PSY117/454
Statistická analýza dat v psychologii
Přednáška 5

Vlastnosti a využití korelace

Pořadová korelace a nezávislost

Robustnost a resistence statistik

Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital.

Aaron Levenstein

Vlastnosti Pearsonova korelačního koeficientu

- Jde o **momentový** koeficient korelace, a tedy je nutná intervalová a vyšší úroveň měření
- Je vhodný pro popis normálně rozložených proměnných (nebo alespoň **stejně rozložených**)
- Vyjadřuje **sílu(těsnost) lineárního** vztahu, tj. jak moc připomíná tvar scatteru štíhlou elipsu, čáru

Co když tyto podmínky nejsou splněny?

AJ: Pearson's product-moment correlation

Pořadová korelace

- Řeší mnohá omezení Pearsonovy *r*
- Čím víc, tím víc/míň nahrazuje ideou shody **pořadí**

Vysoká pozitivní (negativní) korelace pak znamená:

Má-li jeden člověk v jedné proměnné vyšší hodnotu než druhý člověk (tj. nižší pořadí), pak by i v druhé proměnné měl mít ten první vyšší (nižší) hodnotu než druhý.

Kendallův koeficient pořadové korelace tau

známk a M	obvod hlavy	pořadí v M	pořadí v obv. h.	pořadí v M	pořadí v obv. h.	K+, D-
3	**48**	**3**	**3**	**1**	**5**	- - - -
2	**43**	**2**	**2**	**2**	**2**	+ + -
1	**50**	**1**	**5**	**3**	**3**	+ -
4	**49**	**4**	**4**	**4**	**4**	-
5	**40**	**5**	**1**	**5**	**1**	

$$\boldsymbol{\tau = (K\text{-}D) / [N\ (N\ \text{-}1)/2]} = (3\text{-}7)/(5.4/2) = \text{-}4/10 = \text{-}0{,}4$$

AJ: Kendall tau (-b,-c), ranks, rank ordering

Kendallův koeficient pořadové korelace tau

- = přeškálovaná pravděpodobnost, že dva náhodní lidé budou podle obou proměnných shodně(opačně) seřazeni
- <1;-1>
- zachycuje i monotonní nelineární vztah
- díky pořadovému základu není ovlivněno outliery
- kromě pořadové úrovně měření nepředpokládá nic

Modifikace τ_b a τ_c řeší problém shody pořadí (ties).
Podobné: (Goodmanova a Kruskallova) γ a Sommerovo *d*

Spearmanův koeficient pořadové korelace r_s

známka M	obvod hlavy	pořadí v M	pořadí v obv. h.
3	**48**	**3**	**3**
2	**43**	**2**	**2**
1	**50**	**1**	**5**
4	**49**	**4**	**4**
5	**40**	**5**	**1**

$\boldsymbol{r_s}$ **=** Pearsonova ***r*** spočítaná na transformovaných proměnných = -0,6

Spearmanovo r_s (ρ, ró, rho)

r_s – tak na půl cesty mezi r a τ

- Je pořadový a nepředpokládá striktně lineární vztah, ale zohledňuje **velikost odchylek** od ideálního pořadí
- Počítá se jako Pearsonova korelace, ale na pořadích
- Používá se obvykle jako rezistentnější varianta Pearsonovy r, která zachytí i monotónní nelineární vztahy.
  - Je-li $r_s > r$, je možné, že vztah není lineární
- Lze interpretovat r_s^2
- *Vychází obvykle numericky vyšší než tau, ovšem to by nikdy nemělo hrát roli ve vašem rozhodování. V obou případech jde o jiný typ vztahu.*

AJ: Spearman (rank-order correlation) rho,

Vztahy na nominální úrovni

- Korelační koeficienty založené na hodnotě χ^2 počítané nad kontingenční tabulkou.
  - χ^2 je sumou rozdílů mezi získanými četnostmi v kontingenční tabulce a četnostmi, jaké bychom očekávali, kdyby mezi proměnnými nebyl žádný vztah
- Kvůli neexistenci směru mají koeficienty rozsah od 0 (žádný vztah) do 1 (maximálně těsný vztah)
- Větší množství koeficientů se specializovaným užitím
  - Pearsonův kontingenční koeficient
  - Cramerovo *V*
  - r_{ϕ} – koeficient fí (phi)

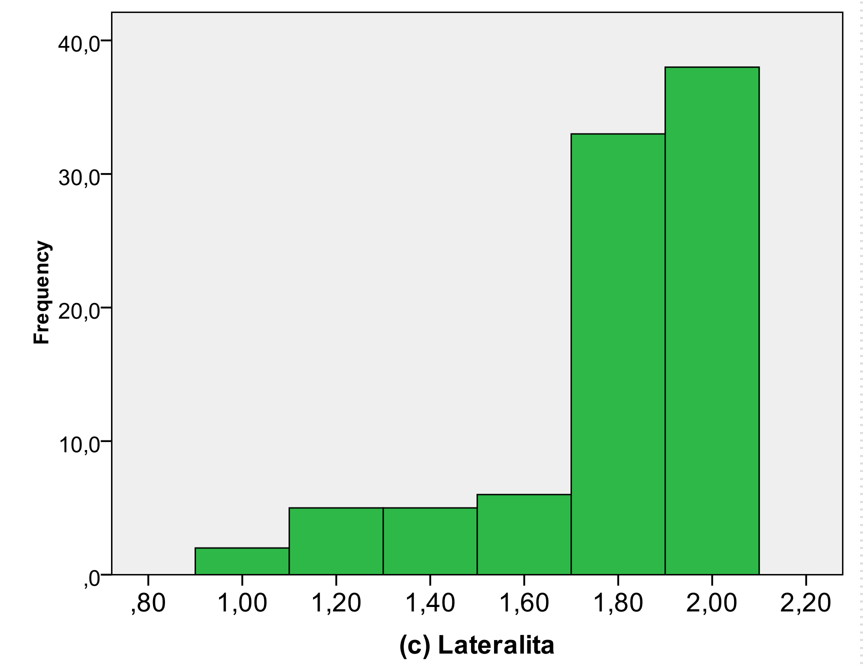
Těmto vztahům se budeme věnovat později.

Konstrukce psychologických škál

Lateralita = součet 5 položek

- psaní, házení, kopání, odraz, zapalování

	M	*SD*
psaní	1,82	0,39
házení	1,93	0,25
kopání	1,89	0,32
odraz	1,53	0,50
sirka	1,82	0,39
(c) Lateralita	**1,80**	**0,25**



Využití korelací v konstrukci psychologických testů

- Položky lze sčítat, pokud spolu korelují.
- Položky korelují, existuje-li společný důvod pro určitý způsob odpovídání na ně – měřená charakteristika.

Jak moc spolu musí korelovat?

$$r_{tt} = \frac{kr_M}{1+(k-1)r_M} \qquad r_{tt} = \frac{k}{k-1}\left(1-\frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_t^2}\right)$$

r_{tt} je vnitřní konzistence, r_M je průměrná korelace mezi položkami, k je počet položek

- při 10 položkách stačí průměrná korelace 0,2

Vnitřní konzistence – **Cronbachovo α** – horní mez reliability

- minimálně 0,7 pro výzkum, 0,9 pro diagnostiku

zpět k *lateralitě*

	psaní	házení	kopání	odraz	sirka
psaní	1,00	0,57	0,57	0,02	0,92
házení	0,57	1,00	0,61	-0,08	0,57
kopání	0,57	0,61	1,00	0,09	0,48
odraz	0,02	-0,08	0,09	1,00	0,02
sirka	0,92	0,57	0,48	0,02	1,00

$r_M = 0{,}379$

$r_{tt} = 0{,}69$ r_{tt}(bez odrazu)=0,87

Jaké statistiky už známe

Četnosti

Popisné statistiky jedné proměnné

- momentové: M, SD, s^2
- pořadové: *min*, *max*, *Md*, Q_1, Q_3, *IQR*, percentily
- kategorické: *Mo*

Ukazatele vztahu mezi dvěma proměnnými

- momentové: Pearsonova r, b
- pořadové: Kendallovo τ, Spearmanova r_s
- kategorické: r_ϕ, Cramerovo V

S jakými předpoklady je spojeno použití těchto statistik?

Co se stane, když nejsou tyto předpoklady splněny?

Předpoklady statistik

Jejich splnění podmiňuje

- matematickou **smysluplnost** výpočtu
  - typicky úroveň měření
- **přesnost**, výpovědní schopnost vypočítané hodnoty
  - typicky tvar rozložení

Při splnění všech předpokladů nese vypočítaná statistika tu informaci, kterou od nich v souladu se statistickou teorií očekáváme.

Statistiky, které nejsou porušením předpokladů příliš ovlivněné, jsou **ROBUSTNÍ**.

  - používáme i pro statistiky s minimálními či žádnými předpoklady.

AJ: assumptions, robust

Co ještě omezuje výpovědní schopnost statistik?

- Odlehlé, extrémní hodnoty
  - Není-li statistika příliš ovlivněna výskytem extrémních hodnot, je **REZISTENTNÍ**
  - Resistenci momentových statistik někdy zvyšujeme ořezáváním extrémů, např. trimmed mean
- Efekt podlahy a stropu
  - snižuje ukazatele variability
  - posunuje ukazatele centrální tendence
  - snižuje korelaci
  - … a nic moc s tím nenaděláme, to je věc metodologie