V.

Analýza rozptylu

ANOVA

ANOVA

F = (vážený rozptyl mezi průměry skupin)/(rozptyl mezi jedinci v téže skupině)

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

$$\mu_{ij} = \mu + \alpha_i$$

- Pevné (fixed effect model) = model I

  normalita reziduí, y

- Náhodné (random effect model) = model II

  normalita reziduí, y a faktory

One-way ANOVA
(jednoduché třídění)

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

$\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$

H_1: alespoň jeden průměr se liší od ostatních

= omnibus test, sleduje pouze porušení globální hypotézy rovnosti průměrů.

Pozorování (č.)	Experimentální situace (výběr) 1	2	…	k	
1	x_{11}	x_{21}	…	x_{k1}	
2	x_{12}	x_{22}	…	x_{k2}	
⋮					
	x_{1n_1}	x_{2n_2}	…	x_{kn_k}	
n_i	n_1	n_2	…	n_k	$n = \sum_{i=1}^{k} n_i$
$x_{i.}$	$x_{1.}$	$x_{2.}$	…	$x_{k.}$	$x_{..} = \sum_{i=1}^{k} x_{i.}$
$\bar{x}_i$	$\bar{x}_1$	$\bar{x}_2$	…	$\bar{x}_k$	$\bar{x} = x_{..}/n$

One-way ANOVA (jednoduché třídění)

Neprůkazný výsledek testu znamená:

1) střední hodnoty se neliší ($\alpha_j = 0$)
2) důsledek chyby II. Druhu

ANOVA pro k=2

odpovídá oboustranné variantě t-testu

One-way ANOVA

$S_T = S_e + S_A$

S_T = celkový součet čtverců
S_e = součet čtverců odchylek uvnitř výběrů
S_A = součet čtverců rozdílů mezi výběry

Zdroj variability	I Součet čtverců	II Stupeň volnosti	III Průměrný součet čtverců	Kritérium F
(1) Experimentální zákroky	$\sum_{i=1}^{k} \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n}$	$k - 1$	III = I/II	Ze sloupce III: (1)/(2)
(2) Reziduální	(2) = (3) — (1)	$n - k$	III = I/II	—
(3) Celkový	$\sum_{i=1}^{k} \sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{..}^2}{n}$	$n - 1$	—	—

One-way ANOVA

Velikost účinku

$$\eta^2 = \frac{S_A}{S_T}$$

Poměr vysvětlené variability k celkové variabilitě

$$\omega^2 = \frac{S_A - (m-1)MS_e}{S_T + MS_e}$$

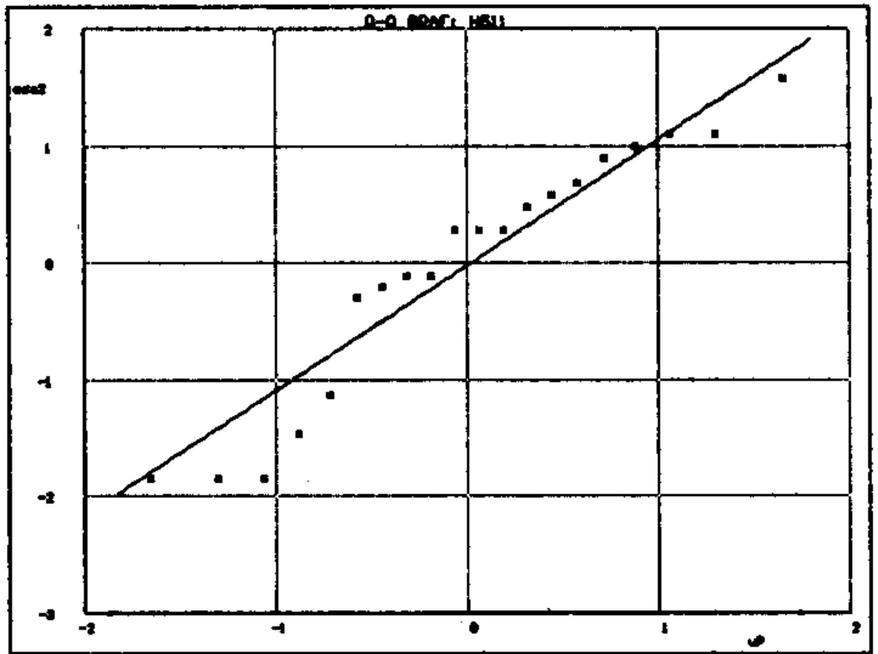
$$MS_e = \frac{S_e}{n-m}$$

Základní předpoklady

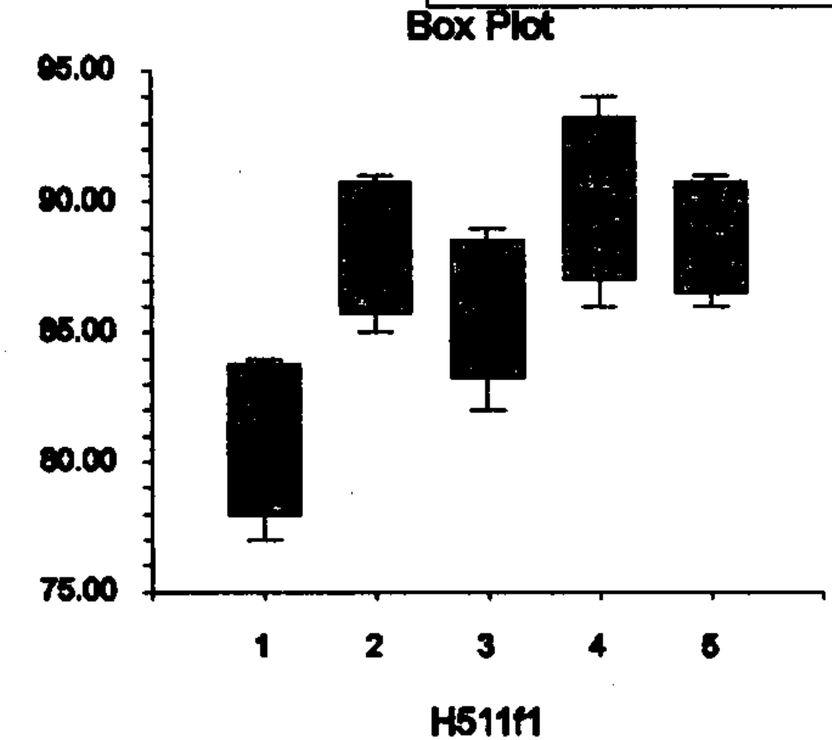
- **<u>Nezávislost měření uvnitř i mezi skupinami.</u>**
- Měření mají normální rozdělení s průměrem $\mu_{i.}$
- Ve všech skupinách mají měření stejný rozptyl kolem průměru.
- Aditivita efektů hladin jednotlivých faktorů (vlivy se sčítají, odchylky od součtu = interakce)

(pomocí testů, graficky)

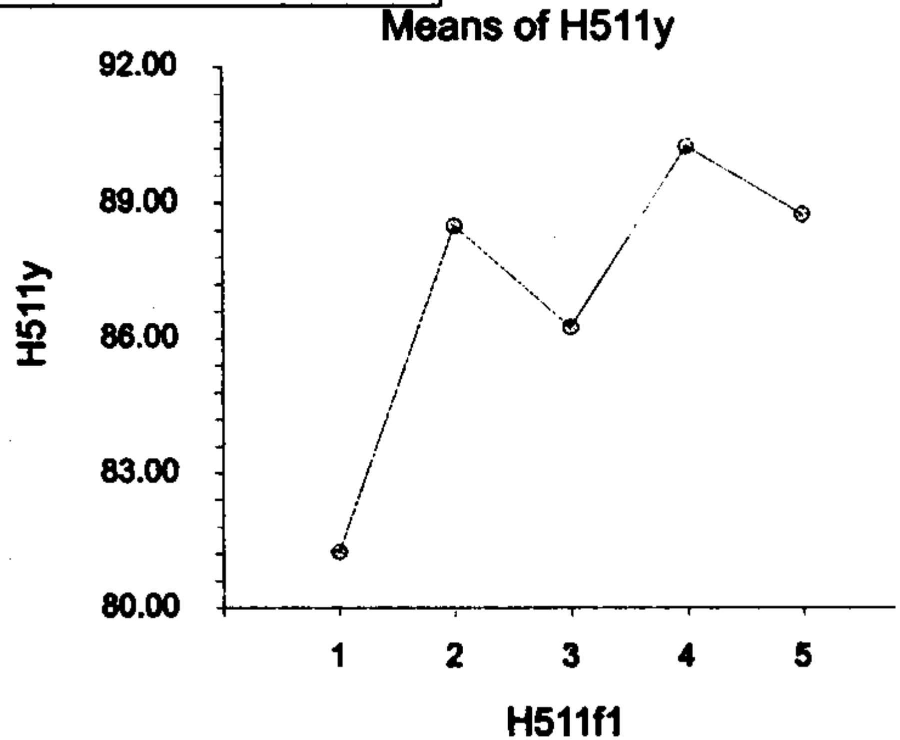
Grafy



2
1
0
-1
-2
-3
-2
-1
0
1
2



Box Plot
95.00
90.00
85.00
80.00
75.00
1
2
3
4
5
H511f1



Means of H511y
92.00
89.00
86.00
83.00
80.00
H511y
1
2
3
4
5
H511f1

Efekty v analýze rozptylu

Aditivní účinek:

Pozorování = vliv faktoru A + celkový průměr + náhod. variabilita

(formální podobnost s lineární regresí!!!)

Multiplikativní účinek:

Pozorování = celkový průměr x vliv faktoru A x náhod. variabilita

Log(pozorování) = log(celkový průměr) + log(vliv faktoru A) + log(náhod. variabilita)

(Tukeyův test neaditivity)

Analýza reziduálních hodnot

- Výpočet reziduí

$$x_{ij} - \overline{x}_j$$

- Grafické znázornění reziduí a jejich absolutních hodnot proti hodnotám faktorů a hodnotám závisle proměnné, zjišťování změn, trendů a konfigurací bodů.

- Ověření normality reziduí: graficky či testem

Analýza reziduálních hodnot

- Transformace dat:
  - Logaritmická (rozptyl se zvětšuje úměrně s průměrem).

    $$X' = \log(X + 1)$$

  - Arcsinová (pro relativní četnosti)

    $$p' = \arcsin \sqrt{p}$$

  - Druhá odmocnina (pro četnosti)

    $$X' = \sqrt{X + 0{,}5}$$

  - Box – Coxova transformace

- Přítomnost odlehlých hodnot = nutno použít neparametrický postup (Kruskal – Wallisův test)

Bartlettův test homogenity rozptylů

Testuje shodu několika rozptylů, měl by předcházet analýzu rozptylu.

$$B = \frac{2{,}3026}{C}\left\{(n = k)\log\frac{\sum_{i=1}^{k}(n_i - 1)s_i^2}{n-k} - \sum_{i=1}^{k}(n_i - 1)\log s_i^2\right\}$$

$$C = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right)$$

Hrubý odhad shody rozptylů

$$\frac{\max s_i}{\min s_i} \le 3$$

Simultánní porovnávání

Plánované srovnání

t-test (modifikace hladiny významnosti podle Bonferroniho): $\alpha' = 0{,}05/k$ k = počet porovnávání

nepříliš vhodný = rostou šířky intervalu spolehlivosti a pravděpodobnost chyby II. druhu!! Aby bylo možno spočítat korekci, už musíme předem vědět které dvojice chceme srovnávat.

Post hoc testy

Fisherův LSD test

Scheffeho test

Simultánní porovnávání

Tukey(ho) test

SNK (Student-Newnam-Keuls) test (modifikace Tukeyho testu, silnější x větší pravděpodobnost chyby I. Druhu)

Duncanův test (vyšší pravděpodobnost chyby I. Druhu, ta se vztahuje na konkrétní pozorování)

Dunnettův test (více pokusů vs. 1 kontrola, podobný Tukeyho testu)

Simultánní porovnávání

POZOR!!!!

Simultánní porovnávání lze použít pouze pro ***model I*** (pevné efekty) !!!!!!!

Pro ***model II*** (náhodné efekty) se někdy odhadují podíly vlivu na varianci (poměr variability uvnitř tříd a mezi třídami)

Kruskal-Wallisův test

= neparametrická varianta jednoduché analýzy rozptylu, kritérium je založeno na rozptylu standardizovaných pořadí.

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1) \qquad \sum_{i=1}^{k} R_i = n(n+1)/2$$

$$H_{kor} = \frac{H}{1 - \frac{\sum_j (t_j^3 - t_j)}{n^3 - n}}$$

Korigovaná hodnota (není-li H signifikantní)

t_j označuje kolikrát se ve smíchaném výběru opakuje j-té pozorování

Simultánní srovnávání: modifikace testu podle Bonferroniho

Jednostranné testování: test Jonckheere - Terpstra

Two-way ANOVA
(dvojné třídění)

H_0:

- $\alpha 1 = \alpha 2 = ... = \alpha i = 0$
- $\beta 1 = \beta 2 = ... = \beta\ i = 0$
- $\alpha 1\ \beta 1 = \alpha 2\ \beta 2 = ... = \alpha i\ \beta\ i = 0$

model I (pevné efekty)
model II (náhodné efekty)
model III (smíšené efekty)

Two-way ANOVA (dvojné třídění)

Blok	Experimentální zákrok							
(i)	1	2	…	k	$x_{i.}$	$\bar{x}_{i.}^2$	$x_{i.}^2/k$	$\sum_j x_{ij}^2$
1	x_{11}	x_{12}	…	x_{1k}	$x_{1.}$	$\bar{x}_{1.}$	$x_{1.}^2/k$	$\sum_j x_{1j}^2$
2	x_{21}	x_{22}	…	x_{2k}	$x_{2.}$	$\bar{x}_{2.}$	$x_{2.}^2/k$	$\sum_j x_{2j}^2$
⋮								⋮
n	x_{n1}	x_{n2}	…	x_{nk}	$x_{n.}$	$\bar{x}_{n.}$	$x_{i.}^2/k$	$\sum_j x_{nj}^2$
$x_{.j}$	$x_{.1}$	$x_{.2}$	…	$x_{.k}$	$x_{..}$		$\frac{1}{k}\sum_{i=1}^{n} x_{i.}^2$	
$\bar{x}_{.j}$	$\bar{x}_{.1}$	$\bar{x}_{.2}$	…	$\bar{x}_{.k}$				
$\frac{x_{.j}^2}{n}$	$\frac{x_{.1}^2}{n}$	$\frac{x_{.2}^2}{n}$	…	$\frac{x_{.k}^2}{n}$	$\frac{1}{n}\sum_{j=1}^{k} x_{.j}^2$			
$\sum_i x_{ij}^2$	$\sum_i x_{i1}^2$	$\sum_i x_{i2}^2$	…	$\sum_i x_{ik}^2$				$\sum_i \sum_j x_{ij}^{2\,2}$

Blok = faktor s náhodným efektem

Two-way ANOVA bez opakování

Jedna hodnota pro každou kombinaci faktorů = není žádná variabilita, tudíž nelze testovat interakcí.

K odhadu celkové variance lze použít pouze odchylky od aditivity.

Friedmanův test

= neparametrická varianta dvoufaktorové analýzy rozptylu, určuje se pořadí hodnot v každém bloku, shodným hodnotám přiřazujeme průměrné pořadí skupiny.

$$H = \frac{12}{nk(k+1)} \sum_{i=1}^{k} R_i^2 - 3n(k+1)$$

$$\sum_{i=1}^{k} R_i = nk(k+1)/2$$

Bloky	Vyšetření			
	1	2	...	k
1	R_{11}	R_{12}	...	R_{1k}
2	R_{21}	R_{22}	...	R_{2k}
⋮				
n	R_{n1}	R_{n2}	...	R_{nk}
Σ	R_1	R_2	...	R_k

Pro malý rozsah výběru – speciální tabulky

Two-way ANOVA s opakováním

- Vyvážený model

  Stejný počet opakování pro každou kombinaci faktorů, nejjednodušší výpočet, největší síla testu pro daný počet pozorování.

- Nevyvážený model

Two-way ANOVA

$$S_T = S_e + S_A + S_B + S_I$$

S_T = celkový součet čtverců
S_e = součet čtverců odchylek uvnitř výběrů
$S_{A,\,B}$ = hlavní efekty faktorů
S_I = efekt interakce

Zdroj variability	I Součet čtverců	II Stupeň volnosti	III Průměrný součet čtverců	Testové kritérium
(1) Bloky	$\frac{1}{k}\sum_{i=1}^{n} x_{i.}^2 - \frac{x_{..}^2}{nk}$	$n-1$	III = I/II	Ze sloupce III: (1) : (3)
(2) Experimentální zákroky	$\frac{1}{n}\sum_{j=1}^{k} x_{.j}^2 - \frac{x_{..}^2}{nk}$	$k-1$	III = I/II	(2) : (3)
(3) Reziduální	*reziduum* = (4) — (1) — (2)	$(n-1)(k-1)$	III = I/II	—
(4) Celkový	$\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}^2 - \frac{x_{..}^2}{nk}$	$nk-1$	—	—

Interakce mezi faktory

Hlavní efekt = přímý efekt faktoru na závisle proměnnou

Interakční efekt = spojený efekt kombinace dvou a více faktorů na závisle proměnnou

Vliv faktorů je neaditivní:

= vliv náhodné variability (pokud interakci a priori zamítáme)

= vliv interakce

Two-way ANOVA

- Podmínky:
- shoda rozptylů
- shoda kovariancí v kovarianční matici různých úrovní faktorů

Podmínka sféricity kovarianční matice
(nárůst chyby I. druhu)

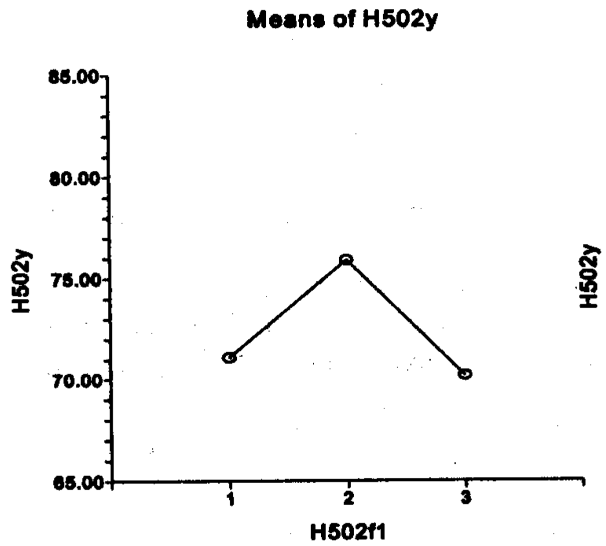
1) F test s Greenhouse-Geisserovou korekcí
2) MANOVA

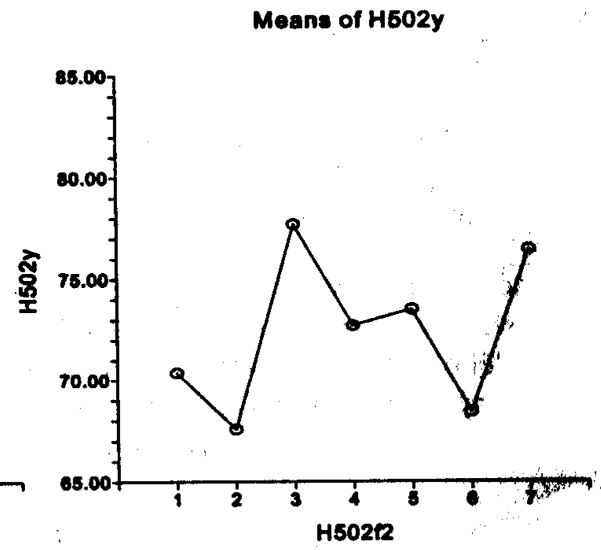
Two-way ANOVA

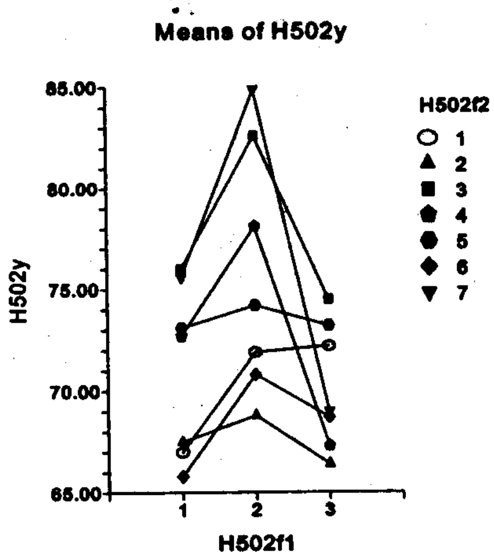
- Mnohonásobné porovnání

  Tukeyův test

  Dunnettův test







Obr. 5.3 Dvoufaktorová analýza rozptylu bez opakování v úloze H5.02:
Nahoře vlevo: Diagram průměrů pro rozličné úrovně faktoru *A* (tři různé kovy slitiny).
Nahoře vpravo: Diagram průměrů pro rozličné úrovně faktoru *B* (7 různých svarů).
Dole: Diagram průměrů pro rozličné úrovně obou faktorů *A* a *B*.

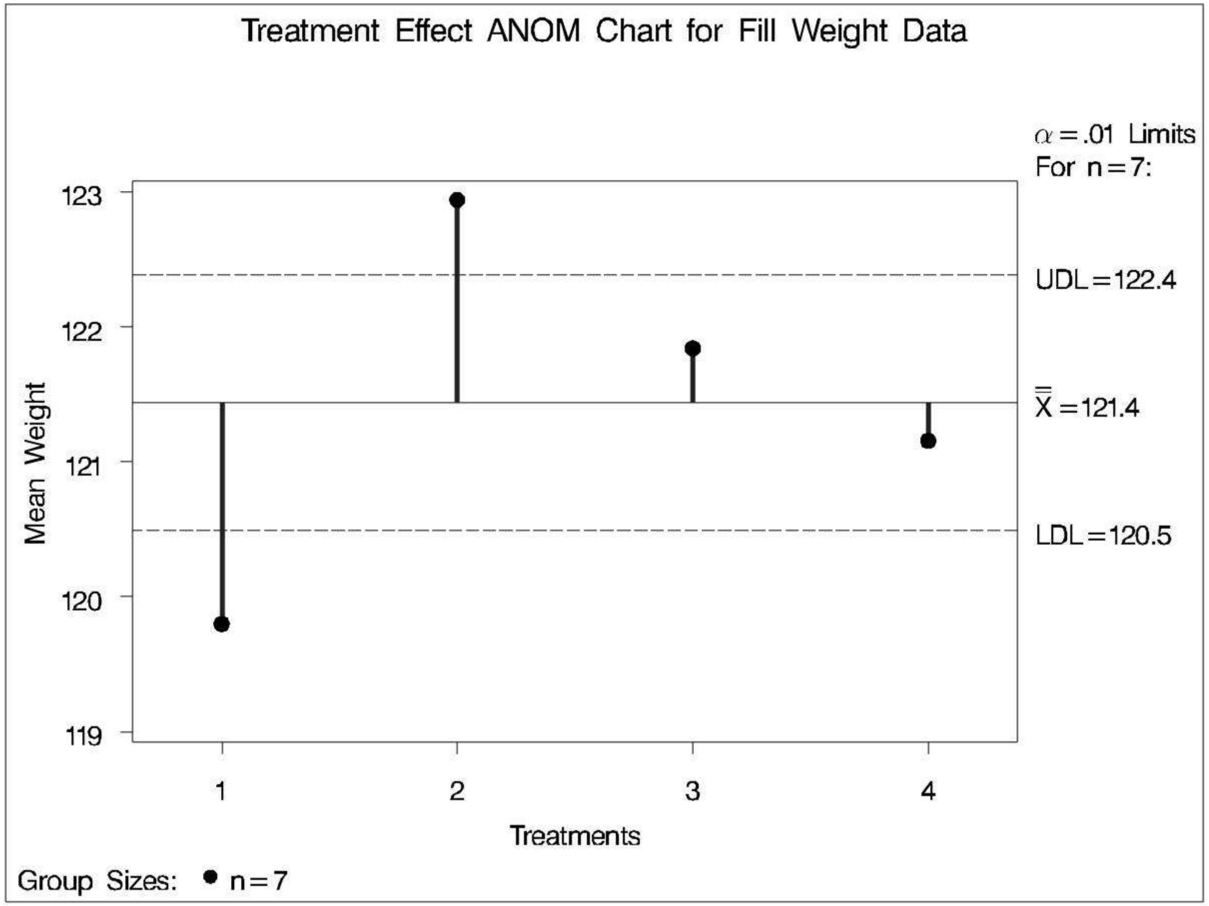
Analysis of means (ANOM)

Hybrid analýzy rozptylu a regulačních diagramů, překročení konfidenčních pásů indikuje statisticky významný vliv daného faktoru. Lze testovat:

Shodu průměrů

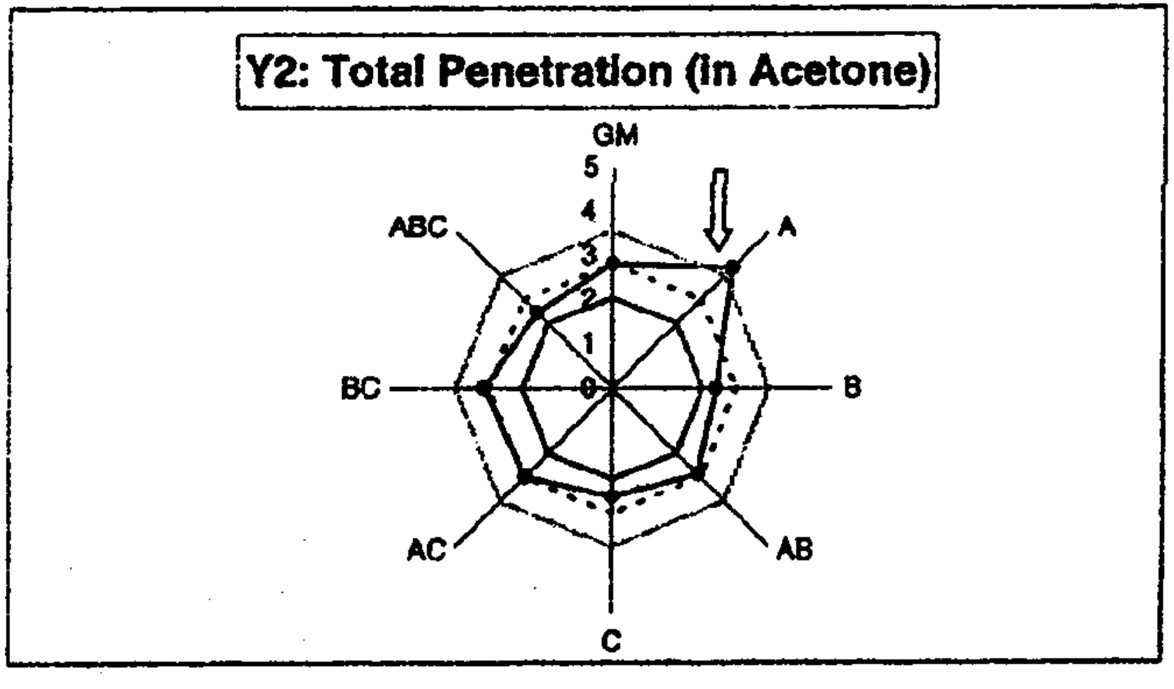
Shodu rozptylů

Shodu korelačních koeficientů



Analysis of means (ANOM)

- HANOM – pro heteroskedastická data
- ANOMR – pro pořadové hodnoty
- Compass plot



Analýza kovariance (ANCOVA)

- Na závisle proměnnou může mít vliv i rušivá proměnná (na intervalové škále), ANCOVA eliminuje její vliv.

- Kombinuje ANOVU s regresní analýzou