

STATISTIKA

VAJE 2

NALOGA 1: Doloži vrste naslednjih statističnih spremenljivk

NUMERIČNE		OPISNE	
ZVEŽNE	DISKRETNE	NOMINALNE	ORDINALNE
starost teža višina temperatura dohodek inteligenci kv. številka čevlja št. pokajenih cig. pH vrednost	št. prijateljev letnik rojstva	vrsta raka krvna skupina smer neba znamka avta kajenje (da/ne)	ocena pri izpitu stopnja izobrazbe

DISCLAIMER: I fucked up, ker sem delil vse z 20, namesto 1%, ker 3 vrednosti manjkajo!

NALOGA 2: V tabeli so podatki za letno količino padavin (v mm) v letu 2014 za 20 meteoroloških postaj v Sloveniji.

a) Doloži frekvenčne razrede in zapiši frekvenčno porazdelitev.

← Frekvenčni razred →

Kadar je možnih vrednosti X veliko Ali ko X zavzame neskončno mnogo vrednosti, jih razdelimo na intervale.

Najlažji je narediti razrede širine tak, da so večkratniki 10. Ker imamo neke tiščne vrednosti, ki so dost bliz, lah damo na 500.

Pomembno je tudi, da najdeš najmanjšo in največjo vrednost, da veš kje začeti in kje končati.

Meteorološka postaja	Količina padavin (mm)
Bilje	2304
Bovec	3959
Brnik, Letališče	1894
Celje, Medlog	1436
Črnomelj, Dobliče	1794
Ilirska Bistrica	2285
Kočevje	1972
Kredarica	2449
Lesce	2114
Ljubljana, Bežigrad	1851
Maribor, Tabor	1233
Murska Sobota, Rakičan	1093
Novo mesto	1482
Portorož, Letališče	1462
Postojna	2069
Rateče	2084
Rogaška Slatina	...
Slap pri Vipavi	...
Šmartno pri Slovenj Gradcu	1498
Velenje	...

min = 1093
max = 3959

$$[1000, 1500) = \{1436, 1498, 1233, 1093, 1482, 1462\}$$

$$[1500, 2000) = \{1894, 1794, 1972, 1851\}$$

$$[2000, 2500) = \{2304, 2285, 2449, 2114, 2069, 2084\}$$

$$[2500, 3000) = \{\}$$

$$[3000, 3500) = \{\}$$

$$[3500, 4000] = \{3959\}$$

Actually, to ni dobr, da sta 2 razreda prazna, zato jih bom razširim na 1000.

$[1000, 2000) = \{1436, 1498, 1233, 1093, 1482, 1452, 1894, 1794, 1972, 1851\} = R_1$
 $[2000, 3000) = \{2304, 2285, 2449, 2114, 2089, 2084\} = R_2$
 $[3000, 4000] = \{3959\} = R_3$

Little better.

← Frekvenčna porazdelitev →

$$f_i = \frac{\#(i\text{-ta vrednost})}{\#(\text{vseh})}$$

kakšen delež predstavlja vrednost med vsemi.

Frekvenčne porazdelitve so torej...

← To so RELATIVNE FREKVENCE!

vsi = 20

R_1 = 10

R_2 = 6

R_3 = 1

$$\longrightarrow f_{R_1}^a = 10/20 = 0,5 \quad f_{R_2}^a = 6/20 = 0,3\bar{3} \quad f_{R_3}^a = 1/20 = 0,05$$

$$f_{R_1}^{\%} = 50\%$$

$$f_{R_2}^{\%} = 33,3\%$$

$$f_{R_3}^{\%} = 5\%$$

← To so FREKVENCE

b) Za vsak razred določi spodnjo in zgornjo mejo, sredino in širino.

Iz vseh podatkov naredim frekvenčno tabelo.

← Spodnja in zgornja meja razreda →

Za $i \in [1, n]$: $x_{i, \max} = x_{i+1, \min}$

← Širina razreda →

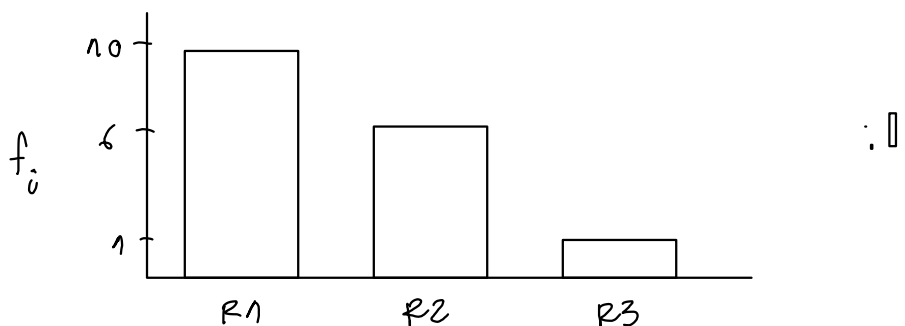
$$d_i = x_{i, \max} - x_{i, \min}$$

← Sredina razreda →

$$x_i = (x_{i, \max} + x_{i, \min}) / 2$$

Količina padavin	$x_{i, \min}$	$x_{i, \max}$	d_i	x_i	f_i	$f_i \%$
$[1000, 2000)$	1000	2000	1000	1500	10	50%
$[2000, 3000)$	2000	3000	1000	2500	6	33,3%
$[3000, 4000]$	3000	4000	1000	3500	1	5%

c) Grafično predstavi frekvenčno porazdelitev s histogramom

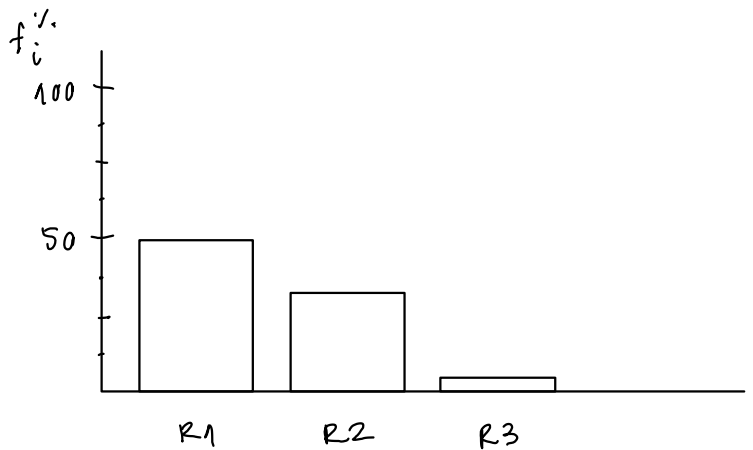


d) Izračunaj relativne frekvence in jih grafično predstavi

← Relativne frekvence →

Delež frekvence enote med vsemi frekvencaми.

$f_{R1}^o = 10/20 = 0,5$	$f_{R1}^{y_i} = 50\%$
$f_{R2}^o = 6/20 = 0,3\bar{3}$	$f_{R2}^{y_i} = 33,3\bar{3}\%$
$f_{R3}^o = 1/20 = 0,05$	$f_{R3}^{y_i} = 5\%$



e) Izračunaj gostoto

$$d_{R1} = 2000 - 1000 = 1000$$

$$d_{R2} = 3000 - 1000 = 1000$$

$$d_{R3} = 4000 - 1000 = 1000$$

f) Izračunaj kumulativne in kumulativne relativne frekvence

← Kumulativne frekvence →

Vsota frekvenc med sabo.

$$F_1 = f_1, F_2 = f_1 + f_2, \dots$$

$$F_n = f_1 + f_2 + \dots + f_n$$

$f_{R1} = 10$	$F_{R1} = 10$
$f_{R2} = 6$	$F_{R2} = 16$
$f_{R3} = 1$	$F_{R3} = 17$

$$F_{R1}^o = 10/20 = 0,5$$

$$F_{R2}^o = 16/20 = 0,8$$

$$F_{R3}^o = 17/20 = 0,85$$

↑
ni prav; glej disclaimer!
moralo bi biti 20/20!

NALOGA 3: Vinogradnik Vinke je na 5 hektarjih vinogradov naredil 14 meritev sladkorja v grozdju (g/cm^3)

16,2 17,1 20,3 18,2 17,9 19,5 19,9 16,8 20,3 19,0 18,5 20,0 19,1 17,8

a) Podatke uredi v ranžirno vrsto in doloži range in relativne range

← Ranžirna vrsta →

Vrednosti x_1, \dots, x_n urejenostne ali številske spremenljivke x urejene po velikosti.

$$16,2_{(1)} \leq 16,8_{(2)} \leq 17,1_{(3)} \leq 17,8_{(4)} \leq 17,9_{(5)} \leq 18,2_{(6)} \leq 18,5_{(7)} \leq 19,0_{(8)} \leq 19,1_{(9)} \leq 19,5_{(10)}$$

$$\leq 19,9_{(11)} \leq 20,0_{(12)} \leq 20,3_{(13)} \leq 20,3_{(14)}$$

te številke so v bistvu rangi elementov.

← Rang →

Položaj elementa v ranžirni vrsti. Z drugimi besedami, št. elementov, ki so manjši od trenutnega, vključno s trenutnim. $R(x)$

$$R(16,2) = 1$$

$$R(16,8) = 2$$

...

$$R(18,5) = 7$$

...

← Relativni rang →

Pove kolikšen delež podatkov je manjših ali enakih trenutnemu

$$r(x) = \frac{R(x)}{n}$$

x = trenutni element

$R(x)$ = rang od x

n = št. elementov

$$n = 14$$

$$r(16,2) = 1/14$$

$$r(16,8) = 2/14$$

$$\dots r(18,5) = 7/14 \dots \text{itn.}$$

b) Izračunaj vrednost od katere je le četrtina vrednosti večja. Poimenujte to vrednost

Ta vrednost se imenuje prvi kvartil in spada med kvantile.

← Kvantil →

Kvantil za določen delež je vrednost, pod katero leži približno dani delež podatkov.

$$\frac{\#(X < Q_\gamma)}{n} \leq \gamma$$

Kvantil za delež $1/4$ = kvartil.

Za $\gamma = 0,25$ oz. $\gamma = 1/4$, je to:

Povezava kvantila in ranga:

$$R = \gamma \cdot n + 0,5$$

$$\frac{\#(X < Q_{0,25})}{n} \leq 0,25$$

delež podatkov manjših od $Q_{0,25}$

Po zgornji enačbi je prvi kvartil torej:

$$R(Q_{0,25}) = 0,25 \cdot 14 + 0,5 = 4$$

Rang prvega kvartila je 4, kar pomeni, da je $Q_{0,25} = 17,8$

c) Izračunaj še ostale kvartile

Podobno poračunamo še za $0,5$ (mediana) in $0,75$ (tretji kvartil).

$$R(Q_{0,5}) = 0,5 \cdot 14 + 0,5 = 7,5 \dots$$

$$R(Q_{0,75}) = 0,75 \cdot 14 + 0,5 = 11 = 19,9$$

To pomeni, da je mediana nekje med 7. in 8. elementom v rangirni vrsti. Vzamemo podatek 1/2 med 7 in 8.

d) Izračunaj tretji decil

$$\frac{18,5 + 19,0}{2} = 18,75$$

← Decili →

Kvantili za deleže $j/100$ ($j=1, \dots, 99$).

$$|18,5 - 19,0| = 0,5 \text{ in } 0,5 \cdot 0,5 = 0,25 \text{ zato } 18,5 + 0,25 = 18,75.$$

Tretji decil $\Rightarrow j=3 \Rightarrow Q_{3/100} = Q_{0,03}$

$$R(Q_{0,03}) = 0,03 \cdot 14 + 0,5 = 0,92 \Rightarrow$$

Tretji decil je nekje med 0 in 1. elementom. Ker 0-ti element ne obstaja, je 3. decil enak prvemu elementu (?)

$$Q_{3/100} = 16,2$$

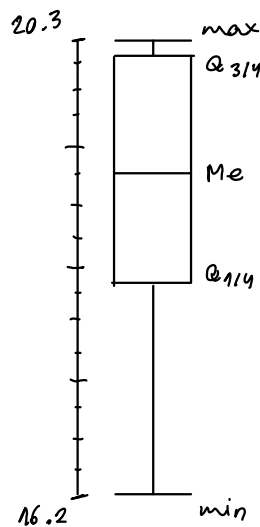
e) Nariši okvir z ročaji (oz. škatla z brki, boxplot)

Če boxplot potrebujemo par vrednosti:

- min
- max
- $Q_{0.25}$
- $Q_{0.75}$
- $Q_{0.5} = \text{Mediana}$

- min = 16.2
- max = 20.3
- $Q_{0.25} = 17.8$
- $Q_{0.75} = 19.9$
- Me = 18.75

⇒



Hell
yeah

NALOGA 4: Podatki 500 ljudi koliko ur na mesec uporabljajo računalnik

a) Položi modusni razred

← Modus →

Vrednost z najvišjo
frekvenco

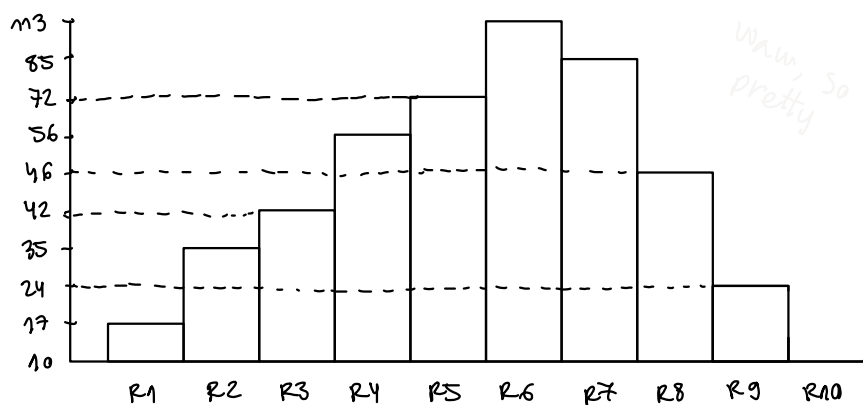
Podatki so razvrščeni v fre-
kvenčne razrede $[0, 9]$, $[10, 19]$,
..., $[90, 99]$. Modus bo torej
razred z najvišjo frekvenco

$$\text{Modus} = [50, 59] = R_6$$

Število ur	Število anketirancev
0 - 9	R_1 17
10 - 19	R_2 35
20 - 29	R_3 42
30 - 39	R_4 56
40 - 49	R_5 72
50 - 59	R_6 113
60 - 69	R_7 85
70 - 79	R_8 46
80 - 89	R_9 24
90 - 99	R_{10} 10

b) Grafično predstavi podatke

Bom naredo histogram števila anketirancev po
frekvenčnih razredih. (neratančen)



Wow, so
pretty

c) Izračunaj kvartile in jih interpretiraj. Grafično oceni še deveti decil D_9

Najprej spet ustvarim razširno vrsto:

- 1 = R_{10}
- 2 = R_1

$$R(Q_{1/4}) = 0.25 \cdot 10 + 0.5 = 3 \Rightarrow Q_{1/4} = 24$$

- 3 = R_9
- 4 = R_2

$$R(Q_{1/2}) = 0.5 \cdot 10 + 0.5 = 5.5 \Rightarrow 50\% \text{ med } 5. \text{ in } 6. (R_3 \text{ in } R_4) \Rightarrow Q_{1/2} = 44$$

- 5 = R_3
- 6 = R_8

$$R(Q_{3/4}) = 0.75 \cdot 10 + 0.5 = 8 \Rightarrow Q_{3/4} = 72$$

- 7 = R_4
- 8 = R_5

- 9 = R_7
- 10 = R_6

NALOGA 5: Neke mase planktona glede na različne spremenljivke.

a) Najvišje izmerjena temperatura v juniju?

Max vrednost na 1. grafu in to je $\approx 21.5^{\circ}\text{C}$

b) Poišči ero porazdelitev, ki je simetrična in ero, ki ni.

Simetrična je npr. temperatura v mesecu juniju (1 graf, J). Nesimetrična pa npr. nano-microplankton biomass v mesecu maju (3 graf, M).

c) Kakšna je bila mediana količine nano-mikroplanktona izmerjena v maju? Kaj pa 3. kvartil?

Mediana grafa 3 za maj je ≈ 220 in 3. kvartil ($Q_{3/4}$) ≈ 400 .

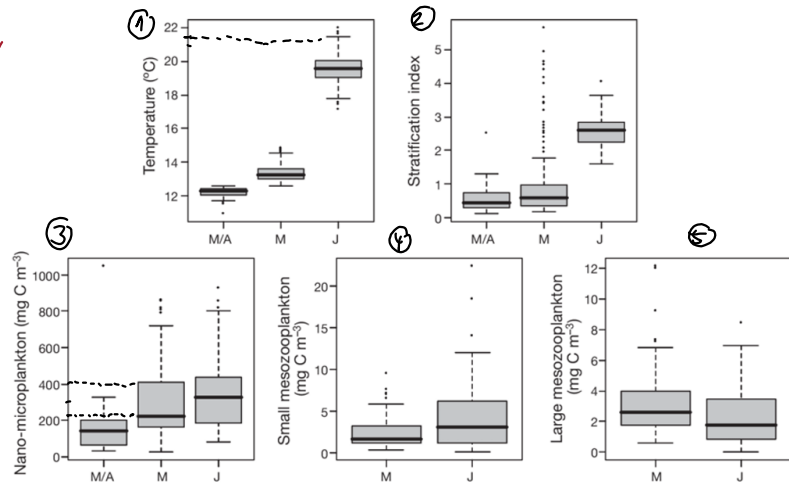
d) Kakšen je medkvartilni razmik pri velikem mezoplanktonu v maju?

Medkvartilni razmik

$$\begin{aligned} Q_{3/4} &\approx 4 \\ Q_{1/4} &\approx 2 \end{aligned} \Rightarrow IQR = 2$$

$$IQR = Q_{3/4} - Q_{1/4}$$

e) Katera porazdelitev je najbolj in katera najmanj razpršena?



Boxplot distributions of physical and biological variables during March/April (M/A), May (M) and June (J). Temperature (°C); stratification index; nano-microplankton biomass (mg C m⁻³); small mesozooplankton biomass (mg C m⁻³); and large mesozooplankton biomass (mg C m⁻³).

