

PSY117/454 Statistická analýza dat v psychologii – přednáška 10

ANALÝZA ROZPTYLU

srovnávání více průměrů

Omezení t-testu

- Umožňuje srovnání pouze 2 průměrů
 - Více skupin (j) \gg mnoho porovnání ($j(j-1)/2$)
- Neúnosný růst pravděpodobnosti chyby 1. typu
 - např. při $\alpha=0,05$ a 20 testech $p=0,68$ (1 nebo více chyb)
 - aplikace binomického rozložení
- Platí to pro jakýkoli statistický test (zejm. korelace)
- Je *zakázáno* provádět velké množství testů na jedněch datech (cca >10)
 - Zneužití se označuje jako rybaření v datech – fishing, capitalizing on chance
- Lze kompenzovat snížením požadované α (Bonferroniho korekce), avšak za cenu značného snížení síly testu ($1-\beta$).

Řešení = ANOVA

- Testuje na více skupinách jen jednu hypotézu:
 - Je někde mezi skupinovými průměry někde rozdíl?
 - Je mezi Pražáky, Brňáky a Ostraváky rozdíl v průměrné lakotě?
 - Je-li odpověď „ano“ ($p < \alpha$), pak se můžeme podívat na jednotlivé rozdíly detailněji (post-hoc testy)
 - Je-li odpověď „ne“ ($p > \alpha$), pak bychom neměli (rybaření)
-

1. terminologická vložka - ANOVA

- ANOVA = ANalysis Of Variance = analýza rozptylu
 - i přes svůj název jde o srovnávání průměrů
 - ANOVA zjišťuje vztah mezi **kategoriální nezávislou a intervalovou závislou**.
 - kategoriální nezávislá = faktor (factor, „-way“)
 - hodnoty kategoriální nez. = úrovně (level)
 - Zjištěný rozdíl = efekt, účinek (effect)
-

Princip ANOVY 1.

□ směrodatná odchylka = MS
= mean square

□ MS_{between} : s spočítaná ze skupinových průměrů, variabilita uvnitř skupiny je ignorována (též MS_A)

$$\square MS_{\text{between}} = SS_{\text{between}} / j - 1$$

□ MS_{within} : variabilita uvnitř skupin ($MS_{e, \text{error}}$)

$$\square MS_{\text{within}} = SS_{\text{within}} / n - j$$

	sk1	sk2	sk3	Celkem
ď1	2	4	6	
ď2	2	4	6	
ď3	2	4	6	
ď4	2	4	6	
ď5	2	4	6	
m	2	4	6	4
s	0	0	0	1,63
			sbw	3,65
			swi	0,00

	sk1	sk2	sk3	Celkem
ď1	0	6	2	
ď2	4	2	10	
ď3	0	6	2	
ď4	4	2	10	
ď5	2	4	6	
m	2	4	6	4
s	1,8	1,8	3,6	3,0
			sbw	3,65
			swi	2,53

	sk1	sk2	sk3	Celkem
ď1	1	4	2	
ď2	3	5	5	
ď3	5	1	3	
ď4	4	2	1	
ď5	2	3	4	
m	3	3	3	3
s	1,4	1,4	1,4	1,41
			sbw	0
			swi	1,41

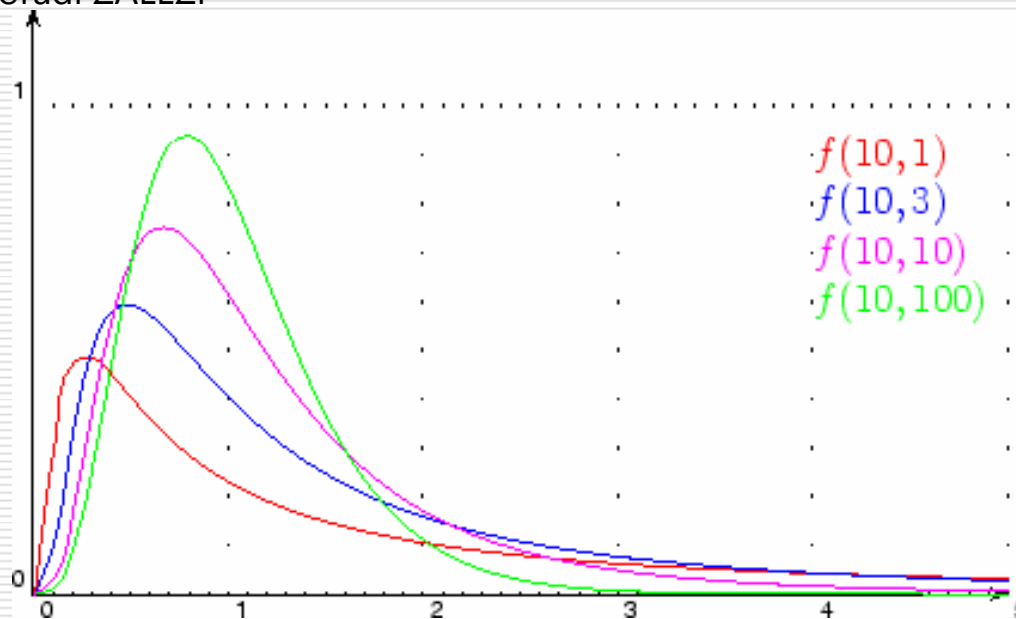
			F	1,443
			$0,95F(2, 12)$	3,885
			p	0,274

Princip ANOVY 2.

- Čím jsou si průměry podobnější, tím je rozptyl mezi skupinami nižší (MS_{between} se blíží 0)
 - Čím nižší je rozptyl uvnitř skupin (MS_{within} se blíží 0), tím průkaznější se průměry mezi skupinami zdají být.
 - Důležitý je poměr těchto dvou odhadů rozptylu: mezi skupinami/uvnitř skupin
 - Tento poměr se nazývá F -poměr
 - Čím vyšší tento poměr je, tím průkaznější jsou rozdíly mezi průměry (rozsah je 0 až ∞)
 - F -poměr má jako výběrová statistika F -rozložení
-

Fisherovo-Snedecorovo F-rozložení

- Podobně jako t -rozložení, je F -rozložení vlastně rodina mnoha rozložení mírně se lišící svým tvarem
- Tato rozložení se liší tentokrát dvěma parametry – stupni volnosti
 - $\nu_1 = \text{počet skupin} - 1$: stupně volnosti čitatele - MS_{between}
 - $\nu_2 = \text{počet lidí} - \text{počet skupin}$: stupně volnosti jmenovatele - MS_{within}
 - na pořadí ZÁLEŽÍ



<http://www.econtools.com/jevons/java/Graphics2D/FDist.html>

Velikost účinku (efektu)

- Podobně jako u regrese chceme vědět, jaká část rozptylu závislé je vysvětlená nezávislou
 - Ekvivalentem R^2 je u anovy η^2 (eta)
 - $\eta^2 = SS_{\text{Between}} / SS_{\text{Total}}$
 - Poněkud přesnější je ω^2
 - Velikost účinku je vždy třeba uvádět
-

Předpoklady ANOVY

- normální rozložení uvnitř skupin
- stejné rozptyly uvnitř skupin - homoscedasticita
- nezávislost všech pozorování

viz Hendl 343

Post-hoc testy

- ❑ Po prokázání „nějakých“ rozdílů mezi průměry obvykle chceme vědět, mezi kterými skupinami konkrétně rozdíly jsou.
 - ❑ K tomu jsou tzv. post-hoc testy
 - ❑ Je-li důležité udržet a pod kontrolou, pak je správnou volbou Scheffeho test
 - ❑ Pokud to není tak kritické, pak je volbou Student-Neuman-Keuls (S-N-K)
 - ❑ Extrémně „dajný“ je LSD a proto se nedoporučuje.
-

Další varianty a rozšíření ANOVy

- ❑ ANOVA pro opakovaná měření (jako párový t-test)
 - ❑ ANOVA s 2 a více faktory
 - ❑ MANOVA – s více závislými proměnnými
 - ❑ To vše v SPSS skryto pod GLM – general linear model
-