



IX. VIACVÝBEROVÉ TESTY

Anotácia

Korelačná analýza je využívaná na vyhodnotenie miery vzťahu dvoch spojitých premenných. Obdobne ako iné štatistické metódy, i korelácie môžu byť parametrické alebo neparametrické

Regresná analýza vytvára model vzťahu dvoch alebo viacerých premenných, teda akým spôsobom jedna premenná (vysvetľovaná) závisí na iných premenných (prediktoroch). Regresná analýza je obdobne ako ANOVA nástrojom na vysvetlenie variability hodnotenej premennej



Korelácia

Na meranie tesnosti lineárneho vzťahu 2 spojitéch premenných

$r = 0 \rightarrow$ nekorelované

$r > 0 \rightarrow$ kladne korelované

$r < 0 \rightarrow$ záporne korelované

H_0 : premenné X, Y sú stochasticky nezávislé náhodné veličiny ($r = 0$)

H_A : premenné X, Y nie sú stochasticky nezávislé náhodné veličiny ($r \neq 0$)

Parametrický korelačný koeficient:

Pearsonov kor. koef. (dvojrozmerné normálne rozloženie)

Neparametrický korelačný koeficient:

Spearmanov (poradový) kor. koef.



Jednoduchá lineárna regrese

Cieľ regresnej analýzy: popísať závislosť hodnôt premennej Y na hodnotách premennej X

Ak nemáme dostatok informácií k teoretickému súboru, snažíme sa odhadnúť typ funkcie pomocou dvojrozmerného diagramu



Anotácia

Analýza rozptylu je základným nástrojom na analýzu rozdielov medzi priemermi v niekoľkých skupinách pacientov.

Základná myšlienka, na ktorej je ANOVA založená, je rozdelenie celkovej variability v dátach (nepoznáme, daná len náhodným rozložením) na časť systematickú (spätú s kategóriami pacientov, vysvetlená variabilita) a časť náhodnú. Pokiaľ systematická, teda nenáhodná a vysvetliteľná časť variability prevažuje, považujeme daný kategoriálny faktor za významný na vysvetlenie variability dát.

Analýza rozptylu vyhodnocuje len celkový vplyv faktoru na variabilitu, v prípade analýzy jednotlivých kategórií je treba využiť tzv. post-hoc testy



ANOVA – analýza rozptylu

Zobecnenie dvojvýberového t-testu

ANOVA je základným nástrojom pre analýzu rozdielov medzi priemermi v niekoľkých skupinách

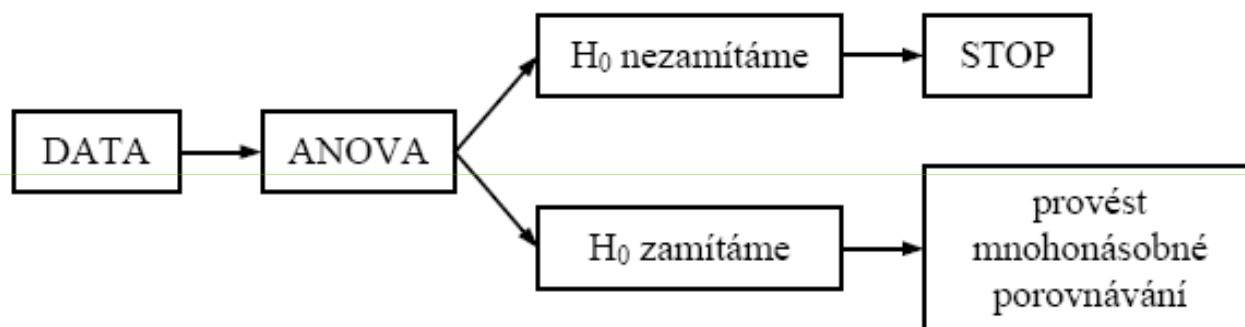
H_0 : všetky stredné hodnoty sú rovnaké

H_A : aspoň jedna dvojica stredných hodnôt sa líši

Predpoklady: normálne rozloženie v skupinách, nezávislosť skupín, zhoda rozptylov (Levenov či Bartlettov test)

Pokiaľ H_0 zamietame na hl. význ. $\alpha \rightarrow$ nás zaujíma, ktorá dvojica stredných hodnôt sa od seba líši

- metódy mnohonásobného testovania (tzv. post hoc testy), napr. Scheffého, Tukeyova metóda



Analýza rozptylu - ANOVA

Predpoklady analýzy rozptylu sú nevyhnutné na dosiahnutie sily testu

Symetrické rozloženie hodnôt a normalita

odchýlok od hodnoteného modelu ANOVA. Veľkú časť dát je možné adekvátne normalizovať použitím logaritmickej transformácie. Predpoklad lognormálnej transformácie môže pochopiteľne byť teoreticky vylúčený u veľa dátových súborov obsahujúcich diskkrétne parametre, kde je indikovaná vhodnosť iného typu transformácie. U asymetricky rozložených a u diskrétnych dát je nutné využiť neparametrickú alternatívu analýzy rozptylu.

Homogenita rozptylu je nutným predpokladom na zmyslupnosť vzájomných porovnaní pokusných variant. Overujeme ju napríklad pomocou Bartlettovho testu. Nehomogenita rozptylu je často vo vzťahu k nenormalite (asymetrii) dát a je možné ju odstrániť vhodnou normalizujúcou transformáciou.

Štatistická nezávislosť rezíduí vyhodnocovaného modelu ANOVA. Pokiaľ odhad a posúdenie korelačných vzťahov medzi pokusnými variantmi nie je priamo predmetom výskumu, je možné ich vplyv na vyhodnotenie odstrániť znáhodnením dát v rámci pokusných variant - teda zmenou poradia na náhodné. Rozsah vplyvu týchto autokorelačných vzťahov musí byť však primárne obmedzený správnosťou experimentálneho usporiadania.

Aditivita ako predpoklad týkajúci sa zložitejších experimentálnych usporiadaní. Exaktné otestovanie aditivity viacerých pokusných faktorov je procedúra pomerne náročná na experimentálny design vyvážený čo do počtu opakovaní. Je rovnako obtiažne testovať interakciu na neštandardných dátach, pretože prípadná transformácia môže zmeniť charakter odchýlok pôvodných dát od hodnoteného modelu ANOVA.

Analýza rozptylu - ANOVA

Obmedzenie aplikácie ANOVA je možné riešiť

Chýbajúce dáta. Vážnym problémom sú chýbajúce údaje o celej skupine kombinácií testovaných látok, napríklad u faktoriálnych pokusov, kedy je znemožnené hodnotenie experimentu ako celku.

Rôzne počty opakovaní Ide o typický jav pre experimentálne dátové súbory. Pri rôznych počtoch opakovaní v experimentálnych variantách sú testy ANOVA citlivejšie na nenormalitu dát. Pokiaľ sú počty opakovaní úplne odlišné (až na rádové rozdiely), je nutné použiť neparametrické techniky alebo analýzu rozptylu nevyvážených pokusov.

Odľahlé hodnoty. Ojedinelé odľahlé hodnoty musia byť pred parametrickou analýzou rozptylu vylúčené.

Nedostatok nezávislosti medzi reziduami modelu. Ide o závažný nedostatok, skresľujúci výsledok F-testu. Veľmi často je táto skutočnosť dôsledkom nesprávneho prevedenia alebo naplánovania experimentu.

Nehomogenita rozptylu. Veľmi častý nedostatok experimentálnych dát, často súvisiaci s nenormalitou rozloženia alebo s odľahlými hodnotami.

Nenormalita dát. I v tomto prípade je možné situáciu upraviť vylúčením odľahlých hodnôt alebo normalizujúcou transformáciou.

Neadditivita kombinovaného vplyvu viacerých pokusných zásahov. Túto situáciu je možné testovať buď špeciálnymi testami aditivity alebo priamo F testom kontrolujúcim významnosť vplyvu interakcie pokusných zásahov. Pri významnej interakcii je nutné preskúmať predovšetkým jej charakter vo vhodnom experimentálnom usporiadaní.



ANOVA – základný výpočet

Základným princípom ANOVY je porovnanie rozptylu pripadajúceho na:

- Rozdelenie dát do skupín (tzv. effect, variance between groups)
- Variabilitu objektov vo vnútri skupín (tzv. error, variance within groups), predpokladá sa, že ide o náhodnú variabilitu (=error)

1. Variabilita medzi skupinami

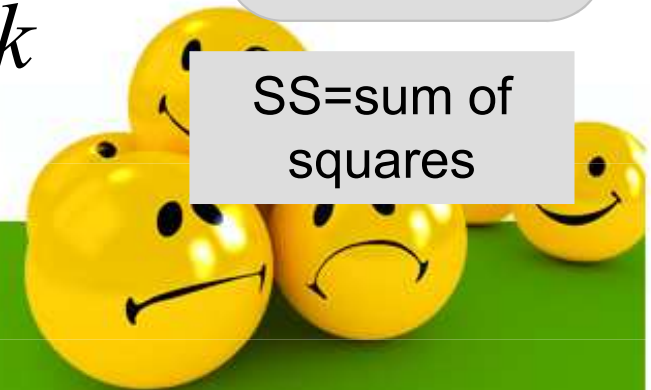
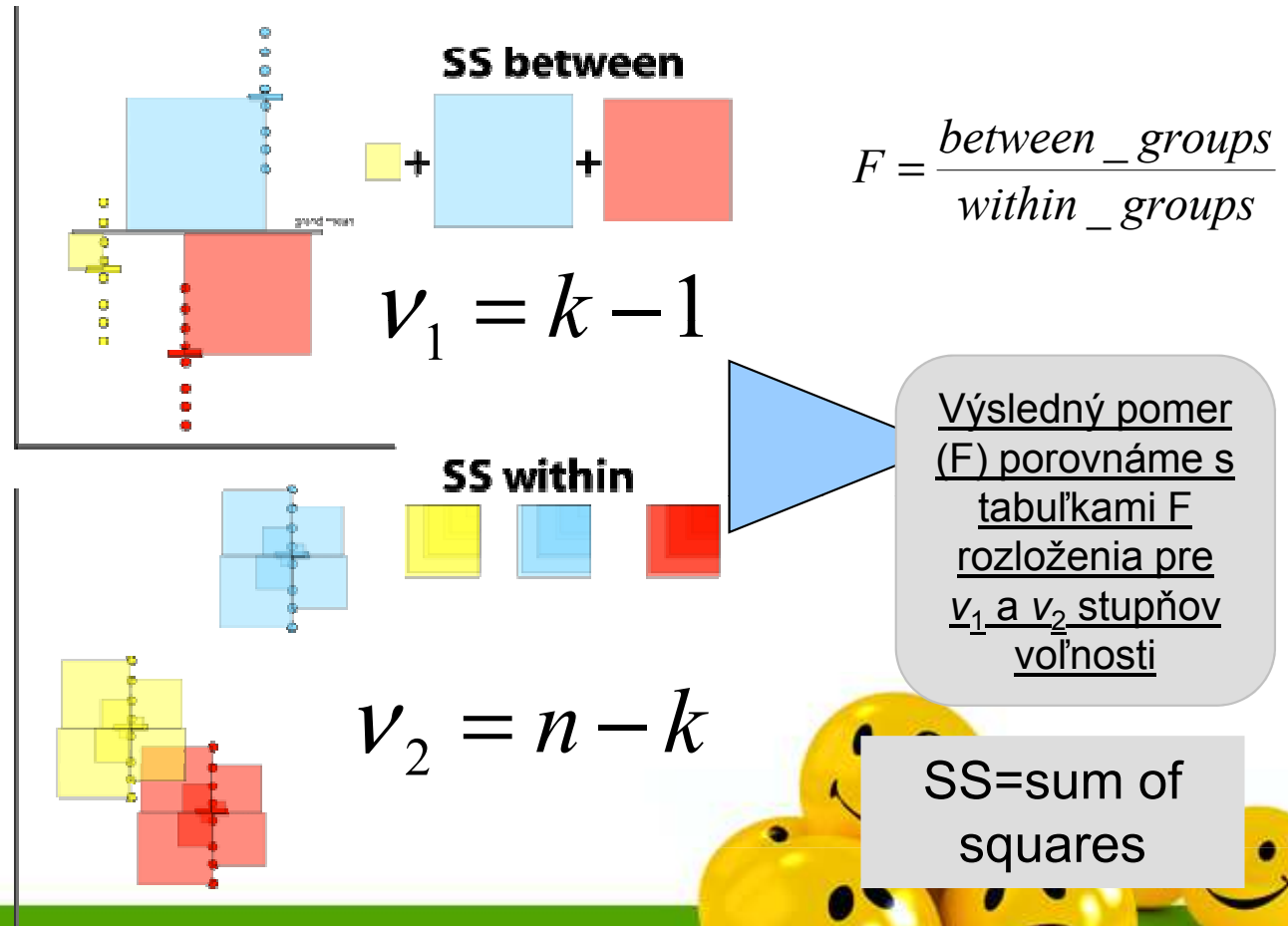
Rozptyl je počítaný pre celkový priemer (tzv. grand mean) a priemery v jednotlivých skupinách dát

Stupne voľnosti sú odvodené od počtu skupín (= počet skupín -1)

2. Variabilita vo vnútri skupín

Rozptyl je počítaný pre priemery jednotlivých skupín a objekty vo vnútri príslušných, celková variabilita je potom sčítaná pre všetky skupiny

Stupne voľnosti sú odvodené od počtu hodnôt (= počet hodnôt - počet skupín)



Zavedieme súčty štvorcov

S_T ... celkový súčet štvorcov (charakterizuje variabilitu jednotlivých pozorovaní okolo celkového priemeru),

- počet stupňov voľnosti $f_T = n - 1$,

S_A ... skupinový súčet štvorcov (charakterizuje variabilitu medzi jednotlivými náhodnými výbermi),

- počet stupňov voľnosti $f_A = r - 1$.

S_E ... reziduálny súčet štvorcov (charakterizuje variabilitu vo vnútri jednotlivých výberov),

- počet stupňov voľnosti $f_E = n - r$.

- Je možné dokázať, že $S_T = S_A + S_E$.



Modely analýzy rozptylu - základný výstup

*Základným výstupom analýzy rozptylu je
Tabuľka ANOVA - frakcionácia komponent rozptylu*

Výsledky výpočtov zapisujeme do tabuľky analýzy rozptylu jednoduchého triedenia.

Zdroj variability	Súčet štvorcov	Stupne voľnosti	podiel	F_A
skupiny	S_A	$f_A = r - 1$	S_A/f_A	$\frac{S_A/f_A}{S_E/f_E}$
reziduálny	S_E	$f_E = n - r$	S_E/f_E	-
celkový	S_T	$f_T = n - 1$	-	-

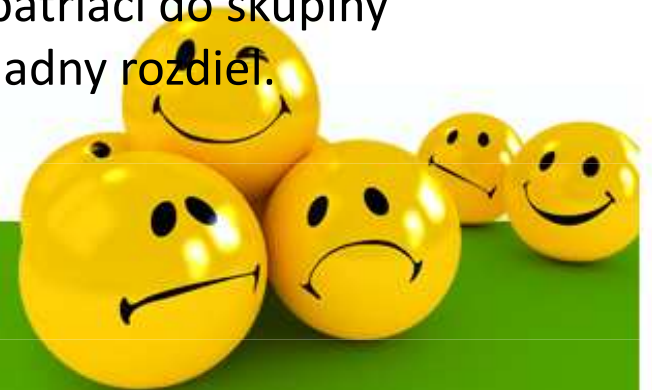
- Silu závislosti náhodnej veličiny X na faktore A môžeme merať pomocou pomeru determinácie: $p^2 = S_A/S_T$. Nadobúda hodnôt z intervalu $\langle 0,1 \rangle$.



Post-hoc metódy: Tukeyova vs. Scheffého metóda

Ak zamietneme na hladine významnosti α hypotézu o zhode stredných hodnôt, chceme zistiť, ktoré dvojice stredných hodnôt sa líšia na danej hladine významnosti α , tj. na hladine významnosti α testujeme $H_0: \mu_l = \mu_k$ proti $H_1: \mu_l \neq \mu_k$ pre všetky $l, k = 1, \dots, r, l \neq k$.

- a) Ak majú všetky výbery rovnaký rozsah p (hovoríme, že triedenie je vyvážené), použijeme Tukeyovu metódu.
- b) Ak nemajú všetky výbery rovnaký rozsah, použijeme Scheffého metódu
- Metódy mnohonásobného porovnávania majú obecnne menšiu silu než ANOVA.
- Môže nastať situácia, kedy pri zamietnutí H_0 nenájdeme metódami mnohonásobného porovnávania významný rozdiel u žiadnej dvojice stredných hodnôt. K tomu dochádza obzvlášť vtedy, keď p -hodnota pre ANOVU je len o málo nižšia než zvolená hladina významnosti. Potom slabší test patriaci do skupiny metód mnohonásobného porovnávania nemusia odhaliť žiadny rozdiel.



Příklad: Anova - One way

Dávka rostlinného stimulantu (0, 4, 8, 12 mg/l)

A = 4 ; n = 8

I. ANOVA

Bartlett's test: P = 0,9847

K-S test: P = 0,482 - 0,6525 pro jednotlivé kategorie

Source	D. f.	SS	MS	F
Between Groups	3	305,8	101,9	8,56
Within Groups	28	322,2	11,9	
Total (corr.)	31	638,0		

II. Multiple Range Test

NKS -test

Level	Average	Homogenous Groups
0	34,8	x
4	41,4	x
12	41,8	x
8	52,6	x



Příklad: Anova - One way

- I. Zásah: 4 klinická stadia virové choroby (napadá kr. buňky)
Sledovaná veličina: aktivita enzymu v těchto krevních buňkách

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

n = 3
 MODEL = ?

	I	II	III	IV
	22,8	16,4	11,2	14,2
	19,4	17,8	18,2	10,1
	12,5	19,1	15,8	12,8
Σ	65,7	53,3	45,2	37,1
průměr	21,9	17,8	15,1	12,4

II.

Source	D.f.	MS	F	P
Between groups	3	49,6	8,39	0,0075
Within groups	8	5,9		
Total (corr.)	11	-		

III. Komponenta rozptylu:

$$\sigma_A^2 \sim S_A^2 = \frac{MS_A - MS_e}{n} = \frac{49,6 - 5,9}{3} = 14,57$$

$$S_A^2 = 2,5 \cdot S_e^2$$

IV.

$$\rho_I \sim r_I = \frac{S_A^2}{S_A^2 + S_e^2} = 0,7142$$



Doporučený postup při provádění analýzy rozptylu:

- a) Ověření normality daných n náhodných výběrů (grafické metody - NP plot, Q-Q plot, histogram, testy hypotéz o normálním rozložení - Lilieforsova varianta Kolmogorovova – Smirnovova testu nebo Shapiroův – Wilkův test).

Doporučuje se kombinace obou způsobů. Závěry učiníme až na základě posouzení obou výsledků.

Obecně lze říci, že analýza rozptylu není příliš citlivá na porušení předpokladu normality, zvláště při větších rozsazích výběrů (nad 20), což je důsledek působení centrální limitní věty. Mírné porušení normality tedy není na závadu, při větším porušení použijeme např. Kruskalův – Wallisův test jako neparametrickou obdobu analýzy rozptylu jednoduchého třídění.

- b) Po ověření normality se testuje homogenitu rozptylů, tj. předpoklad, že všechny náhodné výběry pocházejí z normálních rozložení s tímž rozptylem. Graficky ověřujeme shodu rozptylů pomocí krabicových diagramů, kdy sledujeme, zda je šířka krabic stejná. Numericky testujeme homogenitu rozptylů pomocí Levenova testu.

Slabé porušení homogenity rozptylů nevádí, při větším se doporučuje mediánový test.

- c) Pokud jsou splněny předpoklady normality a homogenity rozptylů, můžeme přistoupit k testování shody středních hodnot. Předtím je samozřejmě vhodné vypočítat průměry a směrodatné odchylky či rozptyly v jednotlivých skupinách.

- d) Dojde-li na zvolené hladině významnosti k zamítnutí hypotézy o shodě středních hodnot, zajímá nás, které dvojice středních hodnot se od sebe liší. K řešení tohoto problému slouží post-hoc metody mnohonásobného porovnávání, např. Scheffého nebo Tukeyova metoda.





**VEĽA ŠŤASTIA NA
SKÚŠKE!**