KGG/STG Statistika pro geografy

6. Úvod do testování statistických hypotéz

Mgr. David Fiedor
23. března 2015

Osnova

Motivace

- pomoc v rozhodování
- objektivní kvantifikace rizik
- ...

$\Rightarrow$ nutnost vypozorování a sesbírání dat

Základní pojmy

Definice

Předpoklad o neznámé vlastnosti základního souboru se nazývá *statistická hypotéza*.

Testováním hypotéz rozumíme rozhodovací postup, který je založen na daném náhodném výběru, s jehož pomocí rozhodujeme.

Charakteristika procesu

- statistiky vypočtené z náhodných výběrů jsou vždy zatíženy určitou variabilitou hodnot - chybou
- rozdíl mezi dvěma výběry nemusí být způsoben skutečným rozdílem parametrů (náhodná chyba)
- náhodná chyba vzniká spolu působením mnoha náhodných vlivů - vliv člověka, okolí při pozorování, způsob měření, . . .

Nulová hypotéza H_0

Při testování hypotéz si vždy musíme dobře rozmyslet předpoklad, který budeme ověřovat. Tento předpoklad nazýváme *nulová hypotéza* H_0.

- velký důraz na správnou formulaci
- obecně formulovaná tak, že nepředpokládá žádný rozdíl, žádnou odlišnost
- opakem je *alternativní* (pracovní) *hypotéza* H_1
- nesmí nastat případ, kdy by neplatila žádná z nich nebo naopak obě (vždy jen jedna)

Hladina významnosti α

Hladinu významnosti α můžeme definovat jako pravděpodobnost, že se zamítne nulová hypotéza, ačkoliv platí.

- často uváděná v procentech
- odpovídá míře ochoty výzkumníka smířit se s výskytem chyby

Chyby I. a II. druhu

1. Nulová hypotéza ve skutečnosti platí a my jsme ji testem nezamítli. V tomto případě je vše v pořádku.
2. Nulová hypotéza ve skutečnosti platí, ale my jsme ji testem zamítli, přestože je pravdivá. V tomto případě se dopouštíme tzv. *chyby I. druhu*, o které jsme se již v závěru předcházející podkapitoly zmínili. Pravděpodobnost této chyby značíme α.

Chyby I. a II. druhu

3. Nulová hypotéza ve skutečnosti neplatí, avšak my jsme ji testem nezamítli, přestože není pravdivá. V tomto případě se dopouštíme tzv. *chyby II. druhu*. Pravděpodobnost této chyby značíme β.
4. Nulová hypotéza ve skutečnosti neplatí a my jsme ji správně testem zamítli. Proto je v tomto případě vše v pořádku.

Chyby I. a II. druhu

Následující tabulka přehledně ilustruje všechny možné situace:

Tabulka: Čtyři možné případy, které mohou nastat při testování hypotéz

	rozhodnutí	
skutečnost	H_0 nezamítáme	H_0 zamítáme
H_0 platí	správné rozhodnutí	chyba I. druhu
H_0 neplatí	chyba II. druhu	správné rozhodnutí

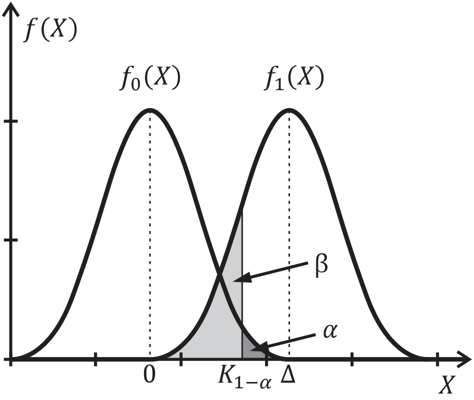
Síla testu

- pravděpodobnost správně nezamítnuté nulové hypotézy (ve skutečnosti platí) je rovna $1 - \alpha$
- čím menší hladinu významnosti α zvolíme, tím menší je pravděpodobnost chyby I. druhu, avšak zároveň tím roste pravděpodobnost chyby II. druhu, a naopak

Definice

Jestliže nulová hypotéza neplatí a my jsme ji zamítli, tak pravděpodobnost tohoto jevu nazýváme *síla testu* a tato pravděpodobnost je rovna $1 - \beta$.

Grafické znázornění chyby I. a II. druhu



Obrázek: Ilustrace vztahu mezi chybou I. a II. druhu

Formulace hypotéz

1. oboustranná alternativa

$$H_0\colon \mu = \mu_0$$

$$H_1\colon \mu \neq \mu_0$$

2. pravostranná alternativa

$$H_0\colon \mu \leq \mu_0$$

$$H_1\colon \mu > \mu_0$$

3. levostranná alternativa

$$H_0\colon \mu \geq \mu_0$$

$$H_1\colon \mu < \mu_0$$

Metody testování statistických hypotéz

- pomocí kritického oboru
- pomocí intervalů spolehlivosti
- pomocí *p*-hodnoty

Obecný postup testování hypotéz si předvedeme na metodě testování pomocí kritického oboru.

Základní pojmy

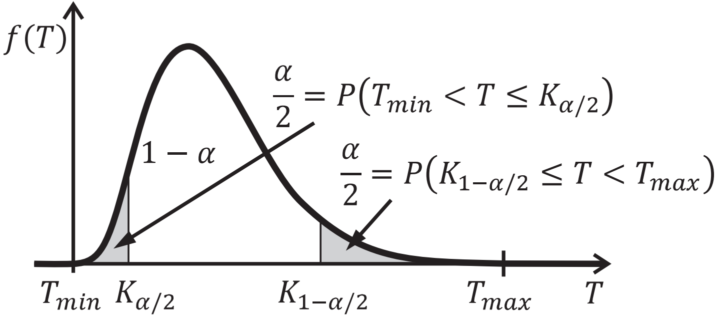
K testování nulové hypotézy H_0 proti alternativní hypotéze H_1 použijeme statistiku T_0, kterou nazýváme *testovací kritérium (testová statistika)*. Jedná se o vhodnou funkci náhodného výběru, která má vztah k nulové hypotéze, a jejíž rozdělení za předpokladu platnosti nulové hypotézy známe.

Množina všech hodnot, kterých může testovací kritérium nabýt, se rozpadá na obor nezamítnutí nulové hypotézy (značíme V) a obor zamítnutí nulové hypotézy, který nazýváme *kritický obor (W)*.

Základní pojmy

Tyto dva obory jsou odděleny *kritickými hodnotami* (pro danou hladinu významnosti α, resp. pro daný rozsah souboru je lze najít ve statistických tabulkách).

Kritický obor



Obrázek: Kritický obor pro oboustrannou alternativu

$$W = \left(T_{min}, K_{\frac{\alpha}{2}} \right\rangle \cup \left\langle K_{1-\frac{\alpha}{2}}, T_{max} \right)$$

Testování hypotéz

Obecný postup testování statistických hypotéz:

1. Formulace problému, sběr dat.
2. Stanovení nulové hypotézy H_0 a alternativní hypotézy H_1.
3. Volba hladiny významnosti α.
4. Volba vhodného testovacího kritéria T_0 a výpočet jeho realizace.
5. Rozhodnutí o nulové hypotéze na základě kritického oboru.
6. Interpretace výsledku testu.

1. Formulace problému, sběr dat

- velmi důležité, nesmí být souhrn několika problémů
- výběr musí být náhodný, reprezentativní, . . .

2. Stanovení nulové hypotézy H_0 a alternativní hypotézy H_1

- oboustranný, či jednostranný test
- nulová hypotéza předpokládá neexistenci rozdílu; náhodný výběr pochází ze základního souboru s předem daným rozdělením; . . .
- alternativní hypotéza je stavěna do protikladu hypotézy nulové - není možná platnost obou zároveň

3. Volba hladiny významnosti α

- hladina významnosti α představuje pravděpodobnost (hodnotu rizika), že zamítneme nulovou hypotézu, i když je správná
- zmenšením pravděpodobnosti dopuštění se chyby I. druhu zvětšujeme pravděpodobnost, že se dopustíme chyby II. druhu
- v praxi nejčastěji volíme hladinu významnosti $\alpha = 0,05$ nebo $\alpha = 0,01$

4. Volba vhodného testovacího kritéria T_0 a výpočet jeho realizace

- každé testovací kritérium T_0 se řídí určitým teoretickým rozdělením, pokud nulová hypotéza platí
- na základě realizací náhodného výběru se vypočítá realizace t_0 testovacího kritéria T_0
- obecný tvar testovacího kritéria

$$\text{testovací kritérium} = \frac{\text{bodový odhad} - \text{hypotetická hodnota}}{\text{směrodatná chyba odhadu}}$$

kde hypotetická hodnota značí hodnotu předpokládanou v nulové hypotéze

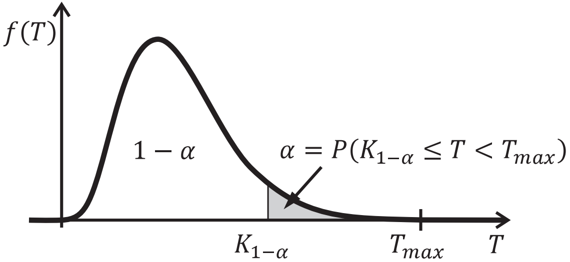
5. Rozhodnutí o nulové hypotéze na základě kritického oboru

Kritický obor W pro oboustrannou alternativu a pro jednostranné varianty testů:

- oboustranná varianta: $W = \left(T_{min}, K_{\frac{\alpha}{2}}\right\rangle \cup \left\langle K_{1-\frac{\alpha}{2}}, T_{max}\right)$,
- pravostranná varianta: $W = \langle K_{1-\alpha}, T_{max})$,
- levostranná varianta: $W = (T_{min}, K_{\alpha}\rangle$,

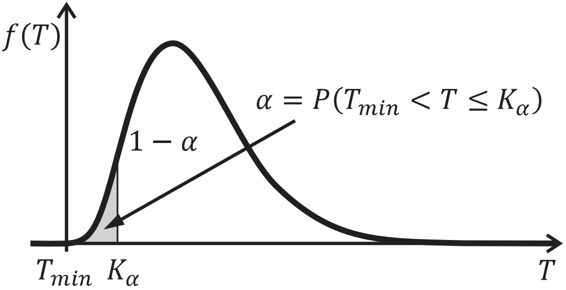
kde T_{min}, resp. T_{max} označuje nejmenší, resp. největší hodnotu testovacího kritéria T_0 a $K_{\frac{\alpha}{2}}$, $K_{1-\frac{\alpha}{2}}$, K_{α}, $K_{1-\alpha}$ značí kvantily rozdělení, jímž se řídí zvolené testovací kritérium při platnosti H_0 (říkáme jim kritické hodnoty)

5. Rozhodnutí o nulové hypotéze na základě kritického oboru



Obrázek: Kritický obor pro pravostrannou alternativu $W = \langle K_{1-\alpha}, T_{max})$

5. Rozhodnutí o nulové hypotéze na základě kritického oboru



Obrázek: Kritický obor pro levostrannou alternativu $W = \langle T_{min}, K_\alpha \rangle$

5. Rozhodnutí o nulové hypotéze na základě kritického oboru

Po vytvoření kritického oboru mohou nastat pouze dvě možnosti:

a) Realizace t_0 testovacího kritéria T_0 leží v kritickém oboru W. V takovém případě zamítáme nulovou hypotézu H_0 na hladině významnosti α a přijímáme alternativní hypotézu H_1.

b) Realizace t_0 testovacího kritéria T_0 neleží v kritickém oboru W. V takovém případě nezamítáme nulovou hypotézu H_0 na hladině významnosti α. To však neznamená, že nulová hypotéza je správná. Pozorovaná data nám pouze poskytla informaci, jež nenasvědčuje tomu, že je nulová hypotéza nepravdivá.

6. Interpretace výsledku testu

- velká obezřetnost, abychom netvrdili to, co se ve skutečnosti neprokázalo
- nezamítnutí nulové hypotézy může být způsobeno dvěma jevy:
  - malý rozsah náhodného výběru
  - skutečná platnost nulové hypotézy

Pomocí intervalů spolehlivosti

Předpokládejme, že testujeme nulovou hypotézu H_0: $\mu = c$ proti oboustranné alternativě:

- sestrojíme pro danou parametrickou funkci její intervalový odhad - získáme $100(1-\alpha)\%$ interval spolehlivosti
- pokryje-li tento interval hodnotu c, pak H_0 nezamítáme na hladině významnosti α
- v opačném případě H_0 na této hladině významnosti α zamítáme

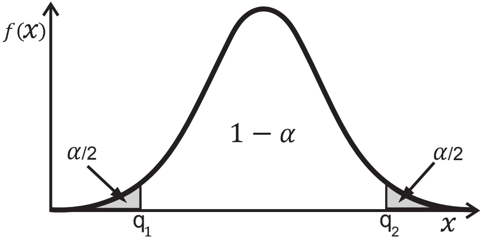
Pomocí intervalů spolehlivosti

Pro test H_0 proti:

- oboustranné alternativě H_1 sestrojíme oboustranný interval spolehlivosti
- levostranné alternativě H_1 sestrojíme pravostranný interval spolehlivosti
- pravostranné alternativě H_1 sestrojíme levostranný interval spolehlivosti

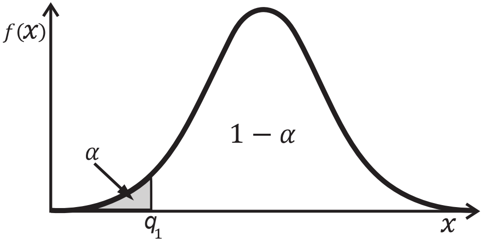
Pomocí intervalů spolehlivosti

V souvislosti s testováním hypotéz budeme pro hranice intervalu spolehlivosti používat symboly d, h.



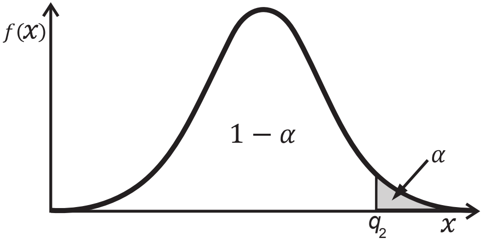
Obrázek: Oboustranný interval spolehlivosti

Pomocí intervalů spolehlivosti



Obrázek: Levostranný interval spolehlivosti

Pomocí intervalů spolehlivosti



Obrázek: Pravostranný interval spolehlivosti

Výpočet p-hodnoty

p-hodnota nám udává nejnižší možnou hladinu významnosti α, při které ještě zamítáme nulovou hypotézu H_0.

Manuální výpočet p-hodnoty:

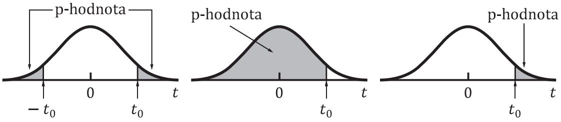
- pro oboustrannou alternativu $p = 2\,min\left\{P\left(T_0 \le t_0\right), P\left(T_0 \ge t_0\right)\right\}$,
- pro levostrannou alternativu $p = P(T_0 \le t_0)$,
- pro pravostrannou alternativu $p = P(T_0 \ge t_0)$,

kde t_0 značí realizaci testovacího kritéria T_0 a P označuje pravděpodobnost daného jevu.

Interpretace *p*-hodnoty

a) Je-li $p \leq \alpha$, pak zamítáme nulovou hypotézu H_0 na hladině významnosti α a přijímáme alternativní hypotézu H_1.

b) Je-li $p > \alpha$, pak nezamítáme nulovou hypotézu H_0 na hladině významnosti α.



Obrázek: Grafické znázornění *p*-hodnoty

Základní rozdělení

- **parametrické** - o parametrech rozdělení základního souboru (průměr, rozptyl, shoda průměrů, . . .)
  - výběr musí pocházet z normálního rozdělení
  - data musí být intervalového nebo poměrového typu
- **neparametrické** - obdoba parametrických
  - používáme v případech, kdy nelze použít test parametrický
  - i pro data ordinálního typu
  - slabší než parametrické testy

Podle počtu výběrů

- jednovýběrové
- dvouvýběrové - zajištěna nezávislost všech veličin
- párové - oba výběry spolu úzce souvisí - např. měření na témž objektu v různém čase (stejný rozsah obou výběrů)

Chybou je především použití dvouvýběrového testu místo párového.

Prostředky grafického ověřování normality

- histogram
- normální pravděpodobnostní graf (N-P plot)
- kvantil-kvantilový graf (Q-Q plot)

N-P plot

- normal probability plot
- na vodorovnou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$
- na svislou osu kvantily u_{α_j}, kde $\alpha_j = \dfrac{3j-1}{3n+1}$

$\Rightarrow$ v případě shodnosti hodnot vynášíme průměrné pořadí

Průměrné pořadí

Příklad

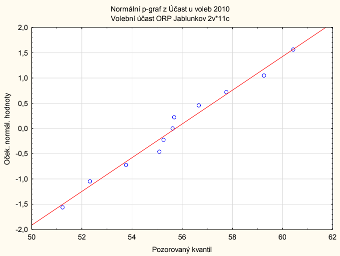
Pětkrát jsme provedli jisté měření s těmito naměřenými hodnotami: 1,8; 2,5; 1,9; 2,1; 1,9. Následující tabulka nám předvede zápis uspořádaných hodnot a jejich pořadí (případně jejich průměrného pořadí, pokud jsou naměřené hodnoty stejné). Po tomto příkladu by již mělo být vše zřejmé.

Tabulka: Tabulka udávající pořadí

uspořádané hodnoty	$1,8$	$1,9$	$1,9$	$2,1$	$2,5$
pořadí	1	2	3	4	5
průměrné pořadí	1	$2,5$	$2,5$	4	5

N-P plot

Pochází-li data z normálního rozdělení, budou uspořádané dvojice $(x_{(j)}, u_{\alpha_j})$ vynesené do N-P plotu ležet na přímce, resp. bude možno jimi proložit přímku tak, že body tohoto grafu budou ležet na přímce, nebo v její těsné blízkosti.



Obrázek: Normální-pravděpodobnostní graf

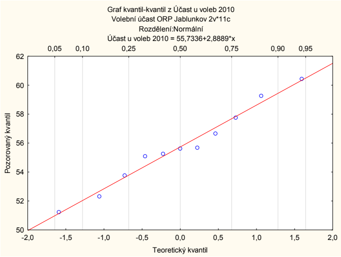
Q-Q plot

- quantile-quantile plot
- na svislou osu vynášíme uspořádané hodnoty $x_{(1)} \leq \cdots \leq x_{(n)}$
- na vodorovnou osu kvantily $K_{\alpha_j}(X)$ vybraného rozdělení, kde $\alpha_j = \dfrac{j - 0,375}{n + 0,25}$

$\Rightarrow$ v případě shodnosti hodnot vynášíme průměrné pořadí

Q-Q plot

Pochází-li data z daného rozdělení, budou uspořádané dvojice $(x_{(j)}, u_{\alpha_j})$ vynesené do Q-Q plotu ležet na přímce, resp. bude možno jimi proložit přímku tak, že body tohoto grafu budou ležet na přímce, nebo v její těsné blízkosti.



Obrázek: Kvantil-kvantilový graf (Volby 2010 v ORP Jablunkov)

Motivace

- nelze u testování opomíjet předpoklady o typu rozdělení - zavádějící výsledky
- grafické ověření nestačí - přílišná subjektivita

Příklady testů normality dat:

- Kolmogorovův-Smirnovův test (K-S test)
- χ^2 test
- Shapiro-Wilkův test (S-W test)

Princip testů

- testujeme hypotézu H_0, která říká, že náhodný výběr $X_1, X_2, \ldots, X_n$ pochází ze základního souboru s normálním rozdělením
- proti alternativě H_1, že výběr pochází ze základního souboru s jiným rozdělením
- vypočítáme testovou statistiku a srovnáme s příslušnou p-hodnotou
- pokud je p-hodnota větší než zvolená hladina významnosti α, pak nulovou hypotézu nezamítáme

$\rightarrow$ neznamená to však, že nulová hypotéza platí!

K-S test

- patří do kategorie testů „dobré shody“ (zjišťování shody mezi teoretickým a empirickým rozdělením)
- zcela obecný a použitelný pro testování jakéhokoliv typu rozdělení
- speciální varianty tohoto testu - Lillieforsova modifikace K-S testu (pouze k ověření normality dat v situacích, kdy neznáme parametry rozdělení a odhadujeme je z dat)
- použitelný pro různé rozsahy výběru, své využití najde především u souborů s malým rozsahem

Princip testu

- založen na výpočtu testovacího kritéria ze srovnání empirické a teoretické distribuční funkce
- testovacím kritériem D v tomto případě rozumíme maximum jejich absolutního rozdílu
- porovnání s kritickou hodnotou, kterou pro daný rozsah n a danou hladinu významnosti α nalezneme v tabulkách
- v případě, že realizace testovacího kritéria D je větší než kritická hodnota D_p, zamítáme nulovou hypotézu

χ^2 test

- patří do kategorie testů „dobré shody" (zjišťování shody mezi teoretickým a empirickým rozdělením)
- testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, X_2, \ldots, X_n$ pochází z rozdělení s danou distribuční funkcí
- použitelný pro testování jakéhokoliv typu rozdělení

Princip testu

- rozdělíme data do r třídicích intervalů $\langle u_j, u_{j+1}\rangle$, $j = 1, \ldots, r$
- zjistíme absolutní četnost n_j j-tého třídicího intervalu a vypočteme pravděpodobnost p_j, že náhodná veličina X se bude realizovat v j-tém třídicím intervalu
- testovací kritérium se za platnosti H_0 řídí rozdělením χ^2 s $r - 1 - p$ stupni volnosti, kde p určuje počet odhadovaných parametrů daného rozdělení – pro normální rozdělení je $p = 2$
- kritické hodnoty lze najít v tabulkách – jsou to kvantily rozdělení $\chi^2(r - 1 - p)$

Omezení a použitelnost testu

- hodnota testovacího kritéria je silně vázána na volbu třídicích intervalů
- neměl by být použit v případě, kdy $np_j < 5$ pro některé $j = 1, \ldots, r$ (raději sloučit intervaly)
- použitelný především pro výběry s větším rozsahem
- od jisté hodnoty rozsahu výběru je dokonce silnějším testem než K-S test

S-W test

- nepatří již do kategorie testů „dobré shody“
- založen na zjištění, zda body v Q-Q plotu jsou významně odlišné od regresní přímky proložené těmito body
- na testovou statistiku tohoto testu můžeme nahlížet jako na korelační koeficient mezi uspořádanými pozorováními a jim odpovídajícími kvantily normovaného normálního rozdělení
- hypotézu o normalitě zamítneme na dané hladině významnosti α, pokud se na této hladině neprokáže korelace mezi daty a jim odpovídajícími kvantily
- především pro soubory s menším rozsahem ($n < 50$)

Děkuji za pozornost...