PSY117/454
Statistická analýza dat v psychologii
Přednáška 10

Statistické testování hypotéz II

Přehled testů, rozdíly průměrů, velikost účinku, síla testu

The great tragedy of Science – the slaying of a beautiful hypothesis by an ugly fact

Thomas Huxley

Základní výzkumné otázky/hypotézy

1. Stanovení hodnoty parametru

   =stanovení intervalu spolehlivosti na μ, σ, ρ, b...

2. Rozdíl mezi skupinami
   - mezi průměry, korelacemi, rozptyly, pravděpodobnostmi, pořadími....
   - např. Muži a ženy se liší v míře úzkostnosti. (oboustranná)
   - např. Muži jsou úzkostnější než ženy. (jednostranná, směrová)

3. Korelace mezi proměnnými
   - korelace, regrese, chí-kvadrát
   - např. Mezi věkem a počtem návštěv lékaře za rok existuje lineární korelace. (oboustranná)
   - např. Mezi věkem a počtem návštěv lékaře za rok existuje pozitivní lineární korelace. (jednostranná)

- 2. lze převést na 3. a naopak
- obecně mluvíme o velikosti efektu/účinku

AJ: difference, association, effect size, two-tailed, one-tailed (directional)

Přehledy statistických testů

- receptář Oseckých
  - třídění podle
    - počtu výběrů – 1 či 2
    - úrovně měření – alternativní, nominální, intervalová
    - typu procedury – interval spolehlivosti, test hypotézy, velikost potřebného výběru
- Hendl – kapitola 12 a str. 235
- online
  - http://www.graphpad.com/www/book/Choose.htm
  - http://www.whichtest.info/index.html
  - http://www.socialresearchmethods.net/selstat/ssstart.htm
  - česky: http://meloun.upce.cz/metody/

Testy na rozdíly středních hodnot

Nominální závislá

- *párový test*: binomický znaménkový test
- *nezávislé skupiny*: chí-kvadrát

Ordinální závislá

- *párový test*: Wilcoxonovo T
- *nezávislé skupiny*: Mann-Whitney U

Intervalová závislá

- *párový test*: párový t-test
- *nezávislé skupiny*:
  - známý rozptyl v populaci: z-test
  - neznámý rozptyl v populaci: t-test pro nezávislé skupiny
    - varianta pro stejné a nestejné rozptyly mezi skupinami

AJ: sign test, chi-square, Wilcoxon T, Mann-Whitney U, paired(-samples) t-test (dependent, repeated measures), one-sample t-test, independent samples t-test

Srovnání 2 nezávislých průměrů: *t*-test

Předpoklady použití … jsou-li výrazně porušeny, volíme raději neparametrický test

- proměnná je v populaci **normálně rozložená** - neřeší se, pokud je $n_1, n_2 > 30$
- homogenita rozptylů (**homoscedascita**), pokud $n_1 \neq n_2$
  - řeší modifikace t-testu pro nestejné rozptyly (6.2.3)
  - testuje se Levenovým testem (od oka $s_1^2/s_2^2 < 2$)
- **nezávislost pozorování** - řeší párový t-test (pro závislé výběry) (6.2.4)

- H_0: $m_1 - m_2 = 0$ (nebo roven konstantě, nebo >/< 0 či *c*) a zvolíme $\alpha = 1\%, 5\%$, nebo 10%
- Rozdíl má

  výběrovou chybu $s_d = \sqrt{\{[((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2))]*[1/n_1+1/n_2]\}}$

  ***t*-rozložení** s n_1+n_2-2 **stupni volnosti** (ν)
- Spočítáme **testovou statistiku** $t = (m_1 - m_2)/s_d$
- Zjistíme jaká je p **(t ≥ |zjištěná hodnota|)** - tabulky, TDIST(t, ν)
- Je-li $p \geq \alpha$, pak H_0 zůstává platná, je-li $p < \alpha$, H_0 zamítáme (a konstatujeme existenci statisticky významného rozdílu).
- Spočítáme Cohenovo d a interval spolehlivosti pro rozdíl průměrů.

Příklad: *t*-test pro nezávislé výběry

- *H*: Muži a ženy se liší v míře úzkostnosti.
  - H_0: $\delta = \mu_m - \mu_ž = 0$
  - nasbíraná data: $m_m=2$; $m_ž = 3$; $s_m=1{,}5$; $s_ž = 1{,}6$; $n_m = n_ž = 20$
  - H_0 budeme testovat na 5% hladině statistické významnosti, $\alpha = 0{,}05$
- Předpoklady splněny >> provádíme *t*-test pro nezávislé výběry (6.2.2)
- $d = m_ž - m_m = 2 - 3 = -1$
- $s_d = \sqrt{\{[((20-1)1{,}5^2+(20-1)1{,}6^2)/(20+20-2))]*[1/20+1/20]\}}=0{,}49$
- rozdíl má *t*-rozložení s $n_1+n_2-2 = 38$ stupni volnosti
- $t = (m_1 - m_2)/s_d = -1/0{,}49 = -2{,}04$
- $p\,(t \geq |-2{,}02|)$ je při $\nu = 38$ rovna 0,048 (TDIST(2,04;38;2)=0,048)
- $p < \alpha$, takže **zamítáme** H_0. Pokud by H_0 platila, zjištěný rozdíl by byl nepravděpodobný.
- 95% interval spolehlivosti: ${}_{0{,}025}t(38)$ = TINV(0,05;38) = 2,02
  $d - 2{,}02 * s_d < \delta < d + 2{,}02 * s_d$, tj. $-1{,}98 < \delta < -0{,}02$
- Cohenovo $d = |-1|/1{,}55 = 0{,}65$, což je středně velký efekt.

Velikost účinku/efektu

- Možnost srovnání mezi studiemi zkoumajícími tutéž výzkumnou otázku pomocí různě operacionalizovaných proměnných
- Možnost srovnání velikosti efektu vyjádřeného různými koeficienty
- Snadnější interpretace

Pro rozdíly středních hodnot

- Cohenovo $d = |m_1 - m_2| / s_{pooled}$; $s_{pooled} = \sqrt{[((n_1 - 1)s_1^2 + (n_2 - 1)s_2^2)/(n_1 + n_2 - 2))]}$
- varianta $d' = |m_1 - m_2| / s_{con}$; $s_{con} = s$ kontrolní skupiny

Pro těsnost vztahu (korelace)

- r a r^2, R^2, η^2(eta), ω^2 – podíl vysvětleného rozptylu závislé proměnné

Indikátory velikosti efektu lze mezi sebou navzájem převádět

- Cohenovo d na r: $r = \sqrt{(d^2/(d^2 + 4))}$
- r na Cohenovo d: $d = 2r/\sqrt{(1 - r^2)}$

AJ: effect size, Cohen's d, strength of association, explained variance

Síla testu

Síla testu (1-β) je pravděpodobnost, že existující rozdíl bude detekován, zjištěn jako statisticky významný.

Záleží na

- skutečné velikosti účinku (*d*, *r*…)
- variabilitě proměnné(ých) - *s*
- velikosti vzorku
- zvoleném riziku chyby I. typu, α: čím nižší je α tím nižší je síla testu
- zvoleném testu (parametrické mají vyšší sílu)

Obvykle toužíme po co nejvyšší síle testu, cca 0,8 a výše.

- Bojujeme o ni velikostí vzorku a kontrolou intervenujících proměnných.

Publikace výsledků testování hypotéz

- Primárně udáváme velikost efektu, nejlépe intervalem spolehlivosti
- Sekundárně udáváme výsledek statistického testování
  - udáváme získanou hodnotu p (Sig.)
  - uvádíme i testovou statistiku (i se stupni volnosti) – $t(\nu)$, $F(\nu_1,\nu_2)$, χ^2, M-W U...
- Interpretujeme nejlépe interval spolehlivosti. Výsledek statistického testování interpretujeme vzhledem k použité nulové hypotéze.

Testy normality rozložení

- Kolmogorov-Smirnov s Lillieforsovou korekcí, Shapiro-Wilks a jiné
- Testují H, že rozložení proměnné se neliší od normálního rozložení
  - jsou to jedny z tzv. **testů dobré shody** (goodness-of-fit tests)
  - testovaná H_0 je shoda; tj. $p<\alpha$ znamená příliš velkou odchylku od normality
- **Běžně se nepoužívají**
  - na malých vzorcích nenormalitu nedetekují (při n=20, 1-β < 0,5)
  - na velkých vzorcích (n > 1000) jsou naopak extrémně přísné
  - *t*-testy a ANOVA jsou proti narušení normality robustní, takže nám obvykle stačí konstatovat unimodalitu bez extrémního zešikmení
  - pro rozhodování mezi použitím parametrických a neparametrických testů volíme spíše **úroveň měření** a výzkumný kontext

AJ: tests of (univariate) normality, goodness-of-fit tests