Cvičení 7 – příklady u tabule

Přehled vzorců pro ANOVu:
Máme $r \geq 3$ nezávislých náhodných výběrů o rozsazích $n_1, \ldots, n_r$, přičemž i-tý náhodný výběr pochází z normálního rozložení $N(\mu, \sigma^2)$, $i = 1, 2, ..., r$. Označme n … celkový rozsah všech r výběrů, $M_{i.}$ … výběrový průměr i-tého výběru, $M_{..}$ … celkový průměr všech r výběrů.

$S_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij} - M_{..})^2$ … celkový součet čtverců, $f_T = n - 1$, $S_A = \sum_{i=1}^{r} n_i (M_{i.} - M_{..})^2$ … skupinový součet čtverců, $f_A = r - 1$, $S_E = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij} - M_{i.})^2$ … reziduální součet čtverců, $f_E = n - r$. Lze dokázat, že $S_T = S_A + S_E$. Poměr determinace: $P^2 = \frac{S_A}{S_T}$.

Testujeme hypotézu H_0: $\mu_1 = \ldots = \mu_r$ proti alternativě H_1: aspoň jedna dvojice středních hodnot je různá. Testová statistika: $F_A = \frac{S_A / f_A}{S_E / f_E} \sim F(f_A, f_E)$, když H_0 platí. Kritický obor: $W = \langle F_{1-\alpha}(f_A, f_E), \infty)$.

Tukeyova metoda: Rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když $\frac{|M_{k.} - M_{l.}|}{\frac{S_*}{\sqrt{p}}} \geq q_{1-\alpha}(r, n-r)$, kde $q_{1-\alpha}(r, n-r)$ jsou kvantily studentizovaného rozpětí.

Scheffého metoda: rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když $|M_{k.} - M_{l.}| \geq S_* \sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right) F_{1-\alpha}(r-1, n-r)}$.

Příklad 1.: (viz př. 8.6.1. ze skript) Jsou známy měsíční tržby (v tisících Kč) tří prodavačů za dobu půl roku.
1. prodavač: 12 10 9 10 11 9
2. prodavač: 10 12 11 12 14 13
3. prodavač: 19 18 16 16 17 15
Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty tržeb všech tří prodavačů jsou stejné. Pokud zamítneme nulovou hypotézu, zjistěte, tržby kterých dvou prodavačů se liší na hladině významnosti 0,05.

Příklad 2.: (viz př. 8.6.3. ze skript) Je dáno 5 nezávislých náhodných výběrů o rozsazích 5, 7, 6, 8, 5, přičemž i-tý výběr pochází z $N(\mu_i, \sigma^2)$, $i = 1, 2, \ldots, 5$. Byl vypočten celkový součet čtverců $S_T = 15$ a reziduální součet čtverců $S_E = 3$. Na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Příklad 3.: (viz př. 8.6.4. ze skript) Je dána tabulka ANOVA. Místo otazníků doplňte chybějící čísla a na hladině významnosti 0,05 testujte hypotézu o shodě středních hodnot.

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F
skupiny	?	2	?	?
reziduální	16,033	?	?	-
celkový	17,301	25	-	-

Přehled vzorců pro test homogenity binomických rozložení:
Máme $r \geq 3$ nezávislých náhodných výběrů o rozsazích $n_1, \ldots, n_r$, přičemž j-tý náhodný výběr pochází z alternativního rozložení $A(\vartheta_j)$, $j = 1, 2, ..., r$.

Testujeme hypotézu H_0: $\vartheta_1 = \ldots = \vartheta_r$ proti alternativě H_1: aspoň jedna dvojice parametrů je různá.

Označme n … celkový rozsah všech r výběrů, $M_* = \frac{\sum_{j=1}^{r} n_j M_j}{n}$ vážený průměr výběrových průměrů. Testová statistika: $Q = \frac{1}{M_*(1-M_*)}\sum_{j=1}^{r} n_j (M_j - M_*)^2 \approx \chi^2(r-1)$, když H_0 platí.

Kritický obor: $W = \langle \chi^2_{1-\alpha}(r-1), \infty)$.

Podmínka dobré aproximace: $n_j M_* > 5$ pro všechna $j = 1, ..., r$. Není-li splněna podmínka dobré aproximace, počítáme testovou statistiku podle vzorce $Q = 4\sum_{j=1}^{r} n_j (A_j - B)^2$, kde $A_j = \arcsin\sqrt{M_j}$, $B = \frac{1}{n}\sum_{j=1}^{r} n_j A_j$.

Mnohonásobné porovnávání: Platí-li nerovnost $|A_k - A_l| \geq \sqrt{\frac{1}{8}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)} \cdot q_{1-\alpha}(r, \infty)$, pak na hladině významnosti α zamítáme hypotézu o shodě parametrů ϑ_k, ϑ_l.

Příklad 4.: (viz př. 8.6.2. ze skript) 104 náhodně vybraných matek bylo dotázáno, zda jejich kojenec dostává dudlík. Zjišťoval se též nejvyšší stupeň dosaženého vzdělání matky.

Vzdělání matky	Počet matek	Počet dětí s dudlíkem
Základní	39	27
Středoškolské	47	34
Vysokoškolské	18	15

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že podíly dětí s dudlíkem nezávisí na vzdělání matky.