

Zpracování náhodného výběru

Zpracování náhodného výběru

- **Př. 1:** Stanovte jednotlivé četnosti a číselné charakteristiky zadaného statistického souboru a nakreslete krabicový graf:
22, 82, 27, 43, 19, 47, 41, 34, 34, 42, 35.
- Statistický soubor seřadíme vzestupně podle velikosti a určíme jednotlivé četnosti.

Zpracování náhodného výběru

v_i	n_i	p_i	m_i	F_i
19	1	1/11	1	1/11
22	1	1/11	2	2/11
27	1	1/11	3	3/11
34	2	2/11	5	5/11
35	1	1/11	6	6/11
41	1	1/11	7	7/11
42	1	1/11	8	8/11
43	1	1/11	9	9/11
47	1	1/11	10	10/11
82	1	1/11	11	1
Σ	11	1		

V tabulce jsou v prvním sloupci uvedeny jednotlivé varianty proměnné, v ostatních sloupcích jednotlivé četnosti.

Zpracování náhodného výběru

- Z tabulky dále vidíme, že:

$$v_{\min} = 19, v_{\max} = 82.$$

- Variační rozpětí R je tedy rovno:

$$R = v_{\max} - v_{\min} = 82 - 19 = 63.$$

- V tomto případě budeme výběrový průměr počítat jako aritmetický průměr:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{11} \cdot (19 + 22 + 27 + 34 + 34 + \dots + 82) \doteq 38,7.$$

Zpracování náhodného výběru

- Modus je ta varianta (příp. varianty) proměnné s největší četností. V našem případě máme modus $Mod = 34$.
- Přistupme nyní ke stanovení kvantilů. Výběrový soubor musíme uspořádat podle velikosti a každému pozorování přiřadíme jeho pořadí. Zaměříme se pouze na dolní kvartil, medián a horní kvartil.

Zpracování náhodného výběru

- Pro pořadí dolního kvartilu platí:

$$z_{0,25} = 11 \cdot 0,25 + 0,5 = 3,25,$$

jelikož pořadí není celé číslo, bude dolním kvartilem aritmetický průměr 3. a 4. hodnoty:

$$x_{0,25} = \frac{27 + 34}{2} = 30,5.$$

x_i	Pořadí
19	1
22	2
27	3
34	4
34	5
35	6
41	7
42	8
43	9
47	10
82	11

Zpracování náhodného výběru

- Mediánem bude hodnota s pořadím:

$$z_{0,5} = 11 \cdot 0,5 + 0,5 = 6,$$

mediánem bude tedy pozorování s pořadím 6,
tedy:

$$x_{0,5} = 35.$$

- Pro horní kvartil platí:

$$z_{0,75} = 11 \cdot 0,75 + 0,5 = 8,75 \Rightarrow x_{0,75} = \frac{42 + 43}{2} = 42,5.$$

Zpracování náhodného výběru

- Pro výběrový rozptyl můžeme psát:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{11-1} \cdot [(19 - 38,7)^2 + \dots + (82 - 38,7)^2] \doteq 284,02.$$

- Výběrovou směrodatnou odchylku stanovíme:

$$s = \sqrt{s^2} = \sqrt{284,02} \doteq 16,85.$$

- Pro variační koeficient platí:

$$V_x = \frac{s}{\bar{x}} = \frac{16,85}{38,7} \doteq 0,44 = 44\%.$$

Zpracování náhodného výběru

- Na základě znalosti dolního a horního kvartilu můžeme stanovit interkvartilové rozpětí:

$$IQR = x_{0,75} - x_{0,25} = 42,5 - 30,5 = 12.$$

- Nyní přistupme ke stanovení mediánu absolutních odchylek od mediánu. Spočítáme tedy absolutní hodnoty odchylek jednotlivých pozorování od mediánu a stanovíme jejich medián.

Zpracování náhodného výběru

x_i	Pořadí	$ x_i - x_{0,5} $	Pořadí
19	1	16	10
22	2	13	9
27	3	8	7
34	4	1	3
34	5	1	2
35	6	0	1
41	7	6	4
42	8	7	5
43	9	8	6
47	10	12	8
82	11	47	11

$$z_{0,5} = 11 \cdot 0,5 + 0,5 = 6,$$

medián absolutních odchylek od mediánu bude tedy hodnota s pořadím 6, tedy:

$$MAD = 8.$$

Zpracování náhodného výběru

- Nyní se pokusíme identifikovat odlehlá pozorování.
- ad 1) Identifikace podle vnitřních hradeb:
 - Dolní hradba $x_{0,25} - 1,5 \cdot IQR = 30,5 - 1,5 \cdot 12 = 12,5$.
 - Horní hradba $x_{0,75} + 1,5 \cdot IQR = 42,5 + 1,5 \cdot 12 = 60,5$.

Vidíme, že mimo vnitřní hradby leží pouze pozorování $x_{11} = 82$, může být tedy identifikováno jako odlehlé.

Zpracování náhodného výběru

- ad 2) Identifikace pomocí z-souřadnice:
 - Spočítejme z-souřadnici pro 1. pozorování:

$$z - \text{souř.}_1 = \frac{19 - 38,7}{16,85} \doteq -1,17 \Rightarrow |z - \text{souř.}_1| \leq 3,$$

můžeme tedy tvrdit, že zleva nebudeme mít žádné odlehlé pozorování.

- Spočítejme z-souřadnici pro 11. pozorování:

$$z - \text{souř.}_{11} = \frac{82 - 38,7}{16,85} \doteq 2,57 \Rightarrow |z - \text{souř.}_{11}| \leq 3,$$

můžeme tedy tvrdit, že ani zprava nebudeme mít žádné odlehlé pozorování.

Zpracování náhodného výběru

- ad 3) Identifikace pomocí $x_{0,5}$ -souřadnice:
 - Spočítejme $x_{0,5}$ -souřadnici pro 1. pozorování:

$$x_{0,5} - \text{souř.}_1 = \frac{19 - 35}{1,483 \cdot 8} \doteq -1,35 \Rightarrow |x_{0,5} - \text{souř.}_1| \leq 3,$$

zleva tedy nebudeme mít žádné odlehlé pozorování.

- Spočítejme $x_{0,5}$ -souřadnici pro 11. pozorování:

$$x_{0,5} - \text{souř.}_{11} = \frac{82 - 35}{1,483 \cdot 8} \doteq 3,96 \Rightarrow |x_{0,5} - \text{souř.}_{11}| > 3,$$

pozorování $x_{11} = 82$ může být identifikováno jako odlehlé.

Zpracování náhodného výběru

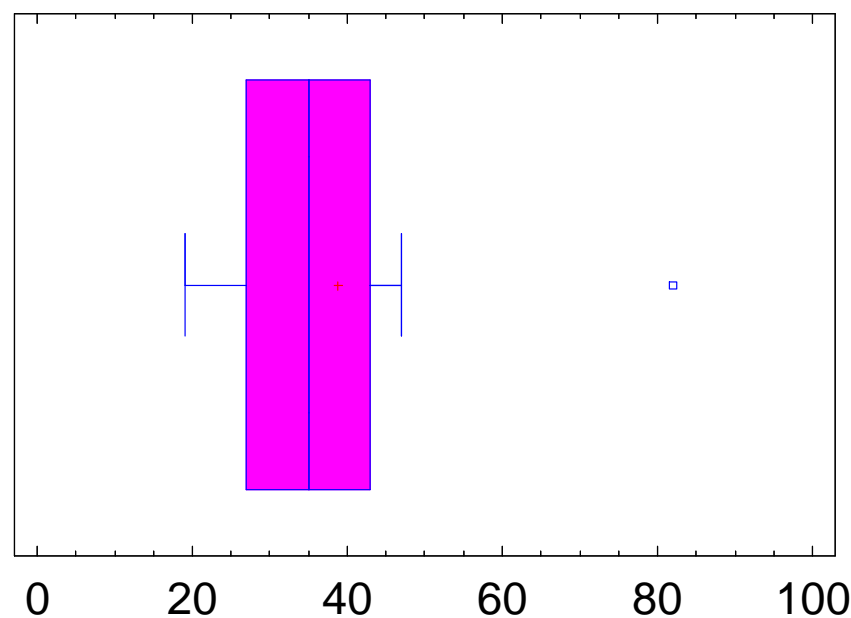
– Ověříme ještě mediánovou souřadnici 10. pozorování:

$$x_{0,5} - \text{souř.}_{10} = \frac{47 - 35}{1,483 \cdot 8} \doteq 1,01 \Rightarrow |x_{0,5} - \text{souř.}_1| \leq 3,$$

10. pozorování již odlehlé není, zprava už žádné další odlehlé pozorování nemáme.

Zpracování náhodného výběru

- Nyní přistupme k nakreslení krabicového grafu. Při jeho sestavě budeme uvažovat s tím, že pozorování 82 jsme identifikovali jako odlehlé.



Zpracování náhodného výběru

- **Př. 2:** V systému hromadné obsluhy byla sledována doba obsluhy v [min]. Rozdělte získaný statistický soubor do tříd, vypočtete základní číselné charakteristiky (výběrový průměr, rozptyl ve vážené formě) a data znázorněte pomocí histogramu.

Zpracování náhodného výběru

- Statistický soubor:

68, 71, 71, 78, 82, 82, 87, 91, 92, 92, 95, 97, 102,
102, 102, 103, 105, 105, 109, 110, 111, 111, 111,
112, 112, 114, 114, 114, 115, 116, 118, 119, 121,
122, 122, 124, 126, 131, 133, 137.

- Rozsah souboru $n = 40$.

Zpracování náhodného výběru

- Ze zadaného statistického souboru vyčteme, že nejnižší varianta proměnné $v_{min} = 68$, nejvyšší varianta proměnné $v_{max} = 137$, pro variační rozpětí tedy platí:

$$R = v_{max} - v_{min} = 137 - 68 = 69.$$

- Počet tříd odhadneme pomocí Sturgesova pravidla:

$$k \approx 1 + 3,3 \cdot \log n = 1 + 3,3 \cdot \log 40 \doteq 6,29 \doteq 6.$$

Zpracování náhodného výběru

- Šířku třídy stanovíme podle vztahu:

$$h \approx \frac{R}{k} = \frac{69}{6} = 11,5 \doteq 12.$$

- Vynásobíme-li počet tříd k se stanovenou šířkou třídy h , dostaneme 72, tedy o 3 více než je variační rozpětí. Nyní musíme vhodně zvolit dolní hranici 1. třídy, zvolme ji o 1,5 nižší než je minimální varianta proměnné v_{min} , tedy 66,5.

Zpracování náhodného výběru

Index třídy	Třída	Třídní znak	Četnost n_i
1	(66,5; 78,5›	72,5	4
2	(78,5; 90,5›	84,5	3
3	(90,5; 102,5›	96,5	8
4	(102,5; 114,5›	108,5	13
5	(114,5; 126,5›	120,5	9
6	(126,5; 138,5›	132,5	3
Σ			40

U jednotlivých tříd volíme interval zleva otevřený, zprava uzavřený, s takovým intervalem pracuje i Excel.

Zpracování náhodného výběru

- Nyní přistoupíme k výpočtu výběrového průměru a rozptylu ve vážené formě:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^k n_i \cdot z_i = \frac{1}{40} \cdot (4 \cdot 72,5 + 3 \cdot 84,5 + \dots + 3 \cdot 132,5) \doteq 105,20,$$

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^k n_i \cdot (z_i - \bar{x})^2 = \frac{1}{40-1} \cdot$$

$$\cdot \left[4 \cdot (72,5 - 105,20)^2 + \dots + 3 \cdot (132,5 - 105,20)^2 \right] \doteq 273,14.$$

Zpracování náhodného výběru

- Jelikož se při výpočtu výběrového průměru a rozptylu ve vážené formě nahrazují všechna pozorování v jedné třídě jednou zástupnou hodnotou – třídním znakem, dopouštíme se samozřejmě nepřesností, viz tabulka.

	Z původních dat	Z roztříděných dat
Výběrový průměr	105,68	105,20
Výběrový rozptyl	299,97	273,14

Zpracování náhodného výběru

