

### 3 Základní číselné charakteristiky

Tabulka 1: Přehled číselných charakteristik podle typu znaku a sledované vlastnosti

	Poloha	Variabilita	Symetrie	Závislost
Nominální	modus	–	–	Cramérův koeficient
Ordinální	medián	interkvartilové rozpětí	–	Spearmanův koef. poř. korel.
Intervalový	aritmetický průměr	rozptyl směrodatná odchylka	koeficient šikmosti koeficient špičatosti	Pearsonův korel. koeficient

#### 3.1 Číselné charakteristiky pro nominální znaky

##### Příklad 3.1. Charakteristika polohy nominálního znaku

Navažme na práci s datasetem 17-anova-newborns-2.txt. V rámci cvičení 2 jsme jako mezivýstup příkladu 2.5 získali kontingenční tabulku simultánních absolutních četností znaků  $X = \text{vzdělání matky}$  a  $Y = \text{porodní hmotnost novorozence}$  (viz tabulka 2). Znaky  $X$  a  $Y$  jsou typickým příkladem znaků nominálního typu. Najděte modus pro znak *vzdělání matky* i pro znak *porodní hmotnost novorozence*.

Tabulka 2: Simultánní absolutní četnosti pro znaky *vzdělání matky* a *porodní hmotnost novorozence*

	nížká	norma	vysoká
ZS	97	312	8
SS	82	346	20
SSm	74	349	12
VS	13	64	4

##### Řešení příkladu 3.1

	nizka	norma	vysoka
ZS	97	312	8
SS	82	346	20
SSm	74	349	12
VS	13	64	4

1  
2  
3  
4  
5

Zaměříme se nejprve na znak  $X = \text{vzdělání matky}$ . Číselná charakteristika *modus* je definována jako nejčetnější varianta sledovaného znaku.

ZS	SS	SSm	VS
417	448	435	81

6  
7

**Interpretace výsledků:** Nejčetnější variantou znaku *vzdělání matky* je ..... ( $n = \dots$ ).  
Nejvíce novorozenců v datovém souboru se narodilo matkám s dokončeným .....

Analogicky nyní najdeme modus znaku  $Y = \text{porodní hmotnost novorozence}$ .

nizka	norma	vysoka
266	1071	44

8  
9

**Interpretace výsledků:** Nejvíce novorozenců v datovém souboru mělo porodní hmotnost .....  
( $n = \dots$ ).



### Příklad 3.2. Charakteristika závislosti mezi dvěma nominálními znaky

Zaměřte se nyní na oba znaky  $X = \text{vzdělání}$  a  $Y = \text{porodní hmotnost novorozence}$  najednou. Určete míru závislosti mezi znaky  $X$  a  $Y$ .

#### Řešení příkladu 3.2

Protože  $X$  a  $Y$  jsou znaky ..... typu, použijeme na určení míry závislosti mezi nimi ..... Stupnice míry závislosti podle hodnoty Cramérova koeficientu je uvedena v tabulce 3

Tabulka 3: Stupnice míry závislosti podle Cramérova koeficientu

Cramérův koeficient	Interpretace
0.0 – 0.1	zanedbatelný stupeň závislosti
0.1 – 0.3	slabý stupeň závislosti
0.3 – 0.7	střední stupeň závislosti
0.7 – 1.0	silný stupeň závislosti

[1] 0.0645725

10

**Interpretace výsledků:** Hodnota Cramérova koeficientu vyšla ..... Mezi vzděláním matky a porodní hmotností novorozence existuje ..... stupeň .....



## 3.2 Číselné charakteristiky pro ordinální znaky

### Příklad 3.3. Základní číselné charakteristiky pro ordinální znak

Načtete datový soubor 17-anova-newborns-2.txt a odstraňte neznámé hodnoty. Zaměřte se pouze na novorozence ženského pohlaví s vysokou porodní hmotností. Zjistěte dimenzi datové tabulky obsahující údaje o těchto novorozencích a vytvořte tabulku základních číselných charakteristik pro znak  $X = \text{počet starších sourozenců}$ .

#### Řešení příkladu 3.3

[1] 14 5

11

Po odstranění neznámých hodnot obsahuje datová tabulka údaje o ..... novorozencích ženského pohlaví s vysokou porodní hmotností, přičemž u každého z těchto novorozenců máme záznamy o ..... znacích.

Znak  $X = \text{počet starších sourozenců novorozence}$  je příkladem ..... dat. V tabulce základních charakteristik budou obsaženy následující charakteristiky: minimální hodnota, dolní kvartil, medián, horní kvartil, maximální hodnota a interkvartilové rozpětí.

1. minimální hodnota  $x_{min} = \dots\dots\dots$
2. dolní kvartil  $x_{0.25}$ 
  - $n = \dots\dots\dots$ ,  $\alpha = \dots\dots\dots$
  - $\alpha \times n = \dots\dots\dots \rightarrow$  je / není celé číslo
  - $x_{0.25} = \dots\dots\dots$

3. medián  $x_{0.50}$

- $n = \dots, \alpha = \dots$
- $\alpha \times n = \dots \rightarrow$  je / není celé číslo
- $x_{0.50} =$

4. horní kvartil  $x_{0.75}$

- $n = \dots, \alpha = \dots$
- $\alpha \times n = \dots \rightarrow$  je / není celé číslo
- $x_{0.75} =$

5. maximální hodnota  $x_{max} = \dots$

6. interkvartilové rozpětí  $IQR = x_{0.75} - x_{0.25} = \dots$

	min	dolní.kv	median	horní.kv	max	IQR
25%	0	0	1	2	3	2

12  
13

**Interpretace výsledků:** Počet starších sourozenců u novorozenců ženského pohlaví s vysokou porodní hmotností se pohyboval v rozmezí  $\dots - \dots$

Dolní kvartil počtu starších sourozenců nabývá hodnoty  $\dots$ , tj.  $\dots$  % novorozenců ženského pohlaví s vysokou porodní hmotností má  $\dots$  starších sourozenců.

Medián počtu starších sourozenců nabývá hodnoty  $\dots$ , tj.  $\dots$  % novorozenců ženského pohlaví s vysokou porodní hmotností má  $\dots$  nebo  $\dots$  starších sourozenců.

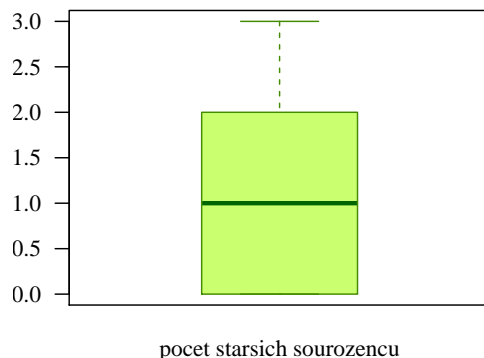
Horní kvartil počtu starších sourozenců nabývá hodnoty  $\dots$ , tj.  $\dots$  % novorozenců ženského pohlaví s vysokou porodní hmotností má  $\dots$ ,  $\dots$  nebo  $\dots$  starších sourozenců. Hodnota interkvartilového rozpětí je rovna  $\dots$



**Příklad 3.4. Krabicový diagram**

Sestrojte krabicový diagram pro znak  $X =$  počet starších sourozenců pro novorozence ženského pohlaví s vysokou porodní hmotností. Zaměřte se na vzhled krabicového diagramu a zamyslete se nad tím, kde je v krabicovém diagramu zobrazen medián, dolní kvartil, horní kvartil a interkvartilové rozpětí.

**Řešení příkladu 3.4**



### Příklad 3.5. Charakteristika závislosti mezi ordinálními znaky

Zaměřme se nyní na oba znaky  $X = \text{počet starších sourozenců}$  a  $Y = \text{porodní hmotnost novorozence}$  najednou. Určete míru závislosti mezi znaky  $X$  a  $Y$  u novorozenců ženského pohlaví.

#### Řešení příkladu 3.5

Znak  $X$  je ..... typu, zatímco znak  $Y$  je ..... typu  $\rightarrow$  ke znaku  $Y$  budeme přistupovat jako ke znaku ..... typu. Ke stanovení míry závislosti použijeme ..... koeficient ..... korelace. Stupnice těsnosti závislosti mezi dvěma znaky podle hodnoty Spearmanova koeficientu pořadové korelace je uvedena v tabulce 4.

Tabulka 4: Stupnice míry závislosti podle Spearmanova a Pearsonova korelačního koeficientu

Spearmanův (Pearsonův) koeficient	Interpretace
0.0	pořadová (lineární) nezávislost
0.0 – 0.1	velmi nízký stupeň závislosti
0.1 – 0.3	nízký stupeň závislosti
0.3 – 0.5	mírný stupeň závislosti
0.5 – 0.7	význačný stupeň závislosti
0.7 – 0.9	vysoký stupeň závislosti
0.9 – 1.0	velmi vysoký stupeň závislosti
1.0	úplná pořadová (lineární) závislost

```
[1] 0.05918883
```

14

**Interpretace výsledku:** Hodnota Spearmanova koeficientu pořadové korelace vyšla ..... Mezi počtem starších sourozenců a porodní hmotností novorozence ženského pohlaví existuje ..... stupeň ..... závislosti. ★

### 3.3 Číselné charakteristiky pro intervalové znaky

#### Příklad 3.6. Základní číselné charakteristiky pro intervalový znak

Načtete datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Zaměřte se pouze na znak  $X = \text{největší šířka mozkovny}$  pro skelety mužského pohlaví. Vytvořte tabulku základních číselných charakteristik pro znak  $X$ .

#### Řešení příkladu 3.6

```
id pop sex skull.L skull.B
1 416 elegant m 188 145
2 417 elegant m 172 139
3 420 elegant m 176 138
```

15  
16  
17  
18

Po odstranění neznámých hodnot obsahuje datová tabulka údaje o ..... skeletech mužského pohlaví. Znak  $X = \text{největší šířka mozkovny}$  pro skelety mužského pohlaví je příkladem ..... dat. V tabulce základních číselných charakteristik budou obsaženy následující charakteristiky: aritmetický průměr, směrodatná odchylka, minimální hodnota, dolní kvartil, medián, horní kvartil, maximální hodnota, interkvartilové rozpětí, koeficient šikmosti a koeficient špičatosti.

1. aritmetický průměr  $m$

•  $m = \frac{1}{n} \sum_{i=1}^n x_i =$

2. rozptyl  $s^2$

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

3. směrodatná odchylka  $s$

•  $s = \sqrt{s^2} =$

4. minimální hodnota  $x_{min} = \dots\dots\dots$

5. dolní kvartil  $x_{0.25}$

•  $n = \dots\dots\dots, \alpha = \dots\dots\dots$

•  $\alpha \times n = \dots\dots\dots \rightarrow$  je / není celé číslo  $\rightarrow$

•  $x_{0.25} =$

6. medián  $x_{0.50}$

•  $n = \dots\dots\dots, \alpha = \dots\dots\dots$

•  $\alpha \times n = \dots\dots\dots \rightarrow$  je / není celé číslo  $\rightarrow$

•  $x_{0.50} =$

7. horní kvartil  $x_{0.75}$

•  $n = \dots\dots\dots, \alpha = \dots\dots\dots$

•  $\alpha \times n = \dots\dots\dots \rightarrow$  je / není celé číslo  $\rightarrow$

•  $x_{0.75} =$

8. interkvartilové rozpětí  $IQR = x_{0.75} - x_{0.25} = \dots\dots\dots$

9. maximální hodnota  $x_{max} = \dots\dots\dots$

10. koeficient šikmosti  $b_1$

•  $b_1 = \dots\dots\dots$

11. koeficient špičatosti  $b_2$

•  $b_2 = \dots\dots\dots$

	m	s	min	dolní.kv	median	horní.kv	max	IQR	šikmost	špicatost
muzi	137.19	4.81	124	134	137	140	149	6	0.08	-0.3

**Interpretace výsledků:** Naměřené hodnoty největší šířky mozkovny pro skelety mužského pohlaví se pohybují v rozmezí ..... – ..... mm. Průměrná hodnota největší šířky mozkovny je ..... mm se směrodatnou odchylkou ..... mm. 25% naměřených hodnot je menších nebo rovných ..... mm, 50% naměřených hodnot je menších nebo rovných ..... mm, 75% naměřených hodnot je menších nebo rovných ..... mm. Interkvartilové rozpětí naměřených hodnot je rovno ..... Hodnota koeficientu šikmosti, ....., ukazuje na ..... zešikmená data (prodloužený ..... konec). Hodnota koeficientu špičatosti, ....., ukazuje na ..... charakter dat. ★

**Příklad 3.7. Charakteristika závislosti pro znaky intervalového typu**

Zaměříme se nyní na znaky  $X =$  největší šířka mozkovny a  $Y =$  největší délka mozkovny pro skelety mužského pohlaví najednou. Určete míru závislosti mezi znaky  $X$  a  $Y$ .

**Řešení příkladu 3.7**

Oba znaky  $X$  a  $Y$  jsou ..... typu. Ke stanovení míry závislosti použijeme ..... korelační koeficient. Stupnice těsnosti závislosti mezi dvěma znaky podle hodnoty Pearsonova korelačního koeficientu je uvedena výše v tabulce 4.

[1] 0.168157

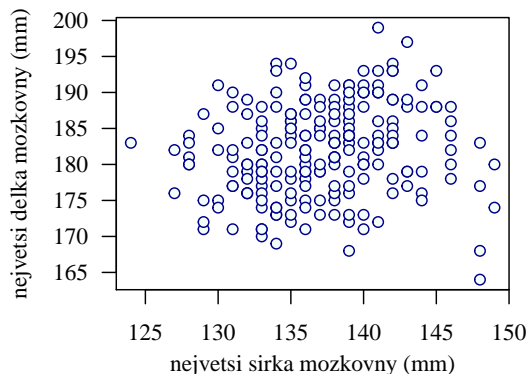
21

**Interpretace výsledků:** Pearsonův korelační koeficient nabývá hodnoty ..... Mezi největší šířkou mozkovny a největší délkou mozkovny pro skelety mužského pohlaví existuje ..... stupeň ..... závislosti. ★

**Příklad 3.8. Dvourozměrný tečkový diagram**

Pro znaky  $X =$  největší šířka mozkovny a  $Y =$  největší délka mozkovny u mužů vykreslete dvourozměrný tečkový diagram.

**Řešení příkladu 3.8**



★