# PSY117/454 Statistická analýza dat v psychologii – seminář 9

Statistické testování hypotéz

# Základní výzkumné otázky/hypotézy

1. Stanovení hodnoty parametru

   =stanovení intervalu spolehlivosti na $\mu$, $\sigma$, $\rho$, $b$…

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

# t-test (6.2) – porovnání 2 průměrů

- Předpoklady použití
  - proměnná je v populaci normálně rozložená
    - t-testu odchylky od normality příliš nevadí – neřeší se
  - homogenita rozptylů (homoscedascita), pokud $n_1 \neq n_2$
    - řeší modifikace t-testu pro nestejné rozptyly (6.2.3)
    - homogenita rozptylů se testuje Levenovým testem (od oka $s_1^2/s_2^2 < 2$)
  - nezávislost pozorování
    - řeší párový t-test (pro závislé výběry) (6.2.4)
- Předpoklady splněny >> provádíme *t*-test pro nezávislé výběry (6.2.2)
- Testujeme, zda rozdíl $m_1 - m_2 = 0$ ($H_0$) (nebo roven konstantě, nebo >/< 0 či *c*)
- Tento rozdíl má výběrovou chybu

  $s_d = \sqrt{\{[((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2))]*[1/n_1+1/n_2]\}}$
- Tento rozdíl má *t*-rozložení s $n_1+n_2-2$ stupni volnosti ($\nu$)
- Vytvoříme 100(1-$\alpha$)% interval spolehlivosti na $m_1 - m_2$
- Pokud je 0 (nebo *c*) v tomto intervalu $H_0$ zůstává platná, v opačném případě ji zamítáme (a konstatujeme existenci statisticky významného rozdílu).
- Spočítáme Cohenovo *d* a uvádíme je spolu výsledky.

# příklad: t-test pro nezáv. výb

- 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
- 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 stupňů volnosti
- 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. $-1{,}98 < \delta < - 0{,}02$
- Stanovený interval neobsahuje 0, takže $H_0$ na 5% hladině statistické významnosti zamítáme. Konstatujeme, že nějaký rozdíl tu p-ně je.
- Alternativně: $t=|d-0|/s_d=-1/0{,}49=2{,}04$; $P(t(38)\geq 2{,}04)$=TDIST(2,04;38;2)=**0,048** což je $<\alpha$
- Cohenovo $d = |-1|/1{,}55 = 0{,}65$ , což je poměrně 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 $r^2$, $R^2$, $\eta^2$(eta), $\omega^2$ – podíl vysvětleného rozptylu závislé p.
- Indikátory velikosti efektu lze mezi sebou navzájem převádět
  - d >> r : $r = \sqrt{(d^2/(d^2 + 4))}$
  - r >> d : $d = 2r/\sqrt{(1 - r^2)}$

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

# 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-$\beta$ < 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

# Publikace výsledků testování hypotéz

- Primárně udáváme velikost efektu, nejlépe intervalem spolehlivosti
- Sekundárně udáváme výsledek statistického testování
  - udáváme nejlépe přesnou hodnotu $p$ (Sig.)
  - uvádíme i testovou statistiku (i se stupni volnosti) – $t$(df), $F$(df1,df2), $\chi^2$, M-W $U$…
- Interpretujeme nejlépe interval spolehlivosti. Výsledek statistického testování interpretujeme vzhledem k použité nulové hypotéze.

# 3. průběžná

- umět spočítat 4 varianty t-testu (jednovýběrový, párový, nezávislý se stejnými a s nestejnými rozptyly), práce s t-rozložením v Excelu (fce TDIST, TINV)
- naučit se přehledy testů: Hendl – kapitola 12 a str. 235
- obligátní terminologie v rozsahu přednášek 7-9a

# Termíny zkoušek

- 5.6. (ve 14:00)
- 19.6.
- 26.6.
- 3.7.

- P21 v 16:00
- limit 30 míst
- cca 60 min