12 12-Kontingenční tabulky

12.1 Testování nezávislosti v kontingenčních tabulkách

- dvourozměrný náhodný výběr rozsahu n, dva nominální znaky X a Y
- znak X $\dots r$ variant; znak Y $\dots s$ variant
- absolutní simultánní četnosti
- marginální absolutní četnosti
- simultánní pst π_{jk}; odhad $p_{jk} = \frac{n_{jk}}{n}$
- marginální pst $\pi_{j.}$, $\pi_{.k}$; odhady $p_{j.} = \frac{n_{j.}}{n}$, $p_{.k} = \frac{n_{.k}}{n}$
- grafické znázornění dat: scatterplot

Testování hypotézy o nezávislosti

- $H_0 : X, Y$ jsou nezávislé
- $H_1 : X, Y$ nejsou nezávislé
- porovnáváme pozorované četnosti n_{jk} s očekávanými četnostmi $\dfrac{n_{j.}n_{.k}}{n}$
- podmínka dobré aproximace: očekávané četnosti musí v 80 % být ≥ 5 a ve zbylých dvaceti % ≥ 2; `chisq.test(data)$expected`
- Testová statistika má tvar:

$$K = \sum_{j=1}^{r}\sum_{k=1}^{s} \frac{(n_{jk} - \frac{n_{j.}n_{.k}}{n})^2}{\frac{n_{j.}n_{.k}}{n}} \sim as.\chi^2((r-1)(s-1))$$

- `chisq.test(data)`
- H_0 zamítáme na hl. význ. α, pokud $K \in W = \langle \chi^2_{1-\alpha}((r-1)(s-1)), \infty)$

Měření závislosti, Cramérův koeficient

- Cramérův koeficient

$$V = \sqrt{\frac{K}{n(m-1)}},$$

  kde $m = \min\{r, s\}$.
- `cramersV` (knihovna `lsr`)

12.2 Čtyřpolní kontingenční tabulky

- náhodné veličiny X, Y mají pouze 2 varianty $\rightarrow$ *čtyřpolní kontingenční tabulka*
- Fisherův faktoriálový (přesný) test
  - $H_0 : X, Y$ jsou nezávislé
  - $H_1 : X, Y$ nejsou nezávislé
  - fisher.test(data) $\rightarrow$ p-hodnota
- **Podíl šancí ve čtyřpolní KT**
  - pokus $\rightarrow$ 2 okolnosti $\rightarrow$ úspěch nebo neúspěch
  - 1.okolnost: podíl počtu úspěchů ku počtu neúspěchů: $\frac{a}{c}$
  - 2.okolnost: podíl počtu úspěchů ku počtu neúspěchů: $\frac{b}{d}$
  - podíl šancí
$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}$$
  - $OR \in \langle 0\,;\,\infty)$ ... nesymetrický interval $\rightarrow \ln OR \in (-\infty\,;\,\infty)$ ... symetrický okolo nuly
  - Závislost X, Y je tím silnější, čím více se OR liší od jedné, nebo čím víc se $\ln OR$ liší od nuly

Testování nezávislosti ve čtyřpolních KT pomocí podílu šancí

- $H_0 : X, Y$ jsou nezávislé ... $\ln o\rho = 0$
- $H_1 : X, Y$ nejsou nezávislé ... $\ln o\rho \neq 0$.
- ověřit podmínky dobré aproximace
- Testovací statistika
$$T_0 = \frac{\ln OR}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}} \sim \text{as}.N(0,1)$$
- kritický obor $W = (-\infty; -u_{1-\alpha/2}\rangle \cup \langle u_{1-\alpha/2}; \infty)$
- $T_0 \in W \rightarrow H_0$ zamítáme na ASYMPTOTICKÉ hladině významnosti α,
- $100(1-\alpha)\%$ asymptotický interval spolehlivosti
$$(d, h) = \left(\ln OR - \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}u_{1-\alpha/2}; \ln OR - \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}u_{\alpha/2}\right).$$
- $0 \notin IS \rightarrow H_0$ zamítáme na asymptotické hl. význ. $\alpha = 0.05$.