

# 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
$\Sigma$	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

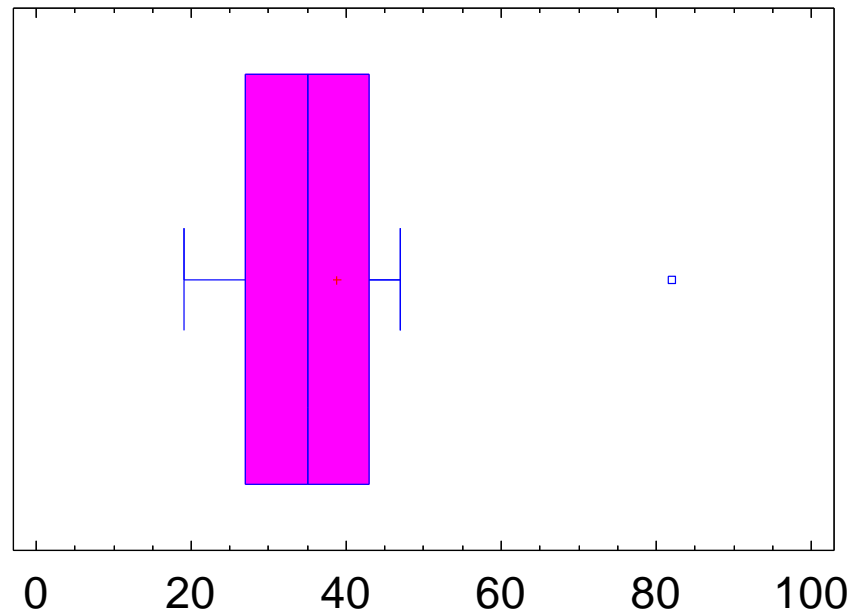
- Ověřme ještě mediánovou souřadnici 10. pozorování:

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

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:** Měřením doby obsluhy v systému hromadné obsluhy byly naměřeny níže uvedené hodnoty v [min]:  
1,2,2,2,3,3,4,5,6,6,7,9,12,21.  
Získaný výběrový soubor zpracujte pomocí metod popisné statistiky.



# Zpracování náhodného výběru

$i$	$x_i$
1	1
2	2
3	2
4	2
5	3
6	3
7	4
8	5
9	6
10	6
11	7
12	9
13	12
14	21
$\Sigma$	83

Index varianty proměnné	Varianta proměnné	Absolutní četnost	Kumulativní četnost	Relativní četnost	Relativní kumulativní četnost
$j$	$x_j$	$n_j$	$m_j$	$p_j$	$F_j$
1	1	1	1	1/14	1/14
2	2	3	4	3/14	2/7
3	3	2	6	1/7	3/7
4	4	1	7	1/14	1/2
5	5	1	8	1/14	4/7
6	6	2	10	1/7	5/7
7	7	1	11	1/14	11/14
8	9	1	12	1/14	6/7
9	12	1	13	1/14	13/14
10	21	1	14	1/14	1
$\Sigma$		14		1	

# Zpracování náhodného výběru

- Míry polohy

– Výběrový průměr:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1+2+2+2+3+3+4+5+6+6+7+9+12+21}{14} =$$
$$= \frac{83}{14} \doteq 5,93 \text{ min.}$$

# Zpracování náhodného výběru

- Modus (varianta proměnné s největší absolutní četností):

$$Mod = 2 .$$

- Výběrové kvantily:

- Dolní kvartil:

$$z_{0,25} = 14 \cdot 0,25 + 0,5 = 4 \Rightarrow x_{0,25} = 2 \text{ min.}$$

- Medián:

$$z_{0,5} = 14 \cdot 0,5 + 0,5 = 7,5 \Rightarrow x_{0,5} = \frac{4+5}{2} = 4,5 \text{ min.}$$

- Horní kvartil:

$$z_{0,75} = 14 \cdot 0,75 + 0,5 = 11 \Rightarrow x_{0,75} = 7 \text{ min.}$$

# Zpracování náhodného výběru

- Míry variability

- Variační rozpětí:

$$R = x_{\max} - x_{\min} = 20 \text{ min.}$$

- Interkvartilové rozpětí:

$$IQR = x_{0,75} - x_{0,25} = 5 \text{ min.}$$

# Zpracování náhodného výběru

– Výběrový rozptyl:

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \\ &= \frac{(1-5,93)^2 + \dots + (21-5,93)^2}{14-1} = \\ &= \frac{366,93}{13} = 28,23 \text{ min}^2. \end{aligned}$$

$i$	$x_i$	$(x_i - x_{\text{prům}})^2$
1	1	24,29
2	2	15,43
3	2	15,43
4	2	15,43
5	3	8,58
6	3	8,58
7	4	3,72
8	5	0,86
9	6	0,01
10	6	0,01
11	7	1,15
12	9	9,43
13	12	36,86
14	21	227,15
$\Sigma$	83	366,93

# Zpracování náhodného výběru

– Výběrová směrodatná odchylka:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \\ = \sqrt{28,23} = 5,31 \text{ min.}$$

– Variační koeficient:

$$V_x = \frac{s}{\bar{x}} = \frac{5,31}{5,93} = 0,90.$$

# Zpracování náhodného výběru

– Medián absolutních odchylek od mediánu:

$$z_{0,5} = 14 \cdot 0,5 + 0,5 = 7,5 \Rightarrow$$

$$\Rightarrow MAD = \frac{2,5 + 2,5}{2} = 2,5 \text{ min.}$$

$i$	$x_i$	$ x_i - x_{0,5} $	$ x_i - x_{0,5} $ seřazené
1	1	3,5	0,5
2	2	2,5	0,5
3	2	2,5	1,5
4	2	2,5	1,5
5	3	1,5	1,5
6	3	1,5	1,5
7	4	0,5	2,5
8	5	0,5	2,5
9	6	1,5	2,5
10	6	1,5	2,5
11	7	2,5	3,5
12	9	4,5	4,5
13	12	7,5	7,5
14	21	16,5	16,5
$\Sigma$	83		

# Zpracování náhodného výběru

- Identifikace odlehlých pozorování:

- Pomocí „vnitřních hradeb“:

Dolní hradba -  $x_{0,25} - 1,5IQR = 2 - 1,5 \cdot 5 = -5,5$  min .

Horní hradba -  $x_{0,75} + 1,5IQR = 7 + 1,5 \cdot 5 = 14,5$  min .

Odlehlým pozorováním by byla hodnota proměnné  $x_{14} = 21$ .



# Zpracování náhodného výběru

– Pomocí z-souřadnice:

$$z\text{-souř}_1 = \frac{x_1 - \bar{x}}{s} = \frac{1 - 5,93}{5,31} = -0,93.$$

$$z\text{-souř}_{14} = \frac{x_{14} - \bar{x}}{s} = \frac{21 - 5,93}{5,31} = 2,83.$$

Žádné pozorování nelze označit jako odlehlé.

# Zpracování náhodného výběru

– Pomocí  $x_{0,5}$ -souřadnice:

$$x_{0,5} - \text{souř}_1 = \frac{x_1 - x_{0,5}}{1,483 \cdot MAD} = \frac{1 - 4,5}{1,483 \cdot 2,5} = -0,94.$$

$$z - \text{souř}_{14} = \frac{x_{14} - x_{0,5}}{1,483 \cdot MAD} = \frac{21 - 4,5}{1,483 \cdot 2,5} = 4,45.$$

$$z - \text{souř}_{13} = \frac{x_{13} - x_{0,5}}{1,483 \cdot MAD} = \frac{12 - 4,5}{1,483 \cdot 2,5} = 2,02.$$

Pozorování  $x_{14} = 21$  lze označit jako odlehlé.

# Zpracování náhodného výběru

## Summary Statistics for Doba obsluhy

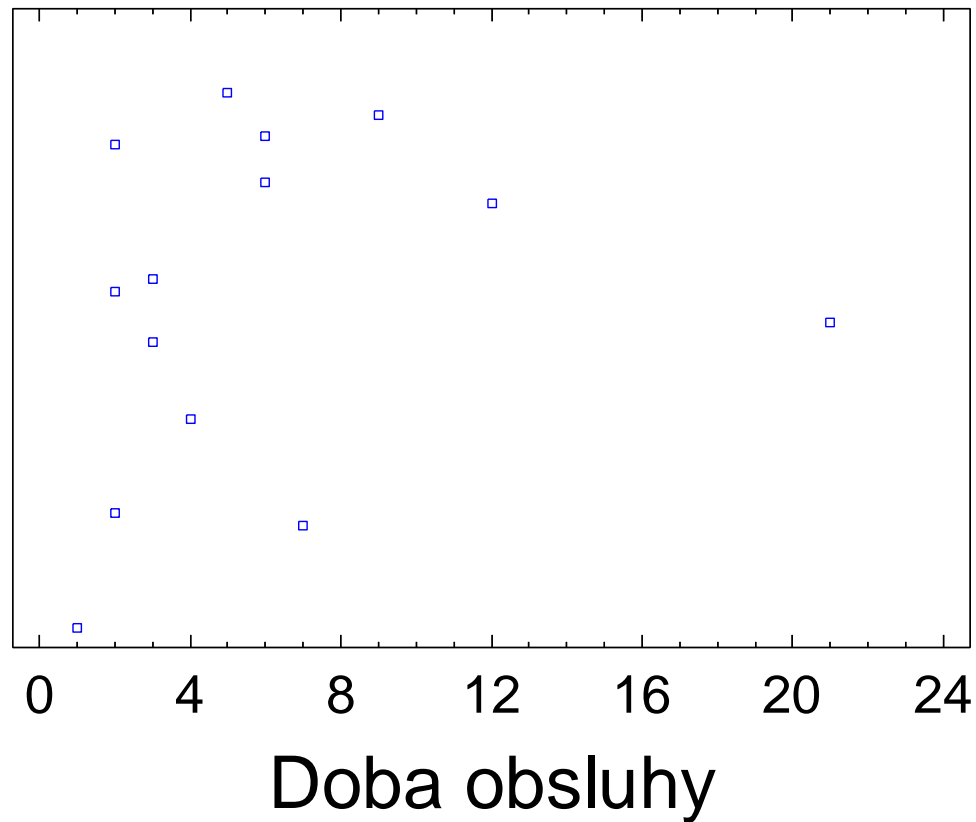
Count = 14  
Average = 5,92857  
Variance = 28,2253  
Standard deviation = 5,31275  
Minimum = 1,0  
Maximum = 21,0  
Range = 20,0  
Std. skewness = 3,03293  
Std. kurtosis = 3,46194

## Percentiles for Doba obsluhy

1,0% = 1,0  
5,0% = 1,0  
10,0% = 2,0  
25,0% = 2,0  
50,0% = 4,5  
75,0% = 7,0  
90,0% = 12,0  
95,0% = 21,0  
99,0% = 21,0

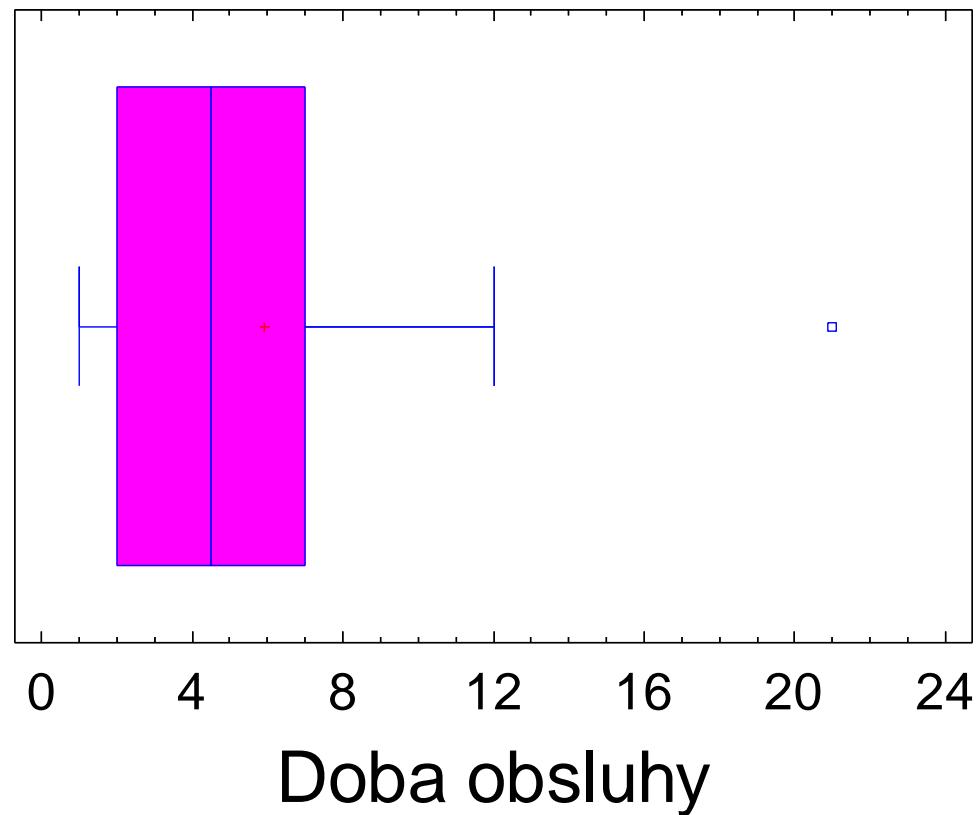
# Zpracování náhodného výběru

## Scatterplot



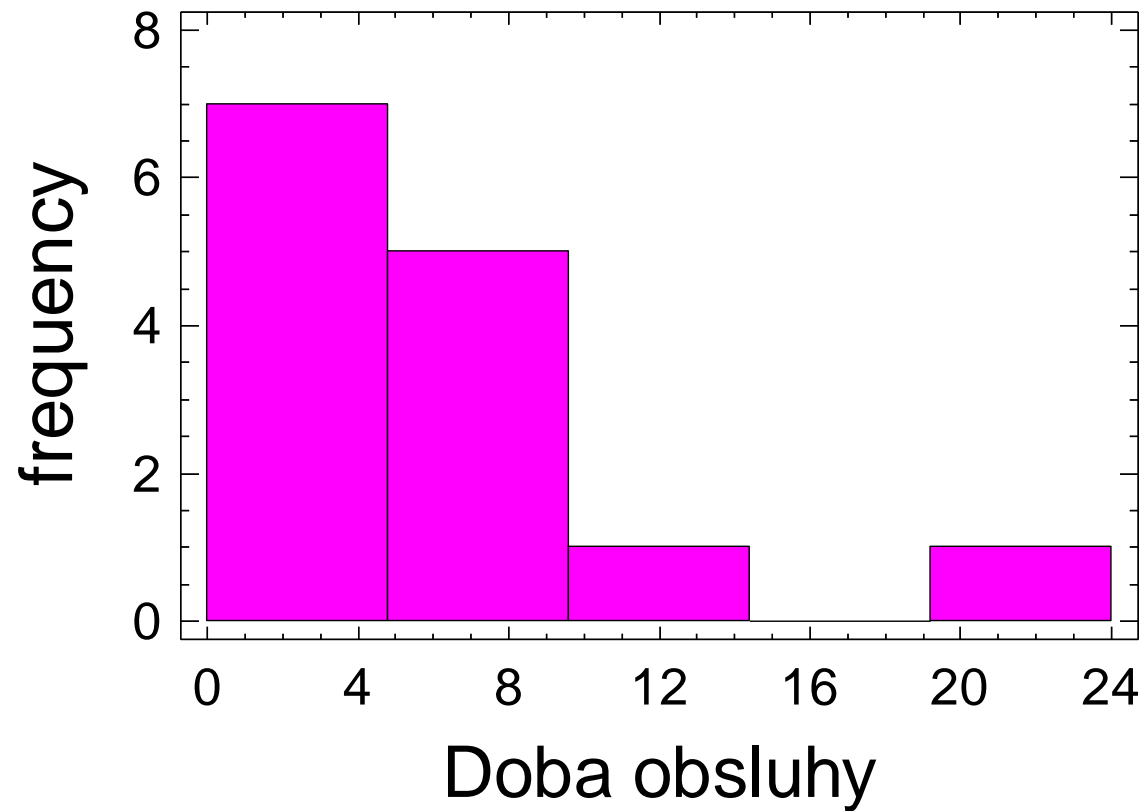
# Zpracování náhodného výběru

## Box-and-Whisker Plot



# Zpracování náhodného výběru

## Histogram



# Zpracování náhodného výběru

- **Př. 3:** 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
$\Sigma$			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 roztržiených dat
Výběrový průměr	105,68	105,20
Výběrový rozptyl	299,97	273,14

# Zpracování náhodného výběru

