9 Jednofaktorová analýza rozptylu – ANOVA

9.1 Testování normality

- Shapirův-Wilkův test ... shapiro.test()
- Lillie-Forsův test ... lillie.test()
- Anderson-Darlingův test ... ad.test(),

9.2 Testování homogenity rozptylů u r náhodných výběrů

- máme $r \geq 2$ náhodných výběrů
- $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2 = \sigma^2$
- H_1 : alespoň jedna dvojice rozptylů se liší

  1. Levenův test
     - testovací statistika založena na odhadech středních hodnot
     - levene.test(y, group, location='mean') knihovna lawstat
  2. Brownův-Forsytův test
     - testovací statistika založena na mediánech
     - používáme, když výběry nejsou normálně rozdělené, ale rozsahy výběrů $n_i > 20$
     - levene.test(y, group, location='median') z knihovny lawstat
  3. Bartlettův test
     - bartlett.test(y, g) knihovna stat
     - používáme, pokud jsou rozsahy všech výběrů ≥ 6

ANOVA funguje dobře i při mírném porušení předpokladu normality nebo shody rozptylů.

9.3 ANOVA - Jednofaktorová analýza rozptylu

- zkoumá závislost intervalové proměnné X na nominální proměnné A
- A ... faktor; varianty A ... úrovně faktoru A
- Má typ potravy pračlověka (A) vliv na šířku stoliček (X)?
- Faktor A má $\mathbf{r} \geq 2$ úrovní $A_1, \ldots A_r$, přičemž i-té úrovni odpovídá n_i pozorování $X_{i1}, \ldots X_{in_i}$. Každý výběr $A_i \sim N(\mu_i, \sigma^2)$.
- důležité pojmy
  - $\mathbf{n} = \sum_{i=1}^{r} n_i$ ... celkový počet pozorování, **r**... počet úrovní faktoru A
  - **celkový součet čtverců $\mathbf{S_T}$**
    - charakterizuje variabilitu jednotlivých pozorování kolem celkového průměru
    - počet stupňů volnosti: $\mathbf{f_T = n - 1}$

- **skupinový součet čtverců $\mathbf{S_A}$**
  - charakterizuje variabilitu mezi jednotlivými náhodnými výběry
  - počet stupňů volnosti: $\mathbf{f_A = r - 1}$
- **reziduální součet čtverců $\mathbf{S_E}$**
  - charakterizuje variabilitu uvnitř jednotlivých výběrů
  - počet stupňů volnosti: $\mathbf{f_E = n - r}$
- $\mathbf{S_T = S_A + S_E}$.
- $\mathbf{f_T = f_A + f_E}$.

Testování hypotézy o shodě středních hodnot

- $H_0 : \mu_1 = \cdots = \mu_r$; střední hodnoty všech výběrů jsou stejné
- $H_1 : \mu_i \neq \mu_j$ pro nějaké i, j; alespoň jedna dvojice středních hodnot se liší.
- Testovací statistika
$$T_0 = \mathbf{F_A} = \frac{S_A/f_A}{S_E/f_E} \sim F(r-1, n-r).$$
- kritický obor $W = \langle F_{1-\alpha}(r-1, n-r), \infty)$
- p-hodnota $= \Pr(T_0 > t_0) =$1-pf(FA, r-1, n-r)=1-pf(FA, fA, fE)
- přehledná tabulka výpočtů:

Zdroj variability	součet čtverců	stupně volnosti	průměrný čtverec	F_A
skupinový	S_A	$f_A = r-1$	S_A/f_A	$\frac{S_A/f_A}{S_E/f_E}$
reziduální	S_E	$f_E = n-r$	S_E/f_E	-
celkový	S_T	$f_T = n-1$	-	-

9.4 Post-hoc metody mnohonásobného porovnávání

- zamítneme-li nulovou hypotézu o shodě středních hodnot, chceme zjistit, která dvojice středních hodnot se od sebe významně liší
- Scheffého metoda
  - vhodná i v případě, že rozsahy všech výběrů nejsou stejné
  - hypotézu $H_0 : \mu_k = \mu_l$ o rovnosti středních hodnot zamítneme na hl. významnosti α, když
$$|M_{k.} - M_{l.}| \geq \frac{S_E}{f_E}\sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right) F_{1-\alpha}(r-1, n-r)}.$$
  - funkce Scheffe(X, group, names, alpha) z RSkriptu AS-funkce.R.
  - metody mnohonásobného porovnávání jsou slabší než ANOVA. Může se stát, že ANOVA zamítne H_0 o shodě středních hodnot ale Scheffého metoda u žádné dvojice významný rozdíl nenajde.