostmi po celi domeni:

$$E(Y) = \int_0^1 cy^2(1 - y)dy = ... = c(\frac{1}{3} - \frac{1}{4})$$

- Vzorčno povprečje je

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Kjer je $\{X_i\}$ množica, ki predstavlja slučajen vzorec slučajne spremenljivke X. Izračunaj $E(\bar{X})$, če veš, da je $E(X) = \mu$.

rešitev: Pričakovana vrednost vzorčnega povprečja lahko ocenimo kot

$$\mu_{\bar{X}} \approx \mu$$

In ker vemo, da je $E(X) = \mu$, potem je tudi $E(\bar{X}) = \mu$.

20. Disperzija slučajne spremenljvike

- Definiraj oziroma pojasni kaj je razpršenost (disperzija) slučajne spremenljivke.

rešitev: Disperzija je raztegnjenost oziroma skrčenost porazdelitve. Pogoste mere za disperzijo so varianca, standardna deviacija in medkvartilni razpon.

- Vržemo kovanec za 20 centov. Kar pade je slučajna spremenljivka X (če pokaže 20, je njena vrednost 20 sicer pa 0). Koliko je disperzija od X?

rešitev:
Uporabimo formulo:

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

Računamo:

$$E(X^2) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 400 = 200$$
$$E(X)^2 = (\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 20)^2 = 100$$

Velja torej:

$$D(X) = 200 - 100 = 100.$$

- Naj bodo X, Y in Z slučajne spremenljivke. Koliko je disperzija od aX, kjer je a realno število in koliko je disperzija od $Y + Z$, če sta slučajni spremenljivki Y in Z neodvisni?

rešitev: Iz lastnosti disperzije se spomnimo, da velja:

$$D(aX) = a^2 D(X).$$

Če sta spremenljivki Y in Z neodvisni, je njuna kovarianca enaka 0. Torej velja aditivno pravilo:

$$D(Y, Z) = D(Y) + D(Z)$$

21. Na spletni strežnik pride v povprečju 100 zahtev na uro, ki so porazdeljene Poissonovo. Ali je časovni razmik (v minutah) med dvema zaporednima zahtevama tudi porazdeljen Poissonovo?

rešitev: Odgovor je Ne. Za opisovanje razmika med dvema dogodkoma, če poznamo povprečno vrednost, kolikokrat na časovno enoto se dogodek zgodi, opisuje eksponentna porazdelitev.

22. Ali obstajata taki slučajni spremenljivki X in Y, da je

$$2K(X, Y) - D(Y) - D(X) = 0?$$

rešitev: Drži. Spomnimo se, da velja:

$$D(X + Y) = D(X) + D(Y) + K(X, Y)$$

Izraz na desni strani je v bistvu enak $-D(X - Y)$, ker je enako 0, ko je $X - Y$ konstanta.

23. Momenti so vzorčne statistike, kar pomeni, da so simetrične funkcije in zato z njimi ne moremo izmeriti asimetrije.

rešitev: To ne drži. TODO

24. ??

25. Ali je pričakovana vrednost poljubne simetrične porazdelitve enaka mediani?

rešitev: Pričakovana vrednost (oz. povprečje) ni nujno enaka mediani. Neeakost lahko opazimo pri asimetrični porazdelitvi, na primer porazdelitvi plač v neki državi.

26. Vzorce velikosti 25 izbiramo iz populacije s povprečjem 40 in standardnim odklonom 7.5. Potem sta pričakovana vrednost in standardni odklon vzorčnih povprečij enaka kateri vrednosti?

rešitev: Pričakovano vrednost vzorcev ocenimo iz pričakovane vrednosti populacije, standardni odklon vzorčnih povprečij pa ocenimo iz standardnega odklona populacije, ki ga delimo s kvadratnim korenom velikosti vzorca.

27. Naj bosta X in Y neodvisni slučajni spremenljivki. Med naslednjimi trditvami samo (c), (f) in (g) niso vedno pravilne:

   (a) $E(X+Y) = E(X) + E(Y)$

   (b) $E(XY) = E(X)E(Y)$

   (c) $D(XY) = D(X)D(Y)$

   (d) $D(X+Y) = D(X) + D(Y)$

   (e) $K(X,Y) = 0$

   (f) $\sigma(XY) = 0$

   (g) $\sigma(X+Y) = \sigma(X) + \sigma(Y)$

   (h) $\rho(X,Y) = 0$

rešitev: Drži. Spomnimo se ustreznih lastnosti iz te snovi.

28. Vržemo pošteno kocko. Izračunaj pričakovano vrednost števila pik.

rešitev: Pričakovano vrednosti izračunamo kot skalarni produkt vektorja, ki vsebuje vse možne vrednosti, ki jih lahko zavzame s.s. z vektorjem, ki na istoležnih mestih vsebuje verjetnosti, da s.s. zavzame to vrednost.
V MATLAB sintaksi:

$$e = x * p'$$

Pričakovana vrednost je torej enaka 3.5.

29. Nek mladi par si želi 3 otroke. Skupaj je 8 možnih razporeditev deklic in dečkov. Dogovorimo se, da oznaka ŽŽM pomeni, da sta prva dva odtroka deklici in tretji deček, ter podobno za ostale možnosti. Vseh 8 razporeditev je približno enako verjetnih.

   Izračunaj povprečno število deklic med temi otroki.

rešitev: Povprečno število deklic dobimo iz pričakovane vrednosti binomske porazdelitve.

$$E(X) = 3 \cdot \frac{1}{2} = 1.5$$

30. Profesor ekonomije ocenjuje svoje študente z ocenami od 0 do 4. Porazdelitev ocen je naslednja:

Razred	Verjetnost
0	0.1
1	0.15
2	0.3
3	0.3
4	0.15

Poišči povprečje (se pravi, sredno vrednost) ocene pri tem predmetu. Izdelaj verjetnostni histogram za porazdelitev ocen in na njem označi srednjo vrednost.

rešitev: Srednjo vrednost dobimo tako, da izračunamo pričakovano vrednost po znanem postopku. Dobimo rezultat 2.25.

31. Ameriška državna loterija je predstavila igro izberi 3, pri kateri ponujajo igralcu več možnih stav. Igralci izberejo trimestno število in stavijo 1$. Loterija vsak večer oznani zmagovalno trimestno število, ki ga izbere slučajno. Če igralec ugane vsa tri števila v poljubnem vrstnem redu, dobi 83.33 $, sicer pa izgubi vplačani dolar.

Poišči povprečni dobitek. (Predpostavi, da igralec izbere število s tremi različnimi števkami.)

rešitev: Vseh možnih kombinacij števk je $\binom{10+3-1}{3}$. Torej je verjetnost, da uganemo števke enaka $\frac{1}{220}$.
Pričakovana vrednost izkupička je torej $-1 + \frac{1}{220} \cdot 83.33 = -0.6212\$$.

32. V večini velikih ameriških mest je organizirana nekakšna ilegalna oblika lota, ki deluje na naslednji način: igralec izbere eno od 1000 trimestnih števil med 000 in 999 in plača posredniku en dolar, da lahko stavi na to število. Vsak dan slučajno izberejo eno trimestno število in če je igralec število uganil, dobi 600$, sicer pa izgubi vplačani dolar.

- Kolikšen je povprečni dobitek?

  rešitev: Verjetnost dobitka je $\frac{1}{10^3}$.
  Torej je povprečni dobitek $-1 + \frac{1}{1000} \cdot 600 = -0.4$

- Janez že veliko let vsak dan stavi na eno število. Kaj pravi zakov velikih števil o Janezovem dobičku, medtem ko Janez nadaljuje s stavami?

rešitev: Zakon pravi, da bo s časom Janezov povprečni izkupiček konvergiral k pričakovani vredosti. Torej bo Janez na dolgi rok vedno izgubljal denar.

33. Ameriška ruleta ima 38 žepkov, označenih z 0, 00 in s števili od 1 do 36. Ko zavrtimo kolo, kroglica z enako verjetnostjo pristane v kateremkoli žepku. Oznake žepkov so razporejene tudi po mizi, na kateri igralci postavljajo svoje stave. Eden od stolpcev vsebuje večkratnike števila 3, torej števila 3, 6,...,36. Igralec položi 1 EUR na stolpec in prejme 3 EUR, če se kroglica ustavi na kateremkoli od teh števil.

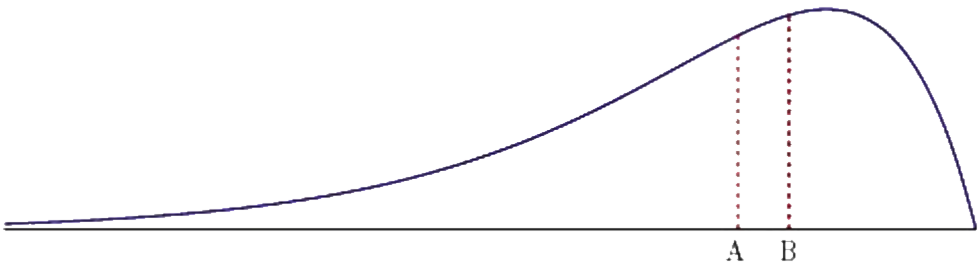
- Kolikšna je verjetnost, da zmaga?

rešitev: Žepkov označenih z večkratniki števila 3 je 12. Vseh možnosti je 38. Torej je verjetnost zmage $\frac{12}{38}$.

- Kolikšen je povprečni dobitek ene igre (na dve decimalki), če upoštevamo, da vsaka igra stane 1 EUR?

rešitev: Povprečni dobitek je $-1 + \frac{12}{38} \cdot 3 = -0.0526$.

34. Na sliki je prikazana nesimetrična verjetnostna porazdelitev. Srednja vrednost in mediana ne sovpadata. Katera od označenih črt je pričakovana vrednost in katera mediana? Odgovor utemelji.



rešitev: Porazdelitev je levo asimetrična. Torej je B mediana in A povprečna vrednost. PAZI v rešitvah je napaka!

35. Pri neki študiji smo izbrali vzorec učencev petega razreda in si zapisali, koliko let šolanja so sčasoma dokončali. Na osnovi te študije lahko podamo naslednji verjetnostni model za leta šolanja, ki jih bo s časoma končal slučajno izbrani petošolec.

Razred	Verjetnost
4	0.010
5	0.007
6	0.007
7	0.013
8	0.032
9	0.068
10	0.070
11	0.041
12	0.752

- Prepričaj se, da ta verjetnostni model zadošča pravilom.

rešitev: V MATLAB vnesemo podatke in nad vektorjem verjetnosti kličemo funkcijo cumsum. Verjetnosti se seštejejo v 1, torej model zadošča pravilom.

- Kateri izidi sestavljajo dogodek "učenec je končal vsaj eno leto srednje šole"? (Srednja šola se začne z devetim letom šolanja.) Kolikšna je verjetnost tega dogodka?

rešitev: Dogodek sestavlja seštevek verjetnosti, da je s.s. zavzela vrednost večjo od 10.

- Izračunaj povprečno število dokončanih let šolanja na dve decimalki natančno.

rešitev: Povprečno število let nam poda pričakovana vrednost, ki jo izračunamo po znanem postopku. Rezultat je 11.25.

36. V igralnicah je zelo popularna igra, imenovana Keno. Kroglice, označene s števili med 1 in 80, premešamo v posebni napravi, medtem ko lahko igralci stavijo tako, da označijo številke na kartici. Nato naključno izberemo 20 kroglic. Spodaj sta dva primera stav. Za vsakega napravi verjetnostni model izidov ter poišči srednjo vrednost in standardni odklon dobitkov.

    - Pri stavi 1 EUR na "Označi eno!" dobimo 3 EUR, če je označeno število med 20 izbranimi, sicer pa izgubimo vloženi evro.

      rešitev: Verjetnost, da bo označeno število med dvajsetimi izbranimi je $\frac{2}{8}$. Torej je pričakovana vrednost $-1 + \frac{2}{8} \cdot 3 = -0.25$

    - Pri stavi 1 EUR na "Označi dve!" dobimo 12 EUR, če sta obe označeni števili med 20 izbranimi. Verjetnost tega dogodka je približno 0.06.

      rešitev: Pričakovana vrednost je torej $-1 + 0.06 \cdot 12 = -0.28$

    - Ali je bolje igrati "Označi eno!" ali "Označi dve!"?

      rešitev: Na dolgi rok je izguba manjša pri "Označi eno!"

37. Obstaja preprost način za izdelavo verjetnostnega modela z dano srednjo vrednostjo μ in standardnim odklonom σ: izida sta samo dva, in sicer $\mu - \sigma$ ter $\mu + \sigma$, vsak pa se pojavi z verjetnostjo 0.5.

    S pomočjo definicij srednje vrednosti in variance verjetnostnega modela pokaći, da je res srednja vrednost takega modela enaka μ in standardni odklon σ.

    rešitev: Velja

    $$E(X) = \frac{\mu - \sigma}{2} + \frac{\mu + \sigma}{2} = \mu s(X) = \sqrt{E(X - \mu)^2} = \sqrt{\frac{\sigma^2}{2} + \frac{\sigma^2}{2}} = \sigma$$

1.7 06 - Binomska in normalna porazdelitev

1. V tovarni pakirajo sladkor v 5kg vreče. Količina sladkorja varira po normalni porazdelitvi in ima pričakovano vrednost 5kg ter standardni odklon 0.05 kg.

   Ali naslednja trditev drži?

   Za računanje verjetnosti dogodkov o težah posameznih vreč bomo uporabili spremenljivko $Z = (X - 5.0)/0.05$, medtem ko bomo za računanje verjetnosti dogodkov o težah vzorčnih povprečij za velikosti n = 25 uporabili spremenljivko $\bar{Z} = (\bar{X} - 5.0)/0.01$.

   rešitev: Trditev je pravilna. Da naravno porazdeljeno slučajno spremenljivko standardiziramo, od nje odštejemo njeno povprečje in jo normaliziramo z njeno standardno deviacijo. Vzorčno povprečje pa iz populacijskega povprečja ocenimo z $\bar{Z} = \frac{\bar{X} - 5.0}{0.01}$. Torej normaliziramo z kvocientom standardne deviacije deljenim z korenom velikosti vzorca, saj velja

   $$\sigma_{\bar{X}} \approx \frac{\sigma}{\sqrt{n}}.$$

2. V nekem podjetju so bile plače zaposlenih porazdeljene normalno s povprečjem 1000 EUR in standardnim odklonom 200 EUR. Ker je bilo podjetje uspešno, se je direktor odločil, da vsem zaposlenim zviša plače za 100 EUR. Ali se je s tem standardni odklon plač zmanjšal?

$Nedri.espremenimovrednostvsemenotampopulacijezaenakovrednost, sespremenilepovprejeinnestandardnioc$

3. Za binomsko porazdelitev $B(200, 0.25)$ izračunaj pričakovano vrednost.

   rešitev:

   Pričakovano vrednost $X \sim B(n, p)$ izračunamo s formulo np.

   Rezultat je torej 50.

4. Naprava izdeluje dele, od katerih je 2% defektnih. Če izbrani naključni vzore petih delov vsebuje dva ali več defektnih delov, je potrebno napravo zaustaviti in poklicati serviserja. Izračunaj verjetnost, da bo potrebno zaustaviti napravo na osnovi omenjenega načrta.

   rešitev: Slučajna spremenljivka, ki meri število defektnih delov v naključnem vzorcu petih delov je porazdeljena binomsko. Verjetnost, da bosta v vzorcu dva ali več defektna dela izračunamo kot:

   $$P(X > 2) = 1 - \binom{5}{1} \cdot 0.02^1 \cdot 0.98^4 - \binom{5}{0} \cdot 0.02^0 \cdot 0.98^5 = 0.004$$

5. Ploščina pod standardizirano normalno krivuljo med $z = 0.0$ in $z = 2.0$ je

   rešitev: Odčitamo od tabele ploščino od $-\infty$ do 2.0 in od te vrednosti odštejemo 0.5. Rezultat je 0.4772.

6. Ali naslednja trditev drži? Bernoullijevi poskusi predstavljajo zaporedje neodvisnih poskusov z dvema rezultatoma, ki ju pogosto poimenujemo uspeh in neuspeh, pri čemer je verjetnost uspeha enaka za vsak poskus.

   rešitev:

   Trditev drži. Primer je npr. metanje kovanca, kjer je uspeh, da npr. pade cifra.

7. Ali naslednja trditev vedno drži? Naj za dogodek A velja $p(A) = \frac{1}{10}$. Potem se bo v 10 poskusih dogodek A zgodil natanko 1-krat.

   rešitev:

   Trditev ne drži vedno, saj se razmerje med uspehi in neuspehi temu razmerju močno približa šele po velikem številu poskusov (zakon velikih števil).

8. Pošteni kovanec vržemo 5-krat.

   - Izračunaj verjetnost, da bo pri vseh 5-ih metih padla cifra.

     rešitev:

     $P(X = 5) = \binom{5}{5} 0.5^5 = 0.0312$

   - Izračunaj verjetnost, da bo padla cifra natanko 2-krat.

     rešitev:

     $P(X = 2) = \binom{5}{2} \cdot 0.5^2 \cdot 0.5^3 = 0.3125$

   - Izračunaj verjetnost, da bo padla cifra več kot 3-krat.

     rešitev:

     $P(X > 3) = \binom{5}{4} 0.5^4 \cdot 0.5^1 + \binom{5}{5} 0.5^5 \cdot 0.5^0 = 0.1875$

   - Izračunaj verjetnost, da bo padla cifra manj kot 2-krat.

     rešitev:

     $P(X > 3) = \binom{5}{0} 0.5^0 \cdot 0.5^5 + \binom{5}{1} 0.5^1 \cdot 0.5^4 = 0.1875$

- Če slučajna spremenljivka spremlja število cifer, koliko je pričakovana vrednost njene porazdelitve?

  rešitev:

  Pričakovana vrednost binomske porazdelitve je $E(X) = np = 5 \cdot 0.5 = 2.5$.

- Če slučajna spremenljivka spremnlja število cifer, koliko je standardni odklon njene porazdelitve?

  rešitev:

  Standardni odklon binomske porazdelitve je $\sigma = \sqrt{npq} = \sqrt{5 \cdot 0.25}$

9. Binomska porazdelitev ima 8 poskusov z verjetnostjo uspeha 0.25 pri vsakem poskusu.

- Izračunaj verjetnost natanko dveh uspehov.

  rešitev:

  $$\binom{8}{2} 0.25^2 \cdot 0.75^6 = 0.311$$

- Izračunaj verjetnost vsaj enega uspeha.

  rešitev:

  $$P(X > 1) = 1 - \binom{8}{0} \cdot 0.75^8 = 0.9$$

- Izračunaj verjetnost več kot 5-ih uspehov.

  rešitev:

  Seštej verjetnosti za $P(X = 6), P(X = 7), P(X = 8)$. (MATLAB: binocdf)

  Rezultat je 0.0042.

- Izračunaj verjetnost največ dveh uspehov.

  rešitev:

  binocdf(2, 8, 0.25)

  rezultat je 0.6785.

- Izračunaj pričakovano vrednost te porazdelitve.

  rešitev:

  Pričakovano vrednost binomske porazdelitve dobimo kot

  $$E(X) = np = 8 \cdot 0.25 = 2$$

- Izračunaj standardni odklon te porazdelitve.

  rešitev:

  Standardni odklon binomsko porazdeljene slučajne spremenljivke izračunamo kot:

  $$\sqrt{\sigma} = \sqrt{npq} = \sqrt{8 \cdot 0.25 \cdot 0.75} = 1.2247$$

10. Ali naslednja trditev drži? Pri binomski porazdelitvi parameter p predstavlja verjetnost, da se bo dogodek zgodil enkrat pri n ponovitvah.

    rešitev: Ne drži. p predstavlja verjetnost, da se pri posamezni ponovitvi poskusa dogodek zgodi.

11. Ali naslednja trditev drži? Za binomsko porazdelitev s fiksno vrednostjo p postaja z naraščanjem velikosti n binomska porazdelitev vse bolj podobna normalni.

rešitev: Drži. To je ena od ključnih značilnosti binomske porazdelitve, ki je uporabna tudi pri računanju.

12. Centralni limitni izrek pravi, da je pričakovana vrednost porazdelitve vzorčnih povprečij blizu povprečju populacije, če je vzorec velik.

rešitev: Ne drži. CLI pravi le, da je porazdelitev standardiziranega vzorčnega povprečja lahko dobro aproksimirana z $N(0,1)$ za npr. $n > 30$.

13. Slučajna spremenljivka X je porazdeljena binomsko: $B(4,p)$, tj. n = 4. Uporabi zgornjo sliko za določitev verjetnosti p:

rešitev: Iz slike lahko vidimo, da velja $\binom{5}{0}(1-p)^5 = 0.4096$
Iz tega izračunamo p, ki je enak 0.2.

14. Naj bo $X \sim B(50, 0.5)$. Predpostavimo, da bomo uporabili normalno aproksimacijo, da bi poiskali p_{30}. Potem je smiselno, da bi približek izračunali s čim?

rešitev: Normalna porazdelitev je lahko uporabljena kot aproksimacija za normalno porazdelitev pod nekaterimi pogoji, in sicer če je n velik in je p blizu $\frac{1}{2}$. V tem primeru velja $X \sim N(np, npq)$
Računali bi $P(29.5 < X < 30.5)$

15. Naj bo X slučajna spremenljivka, ki je porazdeljena $N(0,1)$. Kolikšna je verjetnost $P(X^2 < 1)$?

rešitev: X^2 je porazdeljena po hi-kvadrat porazdelitvi z eno stopnjo prostosti. Verjetnost $P(X^2 < 1)$ odčitamo iz tabele in dobimo rezultat 0.6827.
Matlab ukaz: chi2cdf(1, 1)

16. Kolikšna je verjetnost, da v desetih metih poštenega kovanca pade grb manj kot trikrat?

rešitev: Verjetnost, da v desetih metih poštenega kovanca pade grb manj kot trikrat lahko izračunamo tako, da izračunamo rezultat vsote $P(X=0) + p(X=1) + p(X=2)$. Lahko uporabimo MATLAB ukaz *binocdf(2, 10, 0.5)*. Rezultat je 0.0547.

17. Ana se igra z bratcem Binetom. Najprej Ana brcne žogo proti Binetu, potem pa jo Bine brcne nazaj. Razdalja, ki jo prepotuje žoga, ko jo brcne Ana, je porazdeljena kot $N(10,1)$, ko jo brcne Bine pa $N(5,1)$. Kako je porazdeljena razdalja med začetno in končno lego žoge?

rešitev: Naj je s.s. Z slučajna spremenljivka, ki beleži dolžino med začetno in končno lego žoge. Velja $Z = X - Y$. Velja torej $Z \sim N(\mu_X - \mu_Y, \sigma_X + \sigma_Y)$.

$$Z \sim (5, \sqrt{2})$$

18. V nekem podjetju so bile plače porazdeljene normalno s povprečjem 1000 evrov in standardnim odklonom 200 evrov. Ker je bilo podjetje uspešno, se je direktor odločil, da vsem zviša plače za 10%. Kako so porazdeljene po zvišanju?

rešitev: Če parameter vseh enot v populaciji povečamo za nek delež, se za ta isti delež poveča standardni odklon in povprečje. Velja torej, da so plače po zvišanju porazdeljene kot $N(1100, 220)$.

19. Neko ničelno domnevo želimo testirati s petimi neodvisnimi testi na različnih vzorcih. Približno koliko naj bo stopnja značilnosti pri testih, da bomo z verjetnostjo 0.4 vsaj enkrat zavrnili ničelno domnevo, če je le-ta pravilna?

rešitev: TODO

20. Neko domnevo želimo testirati s tremi neodvisnimi testi na različnih vzorcih. Signifikantnost testa naj bo 0.8. Približno kolikšna je verjetnost, da bomo vsaj enkrat zavrnili ničelno domnevo, če je le-ta pravilna?

rešitev: TODO

21. Vzorce velikosti 25 izbiramo iz populacije s povprečjem 40 in standardnim odklonom 7.5. Koliko sta potem sta pričakovana vrednost in standardni odklon vzorčnih povprečij enaka?

rešitev: Ker imamo znano povprečje in standardni odklon populacije, lahko iz njiju ocenimo pričakovano vrednost in standardni odklon vzorčnih povprečij na naslednji način:

$$\mu_{\bar{X}} \approx \mu$$
$$\sigma_{\bar{X}} \approx \frac{\sigma}{\sqrt{n}}$$

Kjer je n velikost vzorcev.
Iz tega sledi, da je pričakovana vrednost vzorčnih povprečij $\mu_{\bar{X}}40$ in standardni odklon $\sigma_{\bar{X}} = 1.5$.

22. Ali drži naslednja trditev? Centralni limitni izrek pravi, da je porazdelitev vzorčnih povprečij približno normalna, kadar je velikost vzorca n dovolj velika.

rešitev: Da, če je naključni vzorec velikosti n izbran iz populacije s končno pričakovano vrednostjo μ in končno varianco σ. (Na vprašanju je odgovor ne).

23. Šestnajstsedežno letalo je preobremenjeno, če nosi težo večjo od 1440 kg. Vemo, da je teža enega potnika skupaj z njegovo prtljago porazdeljena normalno z upanjem 85 kg in standardnim odklonom 15 kg.

Kolikšen del letal bo preobremenjenih, če potnike izberemo naključno (in so vsi sedeži zasedeni).

rešitev: Porazdelitev vsote neodvisnih normalno porazdeljenih slučajnih spremenljivk lahko zapišemo kot

$$\mu = \sum_{i=1}^{16} \mu_i$$
$$\sigma^2 = \sum_{i=1}^{16} \sigma_i^2$$

Iz tega izračunamo, da je spremenljivka, ki meri vsoto tež potnikov na letalu porazdeljena kot

$$Z \sim N(1360, 60)$$

Izračunamo standardizirano kritično vrednost s.s. (z-score)

$$z = \frac{x - \mu}{\sigma} = 1.3333$$

Iz tabele vidimo, da je verjetnost, da bo standardizirana slučajna spremenljivka zavzela vrednost manjšo kot 1.3333 enaka 0.9082. Da dobimi verjetnost, da zavzame večjo vrednost od te preprosto od 1 odštejemo to vrednost.
Dobimo verjetnost 0.0918, kar je naša rešitev.

24. Če je slučajna spremenljivka X porazdeljena standardno normalno $N(0, 1)$, potem za vsako realno število x velja

$$P(X < x) = 1 - P(X > x)$$

rešitev: Vedno drži, saj meja x razdvaja krivuljo na dva dela, katerih vsota ploščin je enaka 1.

25. Ali je naslednja trditev resnična? Če je slučajna spremenljivka X porazdeljena binomsko $B(n, p)$, potem za vsako celo število k velja:

$$P(X < k) = 1 - P(X > k)$$

rešitev: Ne drži. Morali bi tudi upoštevati možnost, da je X = k.

26. Ali je nasldnja trditev resnična? Zaloga vrednosti standardizirane normalne porazdelitve je $[0, \frac{1}{\sqrt{2\pi}}]$.

rešitev: To je zaloga vrednosti gostote porazdelitve standardizirane normalne porazdelitve, zaloga vrednosti slučajne spremenljivke, ki je porazdeljena normalno pa so v vsakem primeru vsa realna števila. Trditev torej ni resnična.

27. Slučajna spremenljivka X je normalno porazdeljena. Vemo, da je $X + Y \sim N(4, 5)$.

    Kjer je $Y \sim N(3, 4)$ neodvisna od X. Kako je porazdeljena spremenljivka X?

rešitev: Vemo, da velja $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$
Torej rešujemo enačbo $16 + x^2 = 25$, da dobimo σ_X^2. Rešitev je torej $X \sim N(1, 3)$

28. 300-krat vržemo pošten kovanec, meti so neodvisni. Kakšen je najboljši približek za verjetnost, da pade natanko 150 grbov?

rešitev: Uporabimo De Moivrov obrazec. izkaže se, da je najboljši približek 1/20 iz točke c.

29. Za normalno slučajno spremenljivko $X \sim N(\mu, \sigma)$ je $P(X < 3)$ približno 0.16. Kakšne so vrednosti μ in σ?

rešitev:

$$z = \frac{x - \mu}{\sigma}$$

Za x vstavimo 0.16, z vrednost preberemo iz tabele. Preizkusi pare μ in σ, da dobiš najbolj ustrezen rezultat.

30. Bernoullijevo zaporedje neodvisnih poskusov

    - Definiraj zaporedje neodvisnih poskusov.

      rešitev: Zaporedje neodvisnih poskusov je zaporedje poskusov, ki so med sabo neodvisni. Torej rezultat naslednjega poskusa ni odvisen od prejšnega.

    - Kaj je to Bernoullijevo zaporedje neodvisnih poskusov?

      rešitev: Bernoullijevo zaporedje neodvisnih poskusov je zaporedje neodvisnih poskusov, kjer imamo v vsakem poskusu samo dva izključujoča izzida (kanonično 1 in 0). Primer je npr. metanje kovanca.

    - Kaj so to kombinacije in kako izračunamo njihovo število?

      rešitev: Kombinacije so izbori k elementov iz množice n elementov, kjer vrstni red ni pomemben. Število možnih kombinacij velikosti k iz množice velikosti n izračunamo kot $\binom{k}{n} = \frac{n!}{k! \cdot (n-k)!}$.

- Izpelji Bernoullijev obrazec in predstavi dva načina/metodi za njegovo računanje.

rešitev: Bernoullijev obrazec pravimo enačbi:

$$p_n(k) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$

Število načinov izbora ugodnih dogodkov množimo z verjetnostjo, da se je zgodilo k ugodnih dogodkov ter verjetnostjo, da so bili ostali dogodki neugodni.
Če je n velik, lahko uporabimo De Moivrov obrazec. Če je n velik in $p \approx \frac{1}{2}$, lahko uporabimo Laplacov obrazec.

- Podaj Bernoullijev zakov velikih števil

rešitev: Za vsak $\epsilon > 0$ velja:

$$\lim_{n\to\infty} P(|\frac{k}{n} - p| < \epsilon) = 1$$

V limiti, ko se število poskusov približuje neskončnosti, se razmerje ugodnih in vseh dogodkov približuje verjetnosti.

31. Ali je naslednja trditev pravilna? Šestnajstsedežno letalo je preobremenjeno, če nosi težo večjo od 1440 kg. Vemo, da je teža enega potnika skupaj z njegovo prtljago porazdeljena normalno z upanjem 85 kg in standardnim odklonom 15 kg. Če potnike izberemo naključno in so vsi sedeži zasedeni, je delež letal, ki so preobremenjena, enak 0.9082.

rešitev: Glej prejšno vprašanje. Ni res. To je delež letal, ki niso preobremenjena. Da dobimo pravilen delež, moramo to vrednost odšteti od 1.

32. Ali drži naslednja trditev? Zaloga vrednosti normalne porazdelitve $X \sim N(0,1)$ (in ne njene gostote) je $[0, \frac{1}{\sqrt{2\pi}}]$.

rešitev: Ne drži. To vprašanje je že bilo rešeno. Glej nazaj.

33. Kovanec mečemo, dokler trikrat ne pade grb. Katera ocena za verjetnost, da bomo kovanec vrgli vsaj sedemkrat, je najboljša?

rešitev: Ta naloga je že bila rešena. (Negativna binomska porazdelitev...)

34. Približno kakšen je 20. percentil (0.2-kvartil) standardne normalne porazdelitve?

rešitev: V tabelo za CDF standardne normalne porazdelitve poglej, nad katero vrednostjo je še 20% vrednosti. Izkaže se, da je ta vrednost približno 0.85. Zaradi simetričnosti normalne porazdelitve lahko sklepamo, da pod vrednostjo približno -0.85 20% vrednosti.

35. Porazdelitev višin odraslih slovenskih moških je približno normalna s srednjo vrednostjo 175 cm in standardnim odklonom 6 cm. Nariši normalno krivuljo, na kateri sta pravilno označena srednja vrednost in standardni odklon.

rešitev: ...

36. Porazdelitev višin odraslih ameriških moških je približno normalna s srednjo vrednostjo 175 cm in standardnim odklonom 6 cm. Z uporabo pravila 68 - 95 - 99.7 odgovori na naslednja vprašanja o višinah odraslih američanov.

- Med katerima dvema višinama leži srednjih 95 višin?

rešitev: Srednjih 95% višin leži (z uporabo tega pravila) približno v razdalji dveh standardnih deviacij od povprečja. Torej ležijo med višinama 163 in 187.

- Kolikšen delež moških je višji od 190 cm?

rešitev: Uporabimo formulo $z = \frac{x-\mu}{\sigma}$. Ugotovimo, da je oseba z višino 190 cm približno 2.5 standardnih odklonov desno od povprečja. (z vrednost je 2.5).
Iz tabele razberemo, da je verjetnost, da je oseba nad to vrednostjo enaka 0.0062.

- Kolikšen delež moških je nižjih od 170 cm?

rešitev: Uporabimo podobne izračune, da dobimo rezultat $z = 0.8333 \to p = 0.2033$

37. Porazdelitev višin odraslih ameriških moških je približno normalna s srednjo vrednostjo 175 cm in standardnim odklonom 6 cm.

Kje so kvartili porazdelitve?

rešitev: Kvartili porazdelitve so tam, kjer je verjetnost, da s.s. zavzame to vrednost enaka 0.25, 0.5 in 0.75.
Prvi kvartil:

$$x = 6 \cdot -0.67 + 175 = 170.98$$

Drugi kvartil je enak mediani.
Tretji kvartil:

$$x = 6 \cdot 0.67 + 175 = 179.02$$

38. Rezultati standardnega inteligenčnega testa *Wechsler Adult Intelligence Scale* za skupino med 20. in 34. letom starosti so porazdeljeni približno normalno z $\mu = 110$ in $\sigma = 25$.

- Kolikšen delež ljudi iz te skupine doseže rezultate nad 110?

rešitev:

$$Kerjeporazdelitevsimetrina, jihdoseepriblinopolovica.$$

- Kolikšen delež teh ljudi ima rezultat višji od 160?

rešitev:

$$z = \frac{160 - 110}{25} = 2$$

Kolikšen delež jih je oddaljenih za dve stadardni deviaciji od povprečja lahko ocenimo s pravilom 68-95-99.7.
Takih je približno 2.5% oseb.
Oziroma, če uporabimo tabelo z-vrednosti, dobimo delež 2.28% oseb.

- Poišči kvartila te porazdelitve. S preprostimi besedami razloži, kaj nam povesta ti dve števili.

rešitev:

$$Q1 = z \cdot \sigma + \mu = -0.67 \cdot 25 + 110 = 93.25$$

$$Q3 = z \cdot \sigma + \mu = 0.67 \cdot 25 + 110 = 126.75$$

Kaj nam povedo kvartili smo že omenili pri enem od prejšnih vprašanj.

39. Vojska poroča, da je porazdelitev obsega glav vojakov približno normalna s srednjo vrednostjo 58 cm in standardnim odklonom 3 cm.

- Kolikšen delež vojakov ima obseg glave večji od 60 cm?

rešitev: Računamo po znanem postopku:

$$z = \frac{60-58}{3} = 0.6667 \rightarrow p = 0.2546$$

- Vojska želi čelade pripraviti vnaprej in hoče, da bi se prilegale sredinskim 95 vojakom. Preostalim vojakom bodo izdelali čelade posebej. Kateri obsegi so še dovolj majhni ali pa dovolj veliki, da si bodo "prislužili" čelade po naročilu?

rešitev: Po pravilu 68-95-97 lahko čez palec rečemo, da bo meja pri približno dveh standardnih deviacijah stran od povprečja. To je v našem primeru 52 ter 64.

40. Rezultati sprejemnih izpitov SAT so približno normalno razporejeni s srednjo vrednostjo $\mu = 500$ in standardnim odklonom $\sigma = 100$.

- Naključno izberi enega kandidata. Kolikšna je verjetnost, da je rezultat izbranega kandidata večji od 500? Večji od 600? Ker predstavlja vrednost 50 ravno mediano, je verjetnost ravno 0.5. Za verjetnost rezultata večjega of 600 računamo po znanem postopku.

rešitev:

$$z = \frac{600-500}{100} = 1 \rightarrow p = 1 - 0.8413 = 0.1587$$

- Izberi enostavni slučajni vzorec velikosti 4. Kolikšna je verjetnost, da je povprečje njihovih rezultatov večje od 500? Večje od 600?

rešitev: Standardno deviacijo vzorčnih povprečij ocenimo z $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$
Verjetnost, da bo večje od 500 je kar mediana. Verjetnost, da bo večje od 600 je po poračunanem znanem postopku enako 0.0228.

41. Juan v kemijskem laboratoriju izvede nekaj meritev in o rezultatih poroča v svojem laboratorijskem poročilu. Standardni odklon meritev, ki so jih dobili študenti, je $\sigma = 10mg$. Juan ponovi meritve triktrat in izračuna povprečje $\bar{x}$ vseh treh rezultatov.

- Kolikšen je standardni odklon povprečja, ki ga je izračunal Juan? (Se pravi, če bi Juan nadaljeval z izvajanjem meritev in računal povprečja treh zaporednih rezultatov, kolikšen bi bil standardni odklon vseh tako oddaljenih $\bar{x}$?

rešitev: Standardni odklon bi bil okoli $\sigma_{\bar{X}} = \frac{10}{\sqrt{3}} = 5.7735$

- Kolikokrat mora Juan ponoviti meritve, da bi zmanjšal standardni odklon vrednosti $\bar{X}$ na 5? Razloži nekomu, ki ne zna statistike, zakaj je bolje poznati povprečje večih meritev kot pa rezultate ene same.

rešitev:

$$\frac{10}{\sqrt{n}} = 5$$

$$\sqrt{n} = 2$$

$$n = 4$$

Narediti mora torej vsaj 4 meritve.

42. Študentska organizacija namerava vprašati vzorec 50 študentov, če so opazili nove izobraževalne brošure o AIDSu. Zabeležili si bodo delež pozitivnih odgovorov. Njihov svetovalec za statistiko pravi, da bo standardni odklon tega deleža približno 7. Kolikšen bi bil standardni odklon, če bi vzorec vseboval 100 študentov in ne 50?

rešitev: Če vzorec n-krat povečamo, se standardni odklon vzorčnega povprečja $\sqrt{n}$-krat zmanjša.

43. Študentska organizacija namerava vprašati vzorec 50 študentov, če so opazili nove izobraževalne brošure o AIDSu. Zabeležili si bodo delež pozitivnih odgovorov. Njihov svetovalec za statistiko pravi, da bo standardni odklon tega deleža približno 7.

Kako velik vzorec bi morali izbrati, da bi zmanjšali standardni odklon deleža pozitivnih odgovorov iz 7 na 3.5? Pojasni nekomu, ki ne zna statistike, zakaj so večji vzorci pri raziskavah mnenja boljši kot manjši.

rešitev:

$$\frac{7}{\sqrt{n}} = 3.5$$
$$\frac{7}{3.5} = \sqrt{n}$$
$$n = 4$$

Torej bi vzorec morali povečat štirikrat.

44. Koncentracija aktivne sestavine v kapsulah je porazdeljena normalno z $\mu = 10$ in $\sigma = 0.2$.

- Kolikšna je povprečna koncentracija?

  rešitev: Povprečna koncentracija je 10.

- Na katerem intervalu se nahaja srednjih 95 vseh koncentracij?

  rešitev: Z uporabo pravila 68-95-99.7 lahko čez palec rečemo, da se srednjih 95 procentov koncentracij nahaja približno v območju dveh standardnih deviacij stran od povprečja. Torej na intervalu $[9.6, 10.4]$.

- Na katerem intervalu se nahajajo koncentracije srednje polovice vseh kapsul?

  rešitev: Nahajajo se med prvim in tretjim kvartilom.

  Računajmo , kje se nahajata prvi in drugi kvartil.

  $$Q1 = -0.67 \cdot 0.2 + 10 = 9.866$$
  $$Q3 = 0.67 \cdot 0.2 + 10 = 10.134$$

  Torej se nahajajo med tema dvema vrednostima.

- Pri kolikšnem odstotku kapsul je koncentracija aktivne sestavine višja od 10.4?

  rešitev:

  $$z = \frac{10.4 - 10}{0.2} = 2 \rightarrow p = 0.0228 \quad (2.28\%)$$

- Pri kolikšnem odstotku kapsul je koncentracija aktivne sestavine višja od 10.6?

  rešitev:

  $$z = \frac{10.6 - 10}{0.2} = 3 \rightarrow p = 0.0013 \quad (0.13\%)$$

45. Dolžina nosečnosti pri človeku se spreminja po porazdelitvi, ki je približno normalna s srednjo vrednostjo 266 dni in standardnim odklonom 16 dni.

- Med kateri vrednosti pade srednjih 95 vseh nosečnosti?

rešitev: Zanima nas, s katerimi vrednostmi je omejenih 95% nosečnosti. Spomnimo se pravila 68-95-99.7, ki med drugim pravi, da približno 95% vseh vrednosti slučajne spremenljivke leži v razdalji dveh standardnih deviacij of povprečja. Torej sta ti dve števili 234 in 298.

- Kako dolgih je najkrajših 2.5 nosečnosti?

rešitev: Uporabimo formulo

$$z = \frac{x - \mu}{\sigma}$$

$$x = \sigma \cdot z + \mu$$

z razberemo iz tabele. Zanima nas, koliko standardnih deviacij levo od povprečja je 2.5% nosečnosti. Rezultat je -2.81 standardnih deviacij.

Vstavimo vrednosti v formulo in izračunamo rezultat 221 dni.

Do rezultata bi lahko prišli tudi tako, da bi zopet uporabili pravilo 68-95-99.7

46. Decila porazelitve sta točki, pod katerima leži 10 (spodnji decil) oziroma 90 (zgornji decil) vseh vrednosti. Med spodnjim in zgornjim decilom leži 80 vseh podatkov. Pri normalnih porazdelitvah se nahajata na razdalji 1.28 standardnega odklona od srednje vrednosti.

    Koliko točk mora doseči študent, da sodi med zgornjih 10 porazdelitve rezultatov SAT (ki je normalna s srednjo vrednostjo 500 in standardnim odklonom 100)?

    rešitev: Zopet uporabimo formulo

    $$z = \frac{x - \mu}{\sigma}$$

    Iz enačbe izrazimo x in računamo:

    $$x = 100 \cdot 1.28 + 500 = 628$$

    Pisati mora toraj vsaj 628 točk.

47. Dolžina nosečnosti pri človeku se spreminja po porazdelitvi, ki je približno normalna s srednjo vrednostjo 266 dni in standardnim odklonom 16 dni.

    Kako dolgih je najkrajših 10 nosečnosti?

    rešitev: Dolge so manj kot $x = -1.28 \cdot 16 + 266 = 245.52$.

1.8 07 - Večrazsežne porazdelitve

1. Slučajni vektor je porazdeljen diskretno z naslednjo verjetnostno tabelo

$Y \setminus X$	0	1	2
-1	0	$1/4$	0
0	0	$1/4$	$1/4$
1	$1/4$	0	0

Izračunak $K(X, Y)$.

rešitev: Za izračun kovariance uporabimo formulo $K(X,Y) = E(XY) - E(X)E(Y)$.

Najprej izračunajmo vrednosti $E(X)$ in $E(Y)$. Potrebujemo robni porazdelitvi obeh spremenljivk.

$$X \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

$$Y \sim \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

Pričakovane vrednosti $E(X)$ in $E(Y)$ nato izračunamo po znanem postopku in dobimo rezultata $E(X) = 1$ in $E(Y) = 0$

Izračunajmo še E(XY). To vrednost dobimo tako, da vse verjetnosti v tabeli množimo z vrednostmi X in Y za ta par stolpca in vrstice. Dobimo rezultat $E(XY) = -\frac{1}{4}$.

Sedaj lahko izračunamo kovarianco:

$$K(X,Y) = -\frac{1}{4} - 0 = -\frac{1}{4}$$

2. Iz kupa igralnih kart (52) na slepo izberemo eno karto in jo nato vrnemo nazaj. Postopek ponovimo 5-krat.

rešitev: Že rešeno. Glej nazaj.

3. Če v kocki s stranico 1 enakomerno izbiramo slučajne točke, njihove koordinate pa spremlja slučajni vektor (X,Y,Z), potem slučajne spremenljivke (X,Y,Z), niso porazdeljene enakomerno, so pa neodvisne.

rešitev: Že rešeno. Glej nazaj.

4. Naj bosta slučajni spremenljivki X in Y standardno normalno porazdeljeni in neodvisni.

   Potem so zadnje tri trditve pravilne, ostale pa niso pravilne.

rešitev: Že rešeno, glej nazaj.

5. Koliko meri ploščina standardne normalne porazdelitve desno od $z = 2.1$?

rešitev: Preverimo tabelo standardne normalne porazdelitve in od 1 odštejemo ploščino levo od te točke. Rezultat je 0.0179. Pazi, v rešitvah je napaka.

6. Naj bo gostota slučajnega vektorja (X,Y) podana s $p(x,y) = f(x,y)$, kjer je

$$f(x,y) = \begin{cases} cx & \text{če je } 0 \leq x \leq y \text{ in } 0 \leq y \leq 1 \\ 0 & \text{sicer} \end{cases}$$

   Kakšna je vrednost konstante c?

rešitev: Vrednosti se morajo sešteti v 1.

$$\int_0^y \int_0^1 cx \cdot dy \cdot dx = 1$$
$$\int_0^y cx \cdot dx = 1$$
$$\frac{cy^2}{2} = 1$$
$$c = \frac{2}{y^2}$$

Rezultat torej ni nič od naštetega.

7. V tabeli je skupna porazdelitev slučajnih spremenljivk X in Y:

$X \backslash Y$	0	1	2	
-1	0.1	0.2	0.2	
1	0.2	0.1	0.2	

Katera trditev ni resnična (glej Kvizko i3-117)

rešitev: Resnična ni trditev $P(Y = 2|X = 1) = 0.5$. To preverimo tako, da zapišemo porazdelitev s.s. Y za primer $X = 1$. Vse verjetnosti v tej vrstici normaliziramo tako, da jih delimo z robno gostoto verjetnosti v tej vrstici (z 0.5).

8. Kombinatorika in verjetnost

- Definiraj pojem permutacije n elementov ($n!$ različnih jih je) ter pojem permutacije s ponavljanjem (kako so povezane s kombinacijami).

  rešitev: Permutacija je z medsebojnimi zamenjavami preurejeno zaporedje znanega končnega števila elementov (pri tem pa število elementov ostane enako). V matematiki je to bijekcija množice elementov samega v sebe, ki se jo zapiše kot:

  $$\pi : X_n \to X_n$$

- Definiraj Γ funkcijo in podaj čim več njenih vrednosti (vsaj eno celoštevilsko).

  rešitev: Glej zapiske za teoretični del.

- Definiraj nezdružljiva dogodka ter popoln sistem dogodkov in pojasni, kako slednjega uporabimo za klasično definicijo verjetnosti.

  rešitev: Že odgovorjeno - glej nazaj.

- Definiraj vsoto in produkt dogodkov ter utemelji zvezo med verjetnostmi dveh dogodkov, njuno vsoto ter njunim produktom.

  rešitev: Že odgovorjeno - glej nazaj.

- Opiši polinomsko porazdelitev (ali spada med večrazsežne diskretne ali zvezne porazdelitve) ali pa definiraj mediano za slučajno spremenljvko X (to ni mediana zaporedja).

rešitev: Polinomska porazdelitev $P(n; p_1, ..., p_r)$, $\sum p_i = 1$, $\sum k_i = n$ je določena s predpisom:

$$P(X_1 = k_1, ..., X_r = k_r) = \frac{n!}{k_1!...k_r!} p_1^{k_1} ... p_r^{k_r}$$

Koeficient šteje permutacije s ponavljanjem. Za r = 1 dobimo binomsko porazdelitev, tj. $B(n, p) = P(n; p, q)$.

Mediana je vrednost, od katere je polovica vrednosti v populaciji manjših, polovica pa večjih. primer: medianska plača...

9. slučajne spremenljivke

   Ta vprašanja so že bila odgovorjena - glej nazaj.

10. Diskretne slučajne spremenljivke

    Ta vprašanja so že bila odgovorjena - glej nazaj.

11. Zvezne slučajne spremenljivke

    Ta vprašanja so že bila odgovorjena - glej nazaj.

12. Slučajne spremenljivke in neodvisnot

    Ta vprašanja so že bila odgovorjena - glej nazaj.

13. Normalna porazdelitev

    Ta vprašanja so že bila odgovorjena - glej nazaj.

14. Porazdelitvene funkcije vektorjev in srednje vrednosti

    - Definiraj porazdelitveno funkcijo slučajnega vektorja (X, Y) in naštej vsaj tri njene lastnosti.

      TODO

    - Utemelji katere od naslednjih funkcij so porazdelitvene funkcije slučajnega vektorja (X, Y) in katere ne morejo biti:

      (a) $F(x, y) = 1 - e^{-x-y}$, ko sta x in y oba nenegativna, sicer pa je 0,

      (b) $F(x, y) = 1 + e^{-x-y} - e^{-x} - e^{-y}$, ko sta x in y oba nenegativna, sicer pa je 0.

      (c) $F(x, y) = x^2$

      (d) $F(x, y) = x^2 - y^2$

      (e) $F(x, y) = 0$

      TODO

15. Definiraj pričakovano vrednost in standardni odklon slučajne spremenljivke.

    TODO

16. Opiši postopek za standardnizacijo slučajne spremenljivke in zapiši njeno pričakovano vrednost ter njen odklon.

    TODO

17. Opiši binomsko in normalno porazdelitev ter Laplaceov obrazec

TODO

18. Predstavi centralni limitni izrek (CLI)

TODO

19. Opiši večrazsežno normalno porazdelitev

TODO

20. Naj bo X slučajna spremenljivka z $E(X) = \mu$ in $D(X) = \sigma^2$. Za njen slučajni vzorec $\{X_i\}_{i=1}^n$ definiraj vzorčno povprečje $\bar{X}$ in napiši, kaj se dogaja z vzorčnim povprečjem $E(\bar{X})$ in s standardno napako $D(\bar{X})$ z naraščanjem velikosti vzorca.

TODO

21. Ali je funkcija $F(x, y)$, ki je enaka $1+e^{-x-y}-e^{-x}-e^{-y}$, za $x, y \geq 0$, sicer pa 0 porazdelitvena funkcija nekega slučajnega vektorja (X, Y)?

TODO

22. Slučajni vektor (X, Y) je porazdeljen diskretno z naslednjo verjetnostno tabelo (kjer je tiskarski škrat nadomestil število v enem polju z vprašajem).

$Y \setminus X$	-1	1
-1	1/2	0
1	?	1/2

Ali sta slučajni spremenljivki X in Y neodvisni?

Verjetnosti v tabeli se morajo sešteti v 1. To se zgodi le, če je verjetnost, ki jo pokriva vprašaj enaka 0. Spremenljivki ne moreta biti neodvisni, kar lahko sklepamo že ob opazuvanju ničelnih verjetnosti v tabeli.

23. Iz kupa igrali hark (52) na slepo izberemo eno karto in jo nato vrnemo nazaj. Postopek ponovimo 5-krat. Kakšna je verjetnost, da bomo videli dvakrat srce, po enkrat pa pika, križa in karo?

Ta naloga je že bila rešena - glej nazaj.

24. Spodaj so dane vsote po vrsticah in stolpcih dvosmerne tabele z dvema vrsticama in dvema stolpema.

Poišči dva različna nabora števil a, b, c in d, ki bi data enake vsote po vrsticah in stolpcih. Ta primer nam pokaže, da ne moremo razbrati zveze med dvema spremenljivkama le iz posameznih porazdelitev.

Rešitev so npr. vrednosti $a = 10, b = 40, c = 50, d = 0$ ali pa $a = 20, b = 30, c = 40, d = 10$. V splošnem vrednost ene od črk izberemo poljubno in poračunamo ostale.

1.9 08 - Kovarianca

1. Slučajni spremenljivki X in Y sta porazdeljeni enakomerno na intervalu $(0, 1)$. (Ni se težko prepričati, da je $D(X) = \frac{1}{12}$).

   Če velja $K(X, Y) = \frac{1}{24}$, potem je $D(X + 2Y) = \frac{5}{12}$.

   Ne drži. TODO zakaj

2. Za zaporedje parov podatkov so spremenili mere iz čevljev v palce, kar pomeni, da somorali vse podatke pomnožiti z 12. Ali drži, da bo korelacijski koeficient za nove podatke bo 12-krat večji od originalnega?

   Pearsonov korelacijski koeficient meri, kako močna je koreliranost med slučajnimi spremenljivkami. Torej če vse podatke množimo z isto vrednostjo, se bo korelacijski koeficient ohranil. To lahko pokažemo tudi računsko.

3. Ali je korelacija simetrična funkcija?

   Funkcija dveh spremenljivk je simetrična, če velja $f(x, y) = f(y, x)$. Ker je kovarianca simetrična funkcija in ker je produkt standardnih odklonov komutativen, je tudi korelacija simetrična funkcija.

4. Naj bosta X in Y enako porazdeljeni slučajni spremenljivki.

   Čemu je izraz $K(X, Y) - D(X)$ vedno enak?

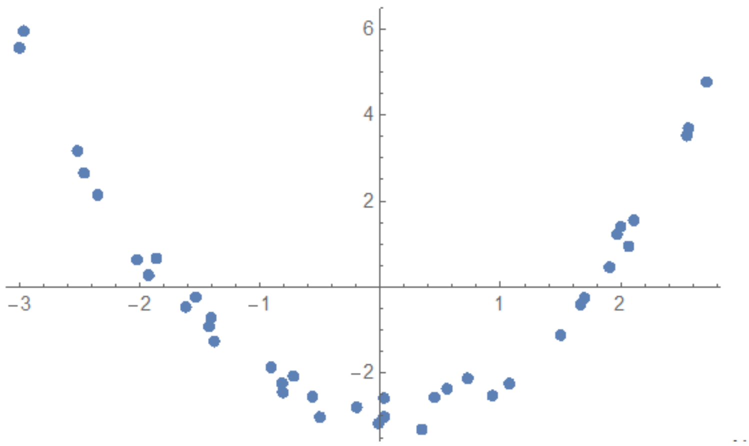
   Izraz je vedno enak $E(XY) - E(X^2)$.

5. Naj bosta X in Y neodvisni slučajni spremenljivki.

   Čemu je izraz $K(X, Y) - D(X)$ vedno enak?

   TODO

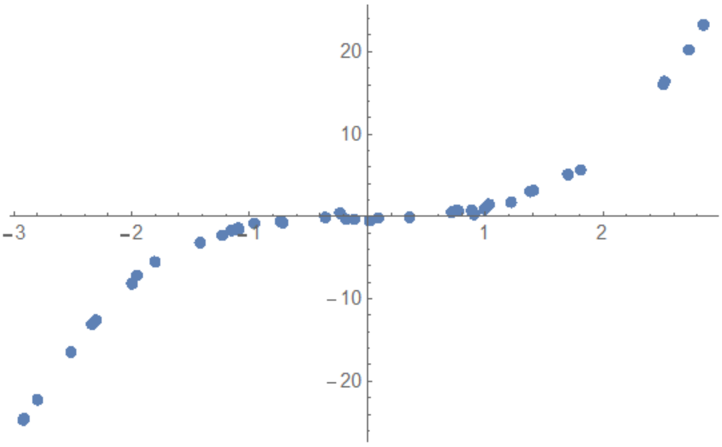
6. Na katere spodnje trditve o naslednjem diagramu lahko odgovorimo pritrdilno?



   1. Spremenljivki X in Y sta neodvisni.
   2. Korelacijski koeficient je majhen.
   3. Zveza med spremenljivkama je linearna.

   rešitev: Ne drži nobena trditev zaradi očitnih razlogov.

7. Na katere trditve o diagramu lahko odgovorimo pritrdilno?



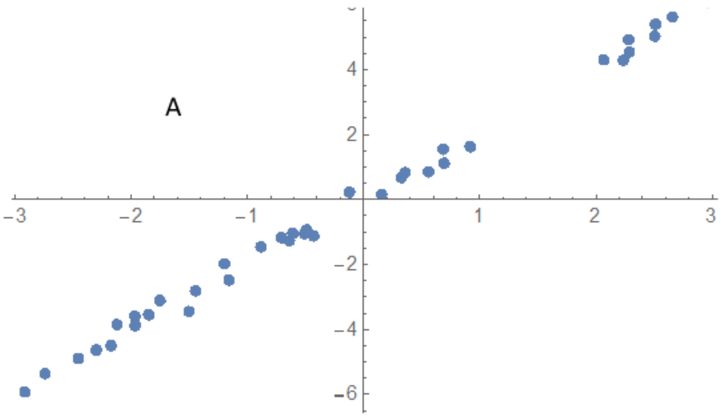
1 Spremenljivki X in Y sta odvisni.

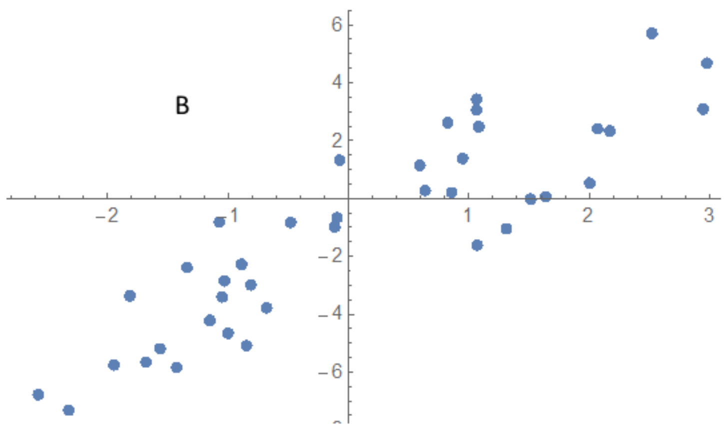
2 Korelacijski koeficient je skoraj nič.

3 Zveza med spremenljivkama je linearna.

rešitev: Pritrdilno lahko odgovorimo na prvo trditev.

8. Na katere trditve o spodnjih diagramih lahko odgovorimo pritrdilno?





1 Za podatke na diagramu A je korelacijski koeficient manjši kot za podatke z diagrama B.

2 Podatki z diagrama B so negativno korelirani, podatki z diagrama A pa pozitivno korelirani.

3 Korelacijski koeficient je v obeh primerih enak 2.

rešitev: Na nobeno ne moremo odgovoriti pritrdilno iz očitnih razlogov.

9. Pri nekih meritvah smo dobili naslednje podatke

Računalništvo	Komunikacije
Ekonomija	Slovenščina
Tuji jeziki	Medicina
Zgodovina	Politologija
Fizika	Statistika
Sociologija	Biologija
Kemija	Matematika
Geografija	Filozofija
Psihologija	

Koliko sta povprečji in korelacijski koeficienti?

rešitev: Povprečje izračunamo po trivialnem postopku in dobimo vrednosti $\bar{x} =$ in $\bar{y} =$
Za izračuna korelacijskega koeficienta uporabimo formulo $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$ in dobimo rezultat 0.9297.

10. Slučajni spremenljivki X in Y sta podani s skupno porazdelitvijo

$X \backslash Y$	-1	1
0	$\frac{1}{8}$	p
1	$\frac{1}{8}$	$\frac{1}{8}$
a	$2p$	$\frac{1}{4}$

Določi p tako, da bo s to tabelo res podana skupna porazdelitev spremenljivk X in Y in določi korelacijski koeficient $\rho(X, Y)$.

rešitev: Vse verjetnosti se morajo sešteti v 1. Zapišemo lahko torej:

$$\frac{5}{8} + 3p = 1$$
$$3p = \frac{3}{8}$$
$$p = \frac{1}{8}$$

Korelacijski koeficient izračunamo po znani formuli. Potrebujemo $E(X)$, $E(Y)$, $E(XY)$, $D(X)$ ter $D(Y)$.

11. Slučajni spremenljivki X in Y sta podani s skupno porazdelitvijo

$X \backslash Y$	-1	1
0	$\frac{1}{2}$	0
1	$\frac{1}{3}$	$\frac{1}{6}$

Določi $K(X, Y)$.

rešitev: Za izračun $K(X, Y)$ potrebujemo vrednosti $E(XY)$, $E(X)$ in $E(Y)$. Ko poračunamo enačbo $K(X, Y) = E(XY) - E(X)E(Y)$ dobimo rezultat $-\frac{1}{12}$.

12. Ali velja naslednja trditev? Če sta slučajni spremenljivki X in Y neodvisni ter imata isto varianco (razpršenost), potem sta slučajni spremenljivki $X + Y$ in $X - Y$ nekorelirani.

rešitev: Intuitivno opazimo, da ta trditev vedno drži.

13. Slučajni spremenljivki X in Y imata isti korelacijski koeficient kot slučajni spremenljivki $2X$ in $3Y$.

rešitev: To vedno drži. Če vsako slučajno spremenljivko pomnožimo s konstantno vrednostjo, se njuna koreliranost ne spremeni.

14. Desetkrat vržemo standardno pošteno kocko. Naj bo X število lihih izidov, Y pa število sodih izidov.

    Katere od naslednjih trditev veljajo?

    - X in Y imata enako porazdelitev.
    - $X \sim B(10, \frac{1}{2})$
    - $E(X) = E(Y)$
    - $X + Y = 10$
    - X in Y sta nekorelirani
    - X in Y imata negativno korelacijo

rešitev: Če poračunamo kovarianco vidimo, da spremenljivki nista nekorelirani. Vse ostale trditve pa držijo in jih je lahko preveriti.

15. Zanima nas odvisnost dosežka na testu s starostjo testiranca. Na vzorcu dobimo naslednje rezultate:

Starost	18	19	20	20	25
Rezultat	99	100	100	101	105

    Kaj lahko povemo o vzorčnem korelacijskem koeficientu?

    Oceni rezultat testiranca, ki je star 23 let.

rešitev: Korelacijski koeficient je med 0 in 1, saj obstaja na prvi pogled opazna pozitivna korelacija. Rezultat testiranca, ki je star 23 let ocenimo kot... TODO (regresijska premica...)

16. Disperzija slučajne spremenljivke

    - Definiraj oziroma pojasni kaj je razpršenost (disperzija) slučajne spremenljivke

      rešitev: Disperzija oz. razpršenost slučajne spremenljivke je mera, ki kvantificira kako je porazdelitev stisnjena ali raztegnjena.

    - Vržemo kovanec za 20 centov. Kar pade je slučajna spremenljivka X (če pokaže 20, je njena vrednost 20, sicer 0). Koliko je disperzija od X?

      rešitev: Uporabimo formulo in izračunamo: $D(X) = E(X^2) - E(X)^2 = 100$.

    - Naj bodo X, Y in Z slučajne spremenljivke. Koliko je disperzija od aX, kjer je a realno število in koliko je disperzija od $Y + X$, če sta slučajni spremenljivki Y in Z neodvisni.

rešitev: Zaradi linearnosti pričakovane vrednosti velja:

$$D(aX) = a^2 D(X)$$

in za neodvisni s.s. Y in Z velja aditivnost:

$$D(X + Z) = D(X) + D(Z).$$

17. Kovarianca

- Definiraj pričakovano vrednost in disperzijo slučajne spremenljivke, nato pa še kovarianco slučajnih spremenljivk.

rešitev: Pričakovana vrednost slučajne spremenljivke je dolgoročna povprečna vrednost vrednosti slučajne spremenljivke. Izračunamo jo tako, da izračunamo skalarni produkt vrednosti v domeni te slučajne spremenljivke z njihovimi verjetnostmi. V primeru zvezne slučajne spremenljivke integriramo konvolucijo vrednosti in verjetnostne funkcije.

- Naj bodo X, Y in Z slučajne spremenljivke. Izračunaj kovarianco $K(aX + bY, Z)$, če poznaš $K(X, Z)$ in $K(Y, Z)$. Koliko je kovarianca $K(Y, Z)$, če poznaš $K(Z, Y)$? Kaj lahko poveš o kovarianci $K(X, Y)$, če veš, da velja $D(X + Y) = D(X) + D(Y)$.

rešitev: Ker za kovarianco velja bilinearnost, lahko zapišemo:

$$K(aX + bY, Z) = aK(X, Z) + bK(Y, Z)$$

Ker je kovarianca simetrična funkcija, velja $K(Y, Z) = K(Z, Y)$.
Če velja, D(X, Y) = D(X) + D(Y), potem lahko iz tega potegnemo implikacijo, da je kovarianca enaka nič. Spremenljivki sta nekorelirani, ampak ne nujno neodvisni.

- X in Y naj bosta slučajni spremenljivki z $D(X) = 2$, $D(Y) = 3$ in $K(X, Y) = -1$. Koliko je a, da velja $K(2X + Y, 3Y - aX) = 0$? Kaj to pomeni za slučajni spremenljivki $2X + Y$ in $3Y - aX$? Kaj pa, če dodatno veš, da sta X in Y porazdeljeni normalno?

rešitev:
Za izračun a upoštevamo lastnost bilinearnosti in zapišemo:

$$\begin{aligned}
K(2X + Y, 3Y - aX) &= 0 \\
2K(X, 3y - aX) + K(Y, 3Y - aX) &= 0 \\
2K(3Y.aX, X) + K(3Y - aX, Y) &= 0 \\
6K(Y, X) - 2aK(X, X) + 3K(Y, Y) - aK(X, Y)
\end{aligned}$$

Upoštevamo, da velja $K(X, X) = D(X)$ ter uporabimo znane vrednosti.
Dobimo, da je $a = 1$.
V splošnem sta $2X + Y$ in $3Y - X$ nekorelirani, če sta X in Y normalni, pa sta tudi neodvisni.
Pomni: Nekorelirani normalno porazdeljeni spremenljivki sta tudi neodvisni.

- Za slučajni spremenljivki X in Y definiraj korelacijski koeficient $r(X, Y)$ ter pojasni kaj veš v primeru, ko je $|r(X, Y)| = 1$.

rešitev: Korelacijski (Pearsonov) koeficient definiramo kot kvocient $r(X,Y) = \frac{K(X,Y)}{\sigma_X \sigma_Y}$
Će je absolutna vrednost korelacijskega koeficienta enaka 1, potem obstaja med s.s. X in Y linearna zveza z verjetnostjo 1.

- Zvezo, ki si jo uporabil v zadnji točki pri (b), utemelji in posploši na slučajne spremenljivke $X_1, ..., X_n$.

rešitev: Po definiciji disperzije velja $D(X,Y) = E(X+Y-E(X,Y))^2 - E(X-E(X)+Y-E(Y))^2$
Oziroma velja:

$$D(X+Y) = D(X) + D(Y) + 2K(X,Y)$$

n
To lahko posplošimo na poljubno število s.s. kot:

$$D(\sum_i X_i) = \sum_i D(X_i) + \sum_{i \neq j} K(X_i, X_j)$$

18. Statistično sklepanje o korelacijski povezanosti

- Kaj je cenilka in kdaj je nepristranska?

rešitev: Vzorčna statistika je poljubna simetrična merljiva funkcija. Cenilka parametra ζ je vzorčna statistika $C = C(X_1, X_2, ..., X_n)$, katere porazdelitveni zakon je odvisen od parametra ζ, njene vrednosti pa ležijo v prostoru parametrov. Cenilka je simetrična funkcija, odvisna je tudi od velikosti vzorca n. Primeri: vzorčna mediana $\tilde{X}$ in vzorčno povprečje $\bar{X}$ sta cenilki za populacijsko povprečje μ; popvravljena vzorčna disperzija S^2 pa je cenilka za populacijsko disperzijo σ^2 pa je cenilka za populacijsko disperzijo σ^2. Cenilka je odvisna od velikosti vzorca.
Cenilka C_n parametra ζ je nepristranska, če je $E(C_n) = \zeta$ (za vsak n); in je asimptotično nepristranska, če je $\lim_{n\to\infty} E(C_n) = \zeta$. Količino $B(C_n) = E(C_n) - \zeta$ imenujemo pristranost (angl. bias) cenilke C_n.

- Definiraj (Pearsonov) koeficient korelacije ρ za dve številski slučajni spremenljivki. Kakšne vrednosti lahko zavzame in kaj se zgodi v primeru, če sta slučajni spremenljivki neodvisni (je to potreben pogoj)?

rešitev:
Pearsonov koeficient korelacije ρ je definiran kot:

$$r_{XY} = \frac{K(X,Y)}{s_X s_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Zavzame lahko vrednosti na intervalu $[-1, 1]$. Če sta slučajni spremenljivki neodvisni, je kovarianca enaka 0 in s tem koeficient korelacije enak 0. Če sta spremenljivki nekorelirani pa še ne pomeni, da sta tudi neodvisni.

- Kaj lahko poveš v primeru, ko ρ doseže največjo oziroma najmanjšo možno vrednost?

  Takrat obstaja med X in Y linearna zveza z verjetnostjo 1.

- Kaj sta prva in druga regresijska funkcija in kje se srečata?

- Opiši preverjanje domneve o linearni povezanosti (će se ne spomniš konkretne testne statistike, napiši vsaj kako se porazdeljuje ter opiši omenjeno porazdelitev).

rešitev: TODO

- Kaj so časovne vrste? Opiši določanje trenda z metodo najmanjših kvadratov.

rešitev:

19. Ali obstajata taki slučajni spremenljivki X in Y, da je $2K(X,Y) - D(Y) - D(X) = 0$

rešitev: Zapišimo:

$$D(X+Y) = D(X) + D(Y) + 2K(X,Y)$$

Opazimo, da je naš izraz enak $-D(X-Y)$, kar je lahko tudi 0, ko je $X - Y$ konstanta.

20. V posodi imamo tri različno težke uteži. Anže in borut vsak naključno izbereta eno utež. Naj bo A teža Anžetove uteži, B pa teža Borutove uteži. Koliko je vrednost $\rho(A,B)$?

rešitev: Korelacija je negativna. Če Anže izbere zelo težko utež, bo Borut izbral lažjo utež in obratno.

21. Naj bosta X in Y neodvisni slučajni spremenljivki. Katere od trditev niso vedno pravilne?

a $E(X+Y) = E(X) + E(Y)$

b $E(XY) = E(X)E(Y)$

c $D(XY) = D(X)D(Y)$

d $D(X+Y) = D(X) + D(Y)$

e $K(X,Y) = 0$

f $\sigma(XY) = 0$

g $\sigma(X+Y) = \sigma(X) + \sigma(Y)$

h $\rho(X,Y) = 0$

rešitev: Ne držijo točke c, f in g.

22. Slučajni spremenljivki sta podani s naslednjo porazdelitvijo:

$X \backslash Y$	-1	1
0	$\frac{1}{2}$	0
1	$\frac{1}{3}$	$\frac{1}{6}$

Kakšna je vrednost $K(X,Y)$?

rešitev:

Kovarianca je definirana kot $K(X,Y) = E(XY) - E(X)E(Y)$.

Izračunajmo robne porazdelitve X in Y, da bomo lahko izračunali pričakovani vrednosti.

$$\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{3}{6} \end{pmatrix}$$

ter

$$\begin{pmatrix} -1 & 1 \\ \frac{5}{6} & \frac{1}{6} \end{pmatrix}$$

Iz tega lahko poračunamo, da velja $E(X) = 0.5$ ter $E(Y) = -\frac{2}{3} = 0.6667$.

Pričakovano vrednost $E(XY)$ izračunamo tako, da množimo vse vrednosti z verjetnostmi. Dobimo rezultat $E(XY) = -\frac{1}{6} = -0.1667$.

Sedaj izračunamo kovarianco kot: $K(X,Y) = -0.1667 - 0.5 \cdot 0.6667 = -0.5$.

23. Multimedijska računalniška igra vsebuje test veščine uporabljanja računalniške miške. Program nariše krog na slučajno izbranem položaju na zaslonu. Igralec se trudi z miško kar najhitreje klikniti kjerkoli v notranjosti kroga. Nov kgor se pojavi takoj, ko je uporabnik kliknil prejšnjega. V spondnji tabeli so podatki o rezultatih, ki jih je dosegel neki igralec, po 20 za vsako roko. Razdalja pomeni razdaljo miškinega kazalca od središča novega kroga v enotah, ki so odvisne od velikosti zaslona. Čas pomeni čas, ki ga je igralec potreboval za naslednji klik (v milisekundah).

čas	Razdalja	Roka
115	190.70	desna
96	138.52	desna
110	165.08	desna
100	126.19	desna
111	163.19	desna
101	305.66	desna
111	176.15	desna
106	162.78	desna
96	147.87	desna
96	271.46	desna
95	40.25	desna
96	24.76	desna
96	104.80	desna
106	136.80	desna
100	308.60	desna
113	279.80	desna
123	125.51	desna
111	329.80	desna
95	51.66	desna
108	201.95	desna
240	19.70	leva
190	138.52	leva
170	165.08	leva
125	126.19	leva
315	163.19	leva
240	305.66	leva
141	176.15	leva
210	162.78	leva
200	147.87	leva
401	271.46	leva
320	40.25	leva
113	24.76	leva
176	104.80	leva
211	136.80	leva
238	308.60	leva
316	279.80	leva
176	125.51	leva
173	329.80	leva
210	51.66	leva
170	201.95	leva

- Sumimo, da je čas odvisen od razdalje. Napravi razsevni diagram časa v odvisnosti od razdalje, pri čemer za vsako roko uporabi drugačne simbole.

  rešitev: MATLAB/R...

- Ali se pozna, da je igralec desničar?

  rešitev: Da je igralec desničar vemo, ker so njegovi reakcijski časi z desno roko bistveno krajši od levih (modre točke imajo manjše x koordinate od rdečih).

- Primerjaj korelaciji pri obeh rokah. Zakaj sta podobni, čeprav je eden od vzorcev precej bolj oster kot drugi?

  rešitev: Korelacija pri desni roki je rD = 0.32. pri levi pa rL = 0.27. Podobno majhni sta, ker so v obeh primerih podatki zelo razpršeni in zelo šibko linearno povezani.

1.10 09 - Centralni limitni izrek

1. V tovarni pakirajo sladkor v 5 kg vreče. Količina sladkorja varira po normalni porazdelitvi in ima pričakovano vrednost 5 kg ter standardni odklon 0.05 kg

   Ali drži naslednja trditev? Za računanje verjetnosti dogodkov o težah posameznih vreč bomo uporabili spremenljivko $Z = (X - 5.0)/0.05$, medtem ko bomo za računanje verjetnosti dogodkov o težah vzorčnih povprečij za velikosti $n = 25$ uporabili spremenljivko $\bar{Z} = (\bar{X} - 5.0)/0.01$.

rešitev: To drži. Prvo spremenljivko uporabimo za računanje verjetnosti iz porazdelitve populacije, ki jo aproksimiramo z normalno porazdelitvijo, drugo spremenljivko pa uporabimo za računanje vrednosti povezanih z vzorčno statistiko vzorčno povprečje. Standardni odklon populacije v tem primeru delimo z kvadratnim korenom velikosti populacije.

2. Ana se igra z bratcem Binetom. Najprej Ana brcne žogo proti Binetu, potem pa jo Bine brcne nazaj. Razdalja, ki jo prepotuje žoga, ko jo brcne Ana, je porazdeljena z $N(10, 1)$, ko jo brcne Bine pa $N(5, 1)$.

   Kako je porazdeljena razdalja med začetno in končno lego žoge?

   rešitev: Porazdeljena je kot razlika danih slučajnih spremenljivk. Torej se povprečji odštejeta, varianci pa seštejeta ($\sqrt{2} \approx 1.414$).

3. Če sta pričakovana vrednost in standardni odklon za neko populacijo zaporedoma enaka 25 in 5, in izbiramo vzorce velikosti $n = 100$, izračunaj standardni odklon za vzorčno povprečje.

   rešitev: Standardni odklon vzorčnega povprečja ocenimo iz standardnega odklona populacijskega povprečja kot:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

   Dobimo rezultat $\frac{5}{10} = 0.5$.

4. Ali velja naslednja trditev? Če seštejemo več zveznih slučajnih spremenljivk z enako porazdelitvijo, potem se po centralnem limitnem izreku njihova vsota porazdeljuje približno normalno.

   rešitev: To ne drži. Centralni limini izrek pravi, da se vzorčne statistike porazdeljujejo približno normalno, ko je velikost vzorca dovolj velika.

5. Ali velja naslednja trditev? Porazdelitev vzorčnih povprečij je normalna za vzorce vseh velikosti, pod pogojem, da je začetna porazdelitev normalna.

   rešitev: To drži. Za vzorce velikosti večje ali enake 30, pa je porazdelitev vzorčnih povprečij normalna tudi če začetna porazdelitev ni normalna.

6. Ali velja naslednja trditev? Porazdelitev vzorčnih povprečij slučajne spremenljivke bo približno normalno porazdeljena za velike vzorce tudi, če naključna slučajna spremenljivka ni normalno porazdeljena.

   rešitev: Zadostiti moramo še pogoju neodvisnosti. Vzorci morajo biti enoastavni naključni vzorci. Če vzorčimo brez vračanja, mora biti velikost vzorca največ 10% velikosti populacije.

7. Ali velja naslednja trditev? Ce ntralni limitni izrek pravi, da je standardni odklon porazdelitve vzorčnih povprečij enak standardnemu odklonu na populaciji, ki ga delimo s korenom velikosti vzorca.

   rešitev: To drži. Če vzorčenje oz. originalna porazdelitev zadosti pogojem za aplikacijo CLI, lahko standardni odklon porazdelitve vzorčnih povprečij ocenimo tako, da standardni odklon populacije delimo s kvadratnim korenom velikosti vzorca.

8. Ali velja naslednja trditev? Po centralnem limitnem izreku je vsota dovolj velikega števila slučajnih spremenljivk približno normalna.

   rešitev: Zadostiti moramo tudi pogoju, da so te slučajne spremenljivke neodvisne.

9. Kaj mora veljati, da je porazdelitev vzorčnih povprečij približno normalna?

rešitev: (multiple choice)

Velikost vzorca mora biti dovolj velika ter pričakovana vrednost in varianca populacije morata biti končni (omejeni).

10. Denimo, da ima slučajna spremenljivka X pričakovano vrednost 0 in standardni odklon 2. Poišči zgornjo mejo $P(|X| \geq 8)$, ki jo dobiš z neenakostjo Čebiševa.

rešitev: TODO

11. Naj bodo $X_1, ..., X_n$ enako porazdeljene slučajne in neodvisne slučajne spremenljivke s končnim upanjem in varianco. V katerem od spodnjih primerov ne moremo uporabiti centralnega limitnega izreka za oceno porazdelitve $S = X_1 + ... + X_n$?

- Spremenljivke X_i so normalno porazdeljene in n je majhen (< 30).
- Spremenljivke X_i so normalno porazdeljene in n je velik (> 30).
- Spremenljivke X_i niso normalno porazdeljene in n je majhen (< 30).
- Spremenljivke X_i niso normalno porazdeljene in n je velik (> 30).

rešitev: Ne moremo uporabiti centralnega limitnega izreka, če spremenljivke X_i niso normalno porazdeljene in je n majhen (< 30).

12. Šestnajstsedežno letalo je preobremenjeno, če nosi težo večjo od 1440 kg. Vemo, da je teža enega potnika skupaj z njegovo prtljago porazdeljena normalno z upanjem 85 kg in standardnim odklonom 15 kg.

Kolikšen delež letal bo preobremenjenih, če potnike izberemo naključno (in so vsi sedeži zasedeni)?

rešitev: Vsota normalnih spremenljivk, ki predstavljajo teže potnikov se porazdeljuje kot

$$S \sim N(1360, 60)$$

Zanima nas verjetnost, da bo slučajna spremenljivka S zavzela vrednost večjo od 1440.
Poiščimo kritično razdaljo od povprečja:

$$z = \frac{1440 - 1360}{60} = 1.3333$$

Preobremenjenost dosežemo torej, ko smo več kot 1.3333 standardni deviacij desno od povprečja. V tabeli pogledamo verjetnost, da slučajna spremenljivka zavzame vrednost bližje oziroma enako oddaljeno povprečju, to vrednost odštejemo od 1 in dobimo rezultat 0.0918.

13. ??

14. Če preučujemo podatke, ki so opisani z normalnim modelom in ne poznamo σ^2 (variance modela), potem lahko uporabimo vzorčno varianco s^2 kot približek za σ^2.

Ali drži naslednje? V tem primeru moramo nadomestiti normalno porazdelitev s Studentovo (t-porazdelitev), če vzorec ni dovolj velik.

rešitev: To drži. t-porazdelitev je sicer na prvi pogled podobna normalni, ampak ima debelejše repe in nam za majhne vzorce, kjer ocenjujemo varianco populacije z varianco vzorca da boljši model porazdelitve kot normalna porazdelitev.

15. Ko preverjamo domnevo in pri tem uporabljamo vrednost π_0 iz ničelne domneve in p kot vzorčni delež, je za oceno stadardne napake porazdelitev vzorčnega deleža najbolje vzeti vrednost $\sqrt{n\pi_0(1-\pi_0)}$.

    rešitev: Za oceno standardne napake porazdelitve vzorčnega deleža je najbolje vzeti vrednost $\frac{\sqrt{p(1-p)}}{n}$.

16. Kdaj je po centralnem limitnem izreku porazdelitev vzorčnih povprečij približno normalna?

    rešitev: (multiple choice)
    To velja, kadar je velikost vzorca dovolj velika in varianca omejena.

17. Momenti, centralni limitni izrek in neenakost Čebiševa

    - Vpelji moment slučajne spremenljivke X reda k glede na točko a (Kjer je k naravno število, a pa realno), pojasni tudi kdaj obstaja.

      rešitev: TODO

    - Definiraj še začetni centralni moment in nato z njimi še pričakovano vrednost in razpršenost ali asimetrijo spremenljivke X.

      rešitev: TODO

    - Vpelji standardizacijo Z slučajne spremenljivke X in izračunaj pričakovano vrednost in razpršenost standardizirane spremenljivke Z.

      rešitev: TODO

    - Predstavi centralni limitni izrek.

      rešitev: TODO

    - Zapiši neenakost Čebiševa, nato pa še z besedami pojasni njen pomen (ali pa podaj kakšno njeno posledico). Dokaži Čebiševo neenakost!

      rešitev: TODO

18. Binomska in normalna porazdelitev ter CLI in vzorčenje

    - Definiraj pričakovano vrednost in standardni odklon slučajne spremenljivke.

      rešitev: TODO

    - Opiši postopek za standardizacijo slučajne spremenljvike in zapiši njeno pričakovano vrednost ter njen odklon.

      rešitev: TODO

    - Opiši binomsko in normalno porazdelitev ter Laplaceov obrazec.

      rešitev: TODO

    - Predstavi centralni limitni izrek (CLI)

      rešitev: TODO

- Opiši večrazsežno normalno porazdelitev.

rešitev: TODO

- Naj bo X slučajna spremenljivka z $E(X) = \mu$ in $D(X) = \sigma^2$. Za njen slučajen vzorec $X_{i}{}_{i=1}^{n}$ definiraj vzorčno povprečje $\bar{X}$ in napiši, kaj se dogaja z vzorčnim povprečjem $E(\bar{X})$ in s standardno napako $D(\bar{X})$ z naraščanjem velikosti vzorca.

rešitev: TODO

19. Šestnajstsedežno letalo je preobremenjeno, če nosi težo večjo od 1440 kg. Vemo, da je teža enega potnika skupaj z njegovo prtljago porazdeljena normalno z upanjem 85 kg in standardnim odklonom 15 kg. Ali velja naslednja trditev? Če potnike izberemo naključno in so vsi sedeži zasedeni, je delež letal, ki so preobremenjena, enak 0.9082.

rešitev: Podobno nalogo smo že rešili - glej nazaj.

20. Naj bodo $X_1, ..., X_n$ enako porazdeljene slučajne in neodvisne slučajne spremenljivke s končnim upanjem in varianco. Ali velja naslednja trditev? Za oceno porazdelitve $S = X_1 + ... + X_n$ ne moremo uporabiti centralnega limtnega izreka (CLI) v primeru, da spremenljivke X_i niso normalno porazdeljene in je n majhen (< 30).

rešitev: To drži. Če je vzorec majhen ne moremo uporabiti CLI, če hkrati tudi vemo, da s.s. niso normalno porazdeljene.

21. Rezultati posameznih kolokvijev so bili porazdeljeno približno normalno z $\mu_1 = 32, \sigma_1 = 15, \mu_2 = 35, \sigma_2 = 10, \mu_3 = 40, \sigma_3 = 20, \mu_4 = 34, \sigma_4 = 10$ in $\mu_5 = 24, \sigma_5 = 5$. Kakšni sta vrednosti μ in σ, ki opredeljujeta povprečno oceno vseh petih kolokvijev?

rešitev: Seštejemo pričakovane vrednosti in seštejemo ter korenimo variance, da dobimo standardno deviacijo.

Za pričakovano vrednost povprečja uporabimo kar povprečje pričakovanih vrednosti s.s., ki so podane.

Naj je $devs = [15, 10, 20, 10, 5]$ vektor standardnih deviacij. Standardno deviacijo porazdelitve povprečja izračunamo kot (MATLAB izraz) *sqrt(sum(devs.$\hat{2}$))/sqrt(length(devs))*

22. Ali velja naslednja trditev? Porazdelitev vzorčnih povprečji je normalna tudi za vzorce majhnih velikosti, pod pogojem, da je začetna porazdelitev tudi normalna.

rešitev: To drži. Vsota normalnih porazdelitev je vedno normalna.

23. Ali velja naslednja trditev? Ko preverjamo domnevo in pri tem uporabljamo vrednosti π_0 iz ničelne domneve in p kot vzorčni delež, je za oceno standardne napake porazdelitve vzorčnega povprečja najbolje vzeti vrednost $\sqrt{\pi_0(\pi_0 - 1)/n}$.

rešitev: Na to vprašanje smo že odgovorili - glej nazaj.

24. Ameriška ruleta ima 38 žepkov, označenih z 0, 00 in s števili od 1 do 36. Ko zavrtimo kolo, kroglica z enako verjetnostjo pristane v kateremkoli od teh žepkov. Oznake žepkov so razporejene tudi po mizi, na kateri igralci postavljajo svoje stave. Eden od stolpcev vsebuje vse večkratnike števila 3, torej števila 3,6,. . . ,36. Igralec položi 1 EUR na stolpec in prejme 3 EUR, če se kroglica ustavi na kateremkoli od teh števil. Povprečni dobitek je seveda negativen - na dolgi rok igralci izgubljajo in igralnica dobiva.

Poišči standardni odklon za ta primer. Koliko igralec izgublja na dolgi rok?

rešitev: Verjetnost, da bo igralec dobil 3 EUR je $\frac{12}{38}$.
Pričakovana vrednost je torej $E(X) = -1 + 3 \cdot \frac{12}{38} = -0.0526$EUR.
Standardni odklon izračunamo kot:

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

Rezultat je 1.39.

25. V večini velikih ameriških mest je organizirana nekakšna ilegalna oblika lota, ki deluje na naslednji način: igralec izbere eno od 1000 trimestnih števil med 000 in 999 in plača posredniku en dolar, da lahko stavi na to število. Vsak dan slučajno izberejo eno trimestno število in če je igralec število uganil, dobi 600$, sicer pa izgubi vplačani dolar.

    Poišči standardni odklon za dobičke posamezne igre. Uporabi centralni limitni izrek, da podaš razpon (srednja vrednost ± 3 standardni odkloni) za dobiček, ki ga ima igralec v enem letu (po 365 igrah) in v 10 letih (po 3650 igrah).

    rešitev: TODO

26. Enostavni slučajni vzorec 400 prebivalcev nekega mesta povprašamo za mnenje in 144 jih je za nov gasilski dom. Kolikšen je približno standardni delež vzorca p za ta primer?

    rešitev: TODO

27. Tovarna izdeluje vreče sladkorja s povprečno težo 5.05 funta in standardnim odklonom 0.05 funta. Slučajno izberemo štiri vreče in izračunamo njihovo povprečno težo. Kolikšen je standardni odklon vzorčne porazdelitve povprečne teže.

28. Vreče sladkorja imajo povprečno težo 5.05 funta in standardni odklon 0.05 funta. Tovarna uporablja 95 kontrolni limiti v kontrolnem diagramu proizvodnega procesa. Izberemo slučajni vzorec 9 vreč in jih stehtamo. Povprečna teža je 5.09 funta.

    Katera od naslednjih trditev je pravilna?

    - Povprečna teža vzorca je ušla nadzoru.

      rešitev: TODO
    - Povprečna teža vzorca je še pod nadzorom.

      rešitev: TODO
    - Nimamo dovolj informacij.

      rešitev: TODO

29. PTC je spojina, ki ima za nekatere močan grenak okus, drugim pa se zdi brez okusa. Sposobnost zaznavanja PTC je dedna. Približno 75% Italijanov lahko zazna PTC. Oceniti želiš delež Američanov, pri katerih je vsaj eden od staršev iz Italije in ki lahko okusijo PTC. Predpostavi, da ocena o 75% deležu Italijanov drži in za to populacijo testiraj 500 ljudi.

    Skiciraj normalno krivuljo, ki prikazuje, kako se bo delež p vseh tistih, ki okusijo PTC, spreminjal pri večkratnem vzorčenju.

rešitev: TODO

30. V eni državi na ameriškem srednjem zahodu ima 84 gospodinjstev za praznike ob koncu leta božično drevesce. V raziskavi slučajni vzorec 400 gospodinjstev vprašamo "Ste letos postavili božično drevesce?"

    Kakšna je vzorčna porazdelitev odgovorov?

    rešitev: TODO

31. Standardni odklon s_p vzorčnega deleža p se spreminja skupaj z dejansko vrednostjo deleća populacije π. Na srečo se ne spreminja veliko, če p ni blizu 0 ali 100. Recimo, da je velikost vzorca n = 500.

    Izračunaj s_p za $\pi = 30, 40, 50, 60$ in 70. Nato izračunaj s_p za $\pi = 0, 10$ in 20. Na katerem delu se s_p spreminja najhitreje, ko se spreminja π? Nariši graf s_p v odvisnosti od π.

    rešitev: TODO

2 Statistika

2.1 10 - Opisna statistika

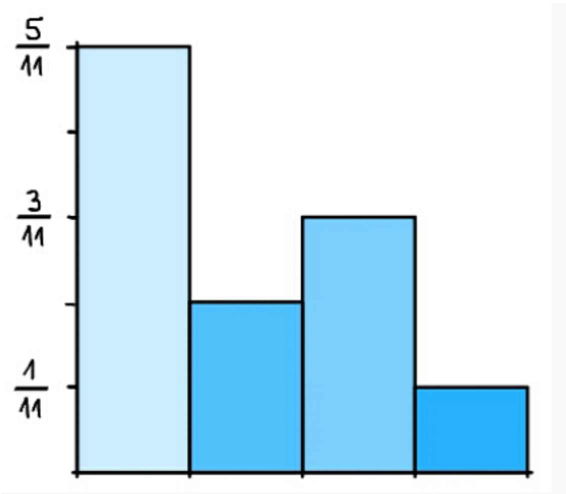
1. Med kakšne spremenljivke uvrščamo slučajne spremenljivko, ki lahko zavzame vrednosti moški in ženska?

    rešitev: Uvrščamo jo med opisne-imenske.

2. Ali je slučajna spremenljivka, ki lahko zavzame vrednosti odličen, pravdober, dober, zadosten in nezadosten je opisno-urejenostna?

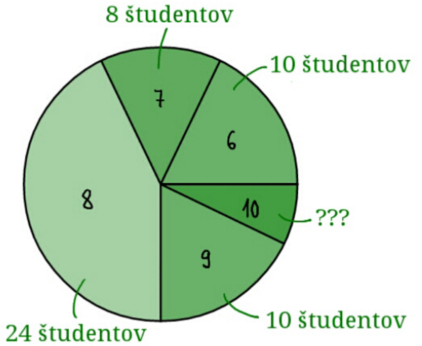
    rešitev: Drži. Imamo 5 opisnih vrednosti, ki jih lahko uredimo po velikosti.

3. Histogram prikazuje, koliko odstotkov vseh točk je v povprečju študent nabral pri vsaki od štirih nalog na 1. kolokviju ViS. Če veš, da so študentje pri zadnji nalogi zbrali 5 točk, koliko točk so v povprečju dosegli na 1. kolokviju na vseh nalogah skupaj?



rešitev: Vidimo, da so v povprečju študentje pri zadnji nalogi dosegli $\frac{1}{11}$ točk, kar znese 5 točk. Iz tega lahko poračunamo, da je povprečno število točk, ki so jih študentje dosegli 55 ($\frac{1}{11} \cdot x = 5$, reši za x).

4. Kolač na sliki prikazuje porazdelitev pozitivnih ocen pri nekem predmetu. Koliko študentov je dobilo oceno 10?



rešitev: Iz grafa vidimo, da 28 študentov predstavlja $\frac{1}{2}$ diagrama. Torej je bilo vseh študentov 56. Vidimo, da delež študentov z oceno 10 predstavlja $\frac{1}{12}$ ploščine tortnega diagrama. Torej je bilo takih študentov $56 \cdot \frac{1}{12} = 4$.

5. Slučajna spremenljivka X je porazdeljena z $N(75, 15.0)$. Za to porazdelitev je 23. centil enak koliko?

rešitev: Za standardizirano normalno porazdeljeno s.s. v tabeli pogledamo, pri kateri vrednosti je meja, kjer je verjetnsot, da se nahaja vrednost s.s. nižje že 23%. To se zgodi pri vrednosti -0.74. Porazdelitev "denormaliziramo" kot $x = \mu + \sigma \cdot z$ in računamo $x = 75 + (-0.74 \cdot 15) = 63.9$.

6. Kako imenujemo značilnosti populacije kot celote?

rešitev: Imenujemo jih statistični parametri. Oceniti jih skušamo iz značilnosti vzorca, ki jih imenujemo statistike.

7. Kaj nam pove število frekvenca?

rešitev: Frekvenca je število pojavitev vrednosti slučajne spremenljivke. Na primer, če deset študentov dobi pri statistiki 80 točk, potem ima vrednost slučajne spremenljivke 80 frekvenco 10.

8. Kakšen prikaz je frekvenčna distribucija?

rešitev: Frekvenčna distribuija je prikaz vrednosti v ustrezni tabeli.

9. Frekvenca posameznega razreda je enaka:

rešitev: Številu vrednost številske spremenljivke, ki spada v tisti razred. Torej moč tega razreda.

10. Širina razreda je enaka:

rešitev: Razliki med zgornjo in spodnjo mejo razreda. Torej kolikšen razmah je med spodnjo in zgornjo mejo. Ta vrednost ni vezana na dejanske vrednosti, ki jih razred zavzema.

11. Frekvenčni kolač (strukturni krog) sestavljajo krožni izseki. Kaj kažejo?

rešitev: Kažejo deleže enot, ki sodijo v posamezne razrede. Deleži se seštejejo v 1.

12. Kdaj koristno uporabljamo uteženo aritmetično sredino?

rešitev: Če statistična spremenljivka zavzame isto vrednost na več enotah populacije. Takrat lahko različnim vrednostim damo različno težo.

13. Varianca, mera za razpršenost posameznih enot, je:

rešitev: Število, ki je enako povprečju kvadratov odklonov posameznih vrednosti statistične spremenljivke od aritmetične sredine. $$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)}{n}$$

14. V razredu s 25 učenci je 8% odličnih, 28% pravdobrih, trije so nezadostni, sedem je zadostnih. Ostali so dobri.

    Izračunajte povprečno oceno razreda in narišite histogram frekvenc ocen.

rešitev: Izračunaj deleže, seštej in deli s številom učencov.

15. V razredu je 25 učencev. Učenka Andreja je računala povprečno število točk pri šolski nalogi. Pri prvem računanju se je zmotila: ni upoštevala svojega dosežka in dobila povprečje 74.5 točk. Ko je napako popravila, je dobila povprečje 75 točk.

    Koliko točk je dosegla Andreja pri šolski nalogi?

rešitev: Rešujemo enačbo: $$\frac{x}{24} = 74.5$$ $$x = 1788\frac{1788 + y}{25} = 75$$ $$y = 25 \cdot 75 - 1788 = 87$$

16. Za izhodišče vzemimo stavek: je zvito kakor kozji rok.

    - V zgornjem stavku opazujte števila črk v posameznih besedah. Izračunajte povprečno število črk v besedah tega stavka in standardno deviacijo števila črk v besedah.

rešitev: Zapišimo število črk v posameznih besedah v obliki vektorja: $$[25553]$$ Povprečno število črk izračunamo po trivialnem postopku in dobimo povprečje 4. Standardno deviacijo izračunamo po formuli $\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$. Dobimo rezultat 1.2649.

    - Opazujte besede v zgornjem stavku še enkrat: tokrat štejte, koliko samoglasnikov vsebuje ta in ona beseda. Ali je razpršenost teh podatkov večja ali manjša od razpršenosti podatkov o dolžini besed iz prejšnega vprašanja?

rešitev: Zapišimo podatke zopet v obliki vektorja: $$[12221]$$ Zopet izračunamo standardo deviacijo in dobimo rezultat 0.4899.

17. V nekem razredu je 30 otrok. Njihovi domovi so od šole oddaljeni največ 25 km. Oddaljenosti so prikazane s frekvenčnim histogramom, v katerem manjka podatek za oddaljenost od 15 km do 20 km. Mejne oddaljenosti 5 km, 10 km,... štejemo v histogramu k manjšim vrednostim, na primer 15 km $\in 10km, 15km$.

- Koliko domov otrok je od šole oddaljenih od 5 km do 10 km?

rešitev: TODO - ni podatkov

- Koliko domov otrok je od šole oddaljenih od 15 km do 20 km?

rešitev: TODO - ni podatkov

- Koliko so domovi otrok povprečno oddaljeni od šole? Izračunajte povprečno vrednost, ki jo lahko dobite iz podatkov v histogramu (v km).

rešitev: TODO - ni podatkov

18. Dva strelca sta vsak dan dvajsetkrat ustrelila v tarčo. Pri vsakem strelu sta lahko dosegla 0 točk, če sta tarčo zgrešila, 1 točko, 2 točki,..., 5 točk, če sta zadela v sredino tarče. Točke prvega so prikazane s tabelo, točke drugega pa s frekvenčnim poligonom.

Kateri strelec ima večje povprečje točk?

rešitev: TODO - ni podatkov

Kateri strelec ima manjši standardni odklon?

rešitev: TODO - ni podatkov

19. V steblo-list diagramu spodaj $1|3$ predstavlja število 13.

Kaj je mediana za zgornje podatke?

rešitev: Vidimo, da so podatki po vrsticah že urejeni naraščujoče. Podatkov je 17, torej vzamemo srednjega (9.). Mediana je torej 29.

20. Koraki statistične analize ter urejanje in prikazovanje podatkov

- Definiraj kvantil, centil, mediano, modus in kumulativo

rešitev: Kvartil je vrednost, od katere četrtina vrednosti zavzame manjšo vrednost. Poznamo tri kvartile, ki razdelijo porazdelitev na 4 dele, ki imajo enako ploščino. Centil je podoben, le da je je meja, kjer je 1% vrednosti manjših. Mediana je vrednost, od katere je polovica podatkov manjša, polovica pa večja. Modus je najpogostejša vrednost, ki se pojavlja med podatki. Komulativa je funkcija, ki sešteva relativne verjetnosti. Npr. pri zvezni porazdelitvi je to funkcija, ki vrne verjetnost, da je s.s. X zavzela vrednost manjšo ali enako x.

- Vpelji frekvenčno porazdelitev in histogram ter opiši postopek za risanje histograma z enako širokimi razredi.

rešitev:
Frekvenčna porazdelitev je seznam, tabela ali graf, ki prikazuje frekvenco različnih rezultatov v vzorcu. Podatke vmestimo v smiselne razrede in jih narišemo.

- Definiraj srednjo vrednost in odklon

rešitev: Srednja vrednost je povprečna vrednost slučajne spremenljivke v neki porazdelitvi. Odklon kvantificira variacijo oziroma disperzijo neke porazdelitve - kako je porazdelitev raztegnjena ali stisnjena.

- Vpelji standardizacijo Z slučajne spremenljivke X in izračunaj srednjo vrednost in odklon od Z (Posamezne korake utemelji).

rešitev:

Slučajno spremenljivko standardiziramo tako, da od nje odštejemo povprečje in delimo z njenim standardnim odklonom.

$$z = \frac{x - \mu}{\sigma}$$

Srednja vrednost je enaka 0, standardni odklon pa je enak 1. (Postopek TODO)

- Ali obstaja kakšna zveza med ogivo in porazdelitveno funkcijo?

  rešitev: Ogiva je na roko narisan graf, ki prikazuje krivuljo porazdelitvene funkcije.

21. Med kakšne spremenljivke uvrščamo slučajno spremenljivko, ki lahko zavzame vrednosti moški in ženska?

    rešitev: Uvrščamo jo med opisne/imenske

22. ??

23. Študije lahko dajo napačne rezultate zaradi mešanja zunanjih vplivov z obravnavanimi terapijami.

    Razloži, kako bi mešanje lahko vplivalo na rezultate v naslednjem primeru: Študentka meni, da bo pitje zeliščnega čaja izboljšalo zdravje pacientov v domu za ostarele. Z nekaj prijatelji redno obiskuje večji dom in streže zeliščni čaj delu varovancev. Drugega dela varovancev ne obiskujejo. Po šestih mesecih ugotovijo, da je bilo skupno število bolniških dni v prvi skupini manjše kot v drugi.

    rešitev: Lahko da je k boljšim rezultatom pripomoglo druženje z varovanci in ne pije čaja. Študija je bila slabo zastavljena, ker pitje čaja ni bila edina razlika med aktivno in kontrolno skupino.

24. Študije lahko dajo napačne rezultate zaradi mešanja zunanjih vplivov z obravnavanimi terapijami.

    Razloži, kako bi mešanje lahko vplivalo na rezultate v naslednjem primeru: Članek v neki reviji je poročal, da so ženske, ki dojijo svoje otroke, bolj povezane z njimi, kot tiste, ki jih hranijo po stekleničkі. Avtorica je od tod sklepala, da ima dojenje pozitiven učinek na materin odnos do otroka.

    rešitev: Ne moremo vedeti kaj je vzork in kaj je posledica.

25. V ZDA so ugotavljali, da ženske volijo demokrate raje kot moški. Politolog je vprašal skupino moških in skupino žensk, za koga so glasovali na zadnjih volitvah.

    Natančno razloži, zakaj to ni eksperimet.

    rešitev: Nobene od skupin nismo podvrgli nobeni terapiji, zgolj opazovali smo rezultate, torej gre za opazovalno študijo, ne pa za eksperiment.

26. Nekateri menijo, da telesna vadba dvige bazični metabolizem za 12 do 24 ur in tako omogoča topljenje maščob tudi po končani vadbi. Raziskovalec prosi udeležence, da več ur hitro hodijo po tekalni stezi. Pri tem meri njihov bazični metabolizem pred, takoj in 12 ur po vadbi.

    Ali je to eksperiment? Zakaj ali zakaj ne?

rešitev: Gre za eksperiment, daj smo skupino posameznikov podvrgli določeni terapiji (vadbi) in potem spremljali njihov odziv (bazični metabolizem). Eksperiment bi lahko izboljšali, če bi uvedli še kontrolno skupino oziroma tem posameznikom ob istih časih izmerili bazični metabolizem tudi ob dneh, ko niso bili podvrženi vadbi.

27. Študija povezave med telesno pripravljenostjo in vodstvenimi sposobnostmi uporabi za vzorec vodstvene delavce srednjih let, ki so se prostovoljno prijavili za program vadbe. Razdelijo jih v dve skupini glede na njihovo telesno pripravljenost, ki jo ocenijo z zdravniškim pregledom. Vsi dobijo psihološki test, ki izmeri njihove vodstvene sposobnosti, nato pa primerjajo rezultate.

    rešitev: Gre za opazovalno študijo, saj posameznikom ne podvržemo nobeni terapiji, ampak obe skupini samo opazujemo in merimo odzive. Pri eksperimentu bi v eno od dveh sicer enakih skupin uvedli neko terapijo (v tem primeru telesno vadbo), potem pa merili njen učinek za vodstvene sposobnosti.

28. Anemija srpastih eritrocitov je dedna bolezen rdečih krvnih celic, ki v ZDA prizadene predvsem Afroameričane. Povzroči lahko močne bolečine in številne zaplete. Leta 1992 je nacionalni zdravstveni inštitut začel raziskavo s hidroksikarbamidom. V raziskavo je bilo vključenih 300 odraslih oseb, ki so imele v preteklem letu vsaj tri napade bolečine, ki jih je povzročila anemija srpastih eritrocitov.

    Zakaj ne bi dobili dobre informacije o učinkovitosti zdravila, če bi dali hidroksikarbamid vsem 300 osebkom?

    Leta 1995 so eksperiment predčasno končali, ker je bilo v skupini, ki je prejemala hidroksikarbamid, pol manj bolečinskih napadov kot v kontrolni skupini. Razloži, zakaj je dal ta eksperiment močan dokaz, da je hidroksikarbamid uspešno zdravilo za anemijo srpastih celic.

    rešitev: Ne bi dobili informacije o tem, ali je morebitne spremembe povzročilo zdravilo ali pa bi nastopile tudi sicer.
    Če zdravilo ne bi pomagalo, bi bil delež napadov v kontrolni skupini blizu deleža napadov v zdravljeni skupini. Ker pa je delež v zdravljeni skupini bistveno manjši, lahko sklepamo, da je to posledica zdravila. Z drugimi besedami, verjetnost (p-vrednost), da bi po naključju bilo pol manj napadov je tako majhna, da so zavrnili ničelno hipotezo, da zdravilo nima učinka.

29. Nekateri finančni svetovalci verjamejo, da grafi preteklih trendov cen napovedujejo, kakšne bodo cene v prihodnosti. Večina ekonomistov se ne strinja. V eksperimentu, ki razišče, kako uporaba grafov vpliva na rezultate, študentje trgujejo z virtualnim denarjem v računalniški simulaciji. Študentov je dvajset, poimenujemo jih $A, B ..., T$ (s črkami angleške abecede). Njihov cilj je zaslužiti čimveč denarja in najboljše nagradimo z manjšimi nagradami. Študentje imajo na razpolago zgodovino predhodnega trgovanja s to valuto, nekateri imajo tudi programsko opremo, ki prikaže trende.

    Opiši načrt tega eksperimenta in uporabi tabelo naključnih števil, kjer to eksperiment zahteva.

rešitev: Študente razdelimo na tiste, ki bodo pri trgovanju uporabljali grafe preteklih trendov, in druge, ki bodo trgovali brez njih. Potem obe skupini pustimo, da trgujeta na istem trgu in na koncu primerjamo npt. povprečne zaslužke v obeh skupinah ali kako drugo smiselno vrednost. Poleg tega lahko npr. eksperiment ponovimo na različnih trgih, da vidimo, če je morda trgovanje s pomočjo trendov bolj učinkovito na trgu s specifičnimi lastnostmi, ali pa najprej razdelimo študente v skupine glede na njihove dosedanje izkušnje in potem znotraj enako uspešnih enim dovolimo uporabo trendov, drugim pa ne. V tem primeru služi druga skupina za kontrolo, študente pa razdelimo v obe skupini s pomočjo tabele naključnih števil.

30. Študentje imajo pri predavanju iz matematike izbiro med običajnim predavanjem in učenjem po lastni presoji. Fakulteta želi primerjati rezultate pri obeh možnostih. Nekdo predlaga, da bi obema skupinama dali enaka končna izpita in primerjali povprečna rezultata v obeh skupinah.

    Razloži, zakaj so mešanja z zunanjimi vplivi rezultati takšne raziskave brez pomena.

    rešitev: Za samostojen študij so se najbrž odločili drugačni študenti. Torej so lahko razlike pri rezultatih posledica drugačnih lastnosti študentov, ki so si izbrali posamezen način učenja.

31. Članek v New England Journal of Medicine, ki predstavlja končne rezultate študije na zdravnikih, se začne z naslednjimi besedami: Študija na zdravnikih je slučajeni, dvojno slepi eksperiment s placebom, s katerim smo želeli ugotoviti, če majhne količine aspirina (325mg vsak drugi dan) zmanjšajo možnost srčnih infarktov in če beta karoten zmanjšuje možnost nastanka rakavih obolenj. Od zdravnikov se pričakuje, da bodo to razumeli.

    Razloži zdravniku, ki ne zna statistike, kaj pomenijo slučajni, dvojno slepi in placebo.

    rešitev: Slučajni pomeni, da smo bolnike, ki jemljejo prvo zdravilo, izbrali naključno, dvojno slepi pomeni, da niti zdravnik, niti bolniki ne vedo, kdo prejema pravo zdravilo in kdo prejema zdravilo brez učinka. Slednje imenujemo placebo.

32. V časopisu prebereš, da so se pri kontroliranih znanstvenih raziskavah nezdravstvene terapije kot sta meditacija in molitev izkazale za učinkovite pri zdravljenju visokega krvnega pritiska, nespečnosti, čirov in astme.

    Razloži, daj je mišljeno s kontroliranimi znanstvenimi raziskavami in zakaj lahko take raziskave pokažejo pozitivne učinke meditacije in molitev.

    rešitev: Pri raziskavi so verjetno opazovali skupino bolnikov, ki je meditirala ali molila, poleg tega pa še kontrolno skupino bolnikov, ki so sicer izhajali iz iste populacije, vendar niso uporabljali nobene nezdravstvene terapije. Raziskava je pokazala, da tovrstne terapije pozitivno učinkujejo na bolezni, ker so le-te povezane z visok

33. Spodaj je seznam 20 pacientov, ki so privolili k sodelovanju v preizkusu kirurškega zdravljenja angine.

    Načrtuj eksperiment, ki bo kirurško zdravljenje primerjal s placebom (lažno operacijo) in uporabi spodno tabelo naključnih števil, da izvedeš ustrezno naključno delitev.

rešitev: Pacientom priredimo števila od 1 do 20. Nato s pomočjo tabele naključnih števil izberemo 10 pacientov, ki bodo podvrženi pravi operaciji, ostali pa dobijo placebo terapijo. Najprej grupiramo številke iz tabele v skupino po dve, nato pa izmed nastalih števil izberemo prvih 10, ki so manjša ali enaka 20. V našem primeru so to 3, 18, 7, 10, 4, 13, 8, 19, 14 in 12. Na koncu zdravljenja pa seveda primerjamo rezultate obeh skupin in vidimo, ali je kirurško zdravljenje uspešno.

34. Spodaj je seznam 20 pacientov, ki so privolili k sodelovanju v preiskusu kirurškega zdravljenja angine. Raziskovalci so zasnovali eksperiment, pri katerem so bolnike razdelili v dve skupini. Bolniki iz prve skupine so bili podvrženi pravi operaciji, ostali pa so dobili placebo terapijo. Raziskovalci pa niso vedeli, da bodo osebki, označeni z zvezdico, med izvajanjem eksperimenta doživeli usodni srčni infarkt. Opazujemo lahko, kako se vzorčna spremenljivost obnaša pri slučajenem eksperimentu, če pogledamo, koliko od teh osem osebkov je bilo v skupini, ki je bila podvržena novi kirurški terapiji.

    Dvajsetkrat naključno izberi 10 osebkov v to skupino in si zapisuj, koliko jih je označenih z zvezdico v vsakem od izborov. Nato nariši histogram, ki prikazuje število izbranih osebkov, ki so doživeli srčni napad. Kolikšno je povprečje po 20 poskusih?

    Razloži, zakaj je bolje uporabiti več tisoč osebkov namesto samo zgornjih dvajset.

rešitev: Osebkom iz tabele priredi števila od 1 do 20. Nato z ukazom sample($1 : 20, 10$) izberemo 10 osebkov, ki bodo prejeli pravo terapijo in preštejemo, koliko med njimi je takih, ki so označeni z zvezdico. To ponovimo dvajsetkrat in nato z ukazom hist narišemo histogram vseh 20 podatkov. Ker je z zvezdico označenih 8 osebkov od 20, mi jih pa izberemo vsakič 10, pričakujemo, da bomo v povprečju izbrali 4 take, ki so označeni z zvezdico.
Pri večjem številu osebkov bo rezultat manj občutljiv na številne zunanje dejavnike, nad katerimi običajno nimamo kontrole (niti vseh ne poznamo). Raziskava lahko namreč da napačne rezultate, če sta deleža osebkov, podvrženih nekem zunanjem dejavniku, med skupinama bistveno različna. Večji kot je vzorec, manjša je verjetnost, da se to zgodi. Denimo, na primer, da iz seznama dveh kadilcev in dveh nekadilcev izberemo dve osebi za kontrolno skupino. Kaj lahko se zgodi, da sta obe izbrani osebi nekadilca (verjetnost za to je 0.25). Če bi v taki raziskavi proučevali vpliv neke vrste vadbe na pljučno kapaciteto, bi verjetno zaključili, da nima čisto nobenega vpliva, čeprav bi morda ta vadba čudovito delovala pri nekadilcih in bi bila neučinkovita le pri kadilcih. Če bi iz seznama 200 kadilcev in 200 nekadilcev izbirali 200 osebkov za kontrolno skupino, bi bila deleža kadilcev v obeh skupinah verjetno zelo podobna, in bi se temu zunanjemu vplivu izognili. V tem primeru bi ugotovili, da vadba pozitivno vpliva na pljučno kapaciteto, ker bi bil rezultat v delovni skupini boljši kot v kontrolni, čeprav seveda ne bi bil tako dober, kot bi bil v primeru, ko bi proučevali zgolj vzorec 400 nekadilcev.

35. S slučajenim primerjalnim eksperimentom so preverjali, če dodatek kalcija v prehrani znižuje krvni pritisk pri zdravih moških. Osebki so 12 tednov prejemali kalcij ali placebo. Raziskovalci so ugotovili, da je bil krvni pritisk v skupini, ki je jemala kalcij, bistveno nižji v primerjavi s kontrolno skupino. Pri tem bistveno pomeni statistično pomembno.

    Razloži, kaj pomeni statistična pomembnost v primeru tega eksperimenta.

rešitev: Krvni pritisk pri pacientih, ki so jemali kalcij, je bil izven intervala, v katerem je 99% ali 95% oziroma nek delež zaupanja vrednosti pritiska pri ostalih bolnikih.

36. Finančni oddelek univerze naredi med študenti anketo o njihovih zaposlitvah in zaslužkih. Poročilo pravi, da so v akademskem letu opazili statistično pomembno razliko v zaslužku med spoloma, pri čemer so moški v povprečju zaslužili več. Nobene razlike ni bilo opaziti med zaslužki črncev in belcev.

    Razloži oba zaključka.

    rešitev: Zaslužki moških so bili nad intervalom, v katerem je bila večina zaslužkov žensk. Zaslužki belcev in črncev pa so bili vsi znotraj istih intervalov.

37. Pogosta oblika neodziva pri telefonskih raziskavah je zvoni, se ne javi. To pomeni, da smo sicer klicali aktivno telefonsko številko, vendar se ni nihče oglasil. Italijanski nacionalni inštitut za statistiko je opazoval neodziv pri raziskavah v italijanskih gospodinjstvih od januarja do velikonočnih praznikov ter preko julija in avgusta. Vsi klici so bili opravljeni med sedmo in deseto zvečer, vendar je bilo v enem od obdobij 21.4% zvoni, se ne javi neodziva, v drugem pa 41.5%.

    Katero od obdobij je po tvojem mnenju imelo večji delež neodziva? Zakaj? Razloži, zakaj so zaradi visokega deleža neodziva rezultati raziskave manj zanesljivi.

    rešitev: Julija in avgusta je več ljudi na počitnicah, v mrzlih mesecih leta pa se več ljudi zadržuje doma, in zato je večji odstotek neodziva nastopil poleti.
    Posledično se v tistem obdobju zmanjša verjetnost, da smo dobili pravi slučajni naključni vzorec.

38. Način zastavljanja vprašanj lahko močno vpliva na izid raziskave. Spodaj sta primera dveh različih formulacij istega vprašanja: 1. Bi morali sprejeti zakon, ki bi preprečil vse možnosti, da bi interesne skupine dajale kandidatom velike količine denarja? 2. Bi morali sprejeti zakon, ki bi interesnim skupinam preprečeval prispevanje h kampanjam, ali imajo skupine pravico prispevati sredstva kandidatom, ki jih podpirajo?

    Pri enem od teh vprašanj je bilo 40% vprašanih za prepoved prispevkov, pri drugem jih je bilo za prepoved 80%. Kateremu vprašanju pripada vsak od teh deležev? Razloži, zakaj sta bila rezultata tako različna.

    rešitev: Prvemu vprašanju pripada 80% delež, drugemu 40%. Drugo vprašanje se da razumeti v smislu pravice do (tudi majhnih) donacij, te pa se večini ne zdijo tako sporne. Prvo vprašanje se da razumeti v smislu prepovedi dajanja podkupnin in preganjanja korupcije, to pa večina podpira.

39. Ali sredinske zavorne luči, ki so obvezne za vse avtomobile, ki so jih v ZDA prodali po letu 1986, res zmanjšajo nevarnost naleta? Slučajeni primerjalni eksperimenti na izposojenih in poslovnih avtomobilih, ki so jih izvedli pred uvedbo obveznih luči, so pokazali, da so se tovrstne nesreče pri uporabi luči zmanjšale za 50. Na žalost pa je uvedba obveznih luči pri vseh avtomobilih pripeljala le do 5 zmanjšanja.

    Razloži, zakaj eksperiment ni realistično posnemal stanja po uvedbi luči.

    rešitev: Pred obvezno uvedbo luči so imeli samo nekateri avtomobili. Zato so luči izstopali in bili bolj opaženi. Predvsem pa je populacija, ki vozi poslovni in izposojeni avtomobili, zelo različna od povprečne splošne populacije.

40. Tvoji bratranci so visoki 101, 91, 46, 96, 76 in 86 cm. Kateri so ubežniki?

    rešitev:

2.2 11 - Vzorčna statistika

1. populacija ima $\mu = x$ in $\sigma = y$.

   Poišči 90. centil za vzorčno porazdelitev, če izbiramo vzorce veliksti 64.

   rešitev:

   Najprej v tabeli pogledamo, koliko standardnih deviacij stran od povprečja se nahaja 90. centil. Vidimo, da se nahaja približno 1.82 standardnih deviacij desno od povprečja. Slučajno spremenljivko denormaliziramo, da dobimo rezultat $x + \frac{1.82}{\sqrt{64}} \cdot y = x + 0.228y$.

2. Če sta pričakovana vrednost in standardni odklon za neko populacijo zaporedoma enaka 25 in 5, in izbiramo vzorce velikosti n = 100, izračunaj standardni odklon za vzorčno povprečje.

   rešitev:

   Standardni odklon ocenimo iz odklona populacije na naslednji način:

   $$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

   Velja torej:

   $$\sigma_{\bar{X}} = \frac{5}{10} = 0.5$$

3. Ali drži naslednja trditev? Če vzorčimo ($n > 1$) iz populacije, ki je porazdeljena normalno s pričakovano vrednostjo 100 in standardnim odklonom 10, potem za vzorčno povprečje $\bar{X}$ velja ???

4. Kaj pomeni vzorčna spremenljivost?

   rešitev: Vzorčna spremenljivost pomeni, da se lahko vrednost vzorčne statistike spreminja, glede na to kateri vzorec izberemo. Zaradi vzorčne spremenljivosti nas potem dejansko zanima porazdelitev vzorčnih statistik, iz katerih potem lahko izvemo marsikaj.

5. Kaj najbolje opiše točkovno cenilko?

   rešitev: Točkovno cenilko najbolje opiše vzorčna statistika, ki jo uporabimo za oceno parametra. Torej statistiko/lastnost vzorcev, iz katere bomo sklepali o lastnostih populacije kot celote.

6. Naj bo X slučajna spremenljivka, ki sledi neko lastnost na izbrani populaciji, katere vrednosti so porazdeljene z $N(78.1, 16.2)$, $\bar{X}$ pa povprečje vzorca velikosti 50 na tej populaciji. Izračunajte naslednji vrednosti ???

7. Preverjali smo ničelno domnevo H_0 napram alternativni H_a s stopnjo tveganja $\alpha = 0.05$. Dobili smo, da je $P - vrednost$ testne statistike enaka 0.01. Kakšen je rezultat testa?

   rešitev: Ker je verjetnost, da bi takšno vrednost testne statistike dobili, če predpostavimo, da hipoteza H_0 drži manjša od $\alpha = 0.05$, ničelno hipotezo zavrnemo in sprejmemo alternativno hipotezo.

8. Ali drži naslednja trditev? Če preučujemo podatke, ki so opisani z normalnim modelom in ne poznamo σ^2 (variance modela), potem lahko uporabimo vzorčno varianco s^2 kot približek za σ^2.

   Vendar pa moramo v tem primeru nadomestiti normalno porazdelitev s Studentovo (t-porazdelitev), če vzorec ni dovolj velik.

rešitev: To drži. Studentova porazdelitev izgleda podobno normalni, ampak ima debelejša repa, saj bi pri malem vzorcu podcenili dejanski parameter σ^2.

9. Ali drži naslednja trditev? Hi-kvadrat test za preverjanje normalne (ali enakomerne) porazdelitve je lahko bodisi enostranski ali dvostranski.

   rešitev: To ne drži. To velja za test z z ali t statistiko.

10. 100 naključno izbranih oseb smo vprašali, ali so zadovoljni z delom vlade. Pritrdilno jih je odgovorilo 35%.

    Kaj je število 35%?

    rešitev: Število 35% je vzorčni delež. Torej predstavlja delež oseb v tem vzorcu, ki so odgovorile pritrdilno (Bernoullijeva porazdelitev in Binomska porazdelitev).

11. Že rešeno, glej nazaj.

12. Vzorčne statistike

    - Pojasni pojem vzorčna statistika.

      rešitev: Vzorčna statistika je neka lastnost vzorca, ki jo opazujemo. Vzorčna statistika je lahko recimo vzorčno povprečje, vzorčni delež, vzorčna mediana... Velik del statistike se ukvarja s tem, kako iz vzorčnih statistik uspešno sklepati o parametrih populacije.

    - Vzorčno povprečje je določeno z zvezo

      $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

      Recimo, da ima spremenljivka X parametra $E(X) = \mu$ in $D(X) = \sigma^2$. Izračunaj $E(\bar{X})$ in $D(\bar{X})$.

      rešitev: Pričakovano vrednost vzorčnih povprečij lahko ocenimo kar s pričakovano vrednost populacije. Varianco vzorčnega povprečja pa lahko ocenimo tako, da varianco populacije delimo z velikostjo vzorca.

    - Iz populacije, porazdeljene normalno $N(\mu, 6)$ vzamemo slučajni vzorec in dobimo

      $$115, 101, 110, 103, 111, 114, 107, 118, a$$

      Kolikšna naj bo vrednost zadnje meritve a, tako da bo vzorčna disperzija čim manjša?

      rešitev: Vrednost zadnje meritve naj je enaka povprečju prejšnih meritev. Lahko tudi zapišemo enačbo za varianco in jo minimiziramo.

    - Kaj se dogaja z vzorčnim povprečjem in s standardno napako z naraščanjem velikosti vzorca (to nam zagotavlja zakon velikih števil)?

      rešitev: Vzorčno povprečje ostaja enako, standardna napaka pa je obratno sorazmerna velikosti vzorca. Intuitivno je vzorčno povprečje porazdeljeno bolj ozko okrog populacijskega/vzorčnega povprečja, saj smo vzeli vzorce, ki vsebujejo velik del populacije. Če bi za vzorec vzeli kar celotno populacijo, bi imeli samo vrednosti vzorčnega povprečja, ki so povsem enake populacijskemu povprečju.

13. Vzorčne statistike in porazdelitve?

- Kakšna je razlika med cenilko in vzorčno statistiko (a najprej definiraj obe)?

  rešitev: Definiciji ... glej nazaj.
  Statistika je funkcija vzorca, cenilka pa je funkcija vzorca, ki je povezana z neko kvantiteto porazdelitve populacije.

- Kako pridemo do porazdelitve hi-kvadrat? Kakšno porazdelitev dobimo, če je n zelo velik?

  rešitev: Hi-kvadrat porazdelitev opisuje porazdelitev vsote kvadratov standardno normalno porazdeljenih slučajnih spremenljivk. Stopnje prostoti pa predstavljajo količino slučajnih spremenljivk, ki jih vsebuje ta vsota. Če je n zelo velik, se porazdelitev približuje normalni porazdelitvi.

- Kdaj rečemo, da je slučajna spremenljivka porazdeljena s Fisherjevo (Snedecorjevo) porazdelitvijo? Če je X porazdeljena s $F(3,6)$, kako je porazdeljena slučajna spremenljivka $\frac{1}{X}$?

  rešitev: Kadar se jo da zapisati kot kvocient dveh neodvisnih slučajnih spremenljivk, ki sta porazdeljeni po hi-kvadrat porazdelitvi.
  Porazdeljena je kot $F(6,3)$.

- Minilo je malo več kot 100 let od kar je Gosset objavil članek o Studentovi t-porazdelitvi z $n-1$ prostostnimi stopnjami. Kako pridemo do nje? Kakšno porazdelitev dobimo če je n zelo velik? Kakšno porazdelitev dobimo, če je $n-1=1$?

  rešitev: Ko n raste, se Studentova t-porazdelitev približuje standardni normalni porazdelitvi. Če je $n-1=1$, potem dobimo Cauchyjevo porazdelitev.

14. Kaj pomeni vzorčna spremenljivost?

    rešitev: Že odgovorjeno - glej nazaj.

15. Kaj najbolje opiše točkovno cenilko?

    rešitev: Že odgovorjeno - glej nazaj.

16. Trgovsko podjetje anketira 50 nakupovalcev, ki jih naključno izbere med 2500 obiskovalci enega od 23 nakupovalnih centrov.

    Kaj sestavlja v tem primeru vzorec?

    rešitev: Vzorec sestavlja 50 izbranih nakupovalcev.

17. V volitvah za župana sodeluje 5 kandidatov in 45 tisoč volilcev. Časopis anketira 750 volilcev, ko zapuščajo volišča. Kaj v tem primeru predstavlja populacijo?

    rešitev: Populacijo predstavlja 45 tisoč volilcev.

18. Pred trgovino z zdravo prehrano izvajamo raziskavo o pozitivnih učinkiv vitaminov.

    Kakšen primer vzorca je to?

    rešitev: To je primer priročnega vzorca.

19. Če ima Anin vzorec mejo napake ± 3, Betin pa pri isti populaciji ± 6, kaj to pomeni?

    rešitev: To pomeni, da je Anin vzorec večji, saj je potem standardna napaka porazdelitve vzorčne statistike manjša.

20. Vsak of petih študentov je izbral vzorec 100 študentov. Vprašali so jih po najljubši vrsti brezalkoholne pijače. Vseh pet rezultatov je bilo različnih. Kaj je razlog za to?

rešitev: Posledica je seveda vzorčna spremenljivost. Vzorčne statistike se pri različnih enostavnih nakl učnih vzorcih lahko spreminjajo.

21. Pri preizkušanju zdravil smo naključno izbrali vzorce za vsako od treh terapij. Niti osebki niti eksperimentator ne vedo, katero terapijo je prejel vsak posameznik.

    Primer kakšnega eksperimenta je to?

rešitev: To je primer slučajnega primerjalnega dvojno slepega eksperimenta.

22. Sociolog želi dobiti mnenja zaposlenih odraslih žensk o sofinanciranju vrtcev. Pridobi si seznam 520 članov lokalnega združenja poslovnih žensk, naključno izbere 100 izmed njih in jim pošlje vprašalnik. V odgovor dovi samo 68 izpolnjenih obrazcev.

    Kaj je v tem primeru populacija in kaj je vzorec?

rešitev: Populacije so zaposlene odrasle ženske. Vzorec so ženske, ki so izpolnile vprašalnik.

23. Gospodinje včasih vlagajo zelenjavo v prazne kozarce majoneze, da prihranijo pri nakupu novih kozarcev. Revijo Organsko vrtnarjenje je zanimalo, kolikšen odstotek teh kozarcev pri tem poči. Zbrali so 100 kozarcev in vanje vložili paradižnik, pri čemer so počili trije kozarci.

    Določi populacijo in vzorec.

rešitev: Populacija so kozarci za majonezo, ki bodo uporabljeni za vlaganje. Vzorec je tistih 100 kozarcev, s katerimi smo izvedli poskus.

24. Nekega ameriškega politika zanima če bo kongres, v katerem je 535 članov, podprl predlagani zakon o omejevanju orožja. Njegovo osebje mu je sporočilo, da so dobili pisma 361 kongresnikov, od katerih jih 323 nasprotuje zakonu.

    Kaj je populacija? Kaj je vzorec? Ali ta vzorec dobro predstavlja populacijo?

rešitev: Populacija je kongres. Vzorec sestoji iz 361 predstavnikov, ki so poslali pisma. Vzorec je verjetno slab, saj je bolj verjetno, da bo pismo poslal nekdo, ki zakonu nasprotuje, kot pa nekdo, ki ga zagovarja. Pričakovali bi torej, da bo delež zagovornikov zakona v kongresu večji kot delež zagovornikov v vzorcu.

25. revija za zdravo prehrano in naravno medicino želi dokazati, da uživanje velikih količin vitaminov pozitivno vpliva na zdravje. Uredniki prosijo bralce, ki so redno jemali vitamine v velikih količinah, da sporočijo svoje izkušnje. Od 2754 prejetih pisem, jih 93% poroča o pozitivnih učinkih.

    Ali je bolj verjetno, da je delež 93% manjši, večji ali enako velik kot delež odraslih, ki bi opazili pozitivne učinke jemanja vitaminov? Zakaj?

rešitev: Delež bo verjetno manjši, saj nismo izbrali ustreznega enostavnega naključnega vzorca. Bralci te revije bodo najbrž bolj nagnjeni k temu mnenju.

26. Leta 1995 je revija USA Weekend objavila oglas, podoben spodnjemu: Ali bi želeli v primeru neozdravljive bolezni imeti pravico do evtanezije. V drobnem tisku je bilo navedeno, da vsak klic stane 50 centov.

    Ali lahko zaupamo rezultatom te ankete? Odgovor utemelji.

rešitev: Prostovoljno bi se odzvali verjetno samo tisti, ki imajo zelo močno mnenje. Posledično to ni ustrezen enostaven naključen vzorec.

27. V televizijski oddaji so izvedli anketo o dobrodelnosti. Prejeli so 50 tisoč klicev in 83% gledalcev je trdilo, da redno darujejo denar in oblačila dobrodelnim organizacijam.

    Razloži, zakaj je ta vzorec skoraj gotovo pristranski.

    V katero smer je pristranski? Se pravi, ali je odstotek ljudi, ki darujejo v humanitarne namene večji ali manjši od 83%, ki jih je dal ta vzorec.

    rešitev: Klicali bodo verjetno samo tisti, ki redno darujejo in se želijo s tem pohvaliti. Dejanski delež, ki bi ga lahko dobili z ustreznim vzorčenjem, je zaradi tega manjši.

28. Mestna policija v Miamiju želi dobiti mnenje črnske skupnosti o delu policije. Sociolog pripravi vprašalnik in izbere vzorec 300 naslovov v pretežno črnskih četrtih. Uniformirani temnopolti policaj obišče vsakega od naslovov in želi govoriti z odraslim članom gospodinjstva.

    Razloži, zakaj pričakuješ, da bodo rezultati te raziskave pristranski. Na kakšen način se bodo razlikovali od resničnega mnenja populacije?

    rešitev: Ker anketira uniformirani policaj, bodo rezultati verjetno ugodnejši za policijo kot bi bili sicer.

29. Podjetje želi razumeti odnos pripadnikov manjšin med vodilnimi uslužbenci do sistema ocenjevanja vodstvenih delavcev. Spodaj je seznam vseh vodilnih uslužbencev, ki pripadajo kakšni od manjšin.

Agarwal	Huang
Anderson	Kim
Baxter	Liao
Brown	Mouring
Bowman	Naber
Castillo	Peters
Cross	Pliego
Dewald	Puri
Fernandez	Richards
Flemming	Rodrigez
Gates	Santiago
Goel	Shen
Gomez	Vega
Hernandez	Wang

    Opiši, kako bi uporabil naslednjo tabelo naključnih števil, da bi s seznama izbral 6 ljudi, ki bodo podrobneje izprašani.

Agarwal	Huang
Anderson	Kim
Baxter	Liao
Brown	Mouring
Bowman	Naber
Castillo	Peters
Cross	Pliego
Dewald	Puri
Fernandez	Richards
Flemming	Rodrigez
Gates	Santiago
Goel	Shen
Gomez	Vega
Hernandez	Wang

rešitev: Najprej imenom iz tabele priredimo števila od 1 do 28, npr. tako, da jih oštevilčimo po stolpcih od leve proti desni, od zgoraj navzdol. Tudi katerikoli drug vrstni red bi bil seveda dober. Iz tabele preberemo naključna dvomestna števila:

73	06	36	36	23
29	38	88	95	07
78	55	36	27	12
89	34	32	04	01

Za pogovor izberemo kandidate, ki so imeli števila 6, 23, 7, 27, 12 in 4.

30. Kakšne vrste programov poujajo fakultete najboljšim maturantom? Odločiš se zbrati informacije s petih naključno izbranih oddelkov.

Računalništvo	Komunikacije
Ekonomija	Slovenščina
Tuji jeziki	Medicina
Zgodovina	Politologija
Fizika	Statistika
Sociologija	Biologija
Kemija	Matematika
Geografija	Filozofija
Psihologija	

Uporabi naslednjo tabelo naključnih števil, da izbereš enostavni slučajni vzorec petih oddelkov s spodnjega seznama.

08900	87788
73717	19287
69954	45917
80026	55598
86757	47905
16890	99047
78249	73739
97076	00525

rešitev: Na voljo imamo 17 oddelkov, ki jih oštevilčimo s števili od 1 do 17. Iz tabele naključnih števil naredimo seznam dvomestnih števil in tista, ki so manjša od 18 uporabimo za izbiro vzorca. Na ta način izberemo oddelke, ki so imeli oznake 8, 17, 2, 5 in 16.

31. Štuentka želi ugotoviti, kaj meijo predavatelji o ustanovitvi nacionalnega odbora za visoko šolstvo, ki bi nadzoroval vse fakultete v državi. Na fakulteti je zaposlenih 380 predavateljev.

- Kaj je v tem primeru populacija?

  rešitev: Populacija so vsi predavatelji (380 predavateljev).

- Natančno razloži, kako bi izbral enostavni slučajni vzorec 50 predavateljev. Uporabi naslednjo tabelo naključnih števil, da izbereš prvih pet oseb tega vzorca.

  rešitev: Predavatelje oštevilčimo s števili od 1 do 50, nato pa s pomočjo tabele naključnih števil izdelamo seznam naključnih dvomestnih števil in izberemo predavatelje iz seznama.

32. Število študentov politologije na Ljubljanski univerzi se je bistveno povečalo, število predavateljev pa je ostalo nespremenjeno. Študentski časopis namerava anketirati 25 od 450 študentov politologije, da bi ugotovili, kaj menijo o velikosti razredov in drugih s tem povezanih problemih. Predlagaš jim enostavni slučajni vzorec.

    Natančno razloži, kako bi izbral(a) ta vzorec. Nato uporabi naslednjo tabelo naključnih števil, da izbereš prvih pet oseb tega vzorca.

    rešitev: Študentom dodelimo števila med 1 in 450, nato pa naključno izberemo 25 od teh števil. Vzorec petih študentov iz zgornje tabele npr. dobimo tako, da števila razdelimo v sklope po 3.

33. Katere od naslednjih trditev o tabeli naključnih števil so pravilne? Odgovore utemelji.

- V vsaki vrstici, ki vsebuje 40 števil, so natanko štiri ničle.

  rešitev: To ni res, v povprečju lahko pričakujemo 4 ničle, seveda pa jih lahko dobimo tudi več ali manj. Število ničel med 40 števili iz tabele naključnih števil je slučajna spremenljivka, ki je porazdeljena binomsko $B(40, \frac{1}{10})$.

- Vsak par števk ima natanko $\frac{1}{100}$ možnosti, da je enak 00.

  rešitev: To drži.

- Skupina 0000 se nikoli ne pojavi v tabeli, ker tako zaporedje ni naključno.

  rešitev: Tudi tako zaporedje se lahko pojavi, in sicer z verjetnostjo $\frac{1}{10^4}$.

34. Gospa Caucus je kandidatka svoje stranke v drugem kongresnem okrožju v Indiani. Stranka želi vedeti, kolikšen delež registriranih volilcev bi glasoval za gospo Caucus, če bi bile volitve jutri. Agencija kontaktira 800 volilcev, od katerih jih 456 izjavi, da bi glasovali za gospo Caucus.

    Kaj je v tem primeru populacija? Kaj je vzorec? Razloži nekomu, ki ne zna statistike, zakaj je vzorec z 800 volilci boljši kot tak z 200 volilci.

    rešitev: Populacija so vsi volivci. Vzorec je 800 kontaktiranih volilcev. Večji vzorec je boljši od manjšega, ker nam da dovolj verodostojne podatke. Z naraščanjem vzorca se namreč manjša interval zaupanja, torej lahko pri isti stopnji zaupanja pri večjem vzorcu pričakujemo manjšo napako. To je posledica dejstva, da je standardni odklon vzorčne statistike obratno sorazmeren kvadratnemu korenu velikosti vzorca n.

35. Izbiranje vzorca s seznama, ki vsebuje le del populacije, je pogost vzrok pristranskosti pri vzorčenju. V vsakem od naslednjih primerov razloži, zakaj bi lahko bil tak vzorec pristranski.

- Da bi dobili javno mnenje o predlogu, ki bi znižal socialno podporo, v agenciji izberejo vzorec tako, da pokličejo s pomočjo računalnika naključno izbrane telefonske številke.

  rešitev: Da bi dobili javno mnenje o predlogu, ki bi znižal socialno podporo, v agenciji izberejo vzorec tako, da pokličejo s pomočjo računalnika naključno izbrane telefonske številke.

- Da bi ugotovila, kako bodo na ta isti predlog odreagirali volilci, članica kongresa pošlje vprašalnik vsem registriranim volilcem v svojem okrožju.

  rešitev: Njeno okrožje morda ni dober primer tipičnega okrožja (morda sodi npr. v posebej prestižno sosesko ali pa revnejši del države).

36. V neki reviji večkrat povabijo bralce, da sodelujejo pri anketah. Ob neki priložnosti so jim postavili naslednje vprašanje: "Če bi se imeli možnost ponovno odločati, bi imeli otroke?" Od skoraj 10 tisoč staršev, ki so se odzvali na anketo, jih je 70% odgovorilo, da ne. Kmalu po tem je nacionalna raziskava postavila enako vprašanje vzorcu 1400 staršev. Pri tem so prejeli 90% pozitivnih odgovorov.

    Kateremu od teh vzorcev gre bolj zaupati? Zakaj?

    rešitev: Drugemu - pri prvem je bilo sodelovnje iniciativa anketiranca, zato so se odzvali le tisti, ki imajo o vprašanju zelo močno mnenje.

37. Janino malo računovodsko podjetje ima 30 strank.

    Namesto Jane anketiraj vzorec petih strank, s katerim bo ugotovila, kako bi lahko povečali zadovoljstvo z njihovimi storitvami. Da bi se izognili pristranskosti, izberemo enostavni slučajni vzorec velikosti 5. Pomagaj si z ukazom sample v programu R.

    rešitev: ??

38. Davčni urad v ZDA namerava pregledati enostavni slučajni vzorec davčnih napovedi za vsako od držav. Med drugim jih zanima delež olajšav. Celotno število napovedi se spreminja od države do države: od preko 13 milijonov v Kaliforniji do manj kot 220 tisoč v Wyomingu.

    Ali se bo meja napake pri ocenjevanju deleža olajšav spreminjala od države do države, če uporabimo enostavni slučajni vzorec 2000 napovedi?

    Ali se bo meja napake spreminjala od države do države, če uporabimo enostavni slučajni vzorec, ki zajema 1% vseh napovedi v vsaki državi? Odgovor utemelji.

    rešitev: Ne.
    Da - velikosti vzorcev se spreminjajo.

39. Na zadnjem koraku pri štetju prebivalstva izberemo naslove znotraj majhnih območij, imenovanih bloki. Uporabimo sistematično slučajno vzorčenje, ki ga bomo predstavili na tem primeru. Recimo, da moramo izbrati 4 naslove izmed 100. Ker je $100/4 = 25$, si lahko mislimo, da je naš seznam sestavljen iz štirih seznamov s po 25 naslovi. Izberemo prvega od petindvajsetih naključno s pomočjo tabele naključnih števil. Vzorec vsebuje ta naslov in

naslove, ki so od njega oddaljeni za 25, 50 in 75 mest. Če na primer izberemo 13, potem vzamemo v vzorec naslove z oznakami 13, 38, 63 in 88.

Kot pri enostavnem slučajnem vzorcu imajo tudi pri sistematičnem slučajnem vzorcu vsi posamezniki enako možnost, da so izbrani. Razloži, zakaj je to res, potem pa še, zakaj kljub temu sistematični slučajni vzorec ni enostavni slučajni vzorec.

rešitev: To ni enostaven slučajni vzorec, ker niso možne izbire vseh kombinacij petih naslovov s seznama.

40. Eksperiment, s katerim naj bi potrdili, da meditacija znižuje stres, je potekal takole: Eksperimentator je anketiral osebke in si zabeležil njihov nivo stresa. Nato so osebke naključno razporedili v dve skupini. Eksperimentator je eno od skupin naučil meditirati, kar so nato počeli en mesec. Drugi skupini so samo rekli, naj se bolj sprostijo. Po enem mesecu je eksperimentator spet izprašal vse osebke in ocenil nivo stresa. Ta je bil nižji v skupini, ki je meditirala. Psihologi trdijo, da so rezultati sumljivi, ker ocenjevanje nivoja stresa ni bilo slepo.

    Razloži, kaj to pomeni, in kako bi to lahko povzročilo pristranskost rezultatov.

    rešitev: Ocenjevanje ni bilo slepo, saj je vsak udeleženec vedel, v kateri skupini je. Tako bi lahko prišlo do placebo efekta, ki bi lahko povzročil pristranskost rezultatov.

41. Ignoriraj praktične težave in moralne pomisleke in osnuj eksperiment, ki bi odgovoril na vprašanje, ali kajenje povzroča pljučnega raka.

    rešitev: V rešitvah piše:
    Izbrali bi dva naključna vzorca ljudi in prvemu vzorcu dali kaditi prave cigarete, drugemu pa cigarete brez tobaka. Primerjali bi deleža ljudi v vzorcih, ki so dobili pljučnega raka.
    Morali bi biti prepričani, da je nadomestek za tobak polnoma neškodljiv za zdravje in ne povzroča plučnega raka.

42. Pri testiranju učinka obstojnih pesticidov bodo raziskovalci 60 dni hranili podgane s hrano, ki bo vsebovala DDT. Nato bodo izmerili njihovo odzivnost živčevja, da bi ugotovili, kakšen je vpliv DDT.

    - Razloži, zakaj bi morali raziskovalci opazovati tudi kontrolno skupino, ki bi dobivala sicer enako, a nekontaminirano hrano.

      rešitev: Zaradi primerjave in morebitnih ostalih faktorjev, ki bi lahko vplivali na rezultate.

    - Recimo, da je na voljo 20 novorojenih podgan. Načrtuj eksperiment in s pomočjo programa R izberi vzorec.

      rešitev:

      Izberemo 10 podgan za terapijo. Ostale podgane damo v kontrolno skupino.

43. continue at FAPP1.10P48

2.3 12 - Intervali zaupanja

1. V tovarni pakirajo sladkor v 5 kg vreče. Količina sladkorja varira po normalni porazdelitvi in ima pričakovano vrednost 5 kg ter standardni odklon 0.05 kg.

Za računanje dogodkov o težah posameznih vreč bomo uporabili spremenljivko $Z = \frac{X-5.0}{0.05}$, medtem, ko bomo za računanje verjetnosti o težah vzorčnih povprečji za velikosti 25 uporabili spremenljivko $\bar{Z} = \frac{\bar{X}-5.0}{0.01}$.

rešitev: To drži. Pomni, da velja $\mu_{\bar{X}} \approx \mu$ in $\sigma_{\bar{X}} \approx \frac{\sigma}{\sqrt{n}}$. Torej pri standardizaciji delimo z kvocientom standardne deviacije in kvadratnim korenom velikosti vzorca.

2. Če vzorčimo iz $N(50, 5)$, izračunaj kakšna mora biti najmanjša velikost vzorca, da bo vsaj 90% vzorčnega povprečja med 48.5 in 51.5.

rešitev:

Najprej izračunamo koliko standardnih deviacij stran od povprečja moramo iti v standardizirani normalni porazdelitvi, da zajamemo 90% vseh vrednosti. V z tabeli pogledamo vrednost za 95% (na vsaki strani bomo pustili 2.5% vrednosti) in dobimo rezultat 1.65.
Sedaj računamo vrednost s.s. ki predstavlja vzorčno povprečje.
Veljati mora:

$$\bar{X} = 50 \pm 1.65 \cdot \frac{5}{\sqrt{n}}$$

Enačbo rešimo za n, vrednost zaokrožimo gor in dobimo rezultat 31.

3. Ali drži naslednja trditev? 95% interval zaupanja, ki ga izračunamo iz vzorca, je večji od 99% intervala zaupanja istega vzorca.

rešitev: To seveda ne drži. 95% interval zaupanja je ožji kot 99% interval zaupanja.

4. Standardni odklon s_p vzorčnega deleža p se spreminja skupaj z dejansko vrednostjo deleža populacije π. Recimo, da je velikost vzorca $n = 1500$. Izračunaj s_p za $\pi = 30, 40, 50, 60$ in 70. Nato izračunaj s_p za $\pi = 0, 10$ in 20. Zanima nas na katerem delu se s_p spreminja najhitreje, ko se spreminja π (v ta namen lahko narišeš graf s_p v odvisnosti od π).

   Ali drži naslednja trditev? Standardni odklon s_p vzorčnega deleža se ne spreminja veliko, če p ni blizu 0 ali 100.

rešitev: Standardni odklon vzorčnega povprečja izračunamo kot $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$. Iz grafa vidimo, da se standardni odklon vzorčnega deleža najbolj spreminja, ko je p blizu 0 ali 100.

5. Ali velja naslednja trditev? 99% interval zaupanja z večjo verjetnostjo vključuje nek parameter populacije kot 95% interval zaupanja.

rešitev: To drži, saj je 99% interval zaupanja širši kot 95% interval zaupanja.

6. Ali drži naslednja trditev? Iz 99% intervala zaupanja $[0.2, 0.4]$ za delež lahko z gotovostjo sklepamo, da populacijski delež leži med 20% in 40%?

rešitev: To ne drži povsem, saj je vseeno možno, da vrednost leži zunaj tega intervala, pa čeprav je bolj malo verjetno.

7. Intervali zaupanja

   - Pojasni razliko med točkovno in intervalno oceno.

- Opiši postopek intervalskega ocenjevanja parametrov in pojasni kaj nam pove koeficient zaupanja (1 - α) (teoretična interpretacija).

  rešitev:

- Na vzorcu velikosti $n = 100$ podjetnikov v majhnih podjetjih v Sloveniji, ki je bil izveden v okviru ankete 'Drobno gospodarstvo v Sloveniji', so izračunali, da je povprečna starost anketiranih podjetnikov $\bar{X}$=40,4 let in standardni odklon $s = 10.2$ let. Pri 5% tveganju želimo z intervalom zaupauja oceniti povprečno starost podjetnikov v majhnih podjetjih v Sloveniji. Tabela porazdelitvene funkcije standardne normalne porazdelitve:

$\Phi(c)$	0,9	0,95	0,975	0,99
c	1,286	1,645	1,96	2,32

- **rešitev:** Zanima nas interval vrednosti vzorčnega povprečja, da bo v 95% vseh vzorcev ta interval vseboval pravo vrednost parametra.

  Najprej v z tabeli pogledamo, koliko standardnih deviacij stran od povprečja moramo iti, da zavzamemo 95% vrednosti v standardizirani normalni porazdelitvi.

  Izkaže se, da moramo iti približno 1.64 standardnih deviacij stran od povprečja.

  Izračunajmo interval:

$$\bar{X} = 40.4 \pm 1.64 \cdot \frac{10.2}{\sqrt{100}}$$

$$[38.73, 42.07]$$

  Ko apliciramo to metodo na vzorcih, bo interval vseboval pravilno vrednost parametra v 95% primerov.

- Opiši ocenjevanje parametrov z majhnimi vzorci (čim več možnosti)

  rešitev: t-statistika... TODO

- Podaj osnovni izrek statistike.

  rešitev: Glej zapiske za teoretični del.

8. Ali velja naslednja trditev? Za populacijo z znanim odklonom σ iščemo interval zaupanja za pričakovano vrednost s stopjo zaupanja $1 - \alpha$. Tudi če za izračun uporabljamo vzorce iste velikosti, širina intervala zaupanja še vedno ni nujno vsakič enaka.

   rešitev: To drži. Stopnja zaupanja samo pomeni, da bo ta metoda pri $1 - \alpha$ vzorcih dala interval, ki bo vseboval pravo vrednost parametra.

9. Ali drži naslednja trditev? 90% interval zaupanja z večjo verjetnostjo vključuje nek parameter populacije kot 95% interval zaupanja.

   rešitev: Ta trditev ne drži, saj je 90% interval zaupanja ožji kot 95% interval zaupanja.

10. Ali drži naslednja trditev? 95% interval zaupanja, ki ga izračunamo iz vzorca, je večji od 99% intervala zaupanja istega vzorca.

    rešitev: Ne drži. 99% interval zaupanja istega vzorca je širši kot 95% interval zaupanja.

11. Eno izmed vprašanj na oftalmologiji (specializacija na očesni kirurgiji, ki skrbi za strukturo in funkcije oči, ter za njihove medicinske oz. kirurške načine zdravljenja) je, ali se sferični lom razlikuje v povprečju med levim in desnim očesom. V študiji so izmerili lom na levo in

desno oko 17 bolnikom. Razlike (desno-levo) v dioptriji so $d_1, d_2, ..., d_{17}$, od koder izračunamo $\sum_{i=1}^{17} d_i = -3.50$ in $\sum_{i=1}^{17} d_i^2 = 19.13$. Ali velja naslednja trditev? 90% interval zaupanja (na dve decimalki natančno) za povprečne razlike (desno - levo) je enak (-0.55, 0.14).

rešitev:

Najprej izračunamo vzorčno povprečje razlik in popravljeno vzorčno varianco:

$$\begin{aligned} \bar{d} &= -0.2059 \\ s_d^2 &= 1.15059 \end{aligned} \tag{1}$$

Za izračun 90% intervala zaupanja zaradi majhnosti vzorca in uporabe standarde deviacije vzorca kot približka za standardno deviacijo populacije uporabimo studentovo t-porazdelitev.

$$-0.2059 \pm 1.746 \cdot \frac{1.07266}{\sqrt{17}} = [-0.66, 0.25]$$

Trditev torej ne drži.

12. Predpostavi, da je neka lastnost na populaciji porazdeljena normalno z varianco 9cm2. Kaj je najmanjša velikost vzorca, ki jo potrebujemo za 95% interval zaupanja (v primeru pričakovane vrednosti te lastnosti), če želimo, da ta interval ni daljši od 1cm? (a) n=1245 (b) n=34 (c) n=95 (d) n=139 Pravilen odgovor je (d).

rešitev: Rešimo enačbo:

$$1.65 \cdot \frac{9}{\sqrt{n}} \leq 0.5$$

za n.

13. Oglaševalska agencija v ZDA izvaja raziskavo, da bi ugotovila, kako ženske reagirajo na različne pridevnike, ki jih lahko uporabimo za opis avtomobilov. Izberejo 600 žensk iz cele države. Vsaki predvajajo seznam pridevnikov kot na primer eleganten ali prestižen. Pri vsakem morajo povedati, kako zaželen se bi jim zdel avto, ki bi ga opisali s tem pridevnikom. Možni odgovori so (1) zelo zaželen, (2) zaželen, (3) neopredeljena ali (4) ne zaželen. Med vprašanimi je 76% odgovorilo, da je avtomobil, ki je opisan kot eleganten, zelo zaželen.

- Kaj je v tej raziskavi populacija?

  rešitev: Populacija so vse odrasle ženske v ZDA.

- Koliko žensk iz vzorca je odgovorilo, da bi bil eleganten avtomobil zelo zaželjen?

  rešitev: Po trivialnem postopku izračunamo rezultat 456.

- Uporabili so slučajni vzorec po vzoru Gallupovih raziskav ($\alpha = 0.05$). Na katerem intervalu se zelo verjetno nahaja resnični odstotek žensk, ki bi menile, da je takšen avtomobil zelo zaželjen?

  rešitev: Računamo interval zaupanja.

  $$\bar{X} = 0.76 \pm 1.96 \cdot \sqrt{\frac{0.76 \cdot 0.24}{600}} = [0.7267, 0.7933]$$

14. V aziskavi mnenja vprašamo 1324 odraslih oseb, če verjamejo, da obstaja življenje na drugih planetih. Med njimi jih 609 odgovori pritrdilno.

Kakšen odstotek vzorca verjame v zunajzemeljsko življenje?

Agencija oznani, da je meja napake pri tej raziskavi $\pm 3\%$. Na katerem intervalu najverjetneje leži delež vseh odraslih, ki verjamejo v življenje na drugih planetih? S kakšno stopnjo tveganja?

rešitev: Odstotek vzorca izračunamo po trivialnem postopku, da dobimo rezultat 46%. Delež leži na intervalu $(0.43, 0.49)$. Stopnjo tveganja izračunamo tako, da pogledamo koliko standardnih deviacij stran ležita mejni vrednosti.

$$z = \frac{0.49 - 0.46}{\sqrt{\frac{0.46 \cdot 0.54}{1324}}} = \frac{0.03}{0.0137} = 2.19$$

Stopnja tveganja je $\alpha = 0.05$.

15. Nacionalne raziskave mnenja kot je na primer Gallupova, običajno vsak teden izberejo vzorec 1500 ljudi.

    - Pri vzorcu te velikosti je običajno meja napake približno ± 3 odstotne točke. Določi interval zaupanja za delež π, če je vzorčni delež enak p.

      rešitev: Delež potem najverjetneje leži na intervalu $p \pm 0.03$.

    - Tik pred predsedniškimi volitvami pa agencije navadno povečajo velikost vzorcev na približno 4000 ljudi. Ali je meja napake zdaj več kot ± 3, manj kot ± 3 ali nespremenjena? Zakaj?

      rešitev: Meja napake je obratno sorazmerna z $\sqrt{n}$, zato je pri večjem vzorcu meja napake manjša.

16. Članek v časopisu poroča, da je v nedavni Gallupovi raziskavi 78% vzorca 1108 odraslih oseb reklo, da verjamejo v obstoj nebes. Samo 60% jih je reklo, da verjamejo v pekel. Članek se zaključi z besedami: Meja napake vzorca je bila 4 odstotne točke.

    rešitev: Ali lahko z gotovostjo trdimo, da med 56% in 64% odraslih verjame v pekel? Odgovor utemelji!

    rešitev: Izračunajmo stopnjo verjetja za ta interval.

    $$z = \frac{0.64 - 0.60}{\sqrt{\frac{0.6 \cdot 0.4}{1108}}} = 2.72$$

    Pogledamo v z-tabelo in vidimo, da lahko to trdimo z približno 99% zaupanjem.

17. Članek v časopisu poroča, da je v nedavni Gallupovi raziskavi 78% vzorca 1108 odraslih oseb reklo, da verjamejo v obstoj nebes. Samo 60% je reklo, da verjamejo v pekel. Članek se zaključi z besedami: Meja napake vzorca je bila 4 odstotne točke.

    Ali lahko z gotovostjo trdimo, da med 56% in 64% odraslih verjame v pekel?

    rešitev: Že rešeno - glej nazaj.

18. Število napak na kvadratni meter preproge se spreminja, srednja vrednost je 1.6, standardni odklon pa 1.2 napake na kvadratni meter. Porazdelitev ne more biti normalna, ker je število napak vedno celo. Inšpektor v vzorec izbere 200 kvadratnih metrov preprog, zabeleži število napak, ki jih je našel v vsakem kvadratnem metru, in izračuna $\bar{x}$, povprečno število napak na kvadratni meter v pregledanem vzorcu. Take preglede opravi večkrat.

Na katerem intervalu leži srednjih 95% vseh $\bar{x}$?

rešitev:

Ker je vzorec velik, se po centralnem limitnem izreku lahko porazdelitev vzorčnih povprečij aproksimiramo z normalno porazdelitvijo.

$$\bar{X} = 1.6 \pm 1.96 \cdot \frac{1.2}{\sqrt{200}} \equiv [1.434, 1.766]$$

19. S slučajnim vzorcem velikosti 2000, ki ga izberemo med prebivalci spodnjega kašlja, ugotovimo, da jih 45 še nikoli ni imelo noric. Poišči 95% interval zaupanja za dejanski delež meščanov, ki še nikoli niso imeli noric.

rešitev: Računamo po znanem postopku:

$$I = 0.0225 \pm 1.96 \cdot \sqrt{\frac{0.0225 \cdot 0.9775}{2000}} = [0.016, 0.029]$$

Pazi, v rešitvah je napaka!

20. Poročilo o raziskavi na vzorcu 1500 odraslih pravi: "S stopnjo zaupanja 95% med 27% in 33% vseh Američanov meni, da so droge največji problem v javnih šolah."

Razloži nekomu, ki ne zna statistike, kaj v tem primeru pomeni "s stopnjo zaupanja 95 %".

rešitev: Z verjetnostjo 95% je dejanski delež vseh Američanov s takim mnenjem med 27 in 33. Bolj natančno to pomeni, da bo 95% vseh naključnih enostavnih vzorcev vsebovalo na svojem intervalu zaupanja, ki ga zgradimo po znani metodi, dejansko pravilno vrednosti parametra populacije, ki nas zanima.

21. Pri Gallupovi raziskavi so vprašali slučajni vzorec 1005 odraslih, če podpirajo dvojezične javne šole ali pa menijo, da bi se morali učenci, ki ne govorijo angleško, vključiti v angleške šole in se naučiti jezika na ta način. Pri tem je bilo 63% za enojezične šole.

V letaku so objavili, da je bila meja napake te raziskave 3%. Razloži, kaj to pomeni.

rešitev: To pomeni, da je dejanski parameter, ki nas zanima z verjetnostjo stopnje zaupanja (pri Gallupu je to 96%) na intervalu med 3% manj in 3% več od navedenih 63%. Bolj natančno to pomeni, da bo okoli 95% intervalov zaupanja, ki jih zgradimo po znani metodi dejansko vsebovalo točni parameter populacije.

22. Pri Gallupovi raziskavi so vprašali slučajni vzorec 1005 odraslih, če podpirajo dvojezične javne šole ali pa menijo, da bi se morali učenci, ki ne govorijo angleško, vključiti v angleške šole in se naučiti jezika na ta način. Ugotovili so, da je 60& vprašanih za enojezične šole.

Podaj 95% interval zaupanja za delež odraslih, ki bi odgovorili, da so za dvojezične šole.

rešitev: Zgradimo interval zaupanja po znanem postopku.

$$I = 0.6 \pm 1.96 \cdot \sqrt{\frac{0.6 * 0.4}{1005}} \equiv [0.5697, 0.6303]$$

23. V ZDA razmišljajo, da bi dodatno omejili število vozil, ki lahko vstopijo v Yellowstonski park. Da bi ocenili odziv javnosti, vprašajo enostavni slučajni vzorec 150 obiskovalcev, če podpirajo takšne ukrepe. Od teh jih 89% odgovori pritrdilno.

Podaj 95% interval zaupanja za delež obiskovalcev parka, ki so za omejitev. Ali lahko s 95% stopnjo zaupanja trdiš, da jih je več kot polovica za omejitev? Odgovor utemelji.

rešitev: Konstruirajmo interval zaupanja po znanem postopku.

$$I = 0.89 \pm 1.96 \cdot \sqrt{\frac{0.89 \cdot 0.11}{150}} \equiv [0.8399, 0.9401]$$

S 95% stopnjo zaupanja lahko trdimo, da to drži, daj interval zaupanja v celoti leži nad 50%.

24. New York Times in CBS News sta izvedla nacionalno raziskavo na 1048 slučajno izbranih najstnikih med 13. in 17. letom. Med njimi jih je 692 imelo televizijski sprejemnik v svoji sobi in 189 jih je kot svoj najljubši program navedlo Fox. Predpostavljali bomo, da je šlo za enostavni slučajni vzorec.

- Podaj 95% interval zaupanja za delež vseh ljudi te starosti, ki imajo v svoji sobi TV sprejemnik, in za tiste, ki najraje gledajo Fox.

  rešitev: Intervala zaupanja izračunamo po znanem postopku.

  $$I = 0.6603 \pm 1.96\sqrt{\frac{0.6603 \cdot 0.3397}{1048}} \equiv (0.631626, 0.688974)$$

  Drugi interval zaupanja izračunamo na enak način.

- Časopisni članek pravi, "Teoretično se v 19 primerih od 20 rezultati ne razlikujejo za več kot 3 odstotne točke od vrednosti, ki bi jo dobili, če bi anketirali vse ameriške najstnike." Pojasni, kako se tvoji izračuni ne ujemajo s to trditvijo.

  rešitev: Članek trdi, da je 95% interval zaupanja širši, kot smo ga mi izračunali.

25. V raziskavo o enakopravnosti žensk v ZDA, ki jo je izvedla televizijska hiša MSNBC, je bilo vključenih 1019 odraslih oseb. Časopisni članek, ki je poročal o raziskavi, je navajal, "Meja napake pri rezultatih je 3 odstotne točke".

- Skupno je 54 vzorca (550 od 1019 ljudi) odgovorilo, da se je na tem področju naredilo dovolj. Poišči 95 interval zaupanja za delež odraslih, ki bi odgovorili pritrdilno. Ali je poročilo o meji napake približno pravilno? (Predpostavi, da je šlo za enostavni slučajni vzorec.)

  rešitev:

  $$I = 0.54 \pm 1.96 \cdot \sqrt{\frac{0.54 \cdot 0.46}{1019}} = (0.509398, 0.570602)$$

  Poročanje je torej pravilno.

- Časopisni članek je trdil, da 65% moških in le 43% žensk meni, da se je na tem področju storilo dovolj. Pojasni, zakaj nimamo dovolj informacij, da bi lahko podali intervale zaupanja posebej za moške in za ženske.

  rešitev: Nimamo natančnih podatkov, koliko moških in koliko žensk je bilo vključenih v raziskavo.

- Ali bi bila meja napake pri 95% intervalu zaupanja za ženske večja, manjša ali enaka 0.03? Zakaj? Navedbe časopisa o meji napake so očitno nekoliko zavajajoče.

rešitev: Načeloma ne vemo, koliko žensk, x, in koliko moških, y, je bilo vključenih v raziskavo, vemo pa, da mora veljati

$$0.43x + 0.65y = 550$$

in

$$x + y = 1019$$

Če rešimo ta sistem enačb, dobimo $x = 510.7$. Seveda je bilo število žensk celo, zato lahko sklepamo, da jih je bilo 511. Razlog, da rezultat ni celo število, je seveda v tem, da sta 65 in 43 zaokroženi vrednosti. Pri tem vprašanju lahko torej predpostaviš, da je bilo v raziskavo vključenih 511 žensk.
Ob predpostavki, da je 219 od 511 žensk odgovorilo pritrdilno, je meja napake, če poračunamo 95% interval zaupanja enak 4.30.

26. Poročilo pravi, da so pri nacionalni raziskavi na 1500 slučajno izbranih odraslih ugotovili, da jih je 43% mnenja, da jim bo šlo naslednje leto slabše. V nadaljevanju poročila je bilo zapisano, da je meja napake 3 odstotne točke s 95% stopnjo zaupanja.

    Kateri od naslednjih virov napak so vključeni v mejo napake iz poročila?

    rešitev: Pri raziskavi so slučajno izbirali telefonske številke, zato so pri tem izpustili ljudi, ki niso imeli telefona. Neodziv: pri nekaterih od teh številk se ni nihče oglasil ali pa je klicani zavrnil sodelovanje. Slučajne variacije pri slučajnem izbiranju telefonskih številk.

27. Poročilo pravi, da so pri nacionalni raziskavi na 1500 slučajno izbranih odraslih ugotovili, da jih je 43 mnenja, da jim bo šlo naslednje leto slabše. V nadaljevanju poročila je bilo zapisano, da je meja napake 3 odstotne točke s 95% stopnjo zaupanja

    Ali bi bila pri 90% intervalu zaupanja na osnovi rezultatov te ankete meja napake večja, manjša ali enaka 3% točkam?

    rešitev: Če bi zmanjšali stopnjo zaupanja, bi se z-vrednost zmanjšala in s tem tudi meja napake ter z njo širina intervala zaupanja.

28. Poročilo pravi, da so pri nacionalni raziskavi na 1500 slučajno izbranih odraslih ugotovili, da jih je 43% mnenja, da jim bo šlo naslednje leto slabše. V nadaljevanju poročila je bilo zapisano, da je meja napake 3 odstotne točke s 95% stopnjo zaupanja. Recimo, da bi pri raziskavi anketirali 1000 ljudi namesto 1500 (in spet ugotovili, da jih je 43% mnenja, da jim bo šlo prihodnje leto slabše).

    Ali bi bila meja napake za 95 interval zaupanja večja, manjša ali enaka 3 odstotnim točkam? Zakaj?

    rešitev: Meja napake bi bila večja, saj se z manjšanjem vzorca poveča standardni odklon deleža, s tem pa tudi meja napake in dolžina intervala zaupanja. Standardni odklon deleža je obratno sorazmeren z kvadratnim korenom velikosti vzorca.

29. Poročilo pravi, da so pri nacionalni raziskavi na 1500 slučajno izbranih odraslih ugotovili, da jih je 43 mnenja, da jim bo šlo naslednje leto slabše. V nadaljevanju poročila je bilo zapisano, da je meja napake 3 odstotne točke s 95 stopnjo zaupanja. Recimo, da smo rezultat 43 dobili

s podobno metodo enostavnega slučajnega vzorčenja za vse odrasle v državi New York (z 18 milijoni prebivalcev) ne pa za celotne ZDA (z 270 milijoni prebivalcev).

Ali bi bila meja napake 95 intervala zaupanja večja, manjša ali enaka 3 odstotnim točkam? Zakaj?

rešitev: Meja napake bi ostala enaka, saj ni odvisna od velikosti populacije.

30. Rezultati študentov na ACT sprejemnih testih v preteklem letu so bili normalno porazdeljeni s srednjo vrednostjo $\mu = 18.6$ in s standardnim odklonom $\sigma = 5.9$.

    Kateri interval vsebuje srednjih 95% rezultatov? Izračunamo povprečje 25 slučajno izbranih rezultatov. Kateri interval vsebuje srednjih 95% povprečij $\bar{x}$?

    rešitev:
    Srednjih 95% rezultatov vsebuje interval

    $$I = 18.6 \pm 1.96 \cdot 5.9 = (7.036, 30.164).$$

    Interval zaupanja s stopnjo zaupanja za porazdelitev vzorčnih povprečij zgradimo po znanem postopku. Ker imamo opravka z majhnim vzorcem, bomo uporabili t-statistiko z 24 stopnjami prostosti.

    $$I = 18.6 \pm 2.064 \cdot \frac{5.9}{\sqrt{25}} \equiv (16.1645, 21.0355)$$

    Srednjih %95 povprečij vsebuje ta interval.

31. Napake pri natančnih merjenjih so velikokrat normalno porazdeljene. Izkušnje kažejo, da se napake pri kontrolnih metodah spreminjajo, ko meritev ponavljamo, v skladu z normalno porazdelitvijo s srednjo vrednostjo 0 (se pravi, da postopek ne bo sistematično precenil ali podcenil dejanske razdalje) in standardnim odklonom 0.03 m. Geodet ponovi vsako meritev trikrat in za končno vrednost uporabi povprečje treh meritev. Napaka pri tej vrednosti je povprečna napaka $\bar{x}$ teh treh zaporednih meritev.

    Kakšna je porazdelitev povprečne napake $\bar{x}$, ko geodet izmeri veliko razdalj? Med katerima vrednostima leži 95% napak?

    rešitev: Povprečna napaka se porazdeljuje kot $N(0, \frac{0.03}{\sqrt{3}}) = N(0, 0.0173)$.
    Meritev ene razdalje lahko obravnavamo kot vzorec s tremi enotami. Od tod pride ocena za parameter standardne deviacije vzorčnih povprečij, ki predstavljajo meritev.
    Interval zaupanja zgradimo po znanem postopku.

    $$I = 0 \pm 1.96 \cdot \frac{0.03}{\sqrt{3}} = (-0.0339482, 0.0339482)$$

32. Pri študiji poklicnih poti upravnikov hotelov so poslali vprašalnike enostavnemu slučajnemu vzorcu 160 hotelov iz velikih ameriških hotelskih verig. Prejeli so 114 odgovorov. Povprečen čas, v katerem je teh 114 upravnikov delalo pri svojem trenutnem podjetju, je bil $\bar{x} = 11.78$ let. Ne poznamo standardnega odklona populacije σ, vzorčni standardni odklon pa je enak $s = 3.2$ leta. Ker je vzorec velik, je s blizu σ.

    Podaj 95% interval zaupanja za povprečno število let, ki so jih upravniki velikih verig preživeli pri svojem trenutnem podjetju.

rešitev: Ker ne poznamo populacijskega standardnega odklona, bomo za približek uporabili vzorčni standardni odklon. Ker je vzorec velik, je s blizu σ in lahko posledično uporabimo z-statistiko. Dobimo rezultat:

$$I = 11.78 \pm 1.96 \cdot \frac{3.2}{\sqrt{114}} = (11.1926, 12.3674)$$

33. Pri laboratorijski tehtnici je standardni odklon $\sigma = 0.001$g pri večkratnih tehtanjih. Predpostavi, da so meritve pri večkratnih tehtanjih normalno porazdeljene s srednjo vrednostjo, ki je enaka dejanski teži tehtanega predmeta. Pri treh tehtanjih nekega primerka smo dobili 3.412 3.414 3.415.

    Podaj 95% interval zaupanja za dejansko težo primerka. Kolikšna sta ocena in meja napake pri tem intervalu?

    Poišči mejo napake za 95% interval zaupanja, če tehtamo vsak primerek dvanajstkrat in ne le trikrat. Prepričaj se, da je tvoj rezultat dvakrat manjši od meje napake iz primera, ko smo vsak primerek tehtali trikrat. Razloži, zakaj lahko že brez računanja ugotovimo, da bo nova meja napake dvakrat manjša.

rešitev:
Ker imamo opravka z majhnim vzorcem, uporabimo t-statistiko.
Izračunamo vzorčno povprečje, ki ga bomo uporabili kot cenilko za populacijsko povprečje.

$$\mu_X = 3.4137$$

Standardni odklon populacije poznamo. Računajmo interval zaupanja po znanem postopku.

$$I = 3.4137 \pm 4.303 \frac{0.001}{\sqrt{3}} = (3.41122, 3.41618)$$

Če bi vsak primerek tehtali 12-krat:

$$I = 3.4137 \pm 4.303 \frac{0.001}{\sqrt{12}} = (3.41246, 3.41494)$$

Kolikokrat se je interval zmanjšal preverimo tako, da delimo razliko med zgornjo mejo in spodnjo mejo drugega intervala z razliko med zgornjo in spodnjo mejo prvega.

34. V spodnji tabeli so rezultati IQ testa 31 učenk 7. razreda iz neke ameriške šole.

114	100	104
89	102	91
114	114	103
105	108	130
120	132	111
128	118	119
86	72	111
103	74	112
107	103	98
96	112	112
93		

Pričakujemo, da bo porazdelitev blizu normalni.

- Napravi stebelni diagram porazdelitve teh 31 rezultatov. Ali na diagramu opaziš kakšne ubežnike, izrazito asimetričnost ali druga nenavadna odstopanja?

  rešitev:

7	2 4
8	6 9
9	1 3 6 8
10	0 2 3 3 3 4 5 7 8
11	1 1 2 2 2 4 4 4 8 9
12	0 8
13	0 2

  Diagram je simetričen, ubežnikov ni.

- Obravnavaj teh 31 deklic kot enostavni slučajni vzorec vseh sedmošolk iz tega okrožja. Predpostavi, da je standardni odklon rezultatov IQ testa v tej populaciji enak $\sigma = 15$. Podaj 95% interval zaupanja za povprečni rezultat v tej populaciji.

  rešitev:

  Interval zaupanja izračunamo po znanem postopku. Ker je vzorec velik in ker poznamo standardno deviacijo populacije, lahko uporabimo z-statistiko.

  Dobimo rezultat $(77.84054, 136.63946)$.

- V resnici pripadajo ti rezultati vsem sedmošolkam ene od šol iz tega okrožja. Natančno razloži, zakaj se ne moremo zanesti na interval zaupanja iz druge točke.

  rešitev:

  Vzorec iz ene šole verjetno ne predstavlja dobro celotne populacije, ker je lahko pristranski.

35. Pri NAEP testu je sodelovalo tudi 1077 žensk med 21. in 25. letom starosti. Njihov povprečni rezultat je bil 275. Predpostavi, da je standardni odklon posameznih rezultatov enak $\sigma = 60$.

- Podaj 95% interval zaupanja za povprečni rezultat μ v populaciji vseh žensk med 21. in 25. letom.

  rešitev: Interval zaupanja konstruiramo po znanem postopku.

$$I = 275 \pm 1.96 \frac{60}{\sqrt{1077}} = (271.417, 278.583)$$

- Predpostavi, da so dobili enak rezultat, $\bar{X} = 275$, za vzorec 250 žensk. Podaj 95% interval zaupanja za srednjo vrednost populacije v tem primeru.

  rešitev:

$$I = 275 \pm 1.96 \frac{60}{\sqrt{250}} = (267.562, 282.438)$$

- Predpostavi, da so pri vzorcu 4000 žensk dobili vzorčno povprečje $\bar{X} = 275$ in še enkrat izračunaj 95 interval zaupanja za μ.

  rešitev:

- Kolikšne so meje napake za vzorce velikosti 250, 1077 in 4000? Kako večja velikost vzorca vpliva na mejo napake za interval zaupanja?

rešitev: Standardni odklon vzorčnega povprečja je obratno sorazmeren s kvadratnim korenom velikosti vzorca.

36. Continue at FAPP4.4P27

2.4 13 - preverjanje domnev

1. Neko ničelno domnevo želimo testirati s petimi neodvisnimi testi na različnih vzorcih.

   Približno kolikšna naj bo stopnja značilnosti pri teh testih, da bomo z verjetnostjo 0.4 vsaj enkrat zavrnili ničelno domnevo, če je le-ta pravilna?

   rešitev: Želimo, da je verjetnost napake 1. tipa takšna, da se bo z verjetnostjo 0.4 zgodila v petih neodvisnih testih.
   Izračunajmo verjetnost, da se v petih poskusih dogodek ne zgodi nikoli z verjetnostjo 0.6.

   $$P(X = 0) = \binom{5}{0} q^5 = 0.6$$

   Rešimo za q, da dobimo $q = 0.9$ in nato za p, da dobimo 0.1.

2. Anketa med 10 študenti računalništva je pokazala, da je povprečen znesek, ki ga v semestru porabijo za prevoz do fakulete enak $\bar{x} = 249$EUR, popravljen standardni odklon pa $s =$ 30EUR. Predpostavljamo, da je ta znesek porazdeljen normalno.

   S pomočjo katere formule bomo interval zaupanja za μ pri stopnji zaupanja 0.95 najbolj pravilno izračunali?

   rešitev:
   Ker je vzorec majhen in ker ocenjujemo populacijsko povprečje s vzorčnim povprečjem, uporabimo t-statistiko.

   $$249 \pm t_{0.975}(10)\frac{30}{\sqrt{10}}$$

3. Izbrali smo vzorec iz populacije s pričakovano vrednostjo $\mu = 18.6$ in za 90% stopnjo zaupanja naračunali interval zaupanja (16.8, 18.2). Kaj to pomeni?

   rešitev: To pomeni, da smo na osnovi naključnega vzorčenja prišli do majhnega vzorčnega povprečja. Da se to zgodi, je približno 10% možnosti. 90% stopnja zaupanja pomeni, da smo prepričani, da bo 90% vseh intervalov izračunanih iz naključnih vzorcev vsebovalo ustrezni parameter.

4. Pri preverjanju domnev z α označimo verjetnost napake prve vrste in z β verjetnost napake druge vrste.

   Ali velja naslednja trditev? Če pri testu povečamo α, se β vedno zmanjša.

   rešitev: Napaka prve vrste je zavrnitev ničelne hipoteze, čeprav ta drži. Napako druge vrste pa storimo, če ne završemo ničelne hipoteze, čeprav ni pravilna.
   Če povečamo verjetnost, da naredimo napako prve vrste, se verjetnost, da bomo naredili napako druge vrste zmanjša, saj smo povečali verjetnost, da bomo hipotezo H_0 zavrnili v vsakem primeru (znižali smo prag zavrnitve).

5. Predpostavljamo, da ima določena lastnost delež $\pi = 0.6$. Lahko bi izbrali podatke, konstruirali 95% interval zaupanja za π in preverili, če je 0.6 na tem intervalu. Alternativno bi lahko preverjali domnevo $H_0 : \pi = 0.6$ proti dvostranski alternativi pri $\alpha = 0.05$.

   Ali velja naslednja trditev? Čeprav sta pristopa različna, vseeno predstavljata podoben vpogled v točnost naše predpostavke.

   rešitev: Da, to drži.

6. Pri preverjanju domev bi radi dosegli, da sta verjetnosti napake 1. vrste in napake 2. vrste kar se da majhni.

   Ali to drži? Čeprav sta si ta cilja v splošnem nasprotujoča, lahko naredimo napredek pri obeh, če nam uspe povečati veliksot vzorca n.

   rešitev: Da. povečanje je ena od stvari, ki jih lahko naredimo (so pod našo kontrolo), da zmanjšamo verjetnosti obeh napak, saj s tem zmanjšamo standardno deviacijo vzorčnega povprečja.

7. P-vrednost za χ^2-test enakomerne porazdelitve je vedno izračunana z uporabo enega repa (intervala), tj. P-vrednost $= P(\chi^2 > T.S.)$, kjer je T.S. vrednost, ki jo izračunamo na osnovi vzorca.

   rešitev: Da. To je značilno za χ^2 test.

8. Ali velja naslednja trditev? Preverjanju domneve za enakost povprečij ne moremo zaupati, če se velikost vzorcev razlikujeta za faktor 5 ali več.

   rešitev: To ne drži vedno. Če sta oba enostavna naključna vzorca neke smiselno velike velikosti, lahko drži.

9. Naloga i2-30

   Preverjaš ničelno domnevo, da je vrednost deleža $\pi = \frac{1}{2}$ proti dvostranski alternativi z uporabo 5% stopnje zaupanja (α). Uporabiš normalno porazdelitev za aproksimacijo vzorčnega deleža p in izračunaš vrednost $z = 2$.

   rešitev: Preverimo z tabelo za verjetnost, da dobimo v standardni normalni porazdelitvi vrednost oddaljeno 2 standardni deviaciji stran od povprečja. Dobimo vrednost 0.9772. Torej je verjetnost, da smo takšno vrednost dobili po naključju in da ničelna hipoteza drži 0.0228.

10. Ali drži naslednja trditev?

    Kritična vrednost, ki jo uporabimo za konstrukcijo kritičnega območja za povprečje populacije, kadar ne poznamo njenega standardnega odklona, je odvisa od stopnje zaupanja in velikosti vzorca.

    rešitev: To drži. Spomnimo se enačbe za konstrukcijo intervalov zaupanja. Uporabimo stopnjo zaupanja ter aproksimiramo populacijski standardni odklon z standardnim odklonom vzorca.

11. Ali velja naslednja trditev?

    P-vrednost (ali ugotovljena stopnja značilnosti za določen statistični test) je verjetnost (ob predpostavki, da drži H_0), da ugotovimo vrednost statistike, ki je vsaj toliko v protislovju s H_0 in podpira H_a, kot tisto, ki je izračunana iz vzorčnih podatkov.

    rešitev: To seveda vedno drži.

12. Na ruleti v Las Vegasu imamo 18 rdečih, 18 črnih in 2 zeleni števki. Zavrtimo jo 228 in uporabimo Hi-kvadrat test, da preverimo domnevo, če so barve porazdeljene tako kot pričakujemo.

rešitev: TODO

13. Ali velja naslednja trditev? Ničelno domnevo moramo zavrniti, kadar je $p - vrednost$ večja od stopnje zaupanja.

rešitev: To seveda ne drži. p-vrednost je verjetnost, da bi ob predpostavki, da H_0 drži, dobili vzorčne statistike, kot smo jih.
Stopnja zaupanja pa je vrednost $1-\alpha$. Ničelno domnevo zavrnemo, če p-vrednost pade zunaj tega območja.

14. Ali velja naslednja trditev? Za preverjanje domneve $H_0 : \mu_1 = \mu_2$ je testna statistika enaka vsoti ustreznih vzorčnih povprečij, tj $\bar{x}_1 + \bar{x}_2$.

rešitev: Ne drži. Testna statistika je enaka razliki vzorčnih povprečij.

15. Ko preverjamo domnevo in pri tem uporabljamo vrednost π_0 iz ničelne domneve in p kot vzorčni delež, je za oceno standardne napake porazdelitve vzorčnega deleža najbolje vzeti vrednost $\sqrt{\frac{p(1-p)}{n}}$.

rešitev: TODO zakaj ne drži?

16. Pri testiranju z enostranskim z-testom je vrednost testne statistike enaka 2.35.

rešitev: Preverimo z-tabelo za delež 99.5%. Izkaže se, da je ta testna statistika še znotraj območja zaupanja.

17. Pri preverjanju domnev z α označimo verjetnost napaka prve vrste in z β verjetnost napake druge vrste. Če pri testu povečamo α, kaj se zgodi z β?

rešitev: Verjetnost napake 2. tipa se zmanjša, saj bomo z zvečanjem α povečamo prag za zavrnitev ničelne hipoteze. Hipotezo bomo tako bolj verjetnost zavrnili v vsakem primeru. S tem se intuitivno zmanjša tudi verjetnost, da jo ne bomo zavrnili, čeprav je napačna.

18. Na vzorcu velikosti 35 dobimo vzorčno povprečje $\bar{x}$ in vzorčni standardni odklon s. Pri preverjanju domneve $H_0 : \mu = \mu_0$ nismo zavrnili. Uporabili smo Z-test in izračunali p vrednost 0.0667. Kaj se zgodi s P-vrednostjo, če jo izračunamo za vzorec velikosti > 35, pri čemer se $\bar{x}$ in s nista spremenila?

rešitev: Z večanjem vzorca se verjetnosti napak 1. in 2. tipa zmanjšajo, saj se zmanjša standardna deviacija vzorčnega povprečja. Torej se P-vrednost zmanjša.

19. Naj bo X enakomerno porazdeljena slučajna spremenljivka z zalogo vrednosti $2, 3, 4, 5$. Enkrat izvedemo poskus. Vrednost X je enaka čemu?

rešitev: Enaka je natanko eni izmed vrednosti v domeni te slučajne spremenljivke (zalogi vrednosti).

20. Plastenke ribezovega soka imajo oznako, da je prostornina tekočine 2 litra. Kupcu se zdi, da je tekočine manj, zato načrtuje testiranje. Kaj bi predstavljalo napako 1. vrste?

rešitev: Kupec zaključi, da vsebujejo plastenke manj kot 2 litra, pri čemer je povprečje v resnici 2 litra. Torej zavrne ničelno domnevo, čeprav je ta v resnici pravilna.

21. Če povečamo stopnjo tveganja α, potem bodo možnosti (1) napake 1. vrste, (2) napake 2. vrste, (3) moč testa (v tem vrstem redu)

rešitev: poveča, zmanjša, poveča

22. Preverjali smo ničelno domnevo H_0 napram alternativni H_a s stopnjo tveganja $\alpha = 0.05$. DObili smo, da je P-vrednost testne statistike enaka 0.01. Kakšen je rezultat testa?

rešitev: Ničelno hipotezo H_0 zavrnemo in sprejmemo alternativno hipotezo H_a.

23. Ali velja naslednja trditev? Hi-kvadrat test za preverjanje normalne (ali enakomerne) porazdelitve je lahko bodisi enostranski ali dvostranski.

rešitev: To ne drži. Hi kvadrat test je le enostranski.

24. Natančnost laboratorijske tehtnice preverjamo z utežjo, ki tehta 1 gram. v štirih poskusih dobimo meritve: 0.96g, 1.02g, 1.01g, 0.98g.

    Pri sotpnji tveganja 0.05 želimo preveriti domnevo, da je tehtnica natančna za maso 1g. Koliko bo kritično območje za test povprečja?

rešitev: Najprej izračunamo povprečje in varianco ter standardno deviacijo meritev.

$$\mu = 0.99$$
$$\sigma = 0.0274$$

Interval zaupanja izračunamo po znanem postopku.

$$I = 0.99 \pm 3.182 \cdot \frac{0.0274}{2} \equiv [0.946407, 1.03359]$$

25. Ali velja naslednja trditev? Če je stopnja značilnosti enaka 0.05 in P-vrednost enaka 0.041, potem ničelno domnevo zavrnemo.

rešitev: Ne drži vedno. ZAKAJ?? TODO

26. Ko preverjamo domnevo in pri tem uporabljamo vrednost π_0 iz ničelne domneve in p kot vzorčni delež, je za oceno standardne napake porazdelitve vzorčnega deleža najbolje vzeti vrednost $\sqrt{n\pi_0(1-\pi_0)}$.

rešitev: To ne drži. Standardno napako porazdelitve vzorčnega deleža ocenimo kot $\sqrt{\frac{p\cdot(1-p)}{n}}$.

27. Ali velja naslednja trditev? Če preverjamo ničelno domnevo $H_0 : \mu = 10$ pri stopnji značilnosti $\alpha = 0.05$ in je povprečje vzorca enako 12, potem H_0 ne bomo nikoli zavrnili, če bomo za alternativno domnevo izbrali $H_a : \mu < 10$.

rešitev: To vedno drži.

28. Proizvajalec trdi, da je (največ) 1% njegovih izdelkov pokvarjenih. Za preizkus te trditve izberemo 55 naključnih izdelkov iz serije velikosti $N = 1000$ (brez vračanja in sprejmemo serijo, če vzorec nima več kot ene pokvarjene enote.

    Zapiši ničelno in alternativno hipotezo.

rešitev: $H_0 : p = 0.01, \quad H_1 : p > 0.01$

29. Ali velja naslednja trditev? Če ničelno domnevo zavrnemo pri stopnji značilnosti $\alpha = 5\%$, potem jo zavrnemo tudi pri stopnji značilnosti $\alpha = 10$.

rešitev: Kritično območje se pri povečanju α poveča. Odgovor je torej vedno da.

30. Ali velja naslednja trditev? Če preverjamo ničelno domnevo $H_0 : \mu = 10$ pri stopnji značilnosti $\alpha = 0.05$ in je vzorčno povprečje enako 12, potem bomo domnevo H_0 vedno zavrnili, če bomo za alternativno domnevo izbrali $H_a : \mu > 10$.

rešitev: Kritično območe je odvisno od stopnje značilnosti α, testna statistika pa še od odklona in velikosti vzorca, saj je enaka $\frac{(\bar{X}-\mu_0)}{\frac{\sigma}{\sqrt{n}}} = \frac{2\sqrt{n}}{\sigma}$. TO nadalje pomeni, da lahko z ustrezno izbiro velikosti vzorca ali pa odklona dosežemo tudi FTR.

31. Preverjanje domnev

- Opiši splošen postopek preverjanja domneve

  rešitev: Glej zapiske za teoretični del.

- Kaj je to zavrnitveni kriterij?

  rešitev: Izberemo stopnjo tveganja α (običajno 10%, 5% ali 1%) glede na to ali gre za enostranski ali dvostranski test določimo kritično območje. Če pade P-vrednost v kritično območje, ničelno hipotezo zavrnemo.

- Kaj je stopnja značilnosti?

  rešitev: Stopnja značilnosti (signifikantnosti) je največji α, ki ga je vodja eksperimenta pripravljen sprejeti (zgornja meja za napako 1. vrste). P-vrednost (ali ugotovljena bistvena stopnja za določeni statistični test) je verjetnost (ob predpostavki, da drži H_0), da ugotovimo vrednost testne statistike, ki je vsaj toliko v protislovju s H_0 in podpira H_a kot tisto, ki je izračunana iz vzorčnih podatkov. Razlaga P-vrednosti:

  – Izberi največjo vrednost za α, ki smo jo pripravljeni tolerirati.

  – Če je P-vrednost testa manjša kot maksimalna vrednost parametra α, potem zavrni ničelno hipotezo.

- Pojasni razliko med napako 1. in 2. vrste.

  rešitev: Pri napaki prve vrste zavrnemo ničelno hipotezo, ki je v resnici pravilna. Verjetnost, da se to zgodi je α. Napaka druge vrste pa je, da ne zavrnemo ničelne hipoteze, ki je v resnici napačna. Verjetnost, da se to zgodi, je β. Če povečamo α, se verjetnost napake druge vrste β zmanjša.

- Slučajna spremenljivka je porazdeljena normalno z $\sigma = 0.3$. Vzorec je:

  rešitev:

  5.93, 6.08, 5.86, 5.91, 6.12.

  Testiraj hipotezo, da je $\mu = 6.15$ proti $\mu < 6.15$ z $\alpha = 0.05$.

rešitev: Ocenino porazdelitev vzorčnih povprečij. Povprečno vrednost ocenimo kar iz vzorčnega povprečja. Vzorčno standardno deviacijo ocenimo iz standardne deviacije populacije, ki jo delimo s kvadratnim korenom velikosti vzorca.

$$\mu_{\bar{x}} = 5.98$$

Izračunajmo interval zaupanja.

$$I = 5.98 \pm 2.353 \cdot \frac{0.3}{\sqrt{5}} \equiv (5.66431, \infty)$$

FTR H_0.

- Slučajna spremenljivka je porazdeljena normalno z $\sigma = 6$. Vzorec je:

  137.4; 140.1; 134.9; 135.9; 140.9; 138.4; 136.5; 137.6 140.3

  Testiraj hipotezo, da je $\mu = 142$ proti $\mu < 142$ z $\sigma = 0.01$

  rešitev: Konstruirajmo interval zaupanja po znanem postopku.

  $$I = 138 \pm 2.896 \cdot \frac{6}{3} \equiv (132.208, \infty)$$

  FTR H_0.

32. Za preverjanje domnev o nekem parametru g poznamo testno statistiko T.S., za katero vemo, da je ob veljavni ničelni domnevi $H_0 : g = g_0$ porazdeljena standardno normalno. Pri določeni vrednosti g_0 smo na enostavnem slučajnem vzorcu dobili vrednost testne statistike $T.S. = -1.75$. Pri stopnji tveganja $\alpha = 0.05$ za vsako alternativno domnevo

    $H1 : g > g_0$

    določite kritično območje (pravilnost tega bom štel za bonus) in preverite veljavnost naslednje trditve. Domneve H_0 ne zavrnemo proti alternativni domnevi H_1.

    rešitev: V z-tabeli preverimo delež porazdelitve, ki se nahaja več kot 1.75 standardnih deviacij levo od povprečja. Dobimo vrednost 0.0401.

33. ??

34. ??

35. Proizvajalec trdi, da je (največ) 1% njegovih izdelkov pokvarjenih. Za preizkus te trditve izberemo 55 naključnih izdelkov iz serije velikosti N=1000 (brez vračanja) in sprejmemo serijo, če vzorec nima več kot ene pokvarjene enote.

    Kateri par domnev dobimo pri tesetu, če s p označimo delež pokvarjenih izdelkov?

    rešitev: Že odgovorjeno - glej nazaj.

36. Ali velja naslednja trditev? Če je P-vrednost za enostranski test v desno 0.042, potem P-vrednost za dvostranski test za isti vzorec ni enaka 0.084.

    rešitev: To drži. P-vrednost ostane enaka. Če enostranski test spremenimo v dvostranski, se spremeni le kritično območje (se prepolovi za enak α).

37. Ko preverjamo domnevo in pri tem uporabljamo vrednost π_0 iz ničelne domneve in p kot vzorčni delež, je za oceno standardne napake porazdelitve vzorčnega deleža najbolje vzeti vrednost $\sqrt{\pi_0(\pi_0-1)/n}$.

rešitev: Na to vprašanje smo že odgovorili - glej nazaj.

38. Ali velja naslednja trditev? Kritična vrednost, ki jo uporabimo za konstrukcijo kritičnega območja za povprečje populacije, kadar ne poznamo njenega standardnega odklona, je odvisna od stopnje zaupanja in velikosti vzorca.

rešitev: To drži. standardni odklon ocenimo iz vzorčnega standardnega odklona in je obratno sorazmeren kvadratnemu korenu velikosti vzorca.

39. Ali velja naslednja trditev? P-vrednost (ali ugotovljena bistvena stopnja značilnosti za določen statistični test) je verjetnost (ob predpostavki, da drži H_0), da ugotovimo vrednost testne statistike, ki je vsaj toliko v protislovju s H_0 in podpira Ha kot tisto, ki je izračunana iz vzorčnih podatkov.

rešitev: To drži (definicija P-vrednosti)

40. Ali drži naslednja trditev? Na ruleti v Las Vegasu imamo 18 rdečih, 18 črnih in 2 zeleni števili. Zavrtimo jo 228 in uporabimo Hi-kvadrat test, da preverimo domnevo, če so barve porazdeljene tako kot pričakujemo. Rdeča števila pričakujemo 108 krat.

rešitev: Izračunamo verjetnost, da se kroglica ustavi na rdečem številu in dobimo rezultat $p = 0.4737$. Izračunamo pričakovano vrednost rdečih števil: $E(X) = 0.4737 \cdot 228 = 108$. Trditev torej drži.

41. Plestenke ribezovega soka imajo oznako, da je prostornina tekočine 2 litra. Kupecu se zdi, da so je tekočine manj, zato načrtuje testiranje. Kaj pomeni pri tem napaka 1. vrste?

rešitev: Kupec zaključi, da vsebujejo plastenke manj kot 2 litra, pri čemer je povprečje v resnici 2 litra. Torej zavrne ničelno hipotezo, čeprav je ta v resnici pravilna.

42. Če povečamo stopnjo tveganja α, potem bodo možnosti (1) napake 1. vrste, (2) napake 2. vrste, (3) moč testa (v tem vrstem redu)...

rešitev: Večje, manjše, večja

43. Na vzorcu preverjamo domnevo s testom T1, ki nam da P-vrednost 8%. Nato na istem vzorcu izvedemo še en test, T2, ki nam vrne P-vrednost 3%. Recimo, da si predpišemo $\alpha = 5$. Ali domnevo na podlagi teh dveh testov zavrnemo?

rešitev: Odgovor je Da, saj je $0.03 < 0.05$.

Psiholog poroča, da je bil "v našem vzrcu etnocentrizen značilno večji med ljudmi, ki so hodili v cerkev, kot med tistimi, ki niso". Pojasni, kaj to pomeni.

rešitev: S statističnim testom smo pokazali, da je verjetnost, da se povprečni stopnji etnocentrizma med obema vzorcema ne razlikujeta, majhna (npr. manjša od 5%). Torej da je zelo majhna verjetnost, da se je to zgodilo po naključju.

44. Cigaretna industrija se je odločila, da morajo izgledati modeli, ki oglašujejo njene izdelke, stari vsaj 25 let. Vendar pa so raziskave pokazale, da potrošniki menijo, da je med temi

modeli veliko mlajših. Spodaj je citat iz raziskave, v kateri so ljudi spraševali, če menijo, da razne znamke cigaret uporabljajo pri oglaševanju modele različnih starosti: Statistična analiza je pokazala, da je vrsta cigaret zelo značilna, kar kaže na to, da povprečna ocenjena starost modelov ni enaka pri obravnavanih 12 znamkah. Kot smo lahko videli, so nekatere znamke, na primer Lucky Strike Lights, Kool Milds in Virginia Slims, uporabljale domnevno mlajše modele. [Vir: M. B. Maziz et al., Perceived age and attractiveness of models in cigarette advertisements, Journal of Marketing 56 (January 1992): 22-37.]

Razloži nekomu, ki ne zna statistike, kaj pomeni ”zelo značilna” in zakaj je to dober dokaz razlik med oglaševalci teh znamk, čeprav so vprašani videli le del oglasov.

rešitev: S statističnim testom smo pokazali, da je verjetnost, da bi bila povprečna ocenjena starost modelov od znamke do znamke enaka, zelo majhna. Torej je zelo malo verjetno, da bi takšne rezultate dobili po naključju.

45. Kako na valjenje jajc vodnega pitona vpliva temperatura kačjega gnezda? Raziskovalci so razdelili novo znesena jajca v tri skupine glede na temperaturo: vroče, nevtralno in mrzlo. V prvi skupini so posnemali toploto, ki jo običajno priskrbi samica pitona, pri tretji skupini pa so simulirali odsotnost samice. V spodnji tabeli so podatki o številih jajc in njihovih usodah. (Vir: R. Shine, T. R. L. Madsen, M. J. Elphick, P. S. Harlow, The influence of nest temperatures and maternal brooding on hatchling phenotypes in water pythons, Ecology 78 (1997): 1713-1721.)

	Št. jajc	Izvaljeni
Hladno	27	16
Nevtralno	56	38
Vroče	104	75

Izdelaj dvosmerno tabelo temperature v odvisnosti od izida ter za vsako skupino izračunaj delež jajc, iz katerih so se izlegli pitoni. Raziskovalci so pričakovali, da bo rzlo okolje zmanjšalo število izvaljenih jajc. Ali podatki podpirajo njihova pričakovanja?

rešitev:

	Neizvaljeni	Izvaljeni	Skupaj
H*ladno*	11	16	27
N*evtralno*	18	38	56
V*roe*	29	75	104

Preveri hipotezo po znanem postopku. Lahko uporabiš tudi Hi-kvadrat test, kjer pričakovane vrednosti dobiš v vrstici Nevtralno.

TODO

46. V študiji vpliva kadilskih navad staršev na navade srednješolcev so raziskovalci anketirali dijake iz osmih srednjih šol v Arizoni. Rezultati so povzeti v spodnji tabeli. (Vir: S. V. Zagona, ed., Studies and Issues in Smoking Behavior, University of Arizona Press, Tucson, 1967, str. 157-180.)

Študent	kadi	ne kadi
Oba starša kadita	400	1380
Eden od staršev kadi	416	1823
Nobeden od staršev ne kadi	188	1168

Opiši povezavo med kadilskimi navadami staršev in njihovih otrok, tako da izračunaš različne deleže in jih primerjaš. Nato povzemi rezultate še s preprostimi besedami.

rešitev: Podobno kot prejšna naloga.

47. Strelno orožje je drugi najpogostejši vzrok smrti iz nezdravstvenih razlogov (prvi so motorna vozila). V tabeli so zbrani podatki za Milwaukee, Wisconsin, med leti 1990 in 1994. Primerjati želimo vrste strelnega orožja, ki so bile uporabljene v umorih in v samomorih. (Vir: S. W. Hargarten et al., Characteristics of firearms involved in fatalities, Journal of American Medical Association, 275 (1996): 42-45.)

    Predpostavljamo, da bodo šibrovke in lovske puške bolj pogosto uporabljene pri samomorih, ker jih ima veliko ljudi doma za lov. Kaj pravi o tem zgornja tabela?

    rešitev: TODO

48. Opiši splošni postopek preverjanja domneve (testiranja hipoteze)

    rešitev: Glej zbirko odgovorov na teoretična vprašanja.

49. Kaj je to zavrnitveni kriterij?

    rešitev: Glej zbirko odgovorov na teoretična vprašanja.

50. Kaj je stopnja značilnosti?

    rešitev: Glej zbirko odgovorov na teoretična vprašanja.

51. Kaj je P-vrednost?

    rešitev: Glej zbirko odgovorov na teoretična vprašanja.

52. Pojasni razliko med napako 1. in 2. vrste.

    rešitev: Glej zbirko odgovorov na teoretična vprašanja.

53. Proizvajalec merilcev električne moči, ki se uporabljajo za uravnavanje pragov energije pri podatkovno-komunikacijskih sistemih, trdi, da ob normalno delujoči proizvodni liniji proizvede največ 10% nedelujočih merilcev. Trgovec je pravkar dobil pošiljko 25 omenjenih merilcev. Recimo, da želi preveriti domnevo $H_0: \quad p = 0.10$ napram $H1: \quad p > 0.10$, kjer je p pravi delež pokvarjenih merilcev. Pri testu uporabi $y \geq 6$ za prag zavrnitve (y je seveda pravo število pokvarjenih merilcev v pošiljki).

    - Določi vrednost stopnje značilnosti testa (signifikantnost) α, ki ji pravimo tudi verjetnost napake 1. vrste.

rešitev: Za prag zavrnitve vzamemo primer, ko je od 25 merilcev več kot šest okvarjenih. Da dobimo stopnjo značilnosti (signifikantnost) testa, izračunajmo verjetnost, da je ob veljavni ničelni hipotezi okvarjenih več kot šest merilcev.

$$X \sim B(25, 0.1)$$

$$P(X \geq 6) = 1 - P(X \leq 5) \approx 1 - 0.9666 = 0.0334$$

- Izračunaj verjetnost napake 2. vrste β v primeru, ko je $p = 0.2$.

rešitev:

Napako 2. vrste bomo zagrešili, če ne bomo zavrnili ničelne hipoteze, čeprav je v resnici napačna.

$$P(X \leq 5|p = 0.20) = 0.6167$$

- Koliko je moč testa za to vrednost p?

rešitev:

Moč testa je verjetnost, da bomo zavrnili ničelno hipotezo, če je ta napačna. To je nasprotni dogodek od dogodka, da sprejmemo napačno ničelno hipotezo. Torej velja, da je moč testa enaka $1 - \beta = 1 - 0.6167 = 0.3833$.

- Izračunaj verjetnost napake 2. vrste β v primeru, ko je $p = 0.4$.

rešitev:

$$P(X \leq 5|p = 0.40) = 0.0294$$

- Koliko je moč testa za to vrednost p?

rešitev: Moč testa je $1 - 0.0294 = 0.9706$

54. S pomočjo računalniške simulacije so raziskovali učinek napak na strojih na performanse proizvodnega sistema (Industrial Engineering, Aug. 1990). Študija se je osredotočila na sistem z enim strojem. Povprečni čas med začetkom dveh procesiranj je 1.25 minute pri konstantnem času procesiranja 1 minuta. Naprava je pokvarjena 10 časa. Po $n = 5$ neodvisnih simulacijah dolžine 160 ur je povprečna produktivnost na teden (40-urni delavnik) $\bar{y} = 1908.8$ izdelkov. Za sistem brez napak je povprečna produktivnost 1920 izdelkov.

    Če predvidevamo, da je standardna deviacija 5 simulacij $s = 18$ izdelkov, preverite domnevo, da je resnična povprečna produktivnost manjša od 1920 izdelkov. Preverjajte pri značilnosti $\alpha = 0.05$.

    rešitev:

    Izračunajmo kritično vrednost t-statistike. Dobimo rezultat $-t_{0.05}(4) = -2.132$

    $$T.S. = \frac{1908.8 - 1920}{\frac{18}{\sqrt{5}}} = -1.39.$$

    Vrednost testne statistike ne pade v kritično območje. Ničelne hipoteze ne moremo zavrniti.

55. Rezultati druge raziskave o nacionalnem zdravju in prehranjevanju v ZDA so pokazali, da imajo ljudje v starosti med pol leta in 74 let v krvi povprečno koncentracijo svinca $14\mu g/dl$ (Analytical Chemistry, feb. 1986). Poleg tega so ugotovili, da imajo črnski otroci v starosti do pet let znatno višje koncentracije od ostalih.

naključnem vzorcu 200 črnskih otrok v starosti pod pet let je bila ugotovljena povprečna koncentracija svinca v krvi $21\mu g/dl$ s standardnim odklonom $10\mu g/dl$. Je to dovolj, da lahko trdimo, da je pravo povprečje pri črnski populaciji pod pet let res večje od $14\mu g/dl$? Preverjajte z vrednostjo $\alpha = 0.01$.

rešitev: Ker je vzorec velik, lahko uporabimo z-statistiko. Iz tabele preberemo kritično vrednost $z = 2.33$. Izračunajmo z-statistiko vzorca:

$$z = \frac{21 - 14}{\frac{10}{\sqrt{200}}} = 9.899$$

Ničelno hipotezo zavrnemo.

56. Dovoljena koncentracija PCB-ja (nevarna substanca) v vodi po standardu, ki ga je postavila EPA, je 5 delcev na milijon. Večji proizvajalec PCB-ja, ki se uporablja za električno izolacijo, spušča manjše količine PCB-ja skupaj z odpadno vodo. Uprava tovarne je izdala navodila, da je treba ustaviti proizvodnjo, če povprečna koncentracija PCB-ja v odplakah preseže 3 delce/milijon. Analiza 50 naključnih vzorcev odpadne vode je pokazala naslednje rezultate (v enotah delcev na milijon):

$\bar{y} = 3.1$ in $s = 0.5$

- Ali so zgornji statistični rezultati analize dovoljšen dokaz da je proizvodnjo potrebno ustaviti? Uporabite $\alpha = 0.01$.

  rešitev: Ker je vzorec velik, lahko uporabimo z-statistiko. Poiščimo kritično vrednost v tabeli: $z = 2.33$.
  $T.S. = \frac{3.1-3}{\frac{0.5}{\sqrt{50}}} = 1.414$
  Ničelne hipoteze ne moremo zavreči.

- Če bi bili vi menedžer tovarne, ali bi uporabili večjo ali manjšo vrednost za α v testu pod prvo točko? Pojasnite!

  rešitev: Ker hoče vodja tovarne narediti zaključek "ničelne predpostavke ne moremo zavreči", si želi male vrednosti za α. To pa naredi zavrnitev ničelne predpostavke za zalo težko nalogo.

57. Vrtanje "globokih lukenj" je družina procesov, ki omogočajo vrtanje lukenj, ki so vsaj desetkrat globlje kot je premer svedra. Eden bistvenih problemov pri globokem vrtanju je zastajanje izvrtanih okruškov v utorih svedra. Izveden je bil eksperiment, s katerim so preučili uspešnost globokega vrtanja v primeru, ko se okruški sprijemajo (Journal of Engineering for Industry, maj 1993). Globina izvrtanih lukenj pri 50 vrtanjih je v povprečju znašala $y = 81.2$mm s standardno deviacijo vzorca $s = 50.2$mm.

Preverjajte domnevo, da se pravo povprečje izvrtanih lukenj μ razlikuje od 75mm. Uporabite stopnjo značilnosti $\alpha = 0.01$

rešitev: Ker imamo velik vzorec (50 lukenj), lahko uporabimo z-statistiko. Iz tabele preberemo kritično vrednost $z = 2.57$ (dvostranski test).

$$T.S. = \frac{81.2 - 75}{\frac{50.2}{\sqrt{50}}} = 0.873319$$

Ničelne hipoteze ne moremo zavreči.

58. Inštitut za okolje in tehnologijo je izdal študijo o onesnaženosti zemlje na Nizozemskem. Skupno so zbrali, posušili in analizirali za prisotnost cianida 72400 gramov vzorcev zemlje. Z infrardečo mikroskopsko metodo so merili koncentracijo cianida v miligramih na kilogram zemlje v vzorcih.

    Povprečna koncentracija cianida v vzorcih je bila $y = 84mg/kg$, standardna deviacija pa $s = 80mg/kg$. Uporabite to informacijo za preverjanje domneve, da je resnično povprečje koncentracije cianida v zemlji na Nizozemskem manjše od $100mg/kg$ pri stopnji značilnosti $\alpha = 0.10$.

    rešitev: $H_0 : \mu = 100mg/kg$, $H_a = \mu < 100mg/kg$.
    Iz tabele preberemo kritično z-vrednost $z = -1.28$
    Računamo testno statistiko:

    $$T.S. = \frac{84 - 100}{\frac{80}{\sqrt{72}}} = -1.697$$

    Ta vrednost je nižja od kritične vrednosti in pade v kritično območje. Ničelno hipotezo zavrnemo. To pomeni, da je dovolj statističnih pokazateljev (za $\alpha = 0.10$), da dejansko povprečje stopnje cianida v zemlji na nizozemskem pade pod 100mg/kg.

59. Ali tekmovanje med različnimi odseki za raziskovanje in razvoj (R&R) v sklopu ameriškega ministrstva za obrambo, ki delajo na istem projektu, izboljša produktivnost? Za odgovor na to vprašanje so raziskali produktivnost pri 58 projektih, dodeljenih večim oddelkom, ter pri 63 projektih, dodeljenih enemu samemu oddelku (IEEE Transactionson Engineering Management, feb. 1990). Glede na rezultate so dobili povprečen faktor uspešnosti pri prvih (tekmovalnih) projektih enak 7.62, pri drugih (netekmovalnih) pa 6.95.

    - Postavite ničelno in alternativno domnevo za odločanje, ali povprečje uspešnosti pri tekmovalnih projektih presega povprečje uspešnosti pri netekmovalnih projektih.

      rešitev: Ničelna hipoteza $\mu_{competitive} - \mu_{noncompetitive} = 0$, alternativna hipoteza $\mu_{competitive} - \mu_{noncompetitive} > 0$.
    - Določite območje zavrnitve pri vrednosti $\alpha = 0.05$.

      rešitev: Zavrni ničelno domnevo, če je T.S. p-vrednost manjša ali enaka 0.05.
    - Izkazalo se je, da pri zgornjem testu p-vrednost leži v intervalu med 0.02 in 0.03. Kakšen je pravilni zaključek?

      rešitev: Zavrnemo ničelno predpostavko, kar pomeni, da je dovolj statističnih pokazateljev (za $\alpha = 0.05$) o tem, da je povprečje produktivnosti pri tekmovalnih projektih dodeljenih več oddelkom večje od povprečja produktivnosti pri projektih dodeljenih enemu samemu oddelku.

60. Inštitut za okolje in prostor je narediš študijo o insekticidih, ki se uporabljajo v nasadu orhidej v dolini San Joaquin v Kaliforniji. Zbrali so vzorce zraka v nasadu in jih testirali vsak dan v obdobju najbolj intenzivnega škropljenja. V spodnji tabeli so prikazane količine oksonov in thionov v zraku (v ng/m^3) in razmerje med količino thionov in oksonov.

Datum	Vreme	Thioni	Oksoni	oksoni/thioni
15. jan.	megla	38.2	10.3	.270
17.	megla	28.6	6.9	.241
18.	megla	30.2	6.2	.205
19.	megla	23.7	12.4	.523
20.	megla	62.3	–	vzorec izgubljen
20.	jasno	74.1	45.8	.618
21.	megla	88.2	9.9	.112
21.	jasno	46.4	27.4	.591
22.	megla	135.9	44.8	.330
23.	megla	102.9	27.8	.270
23.	oblačno	28.9	6.5	.225
25.	megla	46.9	11.2	.239
25.	jasno	44.3	16.6	.375

Primerjaj povprečje razmerij med oksoni in thioni v meglenih in jasnih/oblačnih pogojih v nasadu orhidej z uporabo domneve. Uporabi stopnjo značilnosti $\alpha = 0.05$.

rešitev: TODO

61. Tovarna bi rada ugotovila kateri od naslednjih dveh virov energije – plin in elektrika – bo dal cenejšo energijo. Ena mera za ekonomično proizvodnjo energije (v angl. se imenuje plant investment perdelivered quad) je izračunana tako, da vzamemo vsoto vloženega denarja (dolarji) v določen vir s strani neke tovarne in jo delimo s količino energije (v kvadrilionih britanskih termalnih enot). Manjši kot je kvocijent manj plača tovarna za dobavljeno elektriko. Izbran je bil naključni vzorec 11ih tovarn, ki uporabljata električno energijo in 16ih tovarn, ki uporabljajo plin. Nato je bil izračunan zgoraj omenjeni kvocijent za vsako tovarno in podatki vnešeni v tabele, sledi pa še izpis iz MINITABa za analizo podatkov.

    - Ali ti podatki predstavljajo dovolj dober pokazatelj za stopnjo značilnosti $\alpha = 0.05$, ki kaže na razliko med povprečjem kvocientov, ki uporabljajo plin in tistimi, ki uporabljajo elektriko?

      rešitev: TODO

    - Kakšne predpostavke so potrebne, da bo zgornji postopek veljaven? Preveri, če so te predpostavke smiselno izpolnjene. Kako to vpliva na upravičenost rezultata iz prve točke?

      rešitev: TODO

62. Raziskovalci so naredili eksperiment za določitev vpliva puščavskih granivorjev (živali, ki jedo semena) na gostoto in porazdelitev semen v prsti (Ecology, Dec. 1979). Določene vrste puščavskih glodalcev delajo zaloge semen na površju zemlje, zato je bil eksperiment zasnovan tako, da so raziskovali ali zaradi takšnih zalog semen v povprečju na tistem območju iz semen zraste več mladih rastlinic kot na sosednjih kontrolnih območjih. Locirali so 40 majhnih območij, kjer so glodalci kopičili semena in jih pokrili z mrežo, da so glodalcem preprečili ponoven dostop. Zamrežili so tudi sosednja območja za kontrolo. Potem so opazovali število semen, ki je vzklilo na zamreženih območjih. V spodnji tabeli je povzetek zbranih podatkov.

Ali so rezultati eksperimenta zadosten dokaz (pri stopnji značilnosti $\alpha = 0.05$), da je povprečno število vzkaljenih semen na območjih, kjer so jih kopičili glodalci bistveno večje kot na kontrolnih območjih?

rešitev: TODO

63. Percepcija govora pretežno gluhih oseb se zanaša predvsem na branje z ust, tj., zaznavanje pogovornega jezika z opazovanjem artikuliranih gibov, izrazov na obrazu ter gest sogovornika. Ali se da percepcijo govora izboljšati tako, da bralcu glasu vizulano predstavimo podatek o poudarku zlogov. Da bi raziskali ta fenomen je 10 oseb z normalnim sluhom sodelovalo v eksperimentu pri katerem so morali ustno ponoviti zvočno predvajane stavke, katerih informacijo niso videli na video-monitorju. (Journal of the Acoustical Society of America, Feb. 1986). Stavki so bili prestavljeni osebam pod naslednjima pogojema:

    Preverjajte domnevo, da je povprečje procentov pravilnih vsebin pri pogoju $S+F+A$ presega ustrezno povprečje pod pogoju S. Privzemi $\alpha = 0.05$.

rešitev: TODO

64. Tetraklorodibenzo-p-dioksin (TCDD) je visoko-toksična substanca, ki jo najdemo v industrijskih odpadkih. Znanstveniki so naredili študijo s katero so določili količino TCDD-ja v tkivih volovskih žab, ki živijo na območju Rocky Branch Creek v Arkansasu, za katerega se ve, da je kontaminirano z TCDD (Chemosphere, Feb. 1986). Merili so količino TCDD-ja (v delcih na trilijon) v različnih tkivih štirih samic volovskih žab. Za vsako žabo so izmerili razmerje med količino TCDD-ja v tkivu in v nožni mišici. Relativno razmerje med koncentracijo TCDD-ja v jetrih in jajčnikih s je podana v spremljajoči preglednici.

Žaba	Jetra	Jajčniki
A	11.0	34.2
B	14.6	41.2
C	14.3	32.5
D	12.2	26.2

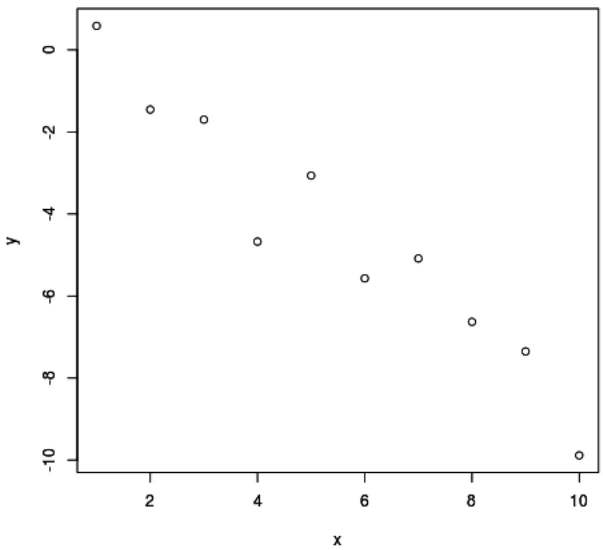
    Raziskovalci po zbranih podatkih sklepajo, da je povprečna koncentracija TCDD-ja v jajčnikih samic volovskih žab večja kot povprečna koncentracija v jetrih. Preverjajte to trditev z uporabo $\alpha = 0.05$.

rešitev: TODO

65. continue at FAPP5.2P13

3 Regresija

1. Na spodnji sliki je razsevni diagram spremenljivk X in Y. Naj bo $r = R(X, Y)$.



Kaj velja za ti dve spremenljivki?

rešitev:
Velja, da je koeficient korelacije $r < 0$ in približno velja $X \approx aX + b$ za $a < 0$. Torej je slučajna spremenljivka Y v približno linearni odvisnosti z slučajno spremenljivko X, kjer je koeficient naklona negativen.

2. Ali velja naslednja trditev? Regresijska premica za množico točk gre vedno skozi $(\bar{x}, \bar{y})$, kjer sta $\bar{x}$ in $\bar{y}$ povprečji za X in Y koordinati teh točk.

rešitev: Ta trditev drži. Obe regresijski premici se sečeta v tej točki, kar se da tudi pokazati iz njunih enačb.

3. Ali velja naslednja trditev? Geometrična interpretacija odmika točke od regresijske premice, je minimalna razdalja med to točko in regresijsko premico.

rešitev: To ni res. Odmik točke od premic je takšen, da je vsota kvadratov odmikov od premice minimalna.

4. Za množico parov (x,y) velja $\bar{x} = 8.0$, $s_x = 4.9$, $\bar{y} = 26.0$, $s_y = 13.2$ in $r = 0.97$. Koliko ima koeficient naklona regresijska premica, ki jo dobimo po metodi najmanjših kvadratov?

rešitev: Enačba prve regresijske premice je

$$y = \bar{y} + (x - \bar{x}) \cdot K(X, Y)/s_x^2$$

korelacijski koeficient pa je enak $r = k(X, Y)/(s_x \cdot s_y)$. Torej je iskani smerni koeficient enak

$$r \cdot \frac{s_y}{s_x} = 0.97 \cdot \frac{13.2}{4.9} = 2.6130$$

5. Ali velj a naslednja trditev? Za množico parov podatkov velja $\bar{x} = 8.0$, $s_x = 4.9$, $\bar{y} = 26.0$, $s_y = 13.2$ in $r = 0.97$. Točka $(8.0, 26.0)$ leži na prvi regresijski premici, ki smo jo izpeljali z metodo najmanjših kvadratov, smernostni koeficient te premice pa je približno 2.6.

rešitev:

Zapišemo enačbo prve regresijske premice in uporabimo dejstvo, da $r(X,Y) = \frac{K(X,Y)}{\sigma_x, sigma_y}$, da izračunamo vrednost $K(X,Y)$. Sedaj imamo dovolj podatkov, da zapišemo enačbo regresijske premice in iz nje po trivialnih postopkih izračunamo odgovor na vprašanje.

Trditev drži.

6. Imamo tabelo, ki prikazuje spol in izbrano pijačo za 100 naključno izbranih oseb, ki izbirajo med napitkoma A in B.

   Kaj je prva ničelna domneva?

   rešitev: Prva ničelna domneva je, da sta spol in izbira pijače neodvisni. Torej da je koeficient β v regresijski premici enak 0.

7. Zanima nas odvisnost dosežka na testu s starostjo testiranca. Na vzorcu dobimo naslednje rezultate:

Starost	18	19	20	20	25
Rezultat	99	100	100	101	105

   Kaj lahko povemo o vzorčnem korelacijskem koeficientu?

   Oceni rezultat testiranca, ki je star 23 let.

   rešitev: Koeficient korelacije izračunamo kot $r(X,Y) = \frac{K(X,Y)}{s_x s_y}$
   Kovarianco izračunamo kot $K(X,Y) = \frac{\sum_i (X-\bar{x})(Y-\bar{y})}{n-1} = 6.25$. Torej velja $r(X,Y) = \frac{6.25}{2.7019 \cdot 2.3452} = 0.986$.
   Izračunamo še povprečje vrednosti s.s. X in povprečje vrednosti s.s. Y in dobimo

$$\mu_x = 20.4 \quad \mu_y = 101$$

   Zapišimo prvo regresijsko premico:

$$Y = \mu_Y + \frac{K(X,Y)}{\sigma_x^2}(X - \mu_X)$$
$$Y = 101 + \frac{6.25}{7.3 \cdot 5.5}(x - 20.4)$$
$$Y = 0.155666x + 97.8244$$

   Starost vstavimo v X in dobimo oceno $Y = 103$.

8. Naj bosta X in Y slučajni spremenljivki. Kaj lahko sklepamo, če sta regresijski premici enaki?

   rešitev:

   Premici se sekata v točki $(\bar{x}, \bar{y})$, zato se ujemata, kadar sta smerna koeficienta enaka.
   Lahko pokažemo, da velja $k(X,Y)^2 = s_x^2 \cdot s_y^2$, kar pomeni, da je po definiciji korelacijskega koeficienta $|r(X,Y)| = 1$.

9. Statistično sklepanje o korelacijski povezanosti in regresija (odprta vprašanja)

   rešitev: Glej zbirko rešitev odgovorov na odprta vprašanja.

10. Pri poskusu dobimo naslednje urejene pare števil (x,y):

$$(1,2), \quad (2,4), \quad (3,6), \quad (3,7)$$

Brez računanja premisli, kateri je najprimernejši opis korelacijskega koeficienta.

rešitev: Vidimo, da z naraščanjem prvega elementa parov v grobem narašča tudi drugi par elementov. Lahko torej rečemo, da je korelacijski koeficient nekaj malega manjši kot 1.

11. Ali velja naslednja trditev? Geometrična interpretacija odmika točke od regresijske premice, je minimalna razdalja med to točko in regresijsko premico.

rešitev: Na to vprašanje smo že odgovorili.

12. Za množico parov (x,y) velja $\bar{x} = 8.0$, $s_x = 4.9$, $\bar{y} = 26.0$, $s_y = 13.2$ in $r = 0.97$. Kakšen smernostni koeficient ima regresijska premica, ki jo dobimo po metodi najmanjših kvadratov?

rešitev: Na zelo podobno vprašanje smo že odgovorili.

13. Za množico parov $(x_i, y_i)_{i=1}^n$ velja $\bar{x} = 8.0$, $s_x = 4.9$, $\bar{y} = 26.0$, $s_y = 13.2$ in $r = 0.97$. Kakšen smernostni koeficient ima regresijska premica, ki jo dobimo po metodi najmanjših kvadratov?

rešitev: Na podobno vprašanje smo že odgovorili.

14. Dnevna poraba ledu y (v funtih) v zabaviščnem parku je povezana z maksimalno temperaturo x (v F). Recimo, da je enačba regresijske premice najmanjših kvadratov $y = 50+20x$. Napovej porabo ledu za dan, ko je maksimalna temperatura 70F.

rešitev:

$$y = 50 + 20 \cdot 70 = 1450$$

15. Morske krave so velika, krotka morska bitja, ki živijo ob obali Floride. Veliko jih ubijejo ali poškodujejo hitri motorni čolni. V spodnji tabeli so podatki o številu registriranih čolnov (v tisočih) in številu morskih krav, ki so jih čolni ubili med leti 1977 in 1990.

leto	št. čolnov	št. ubitih krav
1977	447	13
1978	460	21
1979	481	24
1980	498	16
1981	513	24
1982	512	20
1983	526	15
1984	559	34
1985	585	33
1986	614	33
1987	645	39
1988	675	43
1989	711	50
1990	719	47

Želimo raziskati zvezo med številom motornih čolnov in številom ubitih morskih krav. Katera od spremenljivk je obrazložitvena?

Nariši razsevni diagram. Opiši smer, obliko in moč zveze. Ali opaziš kakšne ubežnike ali druge pomembne nepravilnosti.

rešitev:
Obrazložitvena spremenljivka (tudi neodvisna spremenljivka) je število čolnov. Zanima nas, če je odziva spremenljivka kakorkoli povezana z obrazložitveno spremenljivko.
Na razsevnem diagramu na prvi pogled opazimo močno pozitivno korelacijo. Ubežnik je morda leto 1983.

16. Kako se spreminja poraba goriva, ko se povečuje hitrost? V spodnji tabeli so zbrani podatki za britanski Ford Escort. Hitrost je merjena v kilometrih na uro, poraba pa v litrih na 100 kilometrov. (Vir: T. N. Lam, Estimating fuel consumption from engine size, Journal of Transportation Engineering, 111(1985): 339-357.)

hitrost (km/h)	poraba (l/100km)
10	21.00
20	13.00
30	1.00
40	8.00
50	7.00
60	5.90
70	6.30
80	6.95
90	7.57
100	8.27
110	9.03
120	9.87
130	10.79
140	11.77
150	12.83

Nariši razsevni diagram. Katera od spremenljivk je obrazložitvena?

Opiši vrsto zveze med spremenljivkama. Kako bi opisal smer te zveze? Ali je zveza razumno močna ali precej šibka?

rešitev: Obrazložitvena (tudi neodvisna) spremenljivka je hitrost.
Zveza med spremenljivkama je močna (točke se lepo prilegajo krivulji), vendar ni linearna. Do 60km/h je koreliranost negativna, nato pozitivna.

17. Raziskovalci, ki proučujejo kisli dež, so izmerili kislost padavin v divjini Kolorada preko 150 zaporednih tednov. Kislost merijo v pH. Nižja pH vrednost pomeni večjo kislost. Raziskovalci so s časom opazili linearni vzorec. Poročali so, da se podatkom dobro prilega regresijska premica najmanjših kvadratov z enačbo

$$pH = 5.43 - (0.0053 \cdot st_tednov).$$

(Vir: W. M. Lewis in M. C. Grant, Acid precipitation in the western United States, Science, 207(1980): 176-177.)

Nariši graf te premice. Razloži na preprost način, kaj nam premica pove o spreminjanju pH skozi čas.

Iz premice razberi, kolikšna je bila vrednost pH na začetku opazovanj ($tedni = 0$) in na koncu ($tedni = 150$).

rešitev: Opazimo, da skozi čas pH pada.
Naklon premice je -0.0053. To pomeni, da nam premica napoveduje, da če povečamo neodvisno spremenljivko za en teden, bo odvisna oziroma odzivna spremenljivka padla za približno 0.0053.
Na začetku - za st. tednov vzamemo 0 in dobimo rezultat 5.43. Za 150. teden pa dobimo rezultat 4.635.

18. Asfaltno cestišče se po izdelavi začne sušiti in s časom pridobiva trdnost. Inženirji uporabljajo regresijske premice, da predvidijo, kakšna bo trdnost po 28 dneh (ko bo sušenje končano) na podlagi meritev, ki jih izvedejo po 7ih dneh. Naj bo x moč (v kilogramih na kvadratni centimeter) po 7ih dneh in y moč po 28ih dneh. Iz enega dela meritev so ugotovili, da je enačba regresijske premice najmanjših kvadratov enaka

    Neka nova merjenja po 7ih dneh pokažejo, da je moč $3300kg/cm^2$. Napovej moč tega materiala po 28ih dneh.

    rešitev: Vstavimo podatek v enačbo in dobimo rezultat 4557.

19. V spodnji tabeli so podatki o porabi goriva v odvisnosti od hitrosti za manjši avto...

    Pojasni, zakaj je r majhen, čeprav sta poraba in hitrost močno povezani.

    rešitev: Korelacijski koeficient je majhen, ker hitrost in poraba nista linearno povezni.

20. Recimo, da bi se ženske vedno poročile z moških, ki so dve leti starejši od njih. Kolikšna bi bila v tem primeru korelacija med starostjo moža in žene?

    rešitev: Ker bi bili v tem primeru obe starosti pozitivno linearno poveznai, bi bil korelacijski koeficient enak 1.

21. Archaeopteryx je izumrla zver, ki je imela perje kot ptice ter zobovje in dolgi rep kot plazilci. Znanih je le šest primerkov fosilov. Ker se ti primerki zelo razlikujejo v velikosti, so nekateri znanstveniki mnenja, da gre za različne vrste in ne za posamezne pripadnike iste vrste. Če fosili pripadajo isti vrsti in se razlikujejo v velikosti le zato, ker so eni mlajši od drugih, bi morala obstajati linearna zveza med dolžinami nekega para kosti za vse primerke. Ubežniki bi v tem primeru sugerirali, da gre za drugo vrsto. V naslednji tabeli so podatki o dolžini (v cm) kosti, imenovane femur (gre za eno od kosti noge) in dolžini kosti, imenovane humerus (kost zgornjega dela roke), za pet fosilov, pri katerih sta bili obe kosti ohranjeni. (Vir: M. A. Houck et al., Allometric scaling in the earliest fossil bird, Archaeopteryx lithographica, Science, 247(1990): 195-198.)

Femur	Humerus
38	41
56	63
59	70
64	72
74	84

    Nariši razsevni diagram. Ali meniš, da pripada vseh pet primerkov isti vrsti?

Po definiciji izračunaj korelacijo r. Se pravi, poišči povprečje in standardni odklon dolžin femurjev in dolžin humerusov (za računanje povprečij in standardnih odklonov uporabi kalkulator). Nato izračunaj odklone od povprečja in uporabi formulo za r.

rešitev: Točke na razsevnem diagramu se približno prilegajo premici. Predvidevamo, da bo vseh pet primerkov pripadalo isti vrsti.
Korelacijski koeficient izračunamo na roke s formulo $r(X,Y) = \frac{K(X,Y)}{s_x \cdot s_y}$ ali pa uporabimo MATLAB funkcijo $corrcoef(F,H)$.

22. Prehrambena industrija je prosila skupino 3368 ljudi, da ocenijo število kalorij v večjem številu pogostih vrst hrane. Spodnja tabela prikazuje povprečja njihovih ocen in dejanska števila kalorij.

Hrana	Uganjene	Dejanske
Polnom. mleko	196	159
Špageti z omako	394	163
Makaroni s sirom	350	269
Rezina pšen. kruha	117	61
Rezina belega kruha	136	76
Čokoladna rezina	364	260
Slani krekerji	74	12
Srednje veliko jabolko	107	80
Srednje velik krompir	160	88
Tortica s smetano	419	160

Ocene so veliko previsoke v primeru špagetov in tortice. Obkroži ustrezni točki na svojem razsevnem diagramu. Izračunaj r za ostalih osem vrst hrane, ti dve pa izpusti. Pojasni, zakaj se je r spremenil tako, kot se je.

rešitev: Če odstranimo podatka, ki bistveno odstopata, se r poveča, ker je zdaj linearna zveza močnejša.

23. Močna zveza med dvema spremenljivka ne pomeni vedno, da ena od spremenljivk povzroča spremembe druge. Nekdo ugotovi, "Obstaja močna pozitivna korelacija med številom gasilcev, ki gasijo požar, in škodo, ki jo ta požar povzroči. Če torej h gašenju pokličemo več gasilcev, bo škoda samo še večja."

Pojasni, zakaj je takšno sklepanje napačno.

rešitev: Correlation does not mean causation...

24. Močna zveza med dvema spremenljivka ne pomeni vedno, da ena od spremenljivk povzroča spremembe druge. Raziskave kažejo, da obstaja pozitivna korelacija med velikostjo bolnišnice (merjeno s številom postelj x) in mediano števila dni y, ki jih pacienti preživijo v bolnišnici.

Ali to pomeni, da si lahko skrajšamo število bolnišničnih dni, če za zdravljenje izberemo manjšo bolnišnico? Razloži.

rešitev: Seveda ne, večje bolnišnice sprejmejo tudi hujše primere iz manjših bolnišnic, in ti hujši primeri povečajo povprečno število dni, ki jih pacienti preživijo v bolnišnici.
Glej tudi prejšno vprašanje.