

Příklady k prezentaci na 10. cvičení

Budete-li chtít prezentovat i příkazy v R-studiu, přineste si prosím příslušný kód na flashce, nebo mi ho s předstihem pošlete emailem. Pokud byste v příkladu narazili na nějakou nejasnost, neváhejte se mi ozvat.

Dataset `Muzi43.RData` pochází z dlouhodobého výzkumu provedeného studentkou 2. lékařské fakulty UK na mužích středního věku. Byly zkoumány následující veličiny:

- `ID` = identifikace osoby
- `vyska` = výška postavy
- `vaha` = hmotnost v kg
- `syst1` = měření systolického tlaku v mmHg
- `syst2` = opakované měření systolického tlaku v mmHg
- `chlst` = hladina cholesterolu v mg/100 ml
- `Vino` = zda pije víno
- `cukr` = denní spotřeba cukru (počet kostek)
- `bmi` = body mass index v kg/m^2
- `vek` = věk
- `KOURrisk` = kouření jako rizikový faktor (více než 15 cigaret denně)
- `Skupina` = riziková nebo normální

Tabulku si do R můžete načíst příkazem

```
load("Muzi43.RData")
```

Příklad 1

Ilustrace rozdílu mezi Pearsonovým a Spearmanovým korelačním koeficientem.

- Vytvořte si vektor o alespoň deseti hodnotách. Označme ho jako $\mathbf{vekX} = (x_1, \dots, x_n)$.
- Pomocí vektoru \mathbf{vekX} vytvořte vektor $\mathbf{vekY} = (y_1, \dots, y_n)$ tak, aby \mathbf{vekY} byl lineární funkcí \mathbf{vekX} (tj. složky \mathbf{vekY} budou tvaru $y_i = ax_i + b$ pro nějaké vámi zvolené konstanty a a b). Spočtete Pearsonův i Spearmanův korelační koeficient mezi \mathbf{vekX} a \mathbf{vekY} a výsledek okomentujte.
- Podobně vytvořte vektor \mathbf{vekZ} tak, aby jeho hodnoty byly monotónní (ale nelineární) funkcí hodnot \mathbf{vekX} (připomínám, že monotónní funkce je taková, která je ryze klesající či ryze rostoucí). Spočtete Pearsonův i Spearmanův korelační koeficient mezi \mathbf{vekX} a \mathbf{vekZ} a výsledek okomentujte.

Příklad 2

Pomocí dat Muzi43 ověřte, zda jsou v populaci veličiny hladina cholesterolu a body mass index nezávislé. Uvažujte hladinu $\alpha = 5\%$.

- (a) Formulujte nulovou a alternativní hypotézu.
- (b) Rozhodněte, zda lze předpokládat, že vektor $(bmi, chlst)$ pochází z dvojrozměrného normálního rozdělení.
- (c) Pomocí vhodného testu proveďte test zadané hypotézy. Okomentujte výstup z R-ka. Formulujte svůj závěr.

Příklad 3

Porovnejte hodnotu cholesterolu mužů mezi normální a rizikovou skupinou. Pomocí dat ověřte hypotézu, že muži s rizikovými faktory mají vyšší hladinu cholesterolu. Uvažujte hladinu $\alpha = 5\%$.

- (a) Zapište nulovou a alternativní hypotézu.
- (b) Ověřte předpoklady dvouvýběrového t-testu.
- (c) Proveďte test zadané hypotézy a svůj závěr zformulujte bez použití výrazů „zamítáme/nezamítáme H_0 “.