

# POPISNÉ STATISTIKY — VZTAH DVOU VELIČIN

## 5.3..2013

---

### ÚVODNÍ NASTAVENÍ.

- Otevřete si program R a změňte si pracovní adresář pomocí **File**→**Change working directory** na adresář **statistika**.
  - Přes nabídku **File** →**Load Workspace** si načtět soubor **studenti.RData**, který jste si minule uložili.
  - Spusťte si R-Commander. Měli byste mít k dispozici načtená data včetně všech změn a nových proměnných z minula. Data si aktivujete pomocí tlačítka vedle **Data set**: v horní části okna, po stlačení zvolte data **studenti**.
1. Pomocí **Statistics**→**Summaries** →**Active data set** si nechte vypsat základní popisné statistiky pro všechny veličiny. Připomeňte si, co jednotlivé proměnné znamenají. Ujistěte se, že jsou všechny kategoriální veličiny nastaveny na faktory.
  2. **Popis vztahu kvalitativní a kvantitativní veličiny.** Budeme zkoumat, zda a jak se liší výška mužů a výška žen.
    - (a) Vyberte nabídku **Statistics** →**Summaries** →**Numerical summaries**. Zde zvolte veličinu **vyska** a vyberte, které popisné statistiky si chcete nechat vypsat. Tlačítkem **Summarize by groups** nastavte, že chcete tyto popisné statistiky zvlášť pro muže a ženy. Prohlédněte si, v čem jsou odlišnosti.
    - (b) Graficky lze předchozí čísla znázornit pomocí krabicového grafu. Zvolte **Graphs** →**Boxplot**, zde opět zvolte **Summarize by groups**. Připomeňte si, co boxplot znázorňuje.
    - (c) Dále by nás mohl zajímat histogram výšky pro muže a ženy zvlášť. V nabídce **Graphs**→**Histogram** bohužel není možnost typu “**by groups**”, takže musíme postupovat následovně. Do skriptového okna dopište:

```
library(lattice)
histogram(~vyska|fpohlavi,data=studenti)
```

a oba řádky (postupně nebo najednou) potvrďte pomocí **submit** (nebo klávesovou zkratkou **CTRL + R**).
  3. Stejným postupem si prohlédněte, jestli a jak se liší hmotnost u mužů a žen.
  4. Zjistěte, zda se věk otců liší pro dívky a pro chlapce.
  5. **Změny nastavení v obrázcích.**
    - (a) Znovu si vykreslete krabicové diagramy (boxploty) výšky pro muže a ženy zvlášť. Obrázek upravíme tak, aby byl jinak popsán a jinak barevný. Do posledního řádku skriptu doplníme/upravíme následující volby:
      - Osu  $x$  popíšeme jako “Pohlavi”, změníme tedy `xlab='Pohlavi'` .
      - Osu  $y$  popíšeme jako “Vyska studentu”, změníme tedy `ylab='Vyska studentu'` .
      - Změníme barvy krabic: doplníme `col=c('pink','blue')`.
      - Názvy kategorií na ose  $x$  změníme přidáním `names=c('Devcata','Chlapci')`.

– Celý obrázek popíšeme přidáním `main='Krabicovy diagram'` .  
Nakonec vše potvrdíme pomocí `submit`.

- (b) Podobně lze v obrázcích měnit i velikosti písmen, fonty atd. (nebudeme dělat).
- (c) Seznam všech barev získáte zavoláním `colors()`.
- (d) Podobně si sami upravte obrázek boxplotů hmotnosti a uložte si ho.

6. **Vztah dvou kvantitativních veličin.** Podíváme se na vztah výšky a váhy (společně pro muže i ženy).

- (a) Zvolíme `Graphs` → `Scatterplot`. Zde zvolte výšku a váhu, přičemž uvažte, kterou veličinu dát na  $x$ -ovou a kterou na  $y$ -ovou osu. Dále zde zrušte všechna zaškrtnutá políčka ve sloupci `Options`. Ostatní nastavení neměňte a potvrďte `Ok`. Co lze z obrázku usuzovat? Souvisí hmotnost a výška?
- (b) Projděte si jednotlivé volby v nabídce `Scatterplot` a zkuste je pozměnit, resp. aktivovat. Zejména vyzkoušejte `Least-squares line` a `Smooth line`.
- (c) Vyzkoušejte také volbu `Plot by groups`, kterou v obrázku rozlišíte muže a ženy.
- (d) V předchozím obrázku změňte barvu na růžovou a modrou.

7. Stejným způsobem se podívejte, zda a jak spolu souvisí věk otce a věk matky.

8. Souvisí spolu velikost bot a index BMI?

9. **Vztah dvou kvalitativních veličin.** Budeme se zajímat o to, zda veličina nadváha nějak souvisí s pohlavím.

- (a) Pomocí `Statistics` → `Contingency tables` → `Two-way tables` si nechte vypsát tabulku četností těchto dvou znaků.
- (b) Stejným způsobem si vypište řádkové a sloupcové relativní četnosti. Zamyslete se nad interpretací uvedených hodnot. Je nějaký rozdíl mezi muži a ženami, co se týče nadváhy a podváhy?
- (c) Vykreslíme si sloupcový graf nadváhy zvlášť pro muže a ženy. Do skriptového okna přepište:

```
barplot(table(studenti$nadvaha, studenti$fpohlavi), beside=T, legend=T)
```

- (d) Další možný popisný obrázek si vykreslíme pomocí:

```
plot(studenti$fpohlavi, studenti$nadvaha)
```

Co všechno lze z obrázku vyčíst?

- (e) Změňte pořadí pohlaví a nadváhy. Jak se změní obrázek?

10. Zjistěte, zda nějak souvisí pohlaví a roční období narození.

11. **Uložení práce:** Jestliže jste něco měnili v datech (převedení proměnných na faktory apod.), pak si uložte `File` → `Save R Workspace` pracovní prostředí z R.