9. Analýza rozptylu jednoduchého třídění

9.1. Motivace

9.2. Označení

9.3. Testování hypotézy o shodě středních hodnot

9.4. Testování hypotézy o shodě rozptylů
a) Levenův test
b) Bartlettův test

9.5. Post – hoc metody mnohonásobného porovnávání
a) Tukeyova metoda
b) Scheffého metoda

9.6. Plánované porovnávání - testování významnosti kontrastů

9.7. Porovnávání s kontrolou

9.8. Příklad: U čtyř odrůd brambor (označených symboly A, B, C, D) se zjišťovala celková hmotnost brambor vyrostlých vždy z jednoho trsu. Výsledky (v kg):

odrůda	hmotnost
A	0,9 0,8 0,6 0,9
B	1,3 1,0 1,3
C	1,3 1,5 1,6 1,1 1,5
D	1,1 1,2 1,0

Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota hmotnosti trsu brambor nezávisí na odrůdě. Zamítnete-li nulovou hypotézu, zjistěte, které dvojice odrůd se liší na hladině významnosti 0,05.

Řešení: Data považujeme za realizace čtyř nezávislých náhodných výběrů ze čtyř normálních rozložení se stejným rozptylem. Testujeme hypotézu, že všechny čtyři střední hodnoty jsou stejné.

$n_1 = 4$, $n_2 = 3$, $n_3 = 5$, $n_4 = 3$, $n = 15$, $m_{1.} = 0{,}8$, $m_{2.} = 1{,}2$, $m_{3.} = 1{,}4$, $m_{4.} = 1{,}1$, $m_{..} = 1{,}14$,
$s_1^2 = 0{,}02$, $s_2^2 = 0{,}03$, $s_3^2 = 0{,}04$, $s_4^2 = 0{,}01$,

$$s_*^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 + (n_4-1)s_4^2}{n-4} =$$

$$= \frac{3\cdot 0{,}02 + 2\cdot 0{,}03 + 4\cdot 0{,}04 + 2\cdot 0{,}01}{11} = \frac{3}{110} = 0{,}0\overline{27}$$

$$S_E = (n-4)\cdot s_*^2 = 11\cdot\frac{3}{110} = 0{,}3\,,\ f_E = n-4 = 11$$

$$S_A = n_1(m_{1.} - m_{..})^2 + n_2(m_{2.} - m_{..})^2 + n_3(m_{3.} - m_{..})^2 + n_4(m_{4.} - m_{..})^2 =$$

$$= 4\cdot(0{,}8-1{,}14)^2 + 3\cdot(1{,}2-1{,}14)^2 + 5\cdot(1{,}4-1{,}14)^2 + 3\cdot(1{,}1-1{,}14)^2 = 0{,}816,\ f_A = r-1 = 3$$

$S_T = S_A + S_E = 0{,}816 + 0{,}3 = 1{,}116$, $f_T = n - 1 = 14$

Testová statistika: $F_A = \frac{S_A / f_A}{S_E / f_E} = \frac{0{,}816/3}{0{,}3/11} = 9{,}97$

Kritický obor: $W = \langle F_{0,95}(3{,}11), \infty) = \langle 3{,}59; \infty)$. Protože testová statistika se realizuje v kritickém oboru, H_0 zamítáme na hladině významnosti 0,05.

Výsledky zapíšeme do tabulky ANOVA:

Zdroj variability	Součet čtverců	Stupně volnosti	podíl	F_A
skupiny	$S_A = 0{,}816$	$f_A = 3$	$S_A/3 = 0{,}272$	$\frac{S_A/f_A}{S_E/f_E} = 9{,}97$
reziduální	$S_E = 0{,}3$	$f_E = 11$	$S_E/11 = 0{,}02727$	-
celkový	$S_T = 1{,}116$	$f_T = 14$	-	-

Nyní pomocí Scheffého metody zjistíme, které dvojice odrůd se liší na hladině významnosti 0,05. rovnost středních hodnot μ_k a μ_l zamítneme na hladině významnosti α, když

$$\left|M_{k.} - M_{l.}\right| \geq S_* \sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right) F_{1-\alpha}(r-1, n-r)}\,.$$

Srovnávané odrůdy	Rozdíly $\left|m_{k.} - m_{l.}\right|$	Pravá strana vzorce
A, B	0,4	0,41
A, C	0,6	0,36
A, D	0,3	0,41
B, C	0,2	0,40
B, D	0,1	0,44
C, D	0,3	0,40

Na hladině významnosti 0,05 se liší odrůdy A a C.

9.9. Význam předpokladů v analýze rozptylu