PSY117
Statistická analýza dat v psychologii
Přednáška 9 2016

Statistické testování hypotéz

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

Od vzorku k populaci a zpět

Vzhledem k tomu, jaká nám na vzorku vyšla statistika, jaký je odpovídající populační parametr?

interval spolehlivosti

Pokud předpokládáme, že v populaci je hodnota parametru X, co si myslet o své hypotéze poté, co nám na vzorku vyšlo Y?

statistický test hypotézy

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 a konfrontujme ji 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: $P\,(D\,|H\,)$**
  - *Př.* $D:\ m = 105$ nebo rozdíl mezi statistikou a hypotézou $|m - \mu| = 5$

    $H:\ \mu = 100$
  - $P\,(D\,|H\,)$ je $P\,(m = 105 \mid \mu = 100)$ resp. $P\,(|m - \mu| \geq 5 \mid \mu = 100)$
- Je-li $P\,(D\,|H\,)$ relativně vysoká, je tím hypotéza podpořena.
- Je-li $P\,(D\,|H\,)$ relativně nízká, hypotéza je „činěna méně p-nou“

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

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
  - Fisher (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 negací naší vědecké hypotézy a říkejme jí **nulová H**. Když se nám podaří nulovou H zamítnout, znamená to **podporu** pro naši vědeckou hypotézu.
  - Zamítnutí H_0: $P(D|H_0) <$ **0,05**; **0,01**; 0,001; 0,0001 podle zvyku

Dichotomizace výsledků výzkumu

- Výsledek výzkumu je testováním zredukován na ano-ne

	H_0 podržena $P(D\|H_0)\geq\alpha$	**H_0 zamítnuta** $P(D\|H_0)\geq\alpha$
H_0 pravdivá (žádný efekt)	OK	chyba 1. typu α *(její pravděpodobnost)*
H_0 nepravdivá (efekt)	chyba 2. typu β	OK *P*: **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*.

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

Terminologická vložka

H_0 : **nulová (statistická, testová, testovaná) hypotéza**

- obvykle logická negace (doplněk) vědecké hypotézy

H_1 : **alternativní (vědecká, výzkumná)hypotéza**

- ta, o kterou nám často primárně jde

$P(D\,|H_0)$, podle které rozhodujeme o zamítnutí H_0

- značí se ***p***, též p-value, p-hodnota (nebo v SPSS ***Sig.***, ale to je fuj)
- p-nost chybného zamítnutí H_0 - **chyba prvního typu**
- Je-li stanovena dopředu: **úroveň/hladina statistické významnosti** (průkaznosti), $\boldsymbol{\alpha}$, udává se často v procentech: 5%, 1%
  - 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 **testovou** (nulovou) **hypotézu**, kterou budete testovat (tj. vyvracet) (př. H_0: $\mu = 0$, nebo H_0: $\mu = 6$)
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. $m = 0{,}5$. $P\ (|m| \geq 0{,}5 | \mu = 0)$
   - 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. ($\mu \neq 0$) – oboustranná hypotéza

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
   - $P\ (|t| \geq 2,45\ |\tau = 0)$ = 2*(1–T.DIST(2,45;9;1)) = 0,04 (nebo TDIST(2,45;9;2))
4. $P\ (|m|\geq 2,7|\mu=0) < 0,05$ >> zamítáme H_0 - rozdíl mezi D a H0 je **statisticky významný(průkazný, signifikantní)**

Protože při m =2,7 je velmi málo pravděpodobné, že by rozdíl byl 0, tak nalézáme podporu pro přesvědčení, že μ >0.

Příklad – jednovýběrový *t*-test

Terapie nevhodného chování.

- Rozdíl před-po: m=2,7; s=3,5; N=10
- H : Terapie má efekt. (μ >0) – ***jednostranná*** hypotéza

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
   - $t = (m-\mu)/s_m = 2{,}7/1{,}1 = 2{,}45$
   - $P\ (t \geq 2{,}45\ |\tau = 0)$ = 1–T.DIST(2,45;9;1) = 0,02 (nebo TDIST(2,45;9))
4. $P\ (m \geq 2{,}7 | \mu=0) < 0{,}05$ >> zamítáme H_0 - rozdíl mezi D a H0 je **statisticky významný(průkazný, signifikantní)**

Protože při m =2,7 je velmi málo pravděpodobné, že by rozdíl byl 0 nebo menší, tak nalézáme podporu pro μ >0.

Příklad – jednovýběrový *t*-test

Terapie nevhodného chování.

- Rozdíl před-po: $m=-2{,}7$; $s=3{,}5$; $N=10$
- H : Terapie má efekt. ($\mu>0$) – ***jednostranná*** hypotéza

1. H_0 : Terapie nemá efekt: $\mu = 0$
2. V sociálních vědách běžně $\alpha=0{,}05$
3. $P\ (m\geq-2{,}7|\mu=0) = ?$
   - $s_m=3{,}5/\text{odm}(10)=1{,}1$
   - $t=(m-\mu)/s_m=-2{,}7/1{,}1= 2{,}45$
   - $P\ (t\geq-2{,}45\ |\tau=0) = 1-\text{T.DIST}(-2{,}45;9;1) = 0{,}98$ (nebo 1-TDIST(2,45;9;1))
4. $P\ (m\geq-2{,}7|\mu=0) < 0{,}05$ >> **ne**zamítáme H_0 - rozdíl mezi D a H0 není **statisticky významný(průkazný, signifikantní)**

Protože při $m=-2{,}7$ je pravděpodobnější, že je rozdíl 0, než že je pozitivní, ponecháváme nulovou hypotézu v platnosti.

Jednostranné testy

- Používáme pouze, pokud rozdíl, který by měl opačné znaménko, než čekáme, je bezvýznamný, neinterpretovatelný.

- Obvykle uvažujeme v jednostranných hypotézách, ale testujeme je oboustranně.
- Oboustranné testování je „bezpečná“ volba. Jednostranné obvykle přitahuje žádost o zdůvodnění.

Test signifikance Pearsonova korelačního koeficientu

Pokud H_0: ρ=0, pak

- Z=FISHER(ρ) má normální výběrové rozložení se $s_Z=1/\sqrt{(n-3)}$
- FISHER(r)/s_Z ~ N(0;1)
- $P(D|H_0)$=2*(1 – NORMSDIST(Z/s_Z) pro oboustrannou (non-directional) H_1

Pokud H_0: ρ=c, pak

- D_Z=(FISHER(r)–FISHER(c)) má normální výběrové rozl. se $s_Z=1/\sqrt{(n-3)}$
- D_Z/s_Z ~ N(0;1)
- $P(D|H_0)$=2*(1 – NORMSDIST(D_Z/s_Z) pro oboustrannou (non-directional) H_1

Problémy statistického testování H

- Dichotomizace rozhodnutí
  - stejná *velikost účinku* 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 $P(H|D)$
- Problém nulové hypotézy
  - Test je smysluplný, jen když je nulová hypotéza smysluplná.

Největší problém je tedy formální, bezmyšlenkovité testování.

- Jak z problémů ven?
  - VŽDY se primárně zajímat o velikost účinku (Cohenovo *d*, *r*, R^2, η^2, ω^2)
  - používat intervalové odhady, kdy to jen lze
  - testování hypotéz používat pouze doplňkově

Doporučené čtení

- Cohen.
- ASA statement 2016:
  http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108

Shrnutí

- Statistické testování hypotéz vychází z konstrukce intervalu spolehlivosti pro hypotetizovaný parametr
- Může znamenat (ne)podporu pro hypotézu, nikoli striktně potvrzení/vyvrácení