

STATISTIKA

VAJE 3

NALOGA 1: Izberemo slučajni vzorec 12 krokodilov. Poni so podatki o njihovi dolžini v cm, zaokroženi na eno decimalno mesto.

178,4 165,5 159,4 168,8 175,3 170,0
173,2 175,6 163,3 171,8 179,6 167,4

a) Izračunaj vse mere centralne tendence

← Mere centralne tendence →

- (tehtana) aritmetična sredina
- mediana
- modus

Najlažje bo, da jih spet razvrstim v ranžirno vrsto:

159.4₍₁₎ 163.3₍₂₎ 165.5₍₃₎ 167.4₍₄₎ 168.8₍₅₎
170.0₍₆₎ 171.8₍₇₎ 173.2₍₈₎ 175.3₍₉₎ 175.6₍₁₀₎
178.4₍₁₁₎ 179.6₍₁₂₎

Aritmetična sredina (the normal way):

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2082.9}{12} = 173.575 \approx 173.6$$

Mediano dobim preko one fajne enačbe: $R(Q_{1/2}) = 0.5 \cdot 12 + 0.5 = 6.5$

Modus je vrednost z najvišjo frekvenco; ker se vsi pojavijo enako velikokrat, je vseeno kateri je modus (oz nima smisla) oz. modus ne obstaja?

$$\begin{aligned} R(170.0) &= 6 \\ R(171.8) &= 7 \end{aligned} \Rightarrow Me = \frac{170 + 171.8}{2} = 170.9$$

b) Izračunaj vse mere variabilnosti

← Mere variabilnosti →

Variacijski razmik

$$VR = x_{\max} - x_{\min}$$

Interkvartilni (medkvartilni) razpon

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

Populacijska varianca (u = št. enot v populaciji)

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

Populacijski standardni odklon

$$\sigma$$

Vzorčna varianca (u = velikost vzorca)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Vzorčni standardni odklon

$$s$$

Variacijski razmik: min in max vrednost sta prva in zadnja vrednost v ranžirni vrsti.

$$\begin{aligned} x_{\max} &= 179.6 \\ x_{\min} &= 159.4 \end{aligned}$$

$$VR = 179.6 - 159.4 = 20.2$$

Medkvartilni razpon: najdem prvi in tretji kvartil

$$IQR = 175.5 - 166.5$$

$$R(Q_{1/4}) = 1/4 \cdot 12 + 0.5 = 3.5 \Rightarrow (165.5 + 167.4)/2 = 166.45 \approx 166.5$$

$$R(Q_{3/4}) = 3/4 \cdot 12 + 0.5 = 9.5 \Rightarrow (175.3 + 175.6)/2 = 175.45 \approx 175.5$$

Populacijska varianca: Aritmetično sredino že imam in je $\bar{x} = 173.6$; $n = 12$

$$\sigma^2 = \frac{(159.1 - 173.6)^2 + (163.3 - 173.6)^2 + \dots + (179.6 - 173.6)^2}{12} = 34.6$$

$$s^2 = \frac{(159.1 - 173.6)^2 + (163.3 - 173.6)^2 + \dots + (179.6 - 173.6)^2}{12 - 1} = 7.2$$

Verjetno moram zračunati samo vzorčno varianco, ker imamo **vzorec** 12 krogodilov.

NALOGA 2: Študenti biologije predstavljajo našo populacijo, pisali so izpit. Podani so rezultati v odstotkih urejeni od najslabšega do najboljšega.

a) Izračunaj variacijski in interkvartilni razmik

Fajin je, ker te meen ranžirno vrsto. Variacijski je med min in max, interkvartilni med prvim in tretjim kvartilom.

$$VR = 99 - 12 = 87$$

$$IQR = Q_{3/4} - Q_{1/4} = 58 - 31 = 27$$

1	12, 12, 15, 16, 16, 16, 19, 22, 25, 27,
2	29, 30, 30, 30, 30, 30, 30, 31, 35, 35,
3	36, 37, 39, 39, 39, 40, 40, 41, 44, 44,
4	46, 46, 47, 48, 49, 49, 50, 51, 52, 54,
5	54, 58, 58, 59, 63, 63, 66, 71, 71, 72,
6	72, 72, 74, 77, 81, 84, 85, 86, 86, 89,
7	90, 91, 91, 94, 96, 96, 97, 98, 99, 99.

} $n = 70$

$$\begin{cases} \rightarrow R(Q_{3/4}) = 0.75 \cdot 70 + 0.5 = 53 \Rightarrow Q_{3/4} = 58 \\ \rightarrow R(Q_{1/4}) = 0.25 \cdot 70 + 0.5 = 18 \Rightarrow Q_{1/4} = 31 \end{cases}$$

b) Določi frekvenčno razdelitev v razrede. Širina vseh razredov naj bo enaka

$$\text{Naj bo } d_i = 10: R_1: [10, 20) = \{12, 12, 15, 16, 16, 16, 19\}$$

$$R_2: [20, 30) = \{22, 25, 27, 29\}$$

$$R_3: [30, 40) = \{30, 30, 30, 30, 30, 30, 31, 35, 35, 36, 37, 39, 39, 39\}$$

$$R_4: [40, 50) = \{40, 40, 41, 44, 44, 46, 46, 47, 48, 49, 49\}$$

$$R_5: [50, 60) = \{50, 51, 52, 54, 54, 58, 58, 59\}$$

$$R_6: [60, 70) = \{63, 63, 66\}$$

$$R_7: [70, 80) = \{71, 71, 72, 72, 72, 74, 77\}$$

$$R_8: [80, 90) = \{81, 84, 85, 86, 86, 89\}$$

$$R_9: [90, 100] = \{90, 91, 91, 94, 96, 96, 97, 98, 99, 99\}$$

DISCLAIMER: Zgoraj sem varobe zračunal variance; x_i je SREDINA RAZREDA (15, 25, 35, ..., 95) in ne vrednost na čistem mestu !!!

Fr. razred	f_i
R1	7
R2	4
R3	14
R4	11
R5	8
R6	3
R7	7
R8	6
R9	10

c) Izračunaj tehtano aritmetično sredino in varianco iz frekvenčne porazdelitve

← Tehtana aritmetična sredina →

$$\bar{x} = \frac{f_1 \cdot x_1 + \dots + f_n \cdot x_n}{f_1 + \dots + f_n} = \frac{f_1 \cdot x_1 + \dots + f_n \cdot x_n}{n}$$

n = število elementov

x_i = sredina i -tega razreda ($b-a$)

$$\bar{x} = \frac{7 \cdot 15 + 4 \cdot 25 + \dots + 6 \cdot 85 + 10 \cdot 95}{70} = 54.43$$

Ker imamo populacijo, bom poračunal populacijsko varianco (σ^2).

$$\sigma^2 = \frac{(15 - 54.43)^2 + (25 - 54.43)^2 + \dots + (95 - 54.43)^2}{70}$$

$$\sigma^2 = 85.76$$

d) Položi modus in mediano

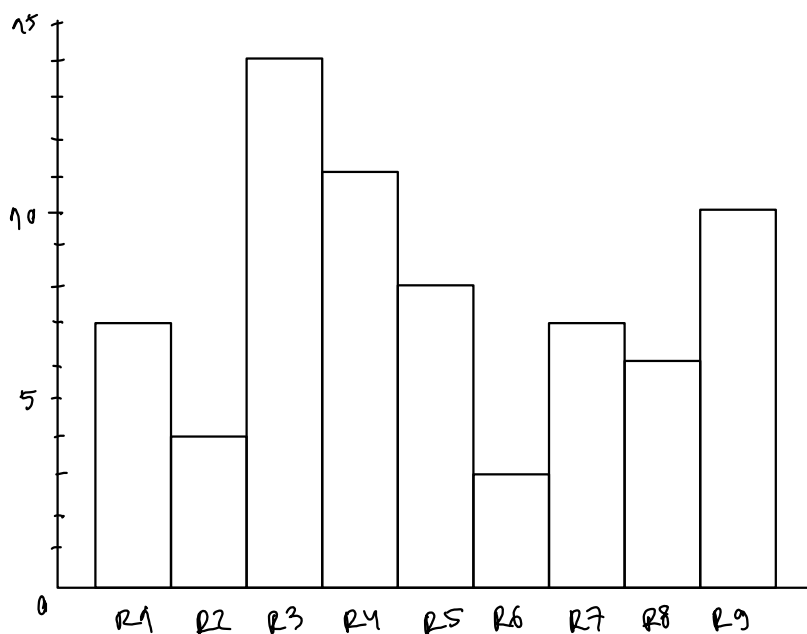
Modus je vrednost z največjo frekvenco (ali je to razred ali specifična vrednost?)

- Modusni razred = R3
- Modusna vrednost = 30 ali 72

$$\text{Mediana: } R(0.5) = 0.5 \cdot 70 + 0.5 = 35.5 \rightarrow \begin{cases} R(39) = 35 \\ R(40) = 36 \end{cases} \Rightarrow Me = 39.5$$

e) Nariši histogram frekvenc. Katere mere je glede na porazdelitev smiselno poročati?

Če je porazdelitev simetrična, je smiselno poročati mere centralne tendence, drugače (asimetrična porazdelitev) pa mere variabilnosti.



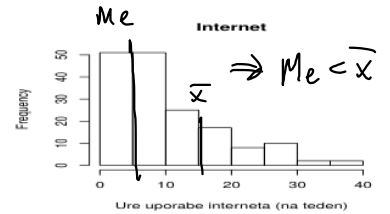
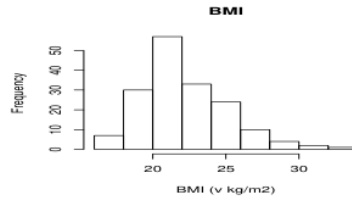
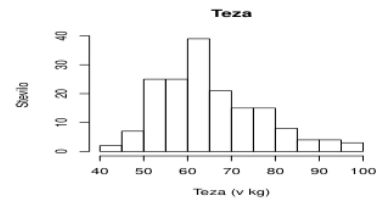
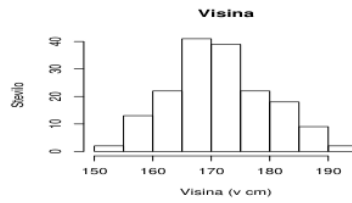
Porazdelitev je precej asimetrična. To lahko vidim tudi numerično:

$$Me = 39.5 \quad \text{in} \quad \bar{x} = 54.43$$

Ko sta si $Me \approx \bar{x}$ lahko rečemo, da je porazdelitev simetrična.

NALOGA 3: Podatki o študentih, ki so obiskovali predmet biostatistika med leti 2008-2010. Katere spremenljivke imajo simetrično (asimetrično) porazdelitev? Za vsake oceni aritmetično sredino in std. odklon, ter povej ali pričakuješ, da je mediana večja/manjša od aritmetične sredine.

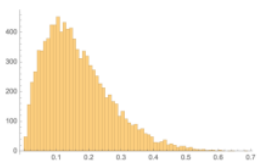
Simetrična	Asimetrična
Višina	Teža
	BMI
	Internet



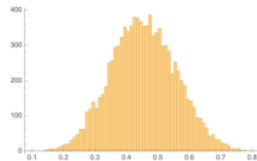
Aritmetična sredina in std. odklon:

	VIŠINA	TEŽA	BMI	INTERNET
\bar{x}	170	70	24	10
σ	majhen	velik	velik	velik

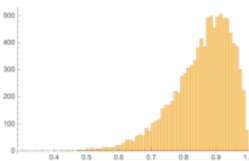
Mediano lahko ocenim po naslednjih dejstvih:



Asimetrična v desno
 $\bar{x} = 0.16$
 $Me = 0.14$
 $\bar{x} > Me$



Približno simetrična
 $\bar{x} = 0.454$
 $Me = 0.452$
 $\bar{x} \approx Me$



Asimetrična v levo
 $\bar{x} = 0.85$
 $Me = 0.87$
 $\bar{x} < Me$

Višina
 Teža
 BMI
 Internet

$Me \approx \bar{x}$
 $Me < \bar{x}$
 $Me < \bar{x}$
 $Me < \bar{x}$

Povejte, katere trditve so pravilne in katere so napačne.

- Teža je porazdeljena simetrično. *Ne*
- Višina je približno normalno porazdeljena. *Da*
- Povprečna in medianska višina sta si zelo podobni. *?* *odvisno od porazdelitve*
- Povprečni BMI je manjši kot medianski BMI. *Ne, obratno.*
- Porazdelitev števila ur tedenske uporabe interneta je negativno asimetrična. *?* *Ja?*
- Standardni odklon višine je približno 20. *Ne, bolj tko 10*
- Povprečje števila ur tedenske uporabe interneta je večje od mediane. *Da*
- Mediana števila ur tedenske uporabe interneta je približno 20 ur. *Ne, tko 5,6*
- Interkvartilni razmik števila ur tedenske uporabe interneta je približno od 15 do 35 ur. *Pa ja...*
- Variacijski razmik višine je od 165 do 185 cm. *Ne, tko 50*
- Približno 50 študentov je tehtalo med 50 in 60 kg. *Da*
- Manj kot 5 študentov je tehtalo več kot 80 kg. *Ne, več*

