

Testování statistických hypotéz

Testování hypotéz je statistická metoda, která určuje, jak pravděpodobná jsou naměřená data v případě platnosti hypotézy, kterou testujeme.

Historické souvislosti

Testování statistické významnosti pomocí testování nulové hypotézy (null hypothesis statistical testing – NHST) je velmi staré. Původním autorem myšlenky je zřejmě John Arbuthnott (1710), který se snažil prokázat Boží prozřetelnost skrze odhalování zákona vyrovnávajícího počet narozených mužů a žen.^{[1] [2]}

Postup

Pro testování vždy vytváříme **dvě hypotézy**. Testovanou, kterou označíme jako **nulovou** a značíme ji H_0 a **alternativní**, kterou značíme H_1 . **Nulová hypotéza H_0** je formulovaná **negativně** („to, co chceme prokázat, neplatí“, tj. **pozorovaný efekt vznikl „náhodou“**) a **alternativní hypotéza H_1** formulovaná jako neplatnost H_0 , tj. že pozorovaný efekt nevznikl jen náhodou.

Nulová hypotéza H_0 bývá formulovaná pomocí rozdělení, nebo pravděpodobnosti...

Cílem testování je H_0 zamítnout a přijmout H_1 .

Příklady formulování hypotéz

- Jev, který chceme prokázat – formulace hypotézy H_0 .
- Účinek léku A se od léku B liší o „x“ – H_0 : střední hodnota veličiny účinku léku A a B je stejná.
- Účinek léku je u diabetiků vyšší – H_0 : střední hodnota veličiny účinku léku u diabetiků a u kontrolní skupiny je stejná.
- Střední doba dožití je po podání léku A vyšší – H_0 : střední doba dožití je po podání léku A stejná jako u kontrolní skupiny.
- Čím vyšší BMI, tím menší střední doba dožití – H_0 : střední doba dožití u skupiny s BMI kolem 20 a u skupiny s BMI kolem 35 je stejná.

Testová statistika

K samotnému testování se využívá tzv. **testová statistika T**. Jedná se o vzorec, funkci dat, která udává, jak pravděpodobná jsou naměřená data, pokud platí nulová hypotéza. **⚠ U testování hypotéz nelze testovou statistiku interpretovat jako pravděpodobnost, s níž platí nulová hypotéza na základě naměřených dat.** Ve frekvenční statistice jsou totiž populační veličiny, o nichž se formulují hypotézy, konstanty, které jsou neměnné, stále (a pokud se nezměří celá populace, tak neznámé). Mluvit o jejich pravděpodobnosti proto nemá smysl^[† 1]. Naopak, mluví se o pravděpodobnosti naměření naměřených dat za předpokladu platnosti nulové hypotézy (testování se tedy provádí svým způsobem naruby).

Nulové rozdělení

Nulové rozdělení je rozdělení testové statistiky při platnosti H_0 .

p-hodnota testu

p-hodnota testu (z anglického *p value*) je pravděpodobnost, že při H_0 by testová statistika T nabyla hodnoty, jaká vyšla z dat, nebo hodnoty ještě extrémnější (mimo interval $<-T, T>$).

Hladina významnosti testu

Hladina významnosti testu se označuje α . (Nejčastěji $\alpha = 0,05 = 5\%$). Je to zvolené číslo z intervalu od 0 do 1, resp. 100 % (čím menší, tím lepší). Pokud je $p < \alpha$, tak platnost H_0 je velmi málo pravděpodobné a potom:

Zamítáme H_0 na hladině významnosti α a přijímáme H_A .

Buď H_0 platí, ale nastala data, která se objevují s **pravděpodobností menší než α** (nastalo něco velmi nepravděpodobného), nebo **vsutku platí H_A** , k čemuž se kloníme.

Výsledek testu je pak **statisticky významný** na hladině α . (Často značeno „*“ pro $\alpha = 0,05$, „**“ pro $\alpha = 0,01$ a „***“ pro $\alpha = 0,001$)

⚠ Pokud $p \geq \alpha$, pak to neznamená, že zamítáme H_A , ale pouze nezamítáme H_0 . Výsledek je pak statisticky nevýznamný (insignifikantní) na hladině α . (Často značeno „NS“.)

Kritická hodnota testu pro hladinu α

Kritická hodnota testu pro hladinu α je hranice mezi statisticky významnou a nevýznamnou hodnotou testové statistiky T .

Neporovnává se p přímo s α , ale pro α se vypočítá kritická hodnota, s níž se porovná samotné T . Jedná se o ekvivalentní porovnání. (Dříve numericky přístupnější — nebylo nutné počítat kvantily.)

Vztah ke konfidenčním intervalům

Konfidenční interval (interval spolehlivosti) pro daný parametr je intervalový odhad nějakého parametru s danou pravděpodobností. Pokud je μ_0 v konfidenčním intervalu na hladině spolehlivosti $1-\alpha$, pak hypotéza $\mu = \mu_0$ se nezamítá na hladině α . Konfidenční interval je složen z hodnot, které test na dané hladině nezamítá.

Statistické chyby

- Chyba prvního typu – zamítneme-li hypotézu, ačkoliv je správná.
- Chyba druhého typu – nezamítneme-li nulovou hypotézu ačkoliv není správná.

Často používané testy

- Studentův t-test
- ANOVA
- χ^2 test

1. V tomto spočívá rozdíl mezi frekvenční a bayesovskou statistikou, v níž populační veličiny nejsou konstantní, ale náhodnými veličinami.

Odkazy

Související články

- Studentův t-test
- ANOVA

Reference

1. ARBUTHNOTT, John. An Argument for Divine Providence, taken from the constant Regularity observ'd in the Births of both Sexes. *Philosophical Transactions of the Royal Society* [online]. 1710, vol. 27, s. 186–190, dostupné také z <<http://www.jstor.org/stable/103111>>.
2. SOUKUP, Petr. Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum - SDA Info* [online]. 2010, roč. 4, no. 2, s. 77–104, dostupné také z <<http://dav.soc.cas.cz/issue/3-data-a-vyzkum-2-2010/6>>. ISSN 2336-2391.

Použitá literatura

- KLASCHKA, Jan. *Testování statistických hypotéz* [přednáška k předmětu Zdravotnická statistika 1,2, obor Všeobecné lékařství, 1. lékařská fakulta Univerzita Karlova]. Praha. 26. 4. 2011.
- BENCKO, Vladimír, et al. *Epidemiologie : výukové texty pro studenty 1. LF UK*. 1. vydání. Praha : Karolinum, 2002. 168 s. ISBN 80-246-0383-7.