Výberové skúmanie

Testovanie štatistických hypotéz

Osnova prednášky

- Základné pojmy
- 4 kroky testovania štatistických hypotéz
- testy hypotéz o parametri rozdelenia jednotiek základného súboru
- test hypotézy o zhode rozdelenia

Základné pojmy

- **Štatistická hypotéza**
  je tvrdenie alebo predpoklad o neznámom parametri rozdelenia štatistických jednotiek v základnom súbore, resp. o type rozdelenia.

- **Testovanie hypotézy**
  je proces, v ktorom na základe skúmania jednotiek z výberového súboru overujeme platnosť alebo neplatnosť hypotézy.

Proces testovania

1^0 Prvý krok: stanovenie nulovej (H_0) a alternatívnej hypotézy (H_1), obsahom ktorých sú predpoklady o veľkosti parametra základného súboru

H_0: $\theta = \theta_0$ H_1: $\theta \neq \theta_0$ (obojstranná alternatíva)

$\theta > \theta_0$ (pravostranná alternatíva)

$\theta < \theta_0$ (ľavostranná alternatíva

kde θ je parameter základného súboru,

θ_0 – jeho predpokladaná hodnota

zhoda rozdelenia empirického a predpokladaného súboru:

H_0: $f(x) = g(x)$ H_1: $f(x) \neq g(x)$

2^0 Druhý krok:
-voľba
-veľkosti **hladiny významnosti α**

α je pravdepodobnosť zamietnutia správnej (pravdivej) H_0
β je pravdepodobnosť nezamietnutia nesprávnej H_0;
$1\text{-}\beta$ - sila testu

Skutočnosť / Úsudok	H_0 je pravdivá	Pravdepodobnosť	H_0 je nepravdivá	Pravdepodobnosť
H_0 sa nezamieta	správne rozhodnutie	$1\text{-}\alpha$	chyba II. druhu	β
H_0 sa zamieta	chyba I. druhu	α	správne rozhodnutie	$1\text{-}\beta$

3^0 Tretí krok:

- určenie testovacej charakteristiky (testovacej štatistiky) a výpočet jej hodnoty
- množina hodnôt, ktoré môže testovacia štatistika nadobudnúť, sa nazýva **výberový priestor**
- výberový priestor je rozdelený na dva disjunktné podpriestory – **obor prijatia** *V a* **kritický obor** *W*

4^0 Štvrtý krok:

- rozhodnutie o výsledku testu

Pre tento krok je potrebné určiť kritickú hodnotu, ktorá vymedzuje **kritický obor**, teda oblasť zamietnutia nulovej hypotézy alebo rozhodnúť podľa ***p*-hodnoty**.

p-hodnota je najnižšia hladina významnosti, na ktorej ešte zamietame nulovú hypotézu

závery:

- ak je hodnota testovacej štatistiky **v kritickom obore**, bola testom preukázaná platnosť hypotézy H_1, **zamietame teda H_0**
- ak je hodnota testovacej štatistiky **v obore prijatia**, **nemôžme zamietnuť hypotézu H_0**

Test hypotézy **o strednej hodnote** základného súboru

1^O H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$ (a); $\mu > \mu_0$ (b); $\mu < \mu_0$ (c)

2^O α

3^O ak σ základného súboru poznáme, testovacou štatistikou je :

$$\boxed{U = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}} \quad (1)$$

ak σ základného súboru nepoznáme, testovacou štatistikou je :

$$\boxed{T = \frac{\overline{X} - \mu_0}{\frac{S_1}{\sqrt{n}}}} \quad (2)$$

4^0 V prípade, že obsahom alternatívnej hypotézy je tvrdenie (a) $\mu \neq \mu_0$,

a

testovacou štatistikou je (1), hypotézu H_0 zamietame,

ak $|U| > u_{1-\alpha/2}$;

ak testovacou štatistikou je (2), hypotézu H_0 zamietame,

ak $|T| > t_{1-\alpha/2,\nu}$; $\nu = n$ -1

resp. $t_{\alpha,\nu}$ (v Exceli)

V prípade, že obsahom alternatívnej hypotézy je tvrdenie (b): $\mu > \mu_0$,

a

testovacou štatistikou je (1), hypotézu H_0 zamietame,

ak $U > u_{1-\alpha}$;

ak testovacou štatistikou je (2), hypotézu H_0 zamietame,

ak $T > t_{1-\alpha,\nu}$,

resp. $t_{2\alpha,\nu}$ (pri hľadaní v Exceli)

V prípade, že obsahom alternatívnej hypotézy je tvrdenie (c) $\mu < \mu_0$,

a

testovacou štatistikou je (1), hypotézu H_0 zamietame,

ak $U < -u_{1-\alpha}$;

ak testovacou štatistikou je (2), hypotézu H_0 zamietame,

ak $T < -t_{1-\alpha,\nu}$

resp. $-t_{2\alpha,\nu}$ (pri hľadaní v Exceli)

Test hypotézy o **relatívnej početnosti** základného súboru

1^O H_0: $\pi = \pi_0$ H_1: $\pi \neq \pi_0$ (a); $\pi > \pi_0$ (b); $\pi < \pi_0$ (c)

2^O α

3^O ak platí podmienka $\longrightarrow$ $n > \dfrac{9}{\pi_0(1-\pi_0)}$

testovacou štatistikou je

$$\boxed{U = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}}$$

4[0] V prípade, že obsahom alternatívnej hypotézy je tvrdenie

(a) $\pi \neq \pi_0$,

hypotézu H_0 zamietame, ak $|U| > u_{1-\alpha/2}$;

(b) $\pi > \pi_0$,

hypotézu H_0 zamietame, ak $U > u_{1-\alpha}$;

(c) $\pi < \pi_0$,

hypotézu H_0 zamietame, ak $U < -u_{1-\alpha}$.

Test hypotézy o **zhode rozdelenia** dvoch základných súborov

1[O] H_0: $f(x) = g(x)$ H_1: $f(x) \neq g(x)$

2[O] α

3[O] Testovacia štatistika $\boxed{\chi^2 = \sum \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}}$

kde n_i sú empirické početnosti

$n.p_i$ sú teoretické početnosti, t.j početnosti za platnosti H_0

4[0] hypotézu H_0 zamietame, ak $\chi^2 > \chi^2_{1-\alpha,\nu}$

resp. $\chi^2 > \chi^2_{\alpha,\nu}$ pri hľadaní kritických hodnôt rozdelenia v Exceli,

$\nu = m - 1 - p - z$,

- m je počet tried, do ktorých sme súbor vytriedili,
- p je počet odhadovaných parametrov rozdelenia
- z je počet zlúčených tried

(zlučujú sa triedy, pri ktorých je $n.p_i \leq 5$)

Ďakujem za pozornosť