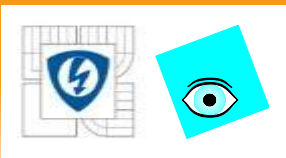


Statistická analýza jednorozměrných dat

Prof. RNDr. Milan Meloun, DrSc.
Univerzita Pardubice, Pardubice

31.ledna 2011

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky.



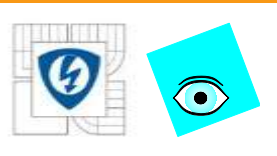
Kapitola 5.1

ANALÝZA ROZPTYLU ANOVA

31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

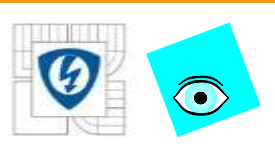




Analýza rozptylu ANOVA

K posouzení významnosti zdrojů variability v datech:

1. vliv přípravy na výsledek (analýzy),
2. vliv přístroje, člověka, experimentu na výsledek,
3. vliv laboratoří na výsledek analýz,
4. vyhodnocování plánovaných experimentů,
5. vliv faktorů A, B, C, ...(například teploty, času, koncentrace) na výsledek (analýzy).



Analýza rozptylu ANOVA

rozptyl = složka objasněná + složka neobjasněná
(známé zdroje variability) (náhodné chyby)

Faktor A je na jistých **úrovních** A_1, A_2, A_3

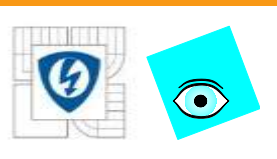
Zdrojem variability měření y_{ij} jsou úrovně A_1, A_2, A_3

$$y_{ij} = \mu_i + \varepsilon_{ij}$$

μ_i je skutečná hodnota a ε_{ij} je j –tá náhodná chyba.

Cíl: který z faktorů $A, B, C, \dots$ má významný vliv na výsledek analýzy (např. reakční výtěžek) μ ?

Hodnotu μ zde odhadujeme $\bar{x}_i, i = 1, \dots, n_i$.



Jednofaktorová analýza rozptylu

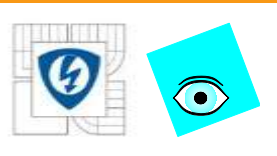
Faktor A má K úrovní $A_1, \dots, A_K$.

Na každé úrovni A_i je provedeno n_i měření $\{y_{ij}\}, j = 1, \dots, n_i$.

Celkový počet měření je $N = \sum_{i=1}^K n_i$.

Uspořádání dat:

	Úroveň faktoru						Celek
	A_1	A_2	...	A_i	...	A_K	
Opakování měření	y_{11}	y_{21}	...	y_{i1}	...	y_{K1}	
	y_{12}	y_{22}	...	y_{i2}	...	y_{K2}	
	...	...	...	...	...	...	
	...	...	...	...	...	...	
	y_{1n_1}	y_{2n_2}	...	y_{in_i}	...	y_{Kn_K}	
Průměry	$\hat{\mu}_1$	$\hat{\mu}_2$	...	$\hat{\mu}_i$	...	$\hat{\mu}_K$	$\hat{\mu}$
Počet	n_1	n_2	...	n_i	...	n_K	N



Sloupcový a celkový průměr, odhad efektů

Sloupcový průměr

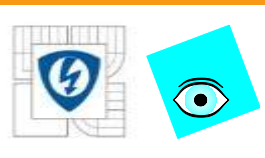
$$\hat{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$$

Celkový průměr

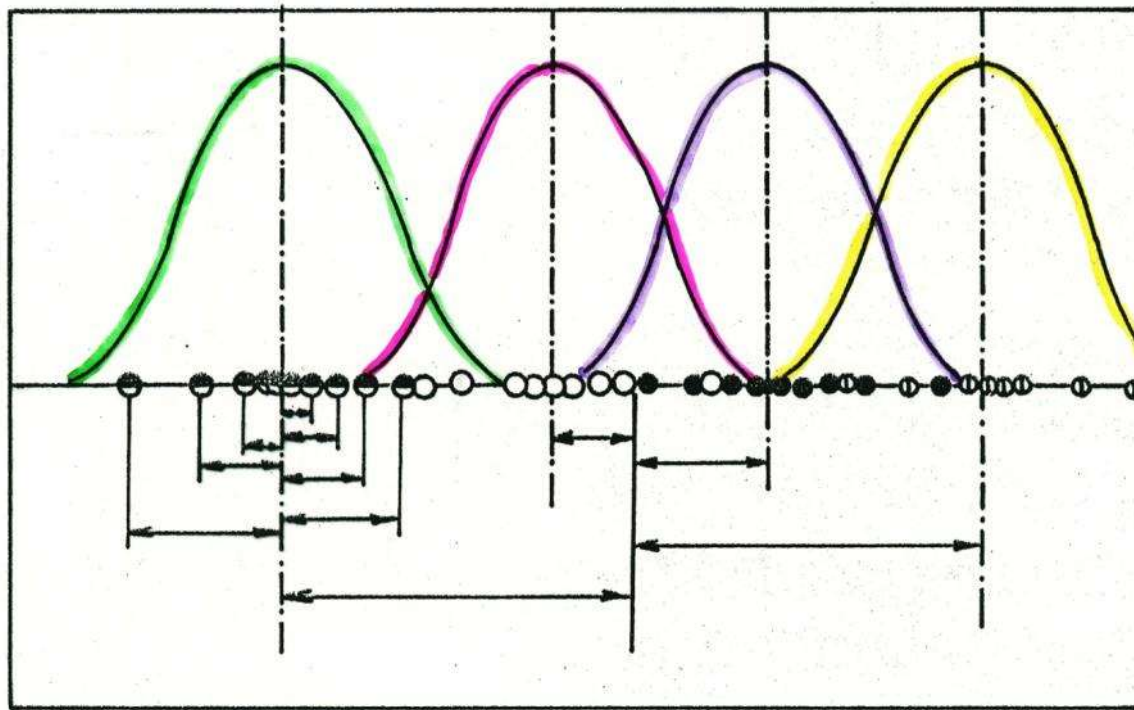
$$\hat{\mu} = \frac{1}{K} \sum_{j=1}^K \hat{\mu}_i$$

Odhad efektů

$$\hat{\alpha}_i = \hat{\mu}_i - \hat{\mu} \text{ za podmínky } \sum_{i=1}^K n_i \alpha_i = 0$$



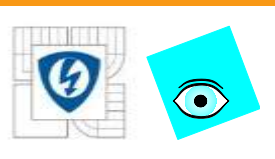
Sloupcový a celkový průměr, odhad efektů



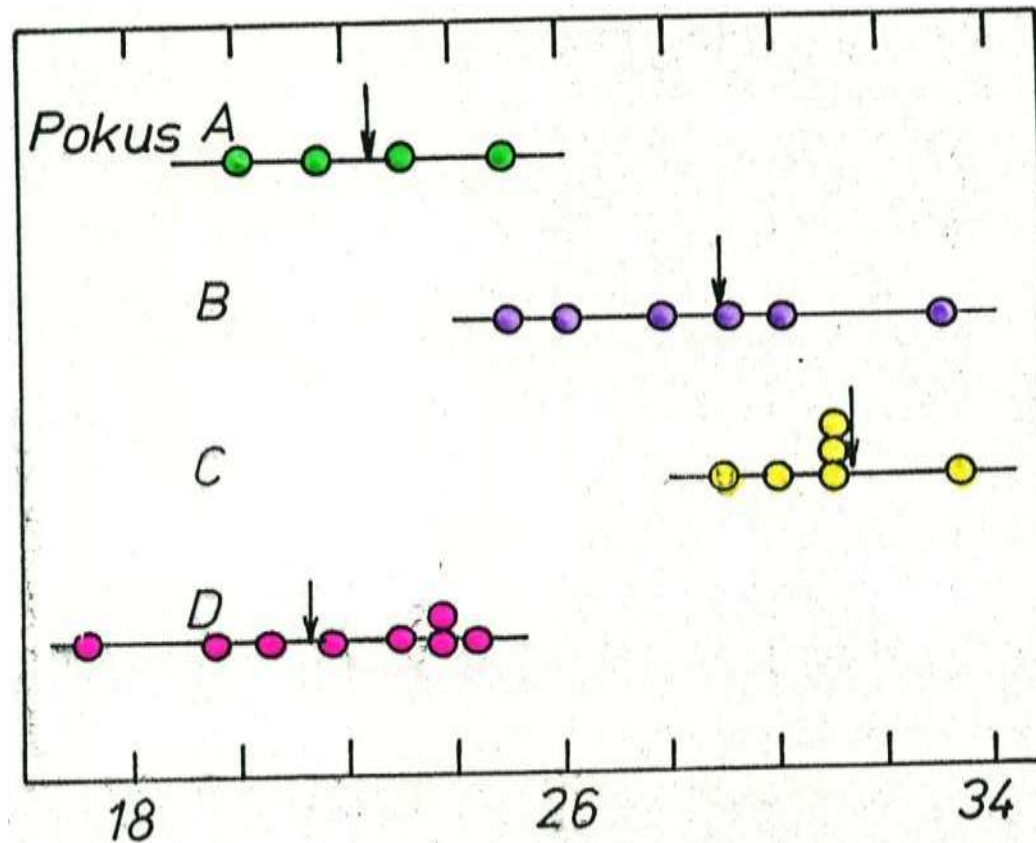
31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





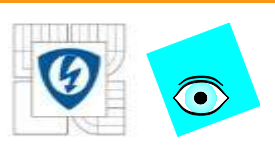
Sloupcový a celkový průměr, odhad efektů



31.1.2011

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ





Modely s pevnými efekty

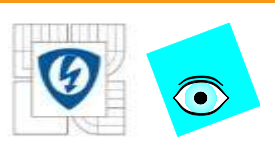
Předpoklad: náhodné chyby ε_{ij} jsou nezávislé náhodné veličiny s normálním rozdělením $N(0, \sigma^2)$.

Součet čtverců odchylek od celkového průměru S_c

$$S_c = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu})^2$$

Ize rozložit na dvě složky

$$S_c = \sum_{i=1}^K \sum_{j=1}^{n_i} ((y_{ij} - \hat{\mu}_i) + (\hat{\mu}_i - \hat{\mu}))^2 = S_A + S_R$$



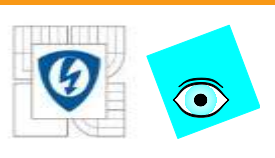
Modely s pevnými efekty

kde S_A je mezi jednotlivými úrovněmi faktoru A

$$S_A = \sum_{i=1}^K n_i (\hat{\mu}_i - \hat{\mu})^2$$

kde S_R je reziduální S_R , tj. uvnitř jednotlivých úrovní faktoru A

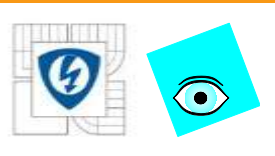
$$S_R = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2$$



Faktor ANOVA u modelu s pevnými efekty

Součet čtverců	Počet st. voln.	Průměrný čtverec	Očekávaná hodnota
Mezi úrovněmi $S_A = \sum_{i=1}^K n_i (\hat{\mu}_i - \hat{\mu})^2$	$K - 1$	$\frac{S_A}{K - 1}$	$\sigma_e^2 + \frac{\sum_{i=1}^K n_i \alpha_i^2}{K - 1}$
Reziduální $S_R = \sum_{i=1}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2$	$N - K$	$\frac{S_R}{N - K}$	σ_e^2
Celkový $S_A + S_R$	$N - 1$	-	-

Hypotéza nulová $H_0: \alpha_i = 0, i = 1, \dots, K$ vs. Alternativní $H_A: \alpha_i \neq 0, i = 1, \dots, K$

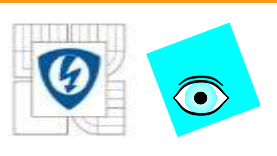


Fisher-Snedecorova F-testu:

$$F_{exp} = \frac{S_A(N - K)}{S_R(K - 1)}$$

Testování:

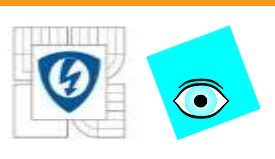
Je-li $F_{exp} > F_{1-\alpha}(K - 1, N - K)$, je nutné H_0 zamítnout a efekty považovat významné.



Příklad: Testování kvality $AgNO_3$ od různých výrobců

U $AgNO_3$ od pěti dodavatelů byla sledována kvalita chemikálie. Z každé láhve byl odebrán počet vzorků ($n_1 = n_3 = 6, n_2 = n_5 = 3, n_4 = 4$).

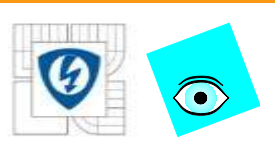
Otázka: Existují významné rozdíly v kvalitě $AgNO_3$ od těchto výrobců.



Pokračování příkladu

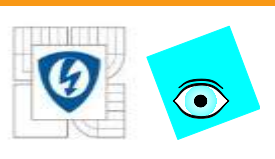
Procentuální obsah chloru při užití $AgNO_3$ od pěti výrobců

Měření	Výrobce				
	V_1	V_2	V_3	V_4	V_5
1	4.40	4.90	5.55	4.45	5.15
2	4.40	4.95	5.10	5.45	6.25
3	5.20	5.40	5.50	4.65	6.14
4	5.45	-	5.98	4.40	-
5	5.80	-	5.60	-	-
6	5.60	-	5.56	-	-

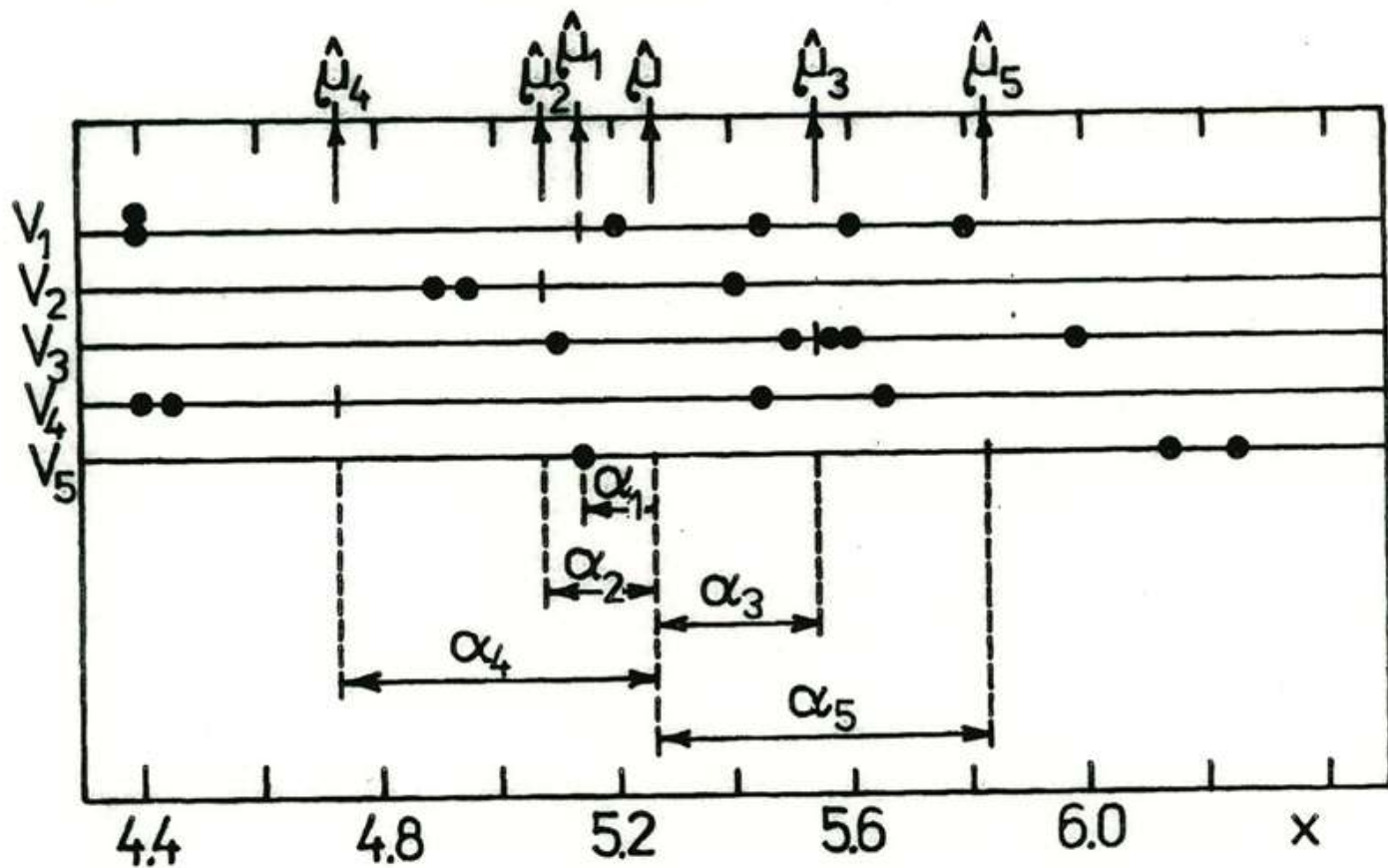


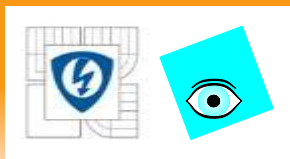
Pokračování příkladu

Řešení: $\hat{\mu} = 5.3715$, $\hat{\mu}_1 = 5.1417$, $\hat{\mu}_2 = 5.0833$,
 $\hat{\mu}_3 = 5.5483$, $\hat{\mu}_4 = 4.7375$, $\hat{\mu}_5 = 5.8467$, $\hat{\alpha}_1 =$
 -0.1298 , $\hat{\alpha}_2 = -0.1882$, $\hat{\alpha}_3 = 0.277$, $\hat{\alpha}_4 = -0.534$,
 $\hat{\alpha}_5 = -0.575$,



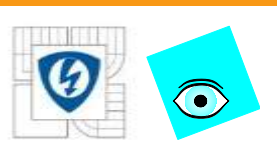
Pokračování příkladu





Tabulka analýzy rozptylu pro jednoduché třídění u modelu s pevnými efekty

Součet čtverců	Počet stupňů volnosti	Průměrný čtverec	Očekávaná hodnota
Mezi úrovněmi S_A	$K - 1$	$\frac{S_A}{K - 1}$	$\sigma_e^2 + \frac{\sum_{i=1}^K n_i \alpha_i^2}{K - 1}$
Reziduální S_R	$N - K$	$\frac{S_R}{N - K}$	σ_e^2
Celkový S_c	$N - 1$	-	-

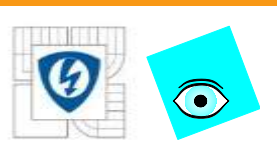


Analýza rozptylu pro obsah $AgNO_3$

Pro $\alpha = 0.05$ je kvantil $F_{0.95}(4,17) = 2.96$.

Součet čtverců	Stupně volnosti	Průměrný čtverec	F_e
Mezi výrobci $S_A = 2.8000$	4	0.6999	3.106
Reziduální $S_R = 3.8322$	17	0.2254	-
Celkový $S_c = 6.632$	20	-	-

Závěr: $F_e > F_{0.95}(4,17)$, a proto H_0 je nutné zamítnout.
Kvalita $AgNO_3$ od pěti výrobců se významně liší.

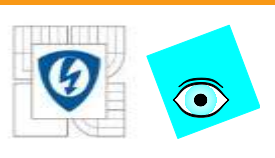


Ověření normality chyb

1. Rankitové grafy
2. Standardizovaná rezidua

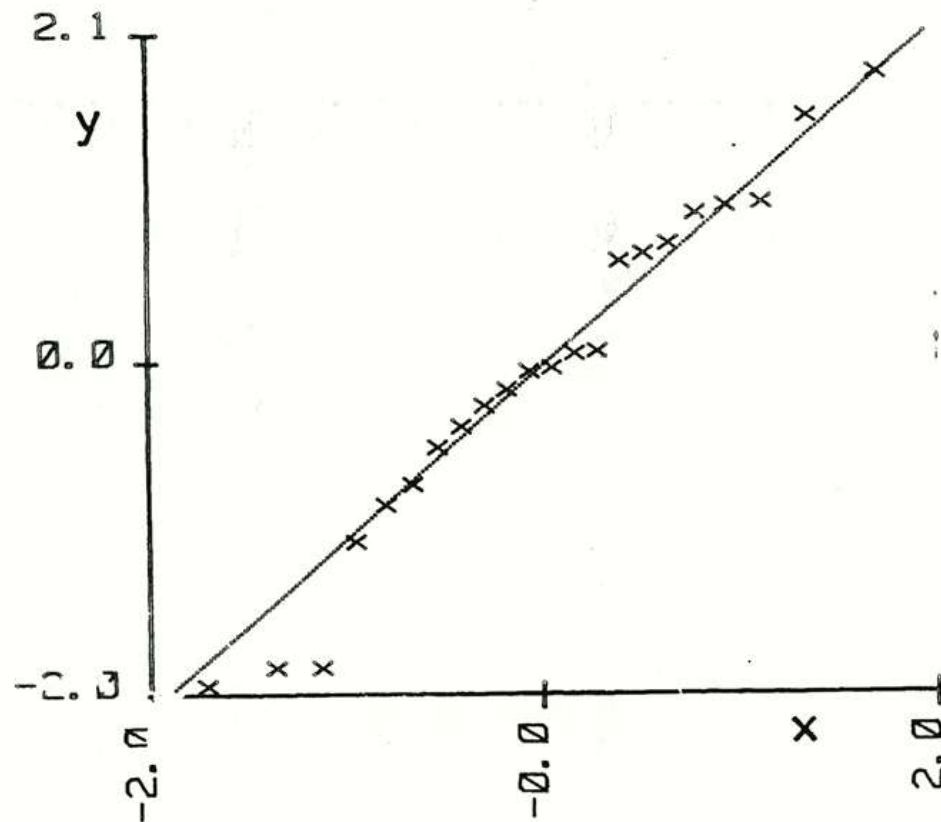
$$\hat{e}_{Si} = \frac{\hat{e}_{ji}}{\hat{\sigma} \sqrt{1 - \frac{1}{n_i}}}$$

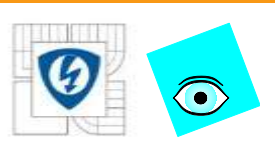
mají přibližně normální rozdělení $N(0,1)$



Indikace $\hat{\varepsilon}_{ij} \approx N(0, \sigma^2)$

Rankitový graf, tzn. přímka s nulovým úsekem a jednotkovou směrnicí





Technika vícenásobného porovnávání

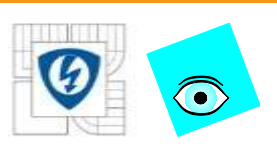
Vliv jednotlivých efektů je významný, tj. rozdíly mezi průměry $\mu_i, \mu_j, i \neq j$ jsou významné.

Scheffého metoda vícenásobného porovnání:

$H_0: \mu_i = \mu_j$ se zamítá pro všechny dvojice (i, j) , pro které platí

$$|\hat{\mu}_i - \hat{\mu}_j| \geq \sqrt{(K - 1) \hat{\sigma}^2 F_{1-\alpha}(K - 1, N_K) \left[\frac{1}{n_i} + \frac{1}{n_j} \right]}$$

pro všechny možné dvojice indexů (i, j) .



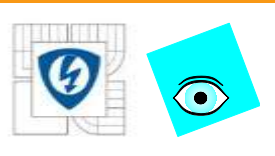
Transformace k zesymetričtění rozdělení

Přiblížení k normalitě transformací, např. posunutou logaritmickou transformací

$$y^* = \ln(y + C)$$

Optimální C se volí tak,

1. aby rezidua byla přibližně symetrická,
2. aby rankitový graf $\hat{e}_{Si}$ byl lineární závislostí.



Vyšetření vybočujících hodnot

Užívají se Jackknife rezidua e_{Jij}

$$\hat{e}_{Jij} = \hat{e}_{sij} \sqrt{\frac{N - K - 1}{N - K - \hat{e}_{sij}^2}}$$

Test: pro $\hat{e}_{Jij}^2 > 10$ lze y_{ij} považovat za silně vybočující.



ANALÝZA ROZPTYLU

Jednofaktorová analýza rozptylu s pevnými efekty

Název: h511

VSTUP

(1) DATA A PODMÍNKY:

Hladina významnosti alfa : 0.050
 Transformace : Ne
 Počet úrovní faktoru A, k : 5
 Total size $n = n[1] + n[2] + \dots + n[k]$: 20
 Jméno výstupního souboru : H511.TXT

(2) OPAKOVANÁ POZOROVÁNÍ NA RŮZNÝCH ÚROVNÍCH FAKTORU A:

Úrovně efektu A

Velikost $n[1] = 4$ $n[2] = 4$ $n[3] = 4$ $n[4] = 4$ $n[5] = 4$

Opakování

1	8.1000E+01	8.5000E+01	8.7000E+01	9.4000E+01	8.8000E+01
2	7.7000E+01	9.1000E+01	8.2000E+01	9.0000E+01	8.6000E+01
3	8.3000E+01	8.8000E+01	8.9000E+01	8.6000E+01	9.1000E+01
4	8.4000E+01	9.0000E+01	8.7000E+01	9.1000E+01	9.0000E+01

VÝSTUP

(1) PRŮMĚRY A EFEKTY ÚROVNÍ:

Celkový průměr = 8.7000E+01
 Reziduální rozptyl = 8.2667E+00

Úroveň	Průměr	Efekt	Hii
1	8.1250E+01	-5.7500E+00	2.5000E-01
2	8.8500E+01	1.5000E+00	2.5000E-01
3	8.6250E+01	-7.5000E-01	2.5000E-01
4	9.0250E+01	3.2500E+00	2.5000E-01
5	8.8750E+01	1.7500E+00	2.5000E-01

(2) DETEKCE VLIVNÝCH BODŮ, (Vybočující a odlehlé body)

Závěr	Odlehlé	Vybočující
	Žádné	Žádné



(3) TABULKA ANOVA:

H₀: Efekty faktorů A jsou nulové, H_A: ... nejsou nulové

Kvantil $F(1-\alpha, k-1, n-k) = 3.056$

Zdroj rozptylu	Stupně volnosti	Součet čtverců	Průměrný čtverec	Testovací kritérium	Závěr H ₀ je	Spočtená hlad.výz.
Mezi úrovněmi	k-1 = 4	1.9800E+02	4.9500E+01	5.988	Zamítnuta	0.004
Rezidua	n-k = 15	1.2400E+02	8.2667E+00			
Celkový	n-1 = 19	3.2200E+02	1.6947E+01			

(4) VÍCENÁSOBNÁ POROVNÁNÍ (Sheffeho procedura):

Hypotéza H ₀	Průměrný rozdíl	Meze dolní	konfidenčního horní intervalu	Závěr
P1 = P2	-7.250	-14.358	-0.142	Zamítnuta
P1 = P3	-5.000	-12.108	2.108	Akceptována
P1 = P4	-9.000	-16.108	-1.892	Zamítnuta
P1 = P5	-7.500	-14.608	-0.392	Zamítnuta
P2 = P3	2.250	-4.858	9.358	Akceptována
P2 = P4	-1.750	-8.858	5.358	Akceptována
P2 = P5	-0.250	-7.358	6.858	Akceptována
P3 = P4	-4.000	-11.108	3.108	Akceptována
P3 = P5	-2.500	-9.608	4.608	Akceptována
P4 = P5	1.500	-5.608	8.608	Akceptována

(5) ZKOUŠKA TRANSFORMACE:

Korelační koeficient, R : -0.253

(Pokud je R blízké nule není transformace nutná.)



Analysis of Variance Report

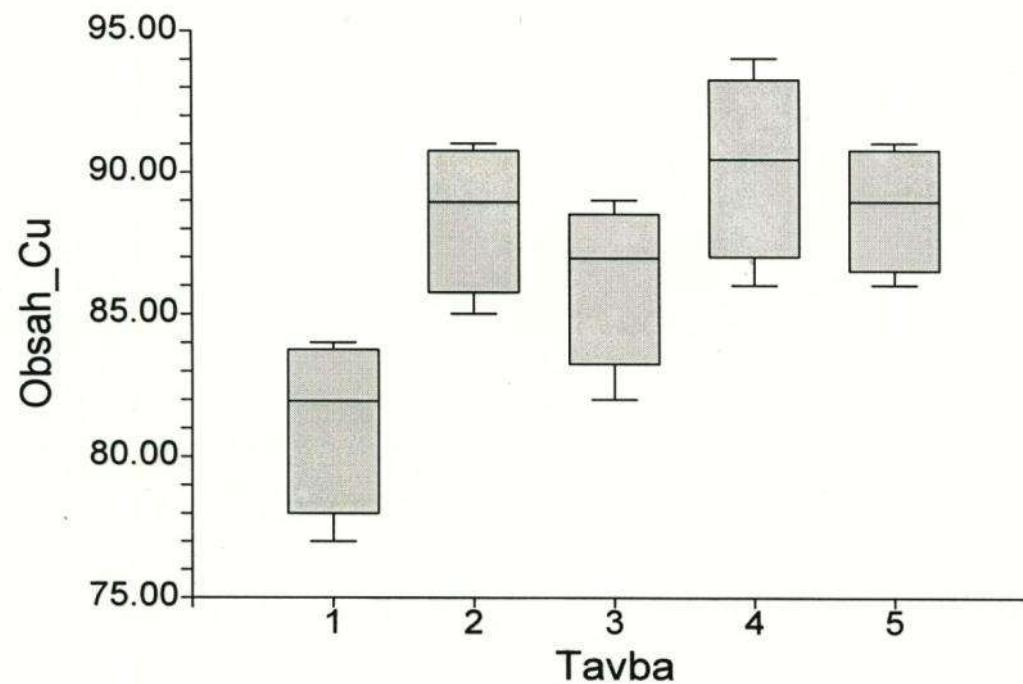
Page/Date/Time 1 04-27-1999 15:07:24
 Database C:\Program Files\NCSS97\Data\H511.S0
 Response Obsah_Cu

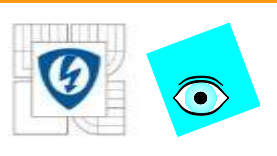
Tests of Assumptions Section

Assumption	Test Value	Prob Level	Decision (0.05)
Skewness Normality of Residuals	-1.1439	0.252672	Accept
Kurtosis Normality of Residuals	-0.9800	0.327102	Accept
Omnibus Normality of Residuals	2.2688	0.321614	Accept
Modified-Levene Equal-Variance Test	0.0765	0.988310	Accept

Box Plot Section

Box Plot





Druhy analýzy rozptylu

Jednofaktorová analýza rozptylu (faktor A)

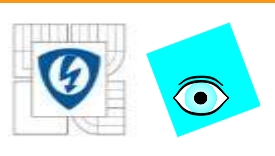
Rozklad $\mu_i: \mu_i = \mu + \alpha_i$

- Na průměr μ ze všech úrovní faktoru A
- Efekt α_i od i -té úrovně faktoru A

Nulová hypotéza $H_0: \mu_1 = \mu_2 = \mu_3$, neboli $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$.

Vícefaktorová analýza rozptylu (faktory A, B, C, ...)

Rozklad $\mu_{ij}: \mu_{ij} = \mu + \alpha_i + \beta_j + \tau_{ij}$, na celkový průměr μ , složky α_i odpovídající vlivu faktoru A, složky β_j odpovídající vlivu faktoru B, a interakce τ_{ij} faktoru A s faktorem B.



Dvoufaktorová analýza rozptylu

V cele je obecně n_{ij} pozorování.

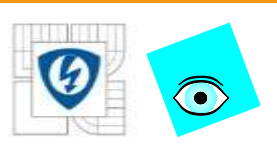
ANOVA bez opakování: v každé cele je jedno pozorování

$$y_{ij} = \mu_{ij} + \varepsilon_{ij}$$

Řádkové efekty α_i , sloupcové efekty β_i , interakce τ_{ij} .

	B ₁	B ₂	...	B _M	
A ₁	.	.	...	.	
A ₂	.		...	.	
.	.	.	...	.	
.	.	.	...	.	
.	.	.	...	.	
A _N	.	.	...	.	

cela A₂ B₂



Modely interakcí

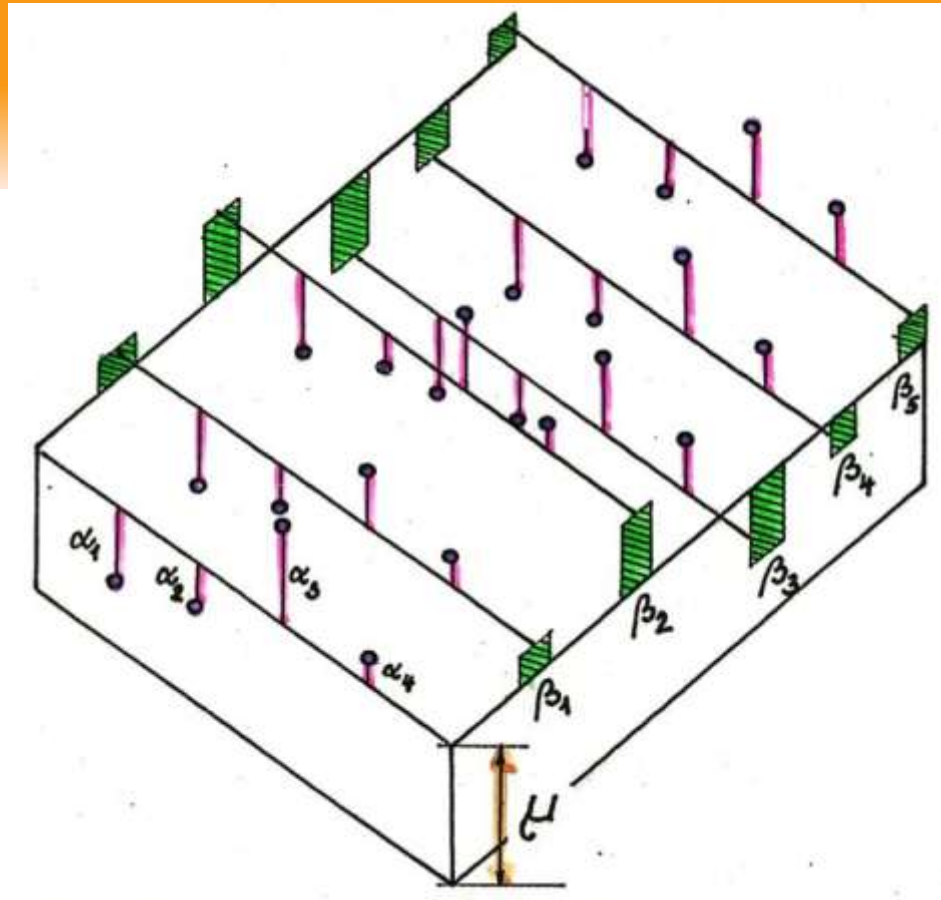
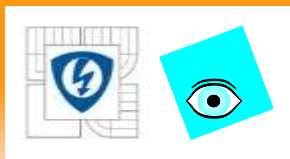
Tukeyův model interakce $\tau_{ij} = C\alpha_i\beta_j$

kde C je konstanta.

Řádkově lineární model interakcí $\tau_{ij} = \gamma_i\beta_jC_R$

Sloupcově lineární model interakcí $\tau_{ij} = \alpha_iC_K\delta_j$

Aditivně multiplikativní model interakcí $\tau_{ij} = \gamma_i\delta_jC_W$

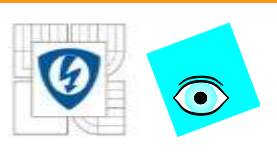


$$H_0: \bigcirc = \textcolor{red}{I} = \textcolor{green}{\text{hatched box}}$$

$$\epsilon_{ij} = \alpha_i = \beta_j$$

$$\text{vs. } H_A: \bigcirc \neq \textcolor{red}{I} \neq \textcolor{green}{\text{hatched box}}$$

$$\epsilon_{ij} \neq \alpha_i \neq \beta_j$$



Modely s pevnými efekty bez opakování

(každá cela: 1 hodnota)

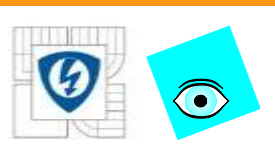
Předpoklady:

1. Náhodné chyby ε_{ij} jsou nezávislé náhodné veličiny s normálním rozdělením $N(0, \sigma^2)$.

2. Omezující podmínky

$$\sum_{i=1}^N \alpha_i = 0; \sum_{j=1}^M \beta_j = 0; \sum_{i=1}^N \tau_{ij} = 0; \sum_{j=1}^M \tau_{ij} = 0$$

3. U modelů bez interakce je $\tau_{ij} = 0$ pro $i = 1, N$, $j = 1, M$.



Odhady parametrů

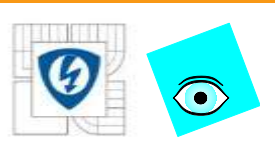
$$\hat{\mu} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M y_{ij}$$

$$\hat{\alpha}_i = \frac{1}{M} \sum_{j=1}^M y_{ij} - \hat{\mu}$$

$$\hat{\beta}_j = \frac{1}{N} \sum_{i=1}^N y_{ij} - \hat{\mu}$$

Výpočet rezidua $\hat{e}_{ij}$ podle $\hat{e}_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$

Výpočet interakce $\tau_{ij} = E(y_{ij}) - \mu - \alpha_i - \beta_j \approx \hat{e}_{ij}$



Tukeyův model interakce

Ze směrnice C přímky grafu neaditivní $\hat{e}_{ij}$ vs. $\hat{\alpha}_i \hat{\beta}_j$ se odhaduje míra interakce

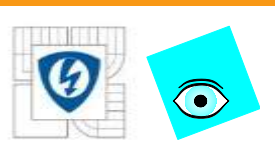
$$\hat{C} = \frac{\sum_{i=1}^N \sum_{j=1}^M \hat{e}_{ij} \hat{\alpha}_i \hat{\beta}_j}{\sum_{i=1}^N \sum_{j=1}^M \hat{\alpha}_i^2 \hat{\beta}_j^2}$$

Nenulová směrnice znamená interakci faktorů a **součet čtverců odchylek Tukeyho interakce** S_T je k testování

$$S_T = \frac{\left(\sum_{i=1}^N \sum_{j=1}^M y_{ij} \hat{\alpha}_i \hat{\beta}_j \right)^2}{\sum_{i=1}^N \sum_{j=1}^M \hat{\alpha}_i^2 \hat{\beta}_j^2}$$



Součet čtverců pro	Stupně volnosti	Průměrný čtverec	Kritérium F
Faktor A $S_A = M \sum_{i=1}^N \alpha_i^2$	N - 1	$M_A = S_A / (N-1)$	$F_A = M_A / M_{AB}$
Faktor B $S_B = N \sum_{j=1}^M \beta_j^2$	M - 1	$M_B = S_B / (M-1)$	$F_B = M_B / M_{AB}$
Interakce (Tukey) $S_T = \dots$	1	$M_T = S_T$	$F_T = M_T / M_E$
Reziduální $S_R = S_{AB} - S_T$	N M - N - M	$M_E = S_R / (NM - N - M)$	-
Celkový $S_C = \sum_{(i)} \sum_{(j)} (\hat{\mu} - y_{ij})^2$	N M - 1	-	-



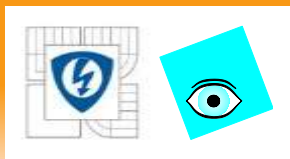
Reziduální součet čtverců bez interakcí

S_{AB} značí reziduální součet čtverců bez interakcí

$$S_{AB} = \sum_{i=1}^N \sum_{j=1}^M (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

a odpovídající průměrný čtverec

$$M_{AB} = \frac{S_{AB}}{(N - 1)(M - 1)}$$



Příklad: Stanovení vody v rozpouštědlech v různých laboratořích

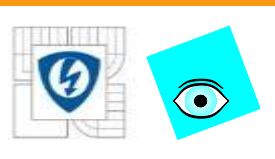
U všech vzorků A_1 , A_2 a A_3 nového rozpouštědla byl ve čtyřech laboratořích B_1 , B_2 , B_3 a B_4 určen obsah vody.

Otázka: jsou významné odchylky v obsahu vody v zadaných vzorcích rozpouštědla a ve výsledcích zvolených laboratoří?

Data: $N = 3$, $M = 4$

Obsah vody [%] v rozpouštědle určený v různých laboratořích

Vzorek	Laboratoř			
	B_1	B_2	B_3	B_4
A_1	1.35	1.13	1.06	0.98
A_2	1.40	1.23	1.26	1.22
A_3	1.49	1.46	1.40	1.35



Pokračování příkladu

Řešení: $\hat{\mu} = 1.2775$;

$\hat{\alpha}_1 = -0.1475$; $\hat{\alpha}_2 = 0$; $\hat{\alpha}_3 = 0.1475$;

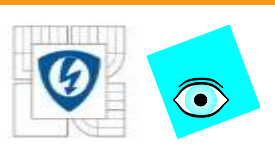
$\hat{\beta}_1 = 0.1358$; $\hat{\beta}_2 = 0.0042$; $\hat{\beta}_3 = -0.0375$ $\hat{\beta}_4 = -0.0942$.

$\hat{C} = -3.532$

$S_T = 0.0156$,

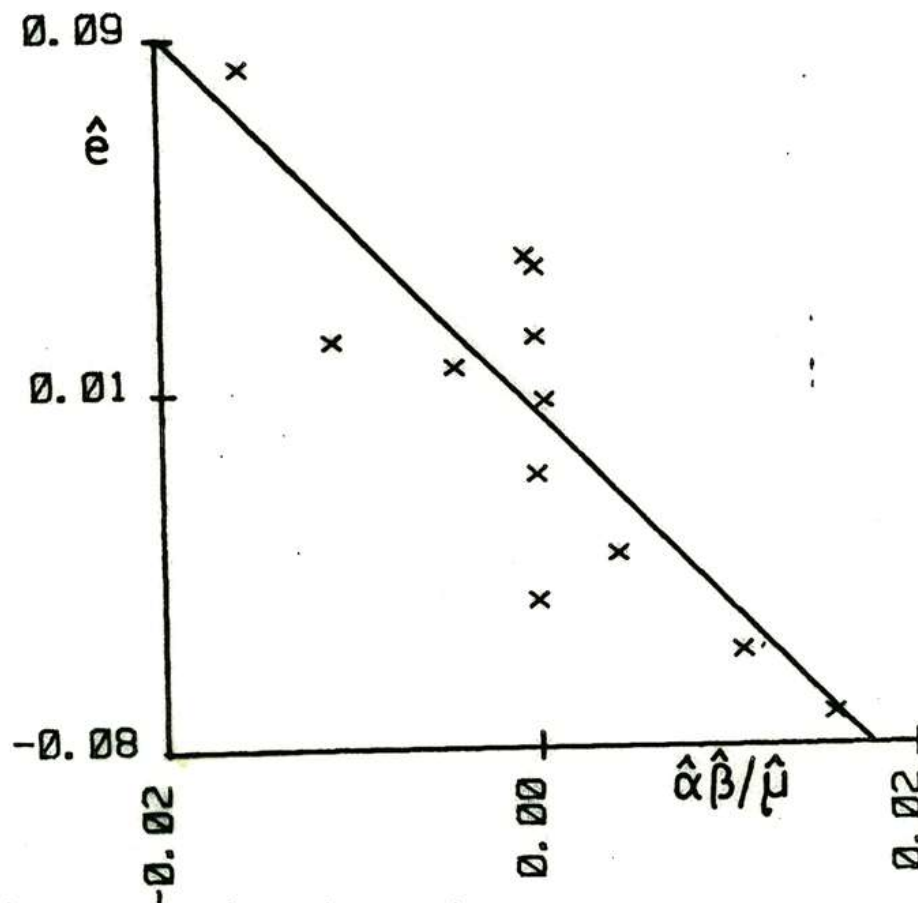
$S_{AB} = 0.02215$,

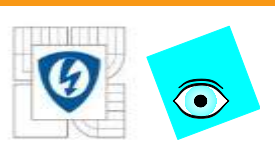
$M_{AB} = 0.003692$.



Pokračování příkladu

Graf neaditivit vykazuje výrazný trend



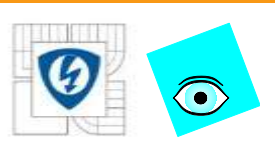


Pokračování příkladu

Analýza rozptylu dat obsahu vody v rozpouštědlech

Součet čtverců pro	Stupně volnosti	Průměrný čtverec	Kritérium F
Vzorky A, $S_A=0.174$	2	0.087	19.64
Laboratoře B, $S_B=0.0862$	3	0.0287	6.49
Interakce Tukey, $S_T= 0.0156$	1	0.0156	3.522
Reziduální, $S_R= 0.0222$	5	0.00443	-
Celkový, $S_C=0.2824$	11	0.02568	-

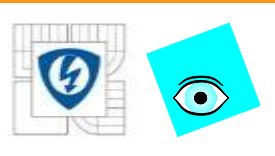
$$F_{0.95} (1, 5) = 6.61, F_{0.95} (2, 5) = 5.79 \text{ a } F_{0.95} (3, 5) = 5.41.$$



Pokračování příkladu

Závěr:

Efekt interakce je nevýznamný a lze použít aditivní model analýzy rozptylu, zatímco efekty vzorků a laboratoří významné jsou.



Vyvážené modely

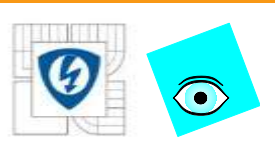
V každé cele je $n_{ij} = n$ pozorování.

Odhadem μ_{ij} jsou aritmetické průměry

$$\hat{\mu}_{ij} = \frac{1}{n} \sum_{k=1}^n y_{ijk} \quad \hat{\mu} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \hat{\mu}_{ij}$$

$$\hat{\alpha}_i = \frac{1}{M} \sum_{j=1}^M \hat{\mu}_{ij} - \hat{\mu}$$

$$\hat{\beta}_j = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_{ij} - \hat{\mu}$$



Vyvážené modely

Odhad reziduí $\hat{e}_{ijk} = y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$

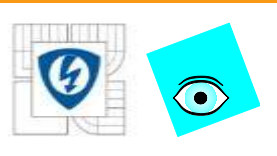
Odhad interakcí $\hat{\tau}_{ijk} = \hat{\mu}_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$

K ověření interakce lze vynášet graf $\hat{\tau}_{ij}$ vs. $\hat{\alpha}_i \hat{\beta}_j$.

Průměrné hodnoty

$$E(M_A) = \sigma^2 + \frac{nM \sum_{i=1}^N \alpha_i^2}{(N-1)\sigma^2} = \sigma^2 + nM\sigma_A^2$$

$$E(M_B) = \sigma^2 + \frac{nM \sum_{j=1}^M \beta_j^2}{(M-1)\sigma^2} = \sigma^2 + nN\sigma_B^2$$

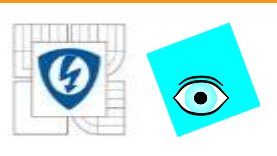


a

$$E(M_{AB}) = \sigma^2 + \frac{n \sum_{i=1}^N \sum_{j=1}^M \tau_{ij}^2}{(N-1)(M-1)\sigma^2} = \sigma^2 + n\sigma_{AB}^2$$

Rozptyl M_R je nevychýleným odhadem σ^2 rozptylu chyb.

Rozptyly σ_A^2 , σ_B^2 a σ_{AB}^2 odpovídají efektům řádků, sloupců a interakcí.



Vyvážené modely

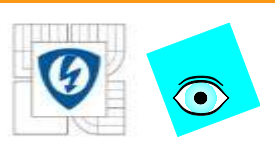
Test: s využitím F_{AB} , F_B a F_A , zda je možné považovat sloupcové a řádkové efekty, resp. interakce za nevýznamné.

$$H_0: \tau_{ij} = 0, i = 1, \dots, N \text{ a } j = 1, \dots, M$$

Je-li $F_{AB} > F_{1-\alpha}\{(N-1)(M-1), MN(n-1)\}$ je H_0 zamítnuta.

$$H_0: \alpha_i = 0, i = 1, \dots, N$$

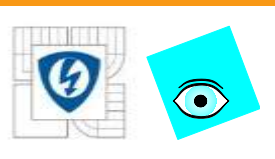
Je-li $F_A > F_{1-\alpha}\{(N-1), MN(n-1)\}$ je H_0 zamítnuta.



Vyvážené modely

$$H_0: \beta_j = 0, j = 1, \dots, M$$

Je-li $F_B > F_{1-\alpha}\{(M-1), M N (n-1)\}$ je H_0 zamítnuta.



Dvoufaktorová ANOVA pro vyvážený experiment

Součet čtverců pro	Stupně volnosti	Průměrný čtverec	Kritérium F
Faktor A $S_A = n M \sum_{i=1}^N \hat{\alpha}_i^2$	$N - 1$	$M_A = \frac{S_A}{N - 1}$	$F_A = \frac{M_A}{M_R}$
Faktor B $S_B = n N \sum_{j=1}^M \hat{\beta}_j^2$	$M - 1$	$M_B = \frac{S_B}{M - 1}$	$F_B = \frac{M_B}{M_R}$
Interakce AB $S_{AB} = n \sum_{i=1}^N \sum_{j=1}^M \hat{\tau}_{ij}^2$	$(N - 1)(M - 1)$	$M_{AB} = \frac{S_{AB}}{(N - 1)(M - 1)}$	$F_{AB} = \frac{M_{AB}}{M_R}$
Reziduální $S_R = \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^n (y_{ijk} - \hat{\mu}_{ij})^2$	$M N n - 1$	$M_R = \frac{S_R}{M N (n - 1)}$	-
Celkový $S_C = \sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^n (y_{ijk} - \hat{\mu})^2$	$M N (n - 1)$	-	-

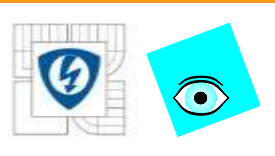


Příklad: Přesnost chromatografického stanovení diethylenglykolu

Tři laboranti A_1 , A_2 a A_3 provádějí dvě opakovaná stanovení diethylenglykolu (DEG) v ethylenglykolu na třech chromatografech B_1 , B_2 a B_3 .

Otázka: má na výsledek analýzy přístroj či laborant

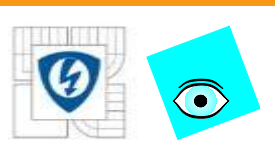
Data: $N = 3$, $M = 3$, $n = 2$.



Pokračování příkladu

Obsahy DEG [%], měření třemi laboranty A na třech přístrojích B

Laborant	Přístroj		
	B ₁	B ₂	B ₃
A ₁	0.110	0.101	0.108
	0.116	0.102	0.109
A ₂	0.112	0.115	0.111
	0.111	0.106	0.109
A ₃	0.114	0.107	0.113
	0.112	0.109	0.110

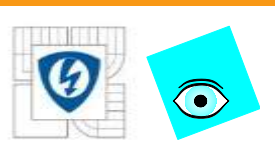


Pokračování příkladu

Řešení:

Součet čtverců pro	Stupně volnosti	Průměrný čtverec	F-kritérium
Laboranty (A): $S_A = 3.811 \cdot 10^{-5}$	2	$1.906 \cdot 10^{-5}$	2.433
Přístroje (B): $S_B = 1.028 \cdot 10^{-4}$	2	$5.139 \cdot 10^{-5}$	6.560
Interakce (AB): $S_{AB} = 6.022 \cdot 10^{-5}$	4	$1.506 \cdot 10^{-5}$	1.922
Reziduální: $S_R = 7.050 \cdot 10^{-5}$	9	$7.833 \cdot 10^{-6}$	-
Celkový	17	-	-

$$F_{0.95}(2,9) = 4.26 \text{ a } F_{0.95}(4,9) = 3.63$$



Pokračování příkladu

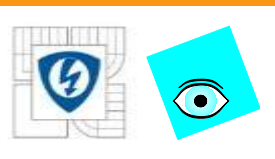
Test:

1. Mají laboranti vliv na výsledek analýz zjistíme vyšetřením nulové hypotézy $H_{01}: \alpha_i = 0$ a zároveň $H_{02}: \tau_{ij} = 0$

$$S_{PA} = \frac{S_A + S_{AB}}{2 + 4} = 1.63 \cdot 10^{-5}$$

a testovací statistika

$$F_{PA} = \frac{1.63 \cdot 10^{-5}}{7.83 \cdot 10^{-6}} = 2.09$$



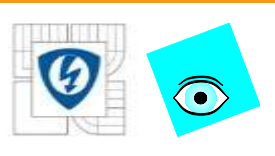
Pokračování příkladu

Jelikož $F_{0.95}(6,9) = 3.373$ je větší než statistika F_{PA} , nemají laboranti významný vliv na výsledek analýz a model ANOVA lze vyjádřit rovnicí

$$y_{ijk} = \mu + \beta_j + e_{ijk}$$

2. Otestujeme $H_0: \beta_j = 0$ a sloučíme příspěvky $S_A + S_{AB}$ s reziduálním součtem čtverců.

$$M_R^* = \frac{S_R + S_A + S_{AB}}{2 + 4 + 9} = \frac{1.688 \cdot 10^{-5}}{15} = 1.12 \cdot 10^{-5}$$



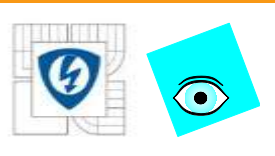
Pokračování příkladu

Testovací statistika $F_B = \frac{M_B}{M_R^*} = 4.58$ je větší než

$F_{0.95}(3,15) = 3.68$ a H_0 je zamítnuta. Vliv faktoru B (přístroje) na výsledek analýzy je na hladině $\alpha = 0.95$ významný.

Střední hodnota průměrného čtverce M_B je $E(M_B) = \sigma^2 + 6\sigma_B^2$. Pro odhad přístrojové chyby platí

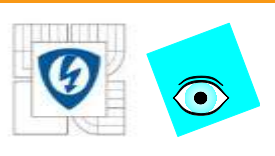
$$\sigma_B^2 + \frac{M_B - \hat{\sigma}^2}{6} = 6.68 \cdot 10^{-6}$$



Závěr:

Na přesnost stanovení diethylenglykolu má statisticky významný vliv pouze použitý chromatograf.

Variabilita způsobaná laboranty je $\sigma^2 = 1.12 \cdot 10^{-5}$,
variabilita způsobená přístroji je $\sigma_B^2 = 6.68 \cdot 10^{-6}$.



Nevyvážené modely

V (i, j) -té cele je n_{ij} pozorování

Přibližný rozklad celkového součtu čtverců

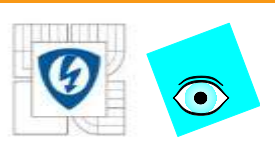
$$\hat{\mu}_{ij} = \frac{1}{n_k} \sum_{k=1}^{n_k} y_{ijk}$$

pro všechny cely.

Reziduální součet čtverců

$$S_R = \sum_{i=1}^N \sum_{j=1}^M \sum_k^{n_k} (y_{ijk} - \hat{\mu}_{ij})^2$$

Pro výpočet dalších složek rozkladu celkového součtu čtverců se používá $\hat{\mu}_{ij}$.



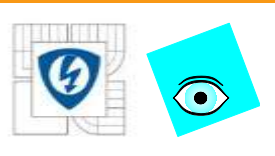
Nevyvážené modely

Jsou určeny z ekvivalentního počtu n^*

$$n^* = \left[\frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \frac{1}{n_{ij}} \right]^{-1}$$

$$S_A = n^* M \sum_{i=1}^N (\hat{\mu}_i - \hat{\mu})^2 \quad \text{s } (N - 1) \text{ stupni volnosti}$$

$$S_B = n^* N \sum_{j=1}^M (\hat{\mu}_j - \hat{\mu})^2 \quad \text{s } (M - 1) \text{ stupni volnosti}$$



Nevyvážené modely

$$S_{AB} = n^* \sum_{i=1}^N \sum_{j=1}^M (\hat{\mu}_{ij} - \hat{\mu}_i - \hat{\mu}_j + \hat{\mu})^2$$

s $(N - 1)(M - 1)$ stupni volnosti.

Je použito

$$\hat{\mu}_i = \frac{1}{M} \sum_{j=1}^M \hat{\mu}_{ij}; \hat{\mu}_j = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_{ij}; \hat{\mu} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M \hat{\mu}_{ij}$$

Testování hypotéz: stejně jako u vyvážených experimentů.