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.txt. V rámci sekce 2 jsme jako mezivýstup příkladu 2.5 získali kontingenční tabulku simultánních absolutních četností znaků $X =$ *vzdělání matky* a $Y =$ *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*

	nízká	norma	vysoká
ZS	75	264	8
SS	79	325	20
SSm	73	341	11
VS	13	63	4

Řešení příkladu 3.1

```
    nizka norma vysoka
ZS     75   264      8
SS     79   325     20
SSm    73   341     11
VS     13    63      4
```

Zaměřme se nejprve na znak $X =$ *vzdělání matky*. Číselná charakteristika *modus* je definována jako nejčetnější varianta sledovaného znaku.

```
 ZS  SS SSm  VS
347 424 425  80
```

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

Analogicky nyní najdeme modus znaku $Y =$ *porodní hmotnost novorozence*.

```
 nizka  norma vysoka
   240    993     43
```

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

Příklad 3.2. Charakteristika závislosti mezi dvěma nominálními znaky
Zaměřte se nyní na oba znaky X = *vzdělání* a Y = *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: Stunice 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] 0.05502639
```
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
Načtěte datový soubor 17-anova-newborns.txt a odstraňte neznámé hodnoty. Zaměřte se na novorozence s vysokou porodní hmotností (větší než 4 200 g). Zjistěte dimenzi datové tabulky s novorozenci s vysokou porodní hmotností. Vytvořte tabulku základních číselných charakteristik pro znak X = *počet starších sourozenců* pro tyto novorozence.

Řešení příkladu 3.3

```
[1] 44  4
```
11

Po odstranění neznámých hodnot obsahuje datová tabulka údaje o .... novorozencích s vysokou porodní hmotností, přičemž u každého novorozence máme záznamy o .... znacích.

Znak X = *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} = ...
2. Dolní kvartil $x_{0.25}$
   - n = ...
   - α = ...
   - $\alpha \times n$ = ... $\rightarrow$ je / není celé číslo

   - $x_{0.25}$ =

3. Medián $x_{0.50}$
    - $n =$ ....
    - $\alpha =$ ....
    - $\alpha \times n =$ .... $\rightarrow$ je / není celé číslo
    - $x_{0.50} =$

4. Horní kvartil $x_{0.75}$
    - $n =$ ....
    - $\alpha =$ ....
    - $\alpha \times n =$ .... $\rightarrow$ je / není celé číslo
    - $x_{0.75} =$

5. Maximální hodnota $x_{max} =$ ....

6. Interkvartilové rozpětí $IQR = x_{0.75} - x_{0.25} =$ ....

```
    min dolni.kv median horni.kv max IQR
25%   0      0.5      1        2   5 1.5
```

Interpretace výsledků: Počet starších sourozenců u novorozenců s vysokou porodní hmotností v datovém souboru se pohyboval v rozmezí .... – ....

Dolní kvartil počtu starších sourozenců u novorozenců s vysokou p.h. v datovém souboru nabývá hodnoty ...., tj. .... % novorozenců v datovém souboru má .... starších sourozenců.

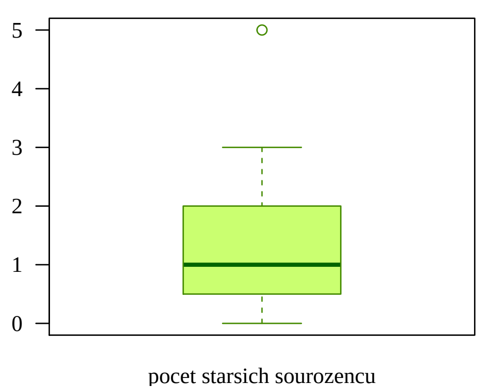
Medián počtu starších sourozenců u novorozenců s vysokou p.h. v datovém souboru nabývá hodnoty ...., tj. .... % novorozenců v datovém souboru má .... nebo .... starších sourozenců.

Horní kvartil počtu starších sourozenců u novorozenců s vysokou p.h. v datovém souboru nabývá hodnoty ...., tj. .... % novorozenců v datovém souboru má ...., ... nebo .... starších sourozenců. Rozsah interkvartilového rozpětí je roven ....

Příklad 3.4. Krabicový diagram

Sestrojte krabicový diagram pro znak X = *počet starších sourozenců* novorozence 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 mezikvartilové rozpětí.

Řešení příkladu 3.4



Příklad 3.5. Charakteristika závislosti mezi ordinálními znaky
Zaměřme se nyní na oba znaky X = *počet starších sourozenců* a Y = *porodní hmotnost novorozence* najednou. Určete míru závislosti mezi znaky X a Y.

Řešení příkladu 3.5
Znak X je .... typu, zatímco znak Y je .... typu → 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

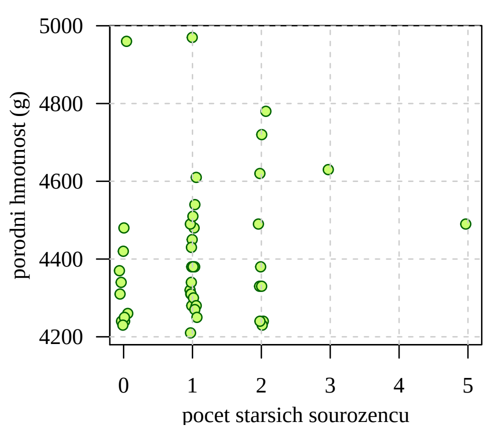
Spearmanův (Pearsonův) koeficient	**Interpretace**
0.0	Pořadová (lineární) nezávislost
0.0 – 0.1	Velmi nízký stupeň závislosti
0.1 – 0.3	Nízký stupeň závislosti
0.3 – 0.5	Mírný stupeň závislosti
0.5 – 0.7	Význačný stupeň závislosti
0.7 – 0.9	Vysoký stupeň závislosti
0.9 – 1.0	Velmi vysoký stupeň závislosti
1.0	Úplná pořadová (lineární) závislost

```
[1] 0.2428544
```
14

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

Příklad 3.6. Dvourozměrný tečkový diagram
Pro znaky X = *počet starších sourozenců* a Y = *porodní hmotnost novorozence* vykreslete dvourozměrný tečkový diagram.

Řešení příkladu 3.6



3.3 Číselné charakteristiky pro intervalové znaky

Příklad 3.7. Základní číselné charakteristiky pro intervalový znak
Načtěte datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Zaměřte se pouze na znak X = *největší šířka mozkovny* pro skelety mužského pohlaví. Vytvořte tabulku základních číselných charakteristik pro znak X.

Řešení příkladu 3.7

```
[1] 216
```

Po odstranění neznámých hodnot obsahuje datová tabulka údaje o .... skeletech mužského pohlaví.

Znak X = *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, mezikvartilové rozpětí, koeficient šikmosti a koeficient špičatosti.

1. Aritmetický průmer 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} = ...
5. Dolní kvartil $x_{0.25}$

   - n = ...
   - α = ...
   - $\alpha \times n$ = ... → je / není celé číslo →

   - $x_{0.25}$ =

6. Medián $x_{0.50}$

   - n = ...
   - α = ...
   - $\alpha \times n$ = ... → je / není celé číslo →

   - $x_{0.50}$ =

7. Horní kvartil $x_{0.75}$
   - $n =$ ....
   - $\alpha =$ ....
   - $\alpha \times n =$ .... $\rightarrow$ je / není celé číslo $\rightarrow$
   - $x_{0.75} =$
8. Maximální hodnota $x_{max} =$ ....
9. Výběrový koeficient šikmosti b_1
   - $b_1 =$ ....
10. Výběrový koeficient špičatosti b_2
    - $b_2 =$ ....

```
          m    s min dolni.kv median horni.kv max iqr sikmost spicatost
muži 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 než .... mm, 50% naměřených hodnot je menších než .... mm, 75% naměřených hodnot je menších než .... mm. Interkvartilové rozpětí naměřených hodnot je rovno .... Hodnota koeficientu šikmosti, ..., ukazuje na .... data (prodloužený .... konec). Hodnota koeficientu špičatosti, ..., ukazuje na .... charakter dat.

Příklad 3.8. Charakteristika závislosti pro znaky intervalového typu
Zaměřme 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 3.8
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
```

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

Příklad 3.9. Dvourozměrný tečkový diagram
Pro znaky X = *největší šířka mozkovny* a Y = *největší délka mozkovny* u mužů vykreslete dvourozměrný tečkový diagram.