

2 Výpočet číselných charakteristik jednorozměrného a dvourozměrného datového souboru

Přehled použitých funkcí: data.frame, apply, library, round, cramersV, read.delim, source, head, names, factor, quantile, boxplot, cor, dotplot, abline, length, mean, var, sqrt, skewness, kurtosis, cbind.

Příklad 2.1. Ve vzorku, který tvořilo 200 studentů (100 mužů a 100 žen) byly standardní dermatoglyfickou metodou snímány dematoglyfy dlaně (Býmová, 1990). Na otiscích bylo hodnoceno zakončení tří hlavních dlaňových linií. Podle vzorce zakončení byly jednotliví studenti rozděleni do tří kategorií: vysoká (Hi), střední (Mi) a nízká (Lo). Současně byla zhodnocena barva vlasů studentů podle standardní Fisher-Sallerové stupnice (Martin a Saller, 1957–1966, s. 391), na základě které byli studenti rozděleni do tří skupin: Světlá (LoH), střední (MH) a tmavá (DaH). K dispozici máme početnosti jedinců v jednotlivých kategoriích, zvlášť pro muže a zvlášť pro ženy.

A) Početnosti v jednotlivých kategoriích pro muže

	Hi	Mi	Lo
LiH	6	6	4
MH	20	15	7
DaH	18	12	12

B) Početnosti v jednotlivých kategoriích pro ženy

	Hi	Mi	Lo
LiH	4	4	6
MH	18	10	10
DaH	12	22	12

A) 1. Určete modus zakončení dlaňových linií a modus barvy vlasů pro muže.

2. Pomocí Cramérova koeficientu stanovte stupeň závislosti mezi zakončením dlaňových linií a barvy vlasů u mužů.

B) 1. Určete modus zakončení dlaňových linií a modus barvy vlasů pro ženy. (*střední; tmavé*)

2. Pomocí Cramérova koeficientu stanovte stupeň závislosti mezi zakončením dlaňových linií a barvy vlasů pro ženy. (*0.179*)

A) 1. Stanovení modu pro muže

```
##   vysoke  stredni   nizke
##      44      33      23
##   svetle  stredni   tmave
##      16      42      42
```

Interpretace: Znak X, zakončení dlaňových linií u mužů, má modus, tj. nejvíce mužů má zakončení dlaňových linií. Znak Y, barva vlasů u mužů, má modus, tj. nejvíce mužů má barvu vlasů

2. Výpočet Cramérova koeficientu

```
## [1] 0.101
```

Interpretace: Cramérův koeficient nabývá hodnoty, tedy mezi zakončením dlaňových linií a barvou vlasů u mužů existuje

Cramérův koeficient	interpretace
0 – 0.1	zanedbatelná závislost
0.1 – 0.3	slabá závislost
0.3 – 0.7	střední závislost
0.7 – 1	silná závislost

B) 1. Stanovení modu pro ženy

##	vysoke	stredni	nizke
##	34	38	28
##	svetle	stredni	tmave
##	16	38	46

Interpretace: Znak X, zakončení dlaňových linií u žen, má modus, tj. nejvíce žen má zakončení dlaňových linií. Znak Y, barva vlasů u žen, má modus, tj. nejvíce žen má barvu vlasů

2. Výpočet Cramérova koeficientu

[1] 0.179

Interpretace: Cramérův koeficient nabývá hodnoty, tedy mezi zakončením dlaňových linií a barvou vlasů u žen existuje

Příklad 2.2. Načtete datový soubor `znamky_me.txt`.

- Pro známky z **angličtiny** a matematiky vypočtete medián, dolní a horní kvartil, kvartilovou odchylku a vytvořte krabicový diagram.
- Vypočtete **Spearmanův korelační koeficient** známek z angličtiny a matematiky pro všechny studenty.

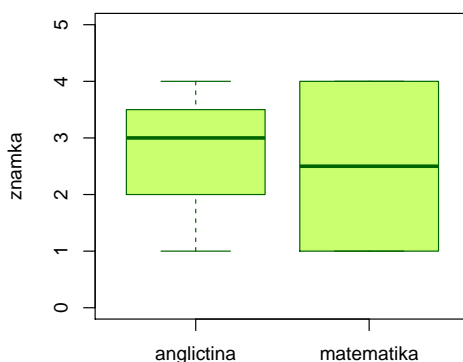
a) ##

	median	kv1	kv3	IQR
## anglictina	3.0	2	3.5	3.0
## matematika	2.5	1	4.0	1.5

Interpretace:

- Hodnota mediánu pro známku z angličtiny je, tedy studentů mělo známku z angličtiny lepší nebo rovnu a studentů mělo známku z angličtiny horší nebo rovnu
- Hodnota dolního kvartilu pro známku z angličtiny je, tedy studentů mělo známku z angličtiny lepší nebo rovnu a studentů mělo známku z angličtiny horší nebo rovnu
- Hodnota horního kvartilu pro známku z angličtiny je, tedy studentů mělo známku z angličtiny lepší nebo rovnu a studentů mělo známku z angličtiny horší nebo rovnu
- Hodnota mediánu pro známku z matematiky je, tedy studentů mělo známku z matematiky lepší nebo rovnu a studentů mělo známku z matematiky horší nebo rovnu
- Hodnota dolního kvartilu pro známku z matematiky je, tedy studentů mělo známku z matematiky lepší nebo rovnu a studentů mělo známku z matematiky horší nebo rovnu
- Hodnota horního kvartilu pro známku z matematiky je, tedy studentů mělo známku z matematiky lepší nebo rovnu a studentů mělo známku z matematiky horší nebo rovnu

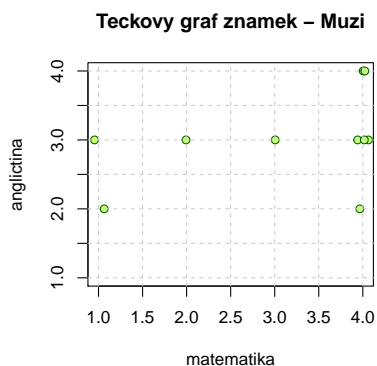
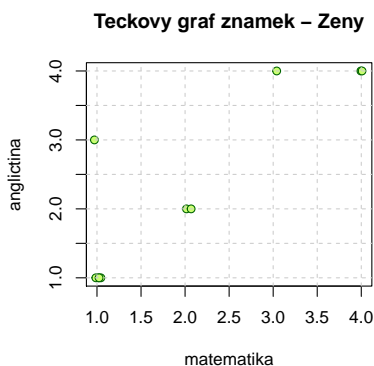
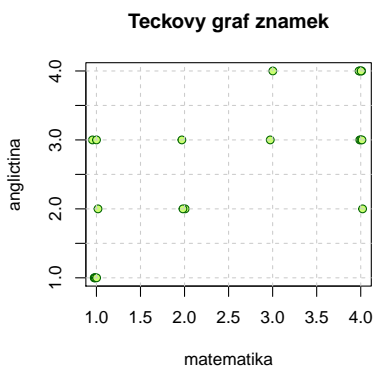
Krabicovy graf – Znamky



```
b) ## [1] "Spearmanuv koeficient - M + F: 0.688442"
## [1] "Spearmanuv koeficient - F: 0.860314"
## [1] "Spearmanuv koeficient - M: 0.373544"
```

Interpretace:

- (a) Hodnota Spearmanova korelačního koeficientu je, tedy mezi známkami z angličtiny a matematiky existuje
- (b) Hodnota Spearmanova korelačního koeficientu je, tedy mezi známkami z angličtiny a matematiky u žen existuje
- (c) Hodnota Spearmanova korelačního koeficientu je, tedy mezi známkami z angličtiny a matematiky u mužů existuje



Význam hodnot Spearmanova (i Pearsonova) koeficientu korelace je popsán v tabulce:

Abs.hod. korel.koef.	Interpretace hodnoty
0	pořadová (lineární) nezávislost
(0; 0.1)	velmi nízký stupeň závislosti
[0.1; 0.3)	nízký stupeň závislosti
[0.30; 0.50)	mírný stupeň závislosti
[0.50; 0.70)	význačný stupeň závislosti
[0.70; 0.90)	vysoký stupeň závislosti
[0.90; 1)	velmi vysoký stupeň závislosti
1	úplná pořadová (lineární) závislost

Příklad 2.3. Vypočítejte medián, dolní kvartil, horní kvartil a kvartilovou odchylku pro známky z **angličtiny**, když víte, že absolutní četnosti známek byly

známka	1	2	3	4
absolutní četnost	4	4	7	5

```
##          median kv1 kv3 IQR
## anglictina      3  2 3.5 1.5
```

Příklad 2.4. Otevřeme datový soubor lebkky.txt.

- Pro největší délku a největší šířku mozkovny mužů vypočítejte aritmetický průměr, rozptyl, směrodatnou odchylku, koeficient variace, šikmost a špičatost.
- Vypočítejte Pearsonův koeficient korelace největší délky a největší šířky mozkovny mužů. Nakreslete dvou-rozměrný tečkový diagram.
- Výpočet základních charakteristik pro **délku** a **šířku** mozkovny mužů

```
##          n  prumer rozptyl sm.odch koef.var sikmost spicatost
## delka 216 182.0324 40.5777  6.3701   3.4994 -0.0551  -0.4511
## sirka 216 137.1852 23.1694  4.8135   3.5087  0.0853  -0.2485
```

Interpretace:

- Aritmetický průměr délky mozkovny mužů je cm se směrodatnou odchylkou cm.
 - Hodnota koeficientu šikmosti pro délku mozkovny mužů ukazuje na zešikmené rozdělení dat. (Data mají prodloužený konec.)
 - Hodnota koeficientu špičatosti pro délku mozkovny mužů ukazuje na rozdělení dat.
 - Aritmetický průměr šířky mozkovny mužů je cm se směrodatnou odchylkou cm.
 - Hodnota koeficientu šikmosti pro šířku mozkovny mužů ukazuje na zešikmené rozdělení dat. (Data mají prodloužený konec.)
 - Hodnota koeficientu špičatosti pro šířku mozkovny mužů ukazuje na rozdělení dat.
- b) Výpočet Pearsonova korelačního koeficientu

```
## [1] 0.168157
```

Interpretace: Hodnota Pearsonova korelačního koeficientu je, tedy mezi délkou mozkovny a šířkou mozkovny u mužů existuje

Tečkový graf delky a sirky lebkky muzů

