PSY117/454
Statistická analýza dat v psychologii
Přednáška 8

Statistické testování hypotéz

Země je kulatá ($p < 0{,}05$).
Jacob Cohen

Hypotézy

- Příklady (statistických) hypotéz
  - H: $\mu = 100$: Populační průměr IQ je roven 100.
  - H: $\sigma = 10$: Populační směrodatná odchylka je 10.
  - H: $\mu_1 - \mu_2 = 0$: Populační průměry μ_1 (psychotici) a μ_2 (zdraví) jsou stejné.
  - H: $\rho_{xy} = 0$: Proměnné X (pití piva) a Y (dominance) spolu nekorelují

- Vezměme si tu první hypotézu konfrontujme s daty:
  - Na vzorku 1000 náhodně vybraných dospělých jsme zjistili průměrné IQ rovné 105 (s =14).

AJ: statistical hypotheses testing, hypothesis, hypothesis supported by data

Statistický test hypotézy

Statistické testování založeno na p-nosti

- Známe-li pravděpodobnostní rozložení statistik můžeme usuzovat, **jak pravděpodobná je určitá výběrová statistika vzhledem k hypotéze: $P\,(D\,|H\,)$**
  - D : např. m=9,78
  - H : např. μ =10, $P\,(D\,|H\,)$ je $P\,(m=9{,}78 \mid \mu=10\,)$
- Je-li $P\,(D\,|H\,)$ vysoká, je tím hypotéza podpořena.
- Je-li $P\,(D\,|H\,)$ nízká, je tím hypotéza „činěna méně p-nou“

- Jak „vysoká$_{\text{nízká}}$“ je vysoká$_{\text{nízká}}$ pravděpodobnost, abychom hypotézu podpořili$_{\text{vyvrátili}}$?

Jak vysoká $P(D|H)$ je nutná k přijetí H?

- Bayesovský přístup – otázka není relevantní
  - s H je spojena určitá p-nost a ta se díky $P(D|H)$ zvyšuje či snižuje
  - Bayesův teorém – $P(H|D) = P(H) * P(D|H) / P(D)$
- Fisher, Pearson, Neyman – otázka je relevantní
  - Popper – princip falzifikace – H nelze potvrdit, pouze vyvrátit
  - My ale nechceme své hypotézy vyvracet, spíš potvrzovat
  - P-N: princip vzájemně se doplňujících konkurenčních hypotéz
    - Vytvořme takovou H, kt. bude logickou negací naší vědecké hypotézy a říkejme jí **nulová H**. Když se nám podaří nulovou H vyvrátit, znamená to **jakousi podporu** pro naši vědeckou hypotézu.
  - Vyvrácení H_0: $P(D|H_0) < \mathbf{0{,}05}$; $\mathbf{0{,}01}$; 0,001; 0,0001 podle zvyku

Terminologická vložka

H_0 : **nulová (statistická) hypotéza**

- logická negace (doplněk) vědecké hypotézy

H_1 : **vědecká, alternativní hypotéza**

- ta, o kterou nám primárně jde; $P\ (H_0 \cup H_1) = 1$

$P\ (D\ |H_0)$, kdy H_0 zamítáme:

- **úroveň/hladina statistické významnosti** (průkaznosti)
- ***α*****,** udává se často v procentech: 5%, 1%
- značí se i ***p*** nebo ***Sig.***
- p-nost chybného zamítnutí H_0 - **chyba prvního typu**
  - chyba, jejíž velikost jsme ochotni tolerovat

Jednostranné vs. **oboustranné** hypotézy

- jednostranné, směrové: $\mu \geq 23$, $\mu \leq 0$, z různých důvodů se jim vyhýbáme
- oboustranné: $\mu = 23$

AJ: null hypothesis, scientific/alternative hypothesis, level of statistical significance, type I error, one-tailed, two-tailed, directional

Postup testování statistické hypotézy

1. Formulujte **statistickou hypotézu**, kterou budete testovat (vyvracet) (H_0: $\mu = 0$)
2. Zvolte **hladinu statistické významnosti**, tj. míru rizika, že dojde k chybě 1. typu (např. $\alpha = 0{,}05$)
3. Hledáme p-nost získání naší výběrové statistiky nebo extrémnější hodnoty, za předpokladu, že H_0 je pravdivá: $P(D|H_0)$, *p*, Sig.
   - cesta vede přes znalost výběrového rozložení statistiky
   - např. $m = 0{,}5$. $P\,(|m|=0{,}5|\mu=0)$
   - obvykle je nutný přepočet na tzv. testovou statistiku, např. *t*, *z*...
4. Vyneseme rozhodnutí o H_0: zamítnutí či přijetí
   - je-li $P(D|H_0) < \alpha$, pak H_0 zamítáme
   - je-li $P(D|H_0) \geq \alpha$, pak H_0 nezamítáme

Příklad – jednovýběrový t-test

Terapie nevhodného chování.

- Rozdíl před-po: m=2,7; s=3,5; N=10
- H : Terapie má efekt. ($\mu \neq 0$)

1. H_0 : Terapie nemá efekt: $\mu = 0$
2. V sociálních vědách běžně α=0,05
3. $P(|m| \geq 2{,}7 | \mu=0) = ?$
   - s_m=3,5/odm(10)=1,1
   - $t = (m-\mu)/s_m$=2,7/1,1= 2,45
   - $P(|t| \geq 2{,}45 | \tau=0)$ = TDIST(2,45;9;2) = 0,04
4. $P(|m| \geq 2{,}7 | \mu=0) < 0{,}05$ >> zamítáme H_0

Protože při m =2,7 je velmi málo pravděpodobné, že by rozdíl byl 0, tak připouštíme, že nějaký rozdíl je.

Dichotomizace výsledků výzkumu

- Výsledek výzkumu je testováním zredukován na ano-ne

	H_0 přijata	**H_0 zamítnuta**
H_0 pravdivá (žádný efekt)	OK	chyba 1. typu α *(její pravděpodobnost)*
H_0 nepravdivá (efekt)	chyba 2. typu β	OK Síla (1-β)

Čím nižší je α, tím vyšší je β. Přesná podoba vztahu závisí na použitém testu. α i β mohou být nízké pouze při vysokých n.

Síla testu viz Hendl 401-411.

AJ: type-I error, type-II error, (statistical) power

Problémy statistického testování H

- Největší problém: dichotomizace
  - stejná velikost efektu dává při různých N jiné rozhodnutí o H_0
  - komplikuje až znemožňuje kumulativní budování znalostní báze
- Problém interpretace
  - p= $P(D\,|H_0)$ a nikoli $P(H\,|D)$
- Jak z jich ven?
  - VŽDY udávat velikost efektu (Cohenovo d, r, R^2, η^2, ω^2)
  - používat intervalové odhady
  - testování hypotéz používat pouze doplňkově