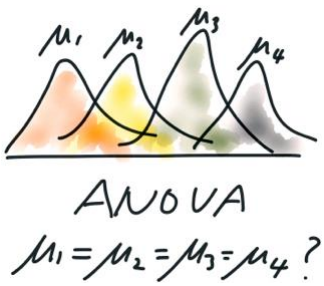
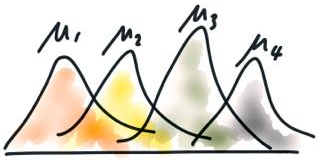
ANALYTICKÁ CHEMOMETRIE

Analýza rozptylu



Patrik Kania
Ústav analytické chemie, VŠCHT Praha

11. 11. 2025



Při použití dvouvýběrového t-testu by bylo potřeba provést šest testů:

$\mu_1 = \mu_2$ $\mu_1 = \mu_3$ $\mu_1 = \mu_4$ $\mu_2 = \mu_3$ $\mu_2 = \mu_4$ $\mu_3 = \mu_4$

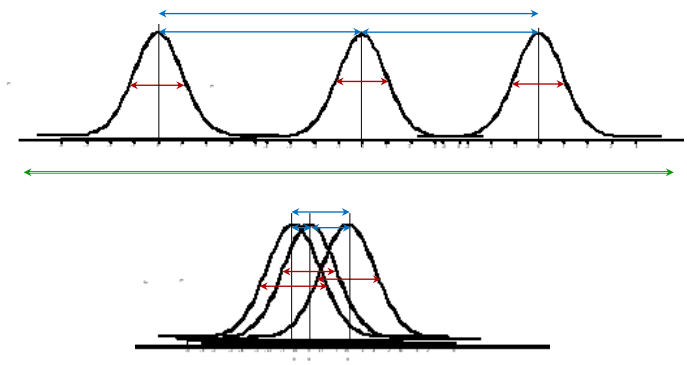
pokud bude každý test proveden na hladině významnosti α = 0,05, pak celková hladina významnosti je

$$\alpha_{\text{celková}} = 1 - (1 - \alpha)^k = 1 - (1 - 0{,}05)^6 = 0{,}265$$

To znamená, že pravděpodobnost falešného závěru vyskočí z 5 na 26,5 %!

Základní předpoklady pro analýzu rozptylu:

- srovnávané výběry pocházejí ze základních souborů s normálním rozdělením
- výběry pocházejí ze základních souborů se shodnými rozptyly
- jednotlivé výběry jsou vzájemně nezávislé
- náhodné chyby e_{ik} jsou náhodné veličiny s N(0, σ^2)



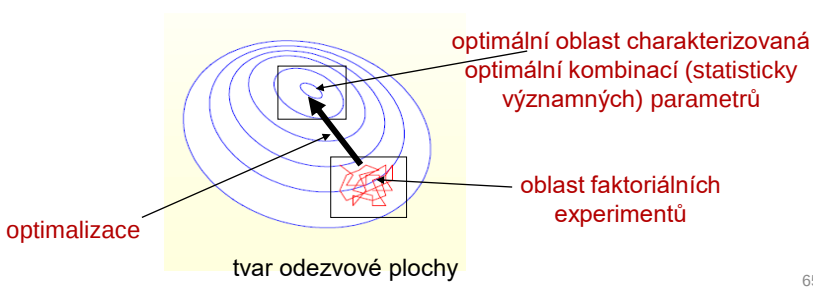
Analýza rozptylu (angl. *analysis of variance*)

- použití v případě testů rovnosti středních hodnot více souborů

  H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

  H_1: existuje minimálně jeden statisticky významný rozdíl mezi dvěma středními hodnotami

- hledání významných faktorů pro optimalizaci



Příklad: Optimalizace separace dvou látek při GC analýze

zkouším měnit: A - teplota při dvou úrovních (1: 70 °C, 2: 100 °C)

B - průtoková rychlost nosného plynu při dvou úrovních (1: 1 ml/s, 2: 5 ml/s)

provedu experiment: např. A_1B_1: udržuji 70 °C a průtok 1 ml/s

A_1B_2: udržuji 70 °C a průtok 5 ml/s

Terminologie faktoriálních experimentů:

parametry - faktory, které testujeme (označení A, B, C, ……)

hodnoty faktorů – úrovně či hladiny (označení A_1, B_2, C_3, …….)

postup - kombinace úrovní faktorů, např. $A_1B_2C_3$

Hodnoty jednotlivých faktorů při matematicko-statistickém vyhodnocení vůbec nepoužíváme!

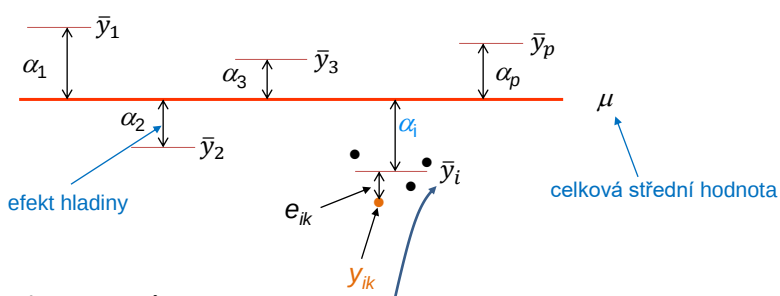
Jednofaktorová analýza rozptylu – s opakováním (angl. *one-way ANOVA*)

model: $y_{ik} = \mu + \alpha_i + e_{ik}$

$i = 1,\ldots, p$ počet úrovní faktoru

$k = 1,\ldots, n$ počet opakování

$N = p \cdot n$ počet pokusů



odhady parametrů:

$$\mu = \frac{\sum_{i=1}^{p}\sum_{k=1}^{n} y_{ik}}{N} \qquad \alpha_i = \frac{\sum_{k=1}^{n} y_{ik}}{n} - \mu \qquad e_{ik} = y_{ik} - \bar{y}_i$$

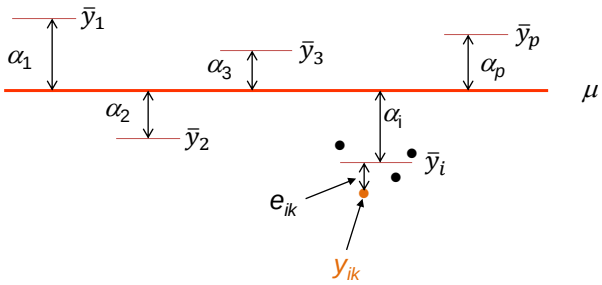
Jednofaktorová analýza rozptylu – s opakováním (angl. *one-way ANOVA*)

H_0: $\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = \bar{y}_i \ldots = \bar{y}_p$

či jinak zapsáno

H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_i \ldots = \alpha_p = 0$

$i = 1,\ldots, p$	počet úrovní faktoru
$k = 1,\ldots, n$	počet opakování
$N = p \cdot n$	počet pokusů



$$S_C = S_A + S_r$$

<u>test:</u> porovnání proměnlivosti dané úrovněmi α_i (meziúrovňová proměnlivost) S_A a reziduální proměnlivosti S_r (vnitroúrovňová proměnlivost)

Test významnosti

H_0: $\sigma_A^2 \leq \sigma_r^2$ H_1: $\sigma_A^2 > \sigma_r^2$

jestliže H_0 platí, není vliv faktoru A

odhad rozptylu způsobeného faktorem A

jednostranný F-test:

$$TK = F = \frac{\frac{S_A}{df_A}}{\frac{S_r}{df_r}}$$

pokud $TK > F_{(\alpha,\, dfA,\, dfr)}$

hypotézu H_0 zamítáme

odhad rozptylu způsobeného náhodnou chybou

df - stupně volnosti

$$S_A = \sum_{i=1}^{p} n_i(\bar{y}_i - \mu)^2 \qquad S_r = \sum_{i=1}^{p}\sum_{k=1}^{n}(y_{ik} - \bar{y}_i)^2 \qquad S_C = \sum_{i=1}^{p}\sum_{k=1}^{n}(y_{ik} - \mu)^2$$

$$df_A = p - 1 \qquad df_r = N - p \qquad df_C = N - 1$$

Jednofaktoriální experiment (plán)

výsledky všech pokusů tvoří experimentální matici **Y**

1 faktor na ***p*** úrovních

úrovně faktoru A	*n* - opakování při *i*-té úrovni faktoru hodnoty pozorování				skupinové průměry	počet
A_1	y_{11}	y_{12}	…..	y_{1n}	$\bar{y}_1$	n_1
……	……	y_{ik}	……	y_{in}	$\bar{y}_i$	n_i
A_p	y_{p1}	y_{p2}	…..	y_{pn}	$\bar{y}_p$	n_p
celkem					μ	N

Příklad: Průběh reakce může být ovlivňován třemi různými biochemickými katalyzátory A, B a C. Zjistěte, zda výtěžek reakce (závisle proměnná) je významně ovlivňován některým z katalyzátorů. Výsledky měření (počet opakování měření je 8):

A: 85 86 83 82 87 90 80 81
B: 87 86 85 93 89 88 86 89
C: 89 85 90 86 83 88 87 91

úrovně faktoru									$\bar{y}_i$	n_i
A	85	86	83	82	87	90	80	81	84,250	8
B	87	86	85	93	89	88	86	89	87,875	8
C	89	87	90	88	83	88	89	91	88,125	8
celkem									86,750	24

$$S_A = \sum_{i=1}^{3} n_i(\bar{y}_i - \bar{y})^2 = 8(84{,}25 - 86{,}750)^2 + 8(87{,}875 - 86{,}750)^2 + 8(88{,}125 - 86{,}750)^2 = 75{,}25$$

$$S_r = \sum_{i=1}^{3}\sum_{k=1}^{8}(y_{iv} - \bar{y}_i)^2 = 169{,}25$$

$$TK = F = \frac{\frac{75{,}25}{3-1}}{\frac{169{,}25}{24-3}} = 4{,}67 \qquad F_{(0{,}05,\,2,\,21)} = 3{,}47$$

$4{,}67 > 3{,}47$ H_0 lze zamítnout

Co s výsledkem analýzy?

H_0 nezamítnuta → konec

H_0 zamítnuta → porovnání efektů → Které skupiny se od sebe statisticky liší?

Testy mnohonásobného porovnávání (angl. *multiple comparison testing*)

- vzájemné

$\bar{y}_1 \longleftrightarrow \bar{y}_2$

$\bar{y}_2 \longleftrightarrow \bar{y}_3$

$\bar{y}_3 \longleftrightarrow \bar{y}_1$

nejpouž. Tukeyho test

Scheffeho test

- s kontrolní skupinou $\bar{y}_k$

$\bar{y}_1 \longleftrightarrow \bar{y}_k$

$\bar{y}_2 \longleftrightarrow \bar{y}_k$

$\bar{y}_3 \longleftrightarrow \bar{y}_k$

nejpouž. Dunnetův test

Tukeyho metoda mnohonásobného porovnávání (angl. *Tukey's test*)

H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$

- patří k nejužívanějším a považuje se také za jednu z nejlepších z hlediska kompromisu síly testu (1-β) a možnosti výskytu chyby α

$$\boxed{TK = q = \frac{|\bar{y}_A - \bar{y}_B|}{SE}}$$

$$SE = \sqrt{\frac{\frac{S_r}{df_r}}{n}}$$

$n_A = n_B = n$

$$SE = \sqrt{\frac{\frac{S_r}{df_r}}{2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$n_A \neq n_B$

pokud $q > q_{\alpha, p, N-p}$ (kritická hodnota studentizovaného rozpětí), potom je rozdíl středních hodnot μ_A a μ_B statisticky významný

p - je počet úrovní faktoru, *N-p* - je rozdíl mezi celkovým počtem měření a úrovní

Kritické hodnoty $q_{\alpha, p, N-p}$ studentizovaného rozdělení

$\alpha = 0{,}05$	*p*													
N-p	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	18	27	32,8	37,1	40,4	43,1	45,4	47,4	49,1	50,6	52	53,2	54,3	55,4
2	6,08	8,33	9,8	10,9	11,7	12,4	13	13,5	14	14,4	14,7	15,1	15,4	15,7
3	4,5	5,91	6,82	7,5	8,04	8,48	8,85	9,18	9,46	9,72	9,95	10,2	10,3	10,5
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,6	7,83	8,03	8,21	8,37	8,52	8,66
5	3,64	4,6	5,22	5,67	6,03	6,33	6,58	6,8	6,99	7,17	7,32	7,47	7,6	7,72
6	3,46	4,34	4,9	5,3	5,63	5,9	6,12	6,32	6,49	6,65	6,79	6,92	7,03	7,14
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,3	6,43	6,55	6,66	6,76
8	3,26	4,04	4,53	4,89	5,17	5,4	5,6	5,77	5,92	6,05	6,18	6,29	6,39	6,48
9	3,2	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98	6,09	6,19	6,28
10	3,15	3,88	4,33	4,65	4,91	5,12	5,3	5,46	5,6	5,72	5,83	5,93	6,03	6,11
11	3,11	3,82	4,26	4,57	4,82	5,03	5,2	5,35	5,49	5,61	5,71	5,81	5,9	5,98
12	3,08	3,77	4,2	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61	5,71	5,8	5,88
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53	5,63	5,71	5,79
14	3,03	3,7	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46	5,55	5,64	5,71
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,2	5,31	5,4	5,49	5,57	5,65
16	3	3,65	4,05	4,33	4,56	4,74	4,9	5,03	5,15	5,26	5,35	5,44	5,52	5,59
17	2,98	3,63	4,02	4,3	4,52	4,7	4,86	4,99	5,11	5,21	5,31	5,39	5,47	5,54
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27	5,35	5,43	5,5
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23	5,31	5,39	5,46
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,9	5,01	5,11	5,2	5,28	5,36	5,43
24	2,92	3,53	3,9	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,1	5,18	5,25	5,32
30	2,89	3,49	3,85	4,1	4,3	4,46	4,6	4,72	4,82	4,92	5,0	5,08	5,15	5,21
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,9	4,98	5,04	5,11
60	2,83	3,4	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81	4,88	4,94	5,0
120	2,8	3,36	3,68	3,92	4,1	4,24	4,36	4,47	4,56	4,64	4,71	4,78	4,84	4,9
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,47	4,55	4,62	4,68	4,74	4,8

Příklad: Průběh reakce může být ovlivňován třemi různými biochemickými katalyzátory A, B a C. Zjistěte, zda výtěžek reakce (závisle proměnná) je významně ovlivňován některým z katalyzátorů. Dále ověřte, zda účinek některého z katalyzátorů se významně odlišuje od ostatních. Výsledky měření (počet opakování měření je 8):

A: 85 86 83 82 87 90 80 81
B: 87 86 85 93 89 88 86 89
C: 89 85 90 86 83 88 87 91

srovnání	rozdíl skupinových ∅	*SE*	*TK*	$q_{0,05;\,3;\,21}$	
A – B	- 3,625	$\sqrt{8{,}060/8} = 1{,}003$	3,614	3,58	
A – C	- 3,875	$\sqrt{8{,}060/8} = 1{,}003$	3,863	3,58	
B – C	- 0,250	$\sqrt{8{,}060/8} = 1{,}003$	0,249	3,58	

úrovně faktoru									$\bar{y}_i$	n_i
A	85	86	83	82	87	90	80	81	84,250	8
B	87	86	85	93	89	88	86	89	87,875	8
C	89	87	90	88	83	88	89	91	88,125	8
celkem									86,750	24

$$S_A = \sum_{i=1}^{3} n_i(\bar{y}_i - \bar{y})^2 = 8(84{,}25 - 86{,}750)^2 + 8(87{,}875 - 86{,}750)^2 + 8(88{,}125 - 86{,}750)^2 = 75{,}25$$

$$S_r = \sum_{i=1}^{3}\sum_{k=1}^{8}(y_{iv} - \bar{y}_i)^2 = 169{,}25$$

$$TK = F = \frac{\frac{75{,}25}{3-1}}{\frac{169{,}25}{24-3}} = 4{,}67 \qquad F_{(0{,}05,\,2,\,21)} = 3{,}47$$

$4{,}67 > 3{,}47$ H_0 neplatí

Příklad: Průběh reakce může být ovlivňován třemi různými biochemickými katalyzátory A, B a C. Zjistěte, zda výtěžek reakce (závisle proměnná) je významně ovlivňován některým z katalyzátorů. Dále ověřte, zda účinek některého z katalyzátorů se významně odlišuje od ostatních. Výsledky měření (počet opakování měření je 8):

A: 85 86 83 82 87 90 80 81
B: 87 86 85 93 89 88 86 89
C: 89 85 90 86 83 88 87 91

srovnání	rozdíl skupinových ∅	*SE*	*TK*	$q_{0,05;\,3;\,21}$	
A – B	- 3,625	$\sqrt{8{,}060/8} = 1{,}003$	3,614	3,58	zamítáme
A – C	- 3,875	$\sqrt{8{,}060/8} = 1{,}003$	3,863	3,58	zamítáme
B – C	- 0,250	$\sqrt{8{,}060/8} = 1{,}003$	0,249	3,58	nezamítáme

Katalyzátor A se statisticky významně odlišuje od katalyzátorů B a C.

Scheffého metoda mnohonásobného porovnávání (angl. *Scheffé test*)

H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$

- slabší než Tukeyho test (vyšší náchylnost k chybě $\beta \rightarrow$ obvykle odhalí méně rozdílů mezi středními hodnotami než Tukeyho test)

$$\boxed{TK = S = \frac{|\bar{y}_A - \bar{y}_B|}{SE}}$$

$$SE = \sqrt{\frac{2\frac{S_r}{df_r}}{n}} \qquad SE = \sqrt{\frac{S_r}{df_r}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$n_A = n_B = n$ $n_A \neq n_B$

pokud $\boldsymbol{S} > \boldsymbol{S_\alpha}$ potom je rozdíl středních hodnot μ_A a μ_B statisticky významný

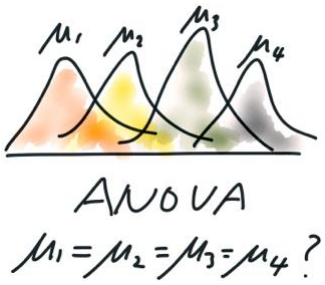
$$S_\alpha = \sqrt{(p-1) \cdot F_{\alpha;p-1;N-p}}$$

výhoda: nepotřebujeme speciální tabulky na kritické hodnoty

Konec

ANALYTICKÁ CHEMOMETRIE

Analýza rozptylu - pokračování



Patrik Kania
Ústav analytické chemie, VŠCHT Praha

11. 11. 2025

Dvoufaktorová analýza rozptylu - bez opakování (angl. *two-way ANOVA*)

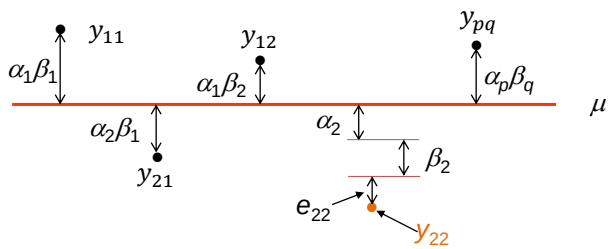
model: $y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$

$i = 1,\dots, p$ počet úrovní faktoru A

$j = 1,\dots, q$ počet úrovní faktoru B

$N = p \cdot q$ počet pokusů y_{ij}

H_0: $\alpha_1 = \alpha_2 \dots = \alpha_p = 0$ a
H_0: $\beta_1 = \beta_2 \dots = \beta_q = 0$



$$S_C = S_A + S_B + S_r$$

<u>test:</u> porovnání proměnlivosti S_A a S_B (meziúrovňové proměnlivosti) dané úrovněmi α_i a β_j dvou faktorů A a B a reziduální proměnlivosti S_r (vnitroúrovňové proměnlivosti)

Celkovou variabilitu lze rozložit:

$$\sum_i^p \sum_j^q (y_{ij} - \mu)^2 = q \cdot \sum_i^p (\bar{y}_i - \mu)^2 + p \cdot \sum_j^q (\bar{y}_j - \mu)^2 + \sum_i^p \sum_j^q (y_{ij} - \bar{y}_i - \bar{y}_j + \mu)^2$$

$$S_C = \sum_i^p \sum_j^q (y_{ij} - \mu)^2 \qquad df_C = p \cdot q - 1$$

$$S_A = q \cdot \sum_i^p (\bar{y}_i - \mu)^2 \qquad df_A = p - 1$$

$$S_B = p \cdot \sum_j^q (\bar{y}_j - \mu)^2 \qquad df_B = q - 1$$

$$S_r = \sum_i^p \sum_j^q (y_{ij} - \bar{y}_i - \bar{y}_j + \mu)^2 \qquad df_r = (p-1)(q-1)$$

Test významnosti

H_0: $\sigma_A^2 \leq \sigma_r^2$ H_0: $\sigma_B^2 \leq \sigma_r^2$

jestliže H_0 platí, není vliv faktoru A nebo faktoru B

jednostranný F-test:

odhad rozptylu způsobeného faktorem A

$$TK = F = \frac{\frac{S_A}{df_A}}{\frac{S_r}{df_r}}$$

odhad rozptylu způsobeného faktorem B

$$TK = F = \frac{\frac{S_B}{df_B}}{\frac{S_r}{df_r}}$$

odhad rozptylu měřené veličiny daný náhodou

pokud $TK > F_{(\alpha,\, dfA,\, dfr)}$ pokud $TK > F_{(\alpha,\, dfB,\, dfr)}$

hypotézu H_0 zamítáme hypotézu H_0 zamítáme

Dvoufaktoriální experiment (plán)

2 faktory na ***p*** a ***q*** úrovních

výsledky všech pokusů tvoří experimentální matici **Y**

úrovně faktoru A/B	B_1	B_2	…..	B_q	skupinové průměry	počet
A_1	y_{11}	y_{12}	…..	y_{1q}	$\bar{y}_{1\cdot}$	q
A_2	y_{21}	y_{22}	…..	y_{2q}	$\bar{y}_{2\cdot}$	q
…..	…..	…..	y_{ij}	y_{iq}	$\bar{y}_{i\cdot}$	q
A_p	y_{p1}	y_{p2}	…..	y_{pq}	$\bar{y}_{p\cdot}$	q
skupinové průměry	$\bar{y}_{\cdot 1}$	$\bar{y}_{\cdot 2}$	$\bar{y}_{\cdot j}$	$\bar{y}_{\cdot q}$	μ	
počet	p	p	p	p		$p \cdot q$

Příklad: Aby byl zjištěn vliv výrobce pivovarských kvasnic na stanovení obsahu sušiny kvasnic, byl proveden pokus, při němž bylo procento sušiny stanoveno na vzorcích od deseti různých výrobců třemi různými analytiky. Získané hodnoty jsou uvedeny v tabulce. Každý analytik provedl na každém vzorku jedno stanovení.

	vzorek (výrobce)									
analytik	1	2	3	4	5	6	7	8	9	10
1	10,1	4,7	3,1	3,0	7,8	8,2	7,8	6,0	4,9	3,4
2	10,0	4,9	3,1	3,2	7,8	8,2	7,7	6,2	5,1	3,4
3	10,2	4,8	3,0	3,1	7,8	8,4	7,9	6,1	5,0	3,3

Zvětšila se proměnlivost dat tím, že stanovení prováděli různí analytici?

Je významný efekt faktoru „výrobce“?

	vzorek (výrobce)										
analytik	1	2	3	4	5	6	7	8	9	10	skupinové průměry
1	10,1	4,7	3,1	3,0	7,8	8,2	7,8	6,0	4,9	3,4	5,900
2	10,0	4,9	3,1	3,2	7,8	8,2	7,7	6,2	5,1	3,4	5,960
3	10,2	4,8	3,0	3,1	7,8	8,4	7,9	6,1	5,0	3,3	5,960
skupinové průměry	10,100	4,800	3,067	3,100	7,800	8,267	7,800	6,100	5,000	3,367	$\mu = 5{,}940$

$$S_A = 10 \cdot \sum_{i=1}^{3} (\bar{y}_i - \mu)^2 = 10 \cdot [(5{,}900 - 5{,}940)^2 + (5{,}960 - 5{,}940)^2 + (5{,}960 - 5{,}940)^2] = 0{,}024$$

$$S_B = 3 \cdot \sum_{j=1}^{10} (\bar{y}_j - \mu)^2 = 3 \cdot [(10{,}100 - 5{,}940)^2 + \quad \dots \quad + (3{,}367 - 5{,}940)^2] = 164{,}372$$

$$S_r = \sum_{i=1}^{3} \sum_{j=1}^{10} (y_{ij} - \bar{y}_i - \bar{y}_j + \mu)^2 = (10{,}1 - 5{,}900 - 10{,}100 + 5{,}940)^2 + \quad \dots$$

$$S_r = S_C - S_A - S_B = \sum_{i=1}^{3} \sum_{j=1}^{10} (y_{ij} - \mu)^2 - S_A - S_B = (10{,}1 - 5{,}940)^2 + \cdots + (3{,}3 - 5{,}940)^2 - S_A - S_B$$

$$= 164{,}532 - 0{,}024 - 164{,}372 = 0{,}136$$

Test významnosti

H_0: $\sigma_A^2 \leq \sigma_r^2$ H_0: $\sigma_B^2 \leq \sigma_r^2$

jestliže H_0 platí, není vliv faktoru A nebo faktoru B

$\alpha_1 = \alpha_2 \ldots = \alpha_p = 0$ $\beta_1 = \beta_2 \ldots = \beta_q = 0$

jednostranný F-test:

$$TK = F = \frac{\frac{0{,}024}{3-1}}{\frac{0{,}136}{(3-1)(10-1)}} = 1{,}59$$

$$TK = F = \frac{\frac{164{,}372}{10-1}}{\frac{0{,}136}{(3-1)(10-1)}} = 2417{,}23$$

$F_{krit(0,05;2;18)} = 3{,}55$

$1{,}59 < 3{,}55$

$\alpha_1 = \alpha_2 \ldots = \alpha_p = 0$ platí

$F_{krit(0,05;9;18)} = 2{,}46$

$2417{,}23 > 2{,}46$

$\beta_1 = \beta_2 \ldots = \beta_q = 0$ neplatí

provedení analýz třemi různými analytiky významně nezvětšilo proměnlivost dat

proměnlivost mezi vzorky je významná faktor „výrobce“ je významný

Tukeyho metoda pro 2F ANOVu bez opakování

- lze použít 1F variantu, ale je třeba upravit příslušné počty měření ***n*** či stupně volnosti ***p*** a ***N-p*** v jednotlivých vzorcích

pro faktor A $SE = \sqrt{\frac{\frac{S_r}{df_r}}{q}}$

pro faktor B $SE = \sqrt{\frac{\frac{S_r}{df_r}}{p}}$

kritická hodnota $q_{\alpha, p, N-p}$ pro faktor A je $q_{\alpha, p, (p-1)(q-1)}$ a pro faktor B je $q_{\alpha, q, (p-1)(q-1)}$

Dvoufaktorová analýza rozptylu - s opakováním

model: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

γ_{ij} - efekty interakce řádkových a sloupcových faktorů
- systematická odchylka od aditivního modelu

$i = 1,\dots, p$	počet úrovní faktoru A
$j = 1,\dots, q$	počet úrovní faktoru B
$k = 1,\dots, n$	počet opakování postupu
$N = p \cdot q \cdot n$	počet pokusů y_{ijk}

H_0: $\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$

H_0: $\beta_1 = \beta_2 = \dots = \beta_q = 0$

H_0: $\gamma_{11} = \gamma_{12} = \dots = \gamma_{pq} = 0$

Celkovou proměnlivost (součet čtverců) lze rozložit na 4 složky:

$$S_C = S_A + S_B + S_{AB} + S_r$$

test: porovnání reziduální proměnlivosti S_r a proměnlivostí S_A, S_B a S_{AB} dané úrovněmi α_i a β_j a efekty interakcí dvou faktorů A a B

Celkovou variabilitu S_C lze rozložit na $S_A + S_B + S_{AB} + S_r$

$$S_C = \sum_i^p \sum_j^q \sum_k^n (y_{ijk} - \mu)^2 \qquad df_C = p \cdot q \cdot n - 1$$

$$S_A = n \cdot q \cdot \sum_i^p (\bar{y}_i - \mu)^2 \qquad df_A = p - 1$$

$$S_B = n \cdot p \cdot \sum_j^q (\bar{y}_j - \mu)^2 \qquad df_B = q - 1$$

$$S_{AB} = n \cdot \sum_i^p \sum_j^q (\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \mu)^2 \qquad df_{AB} = (p-1)(q-1)$$

$$S_r = \sum_i^p \sum_j^q \sum_k^n (y_{ijk} - \bar{y}_{ij})^2 \qquad df_r = p \cdot q \cdot (n-1)$$

Test významnosti

H_0: $\sigma_A^2 \leq \sigma_r^2$ H_0: $\sigma_B^2 \leq \sigma_r^2$ H_0: $\sigma_{AB}^2 \leq \sigma_r^2$

jestliže H_0 platí

není vliv faktoru A faktoru B interakce AB

jednostranný F-test:

$$TK = F = \frac{\frac{S_A}{df_A}}{\frac{S_r}{df_r}}$$

$$TK = F = \frac{\frac{S_B}{df_B}}{\frac{S_r}{df_r}}$$

$$TK = F = \frac{\frac{S_{AB}}{df_{AB}}}{\frac{S_r}{df_r}}$$

pokud $TK > F_{(\alpha, dfA, dfr)}$ pokud $TK > F_{(\alpha, dfB, dfr)}$ pokud $TK > F_{(\alpha, dfAB, dfr)}$

hypotézu H_0 zamítáme hypotézu H_0 zamítáme hypotézu H_0 zamítáme

Příklad: Při stanovení železa spektrofotometrickou metodou byl studován vliv stability wolframové žárovky a vliv osob provádějících stanovení na výsledek stanovení. Testu se účastnili 3 studenti. Každý z nich provedl denně 3 stanovení stejného vzorku a na stejném spektrometru. Testování probíhalo 6 dnů. Získané hodnoty jsou uvedeny v tabulce.

	den					
student	1	2	3	4	5	6
1	1,404 1,346 1,618	1,447 1,569 1,820	1,914 1,477 1,894	1,887 1,485 1,392	1,772 1,728 1,545	1,665 1,539 1,680
2	1,306 1,628 1,410	1,241 1,185 1,546	1,506 1,575 1,649	1,673 1,372 1,114	1,227 1,397 1,531	1,404 1,452 1,627
3	1,932 1,674 1,399	1,426 1,768 1,859	1,382 1,690 1,361	1,721 1,528 1,371	1,320 1,489 1,336	1,663 1,612 1,359

Zjistěte, zda je významná interakce mezi faktorem „student“ a „den“.

Zjistěte, zda je proměnlivost mezi dny významná.

Existují významné rozdíly ve výsledcích studentů?

	den						
student	1	2	3	4	5	6	$\bar{y}_i$
1	1,404 1,346 1,618 ∅ **1,456**	1,447 1,569 1,820 ∅ **1,612**	1,914 1,477 1,894 ∅ **1,762**	1,887 1,485 1,392 ∅ **1,588**	1,772 1,728 1,545 ∅ **1,682**	1,665 1,539 1,680 ∅ **1,628**	1,621
2	1,306 1,628 1,410 ∅ **1,448**	1,241 1,185 1,546 ∅ **1,324**	1,506 1,575 1,649 ∅ **1,577**	1,673 1,372 1,114 ∅ **1,386**	1,227 1,397 1,531 ∅ **1,385**	1,404 1,452 1,627 ∅ **1,494**	1,436
3	1,932 1,674 1,399 ∅ **1,668**	1,426 1,768 1,859 ∅ **1,684**	1,382 1,690 1,361 ∅ **1,478**	1,721 1,528 1,371 ∅ **1,540**	1,320 1,489 1,336 ∅ **1,382**	1,663 1,612 1,359 ∅ **1,545**	1,549
$\bar{y}_j$	1,524	1,540	1,605	1,505	1,483	1,556	1,535

$$S_A = n \cdot q \cdot \sum_{i}^{p} (\bar{y}_i - \mu)^2 = 3 \cdot 6 \cdot \left[\begin{array}{c} (1{,}621 - 1{,}535)^2 + (1{,}436 - 1{,}535)^2 + \\ +(1{,}549 - 1{,}535)^2 \end{array} \right] = 0{,}315$$

	den						
student	1	2	3	4	5	6	$\bar{y}_i$
1	1,404 1,346 1,618 ∅ **1,456**	1,447 1,569 1,820 ∅ **1,612**	1,914 1,477 1,894 ∅ **1,762**	1,887 1,485 1,392 ∅ **1,588**	1,772 1,728 1,545 ∅ **1,682**	1,665 1,539 1,680 ∅ **1,628**	1,621
2	1,306 1,628 1,410 ∅ **1,448**	1,241 1,185 1,546 ∅ **1,324**	1,506 1,575 1,649 ∅ **1,577**	1,673 1,372 1,114 ∅ **1,386**	1,227 1,397 1,531 ∅ **1,385**	1,404 1,452 1,627 ∅ **1,494**	1,436
3	1,932 1,674 1,399 ∅ **1,668**	1,426 1,768 1,859 ∅ **1,684**	1,382 1,690 1,361 ∅ **1,478**	1,721 1,528 1,371 ∅ **1,540**	1,320 1,489 1,336 ∅ **1,382**	1,663 1,612 1,359 ∅ **1,545**	1,549
$\bar{y}_j$	1,524	1,540	1,605	1,505	1,483	1,556	1,535

$$S_B = n \cdot p \cdot \sum_{j}^{q} \left(\bar{y}_j - \mu\right)^2 = 3 \cdot 3[(1{,}524 - 1{,}535)^2 + \cdots + (1{,}556 - 1{,}535)^2] = 0{,}082$$

	den						
student	1	2	3	4	5	6	$\bar{y}_i$
1	1,404 1,346 1,618 ∅ **1,456**	1,447 1,569 1,820 ∅ **1,612**	1,914 1,477 1,894 ∅ **1,762**	1,887 1,485 1,392 ∅ **1,588**	1,772 1,728 1,545 ∅ **1,682**	1,665 1,539 1,680 ∅ **1,628**	1,621
2	1,306 1,628 1,410 ∅ **1,448**	1,241 1,185 1,546 ∅ **1,324**	1,506 1,575 1,649 ∅ **1,577**	1,673 1,372 1,114 ∅ **1,386**	1,227 1,397 1,531 ∅ **1,385**	1,404 1,452 1,627 ∅ **1,494**	1,436
3	1,932 1,674 1,399 ∅ **1,668**	1,426 1,768 1,859 ∅ **1,684**	1,382 1,690 1,361 ∅ **1,478**	1,721 1,528 1,371 ∅ **1,540**	1,320 1,489 1,336 ∅ **1,382**	1,663 1,612 1,359 ∅ **1,545**	1,549
$\bar{y}_j$	1,524	1,540	1,605	1,505	1,483	1,556	1,535

$$S_{AB} = n \cdot \sum_i^p \sum_j^q \left(\bar{y}_{ij} - \bar{y}_i - \bar{y}_j + \mu\right)^2 = 3 \cdot \begin{bmatrix} (1{,}456 - 1{,}621 - 1{,}524 + 1{,}535)^2 + \cdots + \\ +(1{,}545 - 1{,}549 - 1{,}556 + 1{,}535)^2 \end{bmatrix} =$$

$$= 0{,}3933$$

	den						
student	1	2	3	4	5	6	$\bar{y}_i$
1	1,404 1,346 1,618 ∅ **1,456**	1,447 1,569 1,820 ∅ **1,612**	1,914 1,477 1,894 ∅ **1,762**	1,887 1,485 1,392 ∅ **1,588**	1,772 1,728 1,545 ∅ **1,682**	1,665 1,539 1,680 ∅ **1,628**	1,621
2	1,306 1,628 1,410 ∅ **1,448**	1,241 1,185 1,546 ∅ **1,324**	1,506 1,575 1,649 ∅ **1,577**	1,673 1,372 1,114 ∅ **1,386**	1,227 1,397 1,531 ∅ **1,385**	1,404 1,452 1,627 ∅ **1,494**	1,436
3	1,932 1,674 1,399 ∅ **1,668**	1,426 1,768 1,859 ∅ **1,684**	1,382 1,690 1,361 ∅ **1,478**	1,721 1,528 1,371 ∅ **1,540**	1,320 1,489 1,336 ∅ **1,382**	1,663 1,612 1,359 ∅ **1,545**	1,549
$\bar{y}_j$	1,524	1,540	1,605	1,505	1,483	1,556	1,535

$$S_r = \sum_i^p \sum_j^q \sum_k^n (y_{ijk} - \bar{y}_{ij})^2 = (1{,}404 - 1{,}456)^2 + \cdots + (1{,}359 - 1{,}545)^2 = 1{,}231$$

Zjistěte, zda je významná interakce mezi faktorem „student“ a „den“.

$$TK = F = \frac{\frac{0{,}3933}{(3-1)(6-1)}}{\frac{1{,}231}{3 \cdot 6 \cdot (3-1)}} = 1{,}15$$

$F_{krit(0{,}05;10;36)} = 2{,}11$

$1{,}15 < 2{,}11$

H_0: $\gamma_{11} = \gamma_{12} = \ldots = \gamma_{pq} = 0$ platí

interakce je bezvýznamná

Zjistěte, zda je proměnlivost mezi dny nulová.

$$TK = F = \frac{\frac{0{,}082}{(6-1)}}{\frac{1{,}231}{3 \cdot 6 \cdot (3-1)}} = 0{,}48$$

$F_{krit(0{,}05;5;36)} = 2{,}48$

$0{,}48 < 2{,}48$

H_0: $\beta_1 = \beta_2 = \ldots = \beta_q = 0$ platí

faktor „den“ je bezvýznamný

Existují významné rozdíly ve výsledcích jednotlivých studentů?

$$TK = F = \frac{\frac{0{,}315}{(3-1)}}{\frac{1{,}231}{3 \cdot 6 \cdot (3-1)}} = 4{,}60$$

$F_{krit(0{,}05;2;36)} = 3{,}26$

$4{,}60 > 3{,}26$

$\alpha_1 = \alpha_2 = \ldots = \alpha_p = 0$ neplatí

existují statisticky významné rozdíly ve výsledcích studentů

Konec