

PSY117/454 Statistická analýza dat v psychologii – přednáška 12

Analýza nominálních proměnných

Testování hypotézy o četnosti

Závislost nominálních proměnných

Analýza četností hodnot nominální proměnné

- Výzkumné otázky...
 - Liší se významně preference nějakých politických stran?
 - Liší se poměrné zastoupení kuřáků mezi ženami a muži?
 - Souvisí nějak preference politické strany s odhadem měsíčního příjmu respondenta?
 - Otázky směřují
 - buď k rozdílu četností různých jevů v rámci jedné proměnné (četnost různých jevů v jedné populaci),
 - k rozdílu četností jevu mezi různými proměnnými (četnost jevu v různých populacích),
 - Nebo k pravděpodobnosti výskytu dvou (či více) jevů současně.
- Nominální proměnná
 - Též kategoriální, alternativní
 - Zařazení jevu do určité kategorie
 - Jednotlivé kategorie musí být disjunktní – validita
 - Kategorie mohou vzniknout i transformací z proměnné vyššího řádu – kategorizace pořadí, známek ve škole, „nižší úzkost x vyšší úzkost“ atd.
- Klíčová slova
 - Četnost, relativní četnost, očekávaná četnost, rezidua, χ^2 (Chi-kvadrát)

χ^2 – test dobré shody

- Liší se empirické četnosti nějakých jevů od teoreticky očekávaných četností?
 - Házení kostkou – kolikrát padne 1,2,...
 - Preference stran

- $H_0: F(x) = F_0(x)$ vs. $H_1: F(x) \neq F_0(x)$

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$$

- kde k je počet kategorií, n_i pozorovaná četnost v kat. i , p_i je očekávaná četnost
- Rozdělení χ^2 ; stupně volnosti $df = k-1$
- Překoná-li hodnota χ^2 kritickou mez, H_0 zamítáme.
- Pro získání pravděpodobnosti χ^2 CHIDIST(x,volnost); CHIINV(prst, volnost)
- Očekávané četnosti... při uniformním rozložení 1:1:1...; nebo libovolně teoretické (10:24:32...)
- ! N empirických = N očekávaných

Závislost kategoriálních proměnných

- ❑ Jaká je souvislost preference politické strany a úrovně hrubého příjmu voliče?
- ❑ Jaká je pravděpodobnost společného výskytu dvou jevů x a y možných? Podmínka disjunkce!
- ❑ Kontingenční tabulka (crosstab)... $r \times s$
- ❑ Ve těle tabulky jsou četnosti jednotlivých kombinací, v okrajích tzv. marginální četnosti – sumy sloupců nebo řádků. Tedy n_{12} znamená počet osob ve druhém sloupci prvního řádku; počet osob, u nichž nastal jev A_1 a současně B_2 .

	B_1	B_2	...	B_x	Řádkové součty
A_1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
...
A_x	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
Sloupcové součty	$n_{.1}$	$n_{.2}$...	$n_{.s}$	n

Závislost kategoriálních proměnných

- Postup analogický, jako u jednorozměrné verze testu χ^2
- Očekávané četnosti:
- χ^2
- Stupně volnosti: $df = (r-1)*(s-1)$

$$m_{ij} = \frac{n_{i.} \cdot n_{.j}}{n}$$

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{ij} - m_{ij})^2}{m_{ij}}$$

	B ₁	B ₂	...	B _x	Řádkové součty
A ₁	n ₁₁	n ₁₂	...	n _{1s}	n _{1.}
A ₂	n ₂₁	n ₂₂	...	n _{2s}	n _{2.}
...
A _x	n _{r1}	n _{r2}	...	n _{rs}	n _{r.}
Sloupcové součty	n _{.1}	n _{.2}	...	n _{.s}	n

Síla vztahu v kontingenční tabulce

- Koeficient kontingence (Pearson) C_{kor}
- Cramerovo V
- C_{kor} se interpretuje jako Pearsonova korelace, V jako R^2 . Tedy $C_{kor}^2 \approx V$.
- Oba koeficienty v intervalu (0;1) Neindikují však žádným způsobem „směr“ vztahu. Směrů je v kontingenční tabulce mnoho.
- A proto... kontingenční tabulky jsou mnohdy účelné i tehdy, máme-li k dispozici data na vyšší úrovni měření.
 - Nelineární vztahy
 - Možnost výpočtu reziduí: $n_{ij} - m_{ij} = res_i$
 - Součet reziduí v tabulce vždy nula
 - Umožňují zjistit, kde jsou lokalizovány největší odchylky od náhodného rozložení četností v tabulce....
 - V SPSS: Standardizované rezidua (Pearsonova): rozdělení reziduí je normální s odchylkou 1; tedy standardizovaná rezidua s hodnotou $\pm 1,96$ jsou „zajímavá“.

- Hendl str. 297 – 313.