

Vaje za statistiko

2010/11

asistent Emil Polajnar

1. Določi populacijo, spremenljivko in za spremenljivko tip izražanja ter tip merske lestvice.

- a.) V določenem časopisu za nek dan naključno izberemo vzorec člankov. Izračunati hočemo povprečno dolžino članka, ki jo merimo s številom besed.
- b.) Lastnike psov vprašamo po znamki hrane, ki jo najpogosteje kupijo.
- c.) Zanima nas povprečna najvišja temperatura določenega dne v Sloveniji. Naključno izberemo vzorec merilnih postaj.
- d.) Preskuševalec vina na vinskem sejmu dobi vzorec petih kozarcev vina. Kakovost vina oceni na lestvici od 1 do 5.
- e.) Zanimajo nas nakupovalne navade slovenskih študentk. V nakupovalnem središču jih pred trgovino z oblekami povprašamo o njihovi starosti in porabljenem znesku v trgovini.
- f.) Študente na FDV vprašamo po njihovi športni aktivnosti. Na vprašanje o pogostosti ukvarjanja s športom so možni odgovori: nikoli, redko, pogosto in vsak dan.

2. Za neko gimnazijo z 200 dijaki so v naslednji frekvenčni tabeli zbrani podatki o končnem uspehu na maturi. Nariši histogram in poligon za frekvenčno porazdelitev. Kakšne oblike je porazdelitev?

ocena	frekvenca
1	24
2	17
3	71
4	57
5	31

3. V frekvenčni tabeli so zbrani podatki za višino 110 študentov prvega letnika. Nariši histogram in ogivo. Kakšne oblike je porazdelitev?

višina [cm]	frekvenca
145–pod 155	10
155–pod 165	28
165–pod 175	20
175–pod 185	26
185–pod 195	19
195–pod 205	7

4. Podane so naslednje vrednosti.

10 6 24 11 22 14 10 21 6 22

- Izračunaj prvi kvartil.
- Izračunaj osmi decil.
- Izračunaj enaindvajseti centil.

5. Na populaciji izmerimo naslednje vrednosti.

6 17 12 13 3 7 5 6 6 11 20 8 12 5 11

- Izračunaj aritmetično sredino (povprečje).
- Izračunaj mediano.
- Izračunaj modus.
- Izračunaj variacijski razmik.
- Izračunaj povprečni absolutni odklon od mediane.
- Izračunaj kvartilni odklon.
- Izračunaj standardni odklon.
- Izračunaj koeficient variacije.

6. Življenjska doba v mesecih za 50 morskih prašičkov je zbrana v naslednji frekvenčni tabeli. Izračunaj modus in ga grafično preveri. Sestavi še frekvenčno tabelo z razredi širine 20 in ponovno izračunaj modus.

življenjska doba	frekvenca	življenjska doba	frekvenca
20–pod 30	1	20–pod 40	
30–pod 40	4	40–pod 60	
40–pod 50	13	60–pod 80	
50–pod 60	19		
60–pod 70	9		
70–pod 80	4		

7. Vzreditelje mačk smo povprašali, koliko mladičkov je bilo v zadnjem leglu. Podatki so na voljo za populacijo 40 mačjih družin. Izračunaj aritmetično sredino (povprečje), standardni odklon, koeficient asimetričnosti in koeficient sploščenosti.

število mladičkov	frekvenca
1	4
2	6
3	9
4	7
5	5
6	4
7	3
8	2

8. V tabeli so zbrani podatki o številu prebivalcev za 12 največjih mest v Sloveniji (iz popisa leta 2002) in ZDA. Kje je variabilnost števila prebivalstva večja?

Slovenija	prebivalstvo	ZDA	prebivalstvo v milijonih
Ljubljana	258873	New York	8.39
Maribor	93847	Los Angeles	3.83
Celje	37834	Chicago	2.85
Kranj	35587	Houston	2.26
Velenje	26742	Phoenix	1.59
Koper	23726	Philadelphia	1.55
Novo mesto	22415	San Antonio	1.37
Ptuj	18339	San Diego	1.31
Trbovlje	16290	Dallas	1.3
Nova Gorica	13491	San Jose	0.96
Jesenice	13429	Detroit	0.91
Murska Sobota	12437	San Francisco	0.82

9. V nalogi 8 izberimo tretje najbolj naseljeno mesto v Sloveniji in ZDA, torej Celje in Chicago. Katero mesto je relativno bolj poseljeno?

10. Porazdelitev slučajne spremenljivke X je podana z verjetnostno shemo. Določi parameter a , izračunaj matematično upanje in disperzijo.

$$X \sim \begin{pmatrix} 1 & 4 & 7 & 9 \\ 0.4 & a & 0.1 & 0.3 \end{pmatrix}$$

11. V nekem kraju 25% prebivalcev nasprotuje postavitvi novega športnega objekta. Naključno izberemo 5 prebivalcev. Naj bo slučajna spremenljivka število izbranih, ki nasprotujejo postavitvi športnega objekta.

- a.) Kako je porazdeljena slučajna spremenljivka?
- b.) Kakšna je verjetnost, da postavitvi objekta nasprotujejo 4 izmed izbranih prebivalcev?
- c.) Izračunaj matematično upanje?
- d.) Izračunaj disperzijo?

12. Slučajna spremenljivka Z je porazdeljena standardizirano normalno. Izračunaj verjetnosti za naslednje izraze.

- a.) $P(-0.80 < Z < 1.23)$
- b.) $P(Z < 0.55)$
- c.) $P(Z > 1.96)$
- d.) $P(-2.58 < Z < -1.65)$

13. Slučajna spremenljivka Z je porazdeljena standardizirano normalno. Določi vrednost z .

- a.) $P(0 < Z < z) = 0.40$
- b.) $P(Z > z) = 0.60$
- c.) $P(Z < z) = 0.99$
- d.) $P(-z < Z < z) = 0.95$

14. Naj bo slučajna spremenljivka povprečna dnevna temperatura. Podatke imamo na voljo za 310 dni (velikost populacije) na merilni postaji Murska Sobota, in sicer za mesec julij. Predpostavimo, da je porazdelitev normalna, z aritmetično sredino 21.1°C in standardnim odklonom 3.4°C .

- a.) Izračunaj verjetnost in število dni, ko je:
- povprečna dnevna temperatura v juliju manjša od 15°C
 - povprečna dnevna temperatura v juliju med 18°C in 22°C
 - povprečna dnevna temperatura v juliju večja od 20°C
- b.) Določi najnižjo temperaturo za 31 najbolj vročih julijskih dni.
- c.) Določi najvišjo temperaturo za 62 najbolj hladnih julijskih dni.

15. Naj bo slučajna spremenljivka starost prebivalcev Slovenije. V popisu leta 2002 je bilo ugotovljeno, da je povprečna starost prebivalcev Slovenije 39.4 leta s standardnim odklonom 21.5 let. Predpostavimo, da se starost porazdeljuje normalno.

- a.) Na populaciji izberemo vzorec velikosti 100 in izračunamo vzorčno povprečje. Kakšna je standardna napaka vzorčne aritmetične sredine? Kakšna je porazdelitev za vzorčno aritmetično sredino?
- b.) Kakšna je verjetnost, da dobimo vzorec velikosti 100 s povprečno starostjo več kot 42 let?
- c.) Na nekem vzorcu velikosti 100 smo dobili za povprečno starost 38.2 leta in standardni odklon 22 let. Izračunaj 95% interval zaupanja za populacijsko aritmetično sredino. Ali zajamemo pravo populacijsko vrednost?
- d.) Na voljo imamo samo podatke o starosti oseb na vzorcu velikosti 16. Izračunaj povprečno starost in standardni odklon ter izračunaj 95% interval zaupanja za populacijsko aritmetično sredino. Ali zajamemo pravo populacijsko vrednost?

46 6 27 69 44 40 26 37 66 47 36 35 29 22 32 11

- e.) Iz popisa poznamo pravo vrednost za standardni odklon. Najmanj kako velik vzorec moramo izbrati, da bomo pri 5% tveganju za manj kot 1.5 leta oddaljeni od prave populacijske vrednosti?

16. Naj bo slučajna spremenljivka delež mestnega prebivalstva v Sloveniji. V popisu leta 2002 je bilo ugotovljeno, da od skupno 1 964 036 prebivalcev v mestih živi 997 772 prebivalcev.

- a.) Na populaciji izberemo vzorec velikosti 900 in izračunamo vzorčni delež mestnega prebivalstva. Kakšna je standardna napaka vzorčnega deleža? Kakšna je porazdelitev za vzorčni delež?
- b.) Kakšna je verjetnost, da na vzorcu velikosti 900 dobimo delež mestnega prebivalstva, ki je manjši od 50%?
- c.) Na nekem vzorcu velikosti 900 je bilo 491 prebivalcev iz mesta. Izračunaj 99% interval zaupanja za populacijski delež. Ali zajamemo pravi populacijski delež?
- d.) Iz popisa poznamo pravi populacijski delež. Najmanj kako velik vzorec moramo izbrati, da bo pri 1% tveganju razlika med populacijskim in vzorčnim deležem manjša kot 2 odstotni točki?

17. Naj bo slučajna spremenljivka starost prebivalcev Slovenije, pri čemer prebivalce razdelimo po spolu. V popisu leta 2002 je bilo ugotovljeno, da znaša povprečna starost za ženske 41.1 leta s standardnim odklonom 22.3 leta in povprečna starost za moške 37.7 leta s standardnim odklonom 20.5 leta. Predpostavimo, da se starost na obeh populacijah porazdeljuje normalno.

- a.) V vzorec izberemo 400 žensk in 400 moških ter izračunamo razliko vzorčnih aritmetičnih sredin. Kakšna je standardna napaka razlike vzorčnih aritmetičnih sredin? Kakšna je porazdelitev za razliko vzorčnih aritmetičnih sredin?
- b.) Kakšna je verjetnost, da je na vzorcu 400 žensk in 400 moških povprečna starost žensk za manj kot 2 leti večja od povprečne starosti moških?

18. Naj bo slučajna spremenljivka delež prebivalcev z diplomo. Iz popisa 2002 je razvidno, da je v občini Ajdovščina skupaj 15 062 prebivalcev, od katerih jih ima 837 diplom, medtem ko je v občini Murska Sobota 17 382 prebivalcev, od katerih jih ima 1 323 diplom.

- a.) V vsaki občini izberemo vzorec velikosti 225 in izračunamo razliko vzorčnih deležev. Kakšna je standardna napaka za razliko vzorčnih deležev? Kakšna je porazdelitev razlike vzorčnih deležev?
- b.) Kakšna je verjetnost, da na vzorcu velikosti 225 iz vsake občine dobimo razliko deležev, ki je po absolutni vrednosti manjša od 1%?

19. Evropska družboslovna raziskava 2008. V slučajni vzorec je bilo izbranih 196 prebivalcev Slovenije. Eno izmed vprašanj je na lestvici od 1 do 6 merilo strinjanje anketirancev s trditvijo, da jim veliko pomenita bogastvo in denar. Povprečna vrednost odgovora je bila 4.06 s standardnim odklonom 1.21. Ob 5% stopnji značilnosti preveri, ali je povprečna vrednost strinjanja za Slovence večja od 4.

20. Evropska družboslovna raziskava 2008. V slučajni vzorec je bilo izbranih 196 prebivalcev Slovenije. Eno izmed vprašanj je na lestvici od 1 do 6 merilo strinjanje anketirancev s trditvijo, da jim veliko pomenita skrb za naravo in okolje. V vzorec je bilo vključenih 95 moških, kjer je bila povprečna vrednost odgovora 2.15 s standardnim odklonom 0.989, in 101 ženska, kjer je bila povprečna vrednost odgovora 1.94 s standardnim odklonom 1.018. Ob 1% stopnji značilnosti preveri, ali imajo moški in ženske v Sloveniji različno skrb za naravo in okolje.

21. Evropska družboslovna raziskava 2008. V slučajni vzorec je bilo izbranih 196 polnoletnih prebivalcev Slovenije. Eno izmed vprašanj je spraševalo po udeležbi na zadnjih volitvah. Med vsemi anketiranimi je bilo takih 155 oseb. Ali lahko pri 1% stopnji značilnosti trdimo, da je bila udeležba na volitvah večja kot 70%?

22. Evropska družboslovna raziskava 2008. V slučajni vzorec je bilo izbranih 196 prebivalcev Slovenije. Eno izmed vprašanj je bilo, če ima anketiranec mobilni telefon. V vzorec je bilo zajetih 95 moških, od katerih jih je imelo mobitel 88, in 101 ženska, od katerih jih je imelo mobitel 84. Ob 5% stopnji značilnosti preveri domnevo, da sta deleža moških in žensk, ki imajo mobilni telefon, enaka.

23. Proizvajalec žarnic trdi, da je povprečna življenjska doba njegovih žarnic 10 000 ur. Trditev preverimo na vzorcu 9 žarnic. Pri poskusu si zabeležimo čas do odpovedi posamezne žarnice. Ali lahko pri 5% stopnji značilnosti potrdimo proizvajalčevo deklaracijo o povprečni življenjski dobi?

12218.5 109.1 12822.7 3540.5 937.6 4994.9 1202.5 9106.3 7284.2

24. Preveriti želimo kakovost gum dveh proizvajalcev. Za prvega uspemo dobiti vzorec 9 kompletov gum in za drugega vzorec 16 kompletov gum. Preskus opravimo tako, da zabeležimo število prevoženih kilometrov, ko postanejo gume neustrezne za nadaljnjo vožnjo. Ali lahko pri 5% stopnji značilnosti trdimo, da je kakovost proizvajalcev različna?

proizvajalec A:	36630	35211	40322	47362	40970	34096
	40488	30988	35575			
proizvajalec B:	41508	48509	37557	42378	35686	33957
	40468	35395	43416	41444	45070	43573
	46089	42258	40838	44125		

25. V tabeli so zbrani podatki o Nobelovih nagrajencih za določena področja in izbrane države. Zajeta so leta med 1901 in 1950. Upoštevane so nagrade, ki so bile podeljene samo eni osebi. Ali obstaja povezanost med področjem in državo raziskovanja? Kakšna je moč povezanosti? Domnevo o povezanosti preskusi pri 5% stopnji značilnosti.

	Francija	Nemčija	Velika Britanija	ZDA
fizika	3	8	8	6
literatura	6	5	4	4
medicina	4	6	1	3

26. V tabeli so zbrani podatki o površini in največji globini za deset svetovnih jezer. Razvrsti jezera po velikosti in globini ter s pomočjo ustreznega koeficienta ugotovi, ali obstaja povezanost med velikostjo in globino. Ali lahko govorimo o povezanosti tudi na populaciji? Upoštevaj 5% stopnjo značilnosti.

jezero	površina [km ²]	globina [m]
Aralsko jezero	33800	68
Bajkalsko jezero	31500	1741
Gornje jezero (Lake Superior)	82414	406
Huronsko jezero (Lake Huron)	59596	229
Kaspijsko jezero	394299	946
Malavijsko jezero	30044	706
Michigansko jezero (Lake Michigan)	58016	281
Tanganjiško jezero	32893	1435
Veliko medvedje jezero	31080	82
Viktorijino jezero	69485	82

27. V tabeli so zbrani podatki za potrese z največ žrtvami v danem letu. Razvrsti potrese po magnitudi in po številu žrtev ter s pomočjo ustreznega koeficienta ugotovi, ali obstaja povezanost med močjo potresa in številom žrtev. Ali lahko govorimo o povezanosti tudi na populaciji? Upoštevaj 10% stopnjo značilnosti.

leto potresa	magnituda	število žrtev
2001	7.7	20023
2002	6.1	1000
2003	6.6	31000
2004	9.1	227898
2005	7.6	80361
2006	6.3	5749
2007	8.0	514
2008	7.9	87587
2009	7.5	1117
2010	7.0	222570

28. Indeks CPI (Corruption Perceptions Index, www.transparency.org) meri zaznavnost korupcije v javnem sektorju na lestvici od 0 do 100 (višja vrednost pomeni manjšo korupcijo). V tabeli so zbrani podatki za indeks CPI in BDP na prebivalca v 1000 US\$ za vzorec 15 držav v letu 2009.

država	CPI	BDP
Avstralija	87	38.8
Bocvana	56	13.1
Bolgarija	38	12.6
Čad	16	1.6
Čile	67	14.7
Gambija	29	1.4
Grčija	38	32.1
Mavretanija	25	2.1
Nemčija	80	34.1
Pakistan	24	2.6
Panama	34	11.9
Peru	37	8.6
Poljska	50	17.9
Rusija	22	15.1
Somalija	11	0.6

- Ali obstaja linearna povezanost med spremenljivkama?
- Ali lahko pri 1% stopnji značilnosti linearno povezanost posplošimo na celotno populacijo držav (torej na ves svet)?
- Nariši razsevni diagram. Kaj lahko povemo o vplivu BDP na indeks CPI?
- Izračunaj regresijsko premico in interpretiraj rezultat.
- S pomočjo regresijske premice napovej indeks CPI za Slovenijo, ki je v letu 2009 imela BDP na prebivalca \$27900. Po podatkih iz raziskave je indeks CPI za Slovenijo 66.
- Pri 5% stopnji značilnosti preveri domnevo, ali BDP pozitivno vpliva na indeks CPI.
- Kolikšen del variabilnosti v indeksu CPI uspemo pojasniti s spremenljivko BDP?
- Izračunaj standardno napako ocene.

Rešitve

4.

a.) $P = 0.25, R = 3, Q_1 = 10$

b.) $P = 0.80, R = 8.5, x = 22$

c.) $P = 0.21, R = 2.6, x = 8.4$

5.

a.) $\sum x_i = 142, \mu = 9.47$

b.) $\text{Me} = 8$

c.) $\text{Mo} = 6$

d.) $R = 17$

e.) $\sum |x_i - \text{Me}| = 58, AD_{\text{Me}} = 3.87$

f.) $Q_1 = 6, Q_3 = 12, Q = 3$

g.) $\sum (x_i - \mu)^2 = 323.73, \sigma^2 = 21.582, \sigma = 4.65$

h.) $\text{KV} = 0.49$

6. prvotna tabela $\text{Mo} = 53.75$, tabela s tremi razredi $\text{Mo} = 51.74$

7. $\sum f_i x_i = 157, \mu = 3.925, \sum f_i (x_i - \mu)^2 = 148.775, \sigma^2 = 3.719, \sigma = 1.9286,$
 $\sum f_i (x_i - \mu)^3 = 114.49, g_1 = 0.399$ (asimetrična v desno), $\sum f_i (x_i - \mu)^4 = 1282.33,$
 $g_2 = -0.6826$ (sploščena)

8. $\mu_{SLO} = 47751, \sigma_{SLO} = 67118, \text{KV}_{SLO} = 1.41, \mu_{ZDA} = 2.262, \sigma_{ZDA} = 2.032,$
 $\text{KV}_{ZDA} = 0.898$, večja je variabilnost med slovenskimi mesti

9. $z_{SLO} = -0.148, z_{ZDA} = 0.289$, relativno bolj je poseljen Chicago

10. $a = 0.2, E(X) = 4.6, E(X^2) = 32.8, D(X) = 11.64$

11.

- a.) $X \sim \text{bin}(5, 0.25)$
- b.) $P(k = 4) = 0.015$
- c.) $E(X) = 1.25$
- d.) $D(X) = 0.9375$

12.

- a.) $P(-0.8 < Z < 1.23) = 0.2881 + 0.3907 = 0.6788$
- b.) $P(Z < 0.55) = 0.5 + 0.2088 = 0.7088$
- c.) $P(Z > 1.96) = 0.5 - 0.4750 = 0.025$
- d.) $P(-2.58 < Z < -1.65) = 0.4951 - 0.4505 = 0.0446$

13.

- a.) $P(0 < Z < z) = 0.40 \Rightarrow z = 1.28$
- b.) $P(Z > z) = 0.60 \Rightarrow z = -0.25$
- c.) $P(Z < z) = 0.99 \Rightarrow z = 2.33$
- d.) $P(-z < Z < z) = 0.95 \Rightarrow z = 1.96$

14.

- a.) $P(T_{pop} < 15) = 0.0367$ (11 dni), $P(18 < T_{pop} < 22) = 0.4212$ (131 dni),
 $P(T_{pop} > 20) = 0.6255$ (194 dni)
- b.) 31 najbolj vročih dni ima najnižjo temperaturo $T = 25.5$
- c.) 62 najbolj hladnih dni ima najvišjo temperaturo $T = 18.2$

15.

- a.) $\text{SE}(\bar{X}) = 2.15$, $\bar{X} \sim N(39.4, 2.15)$
- b.) $P(\bar{X} > 42) = P(Z > 1.21) = 0.5 - 0.3869 = 0.1131$
- c.) $z_{\alpha/2} = 1.96$, $P(33.89 < \mu < 42.51) = 0.95$
- d.) $\bar{x} = 35.81$, $s = 16.85$, $t_{\alpha/2}(15) = 2.13$, $P(26.84 < \mu < 44.78) = 0.95$
- e.) $n \geq 790$

16.

- a.) $\pi = 0.508$, $SE(p) = 0.0167$, $p \sim N(0.508, 0.0167)$
- b.) $P(p < 0.50) = P(Z < -0.48) = 0.5 - 0.1844 = 0.3156$
- c.) $p = 0.546$, $SE(p) = 0.0166$, $z_{\alpha/2} = 2.58$, $P(0.503 < \pi < 0.589) = 0.99$
- d.) $n \geq 4160$

17.

- a.) $SE(\bar{X}_F - \bar{X}_M) = 1.52$, $\bar{X}_F - \bar{X}_M \sim N(3.4, 1.52)$
- b.) $P(\bar{X}_F - \bar{X}_M < 2) = P(Z < -0.92) = 0.5 - 0.3212 = 0.1788$

18.

- a.) $\pi_{MS} = 0.076$, $\pi_A = 0.056$, $SE(p_{MS} - p_A) = 0.023$, $p_{MS} - p_A \sim N(0.02, 0.023)$
- b.) $P(|p_{MS} - p_A| < 0.01) = P(-0.01 < p_{MS} - p_A < 0.01) = P(-1.30 < Z < -0.44) = 0.4032 - 0.1700 = 0.2332$

19. $H_A : \mu > 4$, $z_{\alpha} = 1.65$, $SE(\bar{X}) = 0.086$, $z_e = 0.70$

20. $H_A : \mu_M - \mu_F \neq 0$, $z_{\alpha/2} = 2.58$, $SE(\bar{X}_M - \bar{X}_F) = 0.143$, $z_e = 1.47$

21. $H_A : \pi > 0.70$, $z_{\alpha} = 2.33$, $p = 0.79$, $SE(p) = 0.033$, $z_e = 2.73$

22. $H_A : \pi_M - \pi_F \neq 0$, $z_{\alpha/2} = 1.96$, $p_M = 0.93$, $p_F = 0.83$, $p = 0.88$, $SE(p_M - p_F) = 0.046$, $z_e = 2.17$

23. $H_A : \mu \neq 10000$, $t_{\alpha/2}(8) = 2.31$, $\bar{x} = 5802$, $s = 4829$, $SE(\bar{X}) = 1610$, $t_e = -2.61$

24. $H_A : \mu_A - \mu_B \neq 0$, $t_{\alpha/2}(23) = 2.07$, $x_A = 37960$, $s_A = 4847$, $x_B = 41392$, $s_B = 4033$, $s = 4334$, $SE(\bar{X}_A - \bar{X}_B) = 1806$, $t_e = -1.90$

25. $H_A : \chi^2 > 0$, $\chi^2_{1-\alpha}(6) = 12.59$, $\chi_e^2 = 5.3866$, $k = 3$, $\alpha = 0.2155$, $C_{pop} = 0.3570$

26. $H_A : \rho_S \neq 0$, $t_{\alpha/2}(8) = 2.31$, $\sum d_i^2 = 179.5$, $r_S = -0.0879$, $t_e = -0.2495$

27. $H_A : \rho_S \neq 0$, $t_{\alpha/2}(8) = 1.86$, $\sum d_i^2 = 112$, $r_S = 0.3212$, $t_e = 0.9594$

28.

a.) $\bar{x} = 13.81333$, $s_x = 12.42991$, $\bar{y} = 40.93333$, $s_y = 22.79871$, $r_{xy} = 0.8007793$

b.) $H_A : \rho \neq 0$, $t_{\alpha/2}(13) = 3.01$, $t_e = 4.820444$

d.) $Y' = 20.6447 + 1.4688X$

e.) $Y'_{SLO} = 61.6$

f.) $H_A : \beta > 0$, $t_{\alpha}(13) = 1.78$, $t_e = 4.820444$

g.) $R = 0.6412475$

h.) $s_e = 13.65550$