

# 1 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 1.1.** U 100 náhodně vybraných domácností byl zjišťován způsob zásobování bramborami (znak X, varianty 1 = vlastní sklep, 2 = jinde, 3 = nákup) a bydliště (znak Y, varianty 1 = velké město, 2 = malé město, 3 = vesnice).

	velké město	malé město	vesnice
vlastní sklep	13	15	14
jinde	11	7	2
nákup	19	9	10

- a) Pro oba znaky určíme modus.
- b) Vypočteme Cramérův koeficient znaků X, Y.

a) Stanovení modu

```
##      velke.mesto male.mesto vesnice
## sklep      13      15      14
## jinde      11       7       2
## nakup      19       9      10
```

```
## sklep jinde nakup
##    42    20    38
```

```
## velke.mesto male.mesto vesnice
##           43          31          26
```

Znak X má modus 1, tj. nejvíce domácností skladuje brambory ve vlastním sklepe a znak Y má také modus 1, tj. nejvíce domácností bydlí ve velkém městě.

b) Výpočet Cramérova koeficientu

Hodnotu Cramérova koeficientu vypočítáme pomocí funkce `cramersV`, která je součástí knihovny `lsr`. Nejprve tedy musíme nainstalovat tuto knihovnu (`Packages` → `Install` → `lsr` → `Install`) a následně ji načíst (`library(lsr)`). Teprve potom můžeme funkci `cramersV()` použít na naši datovou tabulku a Cramérův koeficient dopočítat.

```
## [1] 0.179
```

Cramérův koeficient nabývá hodnoty 0.179, tedy mezi způsobem zásobování bramborami a bydlištěm domácnosti existuje jen slabá závislost - viz následující tabulka:

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

**Příklad 1.2.** Otevřeme datový soubor znamky.txt.

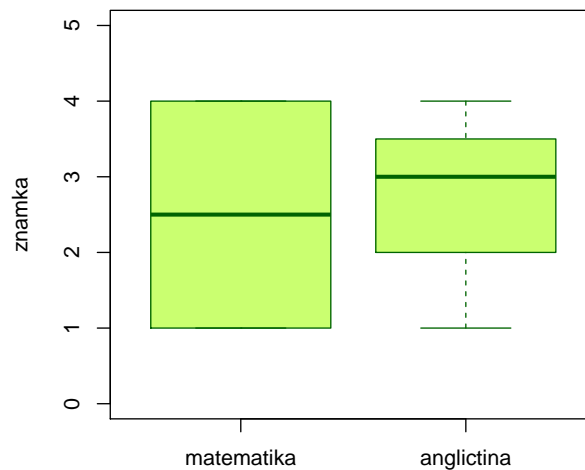
- Pro známky z matematiky a angličtiny vypočteme medián, dolní a horní kvartil, kvartilovou odchylku a vytvoříme krabicový diagram.
- Vypočteme Spearmanův korelační koeficient známek z matematiky a angličtiny pro všechny studenty, pak samostatně pro muže a samostatně pro ženy. Získané výsledky budeme interpretovat.

```
a) ## V1 V2 V3
## 1 2 2 0
## 2 1 3 1
## 3 4 3 1
## 4 1 1 0
## 5 1 2 1
## 6 4 4 1
```

```
## matematika anglictina pohlavi
## 1 2 2 zena
## 2 1 3 muz
## 3 4 3 muz
## 4 1 1 zena
## 5 1 2 muz
## 6 4 4 muz
```

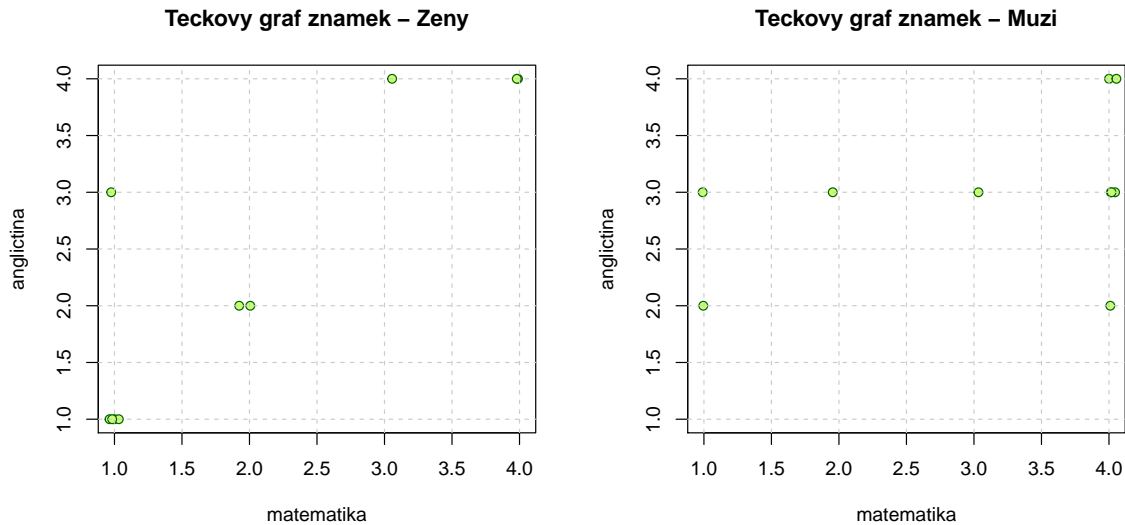
```
## median kv1 kv3 IQR
## matematika 2.5 1 4.0 3.0
## anglictina 3.0 2 3.5 1.5
```

**Krabicovy graf dvou promennych**



```
b) ## [1] "Spearmanuv koeficient pro vsechny = 0.6884"
## [1] "Spearmanuv koeficient pro zeny = 0.8603"
## [1] "Spearmanuv koeficient pro muze = 0.3735"
```

Vidíme, že nejsilnější přímá pořadová závislost mezi známkami z matematiky a angličtiny je u žen,  $r_S = 0.86$ . U mužů je tato závislost mnohem slabší,  $r_S = 0.37$ . U žen tedy dochází k tomu, že se sdružují podobné známky z obou předmětů, zatímco u mužů se projevuje spíše tendence k různým známkám. Je to zřetelně vidět na dvourozměrných tečkových diagramech.



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

Podle výše uvedené tabulky existuje mezi známkami z matematiky a známkami z angličtiny význačný stupeň přímé pořadové závislosti ( $r_S = 0.69$ ), dále v případě žen existuje mezi známkami z matematiky a z angličtiny vysoký stupeň přímé pořadové závislosti ( $r_S = 0.86$ ), zatímco u mužů existuje mezi známkami z matematiky a z angličtiny pouze mírný stupeň přímé pořadové závislosti ( $r_S = 0.37$ ).

**Příklad 1.3.** Otevřeme datový soubor lebky.txt.

- Pro největší délku a největší šířku mozkovny mužů vypočteme aritmetický průměr, rozptyl, směrodatnou odchylku, koeficient variace, šikmost a špičatost.
- Vypočítejte Pearsonův koeficient korelace největší délky a největší šířky mozkovny mužů. Dále vypočtete kovarianci těchto dvou znaků a nakreslete dvourozměrný tečkový diagram.

```
a) ##   delka  sirka  pohlavi
## 1   188   145   muž
## 2   172   139   muž
## 3   176   138   muž
## 4   184   128   muž
## 5   183   139   muž
## 6   177   143   muž
```

```
##      n  prumer rozptyl sm.odch koef.var sikmost spicatost
## 1 216 182.0324 40.5777 6.3701 3.4994 -0.0551 -0.4511
```

Analogický postup zvolíme pro výpočty základních charakteristik pro šířku mozkovny mužů. Výsledné charakteristiky pro obě proměnné sloučíme do jedné tabulky.

```
##      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
```

## b) Výpočet Pearsonova korelačního koeficientu

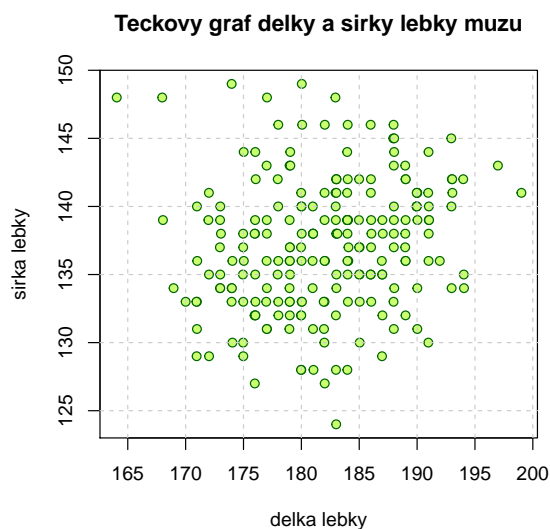
```
## [1] 0.168157
```

Vidíme, že mezi délkou mozkovny a šířkou mozkovny u mužů existuje nízký stupeň přímé lineární závislosti.

## Výpočet kovariance

```
## [1] 5.156
```

## Tečkový diagram



Vzhledu diagramu potvrzuje naše zjištění, že mezi délkou a šířkou mozkovny u mužů existuje nízká přímá lineární závislost.