

# Testovanie štatistických hypotéz

## *(Statistical hypotheses testing)*

Testovanie štatistických hypotéz (TŠH) je klasická súčasť matematickej štatistiky, používaná v takmer všetkých technických, prírodných a spoločenských vedách, v medicínskom výskume, marketingovom výskume...

Pochopiť TŠH žiaľ nie je úplne jednoduché. Navyše, existuje obrovské množstvo "inštancií" štatistických testov, ktoré sa používajú v rôznych oblastiach, každý z nich má svoju (často komplikovanú) „teóriu“ aj „prax“. Tu si vysvetlíme len základné všeobecné princípy TŠH.

### *Parametrický pravdepodobnostný model*

Základom TŠH je trieda pravdepodobnostných rozdelení (*parametric family*) na pozorované dáta, a.k.a. parametrický pravdepodobnostný model (*parametric probabilistic model*).

Pravdepodobnostný model je pre TŠH nutný, pretože ŠH je priamo formulovaná v reči pravdepodobnostného modelu.

Budeme uvažovať len parametrické triedy pravd. modelov pre spojité kvantitatívne dáta. Analogické princípy TŠH platia aj pre iné triedy pravdepodobnostných modelov, v rámci takzvanej „frekventistickej“ paradigmy (nie „Bayesovskej“):

$$\{ f(., \theta) \mid \theta \in \Theta \}$$

- $\Theta \subseteq R^m$  je “parametrický priestor” (*parametric space*)
  - $f(., \theta)$  je hustota na  $R^n$  pre každé  $\theta \in \Theta$ .
- Predpoklad: Existuje „skutočný“ parameter (*true parameter*)  $\theta^* \in \Theta$ , taký, že pozorované dáta sú (alebo budú) realizáciou náh. premennej (alebo náh. vektora) s hustotou  $f(., \theta^*)$ .

### *Nulová a alternatívna hypotéza*

- Parametrický priestor si rozdelíme na dve disjunktné podmnožiny

$$\Theta = \Theta_0 \cup \Theta_1$$

- Nulová a alternatívna hypotéza

$$H_0: \theta^* \in \Theta_0 \quad vs. \quad H_1: \theta^* \in \Theta_1$$

## Test štatistickej hypotézy

Test štatistickej hypotézy je funkcia z množiny potenciálnych dát (typicky z  $R^n$ ) do množiny

$\{\text{"zamietam"}, \text{"nezamietam"}\}$  ( $\{\text{"reject"}, \text{"don't reject"}\}$ ).

Presnejšie, do množiny {"zamietam  $H_0$  na hladine významnosti  $\alpha$ ", "nezamietam  $H_0$  na hladine významnosti  $\alpha$ "}; po anglicky {"I reject  $H_0$  at the significance level  $\alpha$ ", "I don't reject  $H_0$  at the significance level  $\alpha$ "}, kde  $\alpha \in (0,1)$  je istý parameter samotného testu (niečo úplne iné než  $\theta$ ).

**Príklad: Test hypotézy, že stredná hodnota náhodného výberu z normálneho rozdelenia je zadané číslo  $\mu_0$ :**

- *Parametrický pravdepodobnostný model na dáta:*

Realizácie nezávislých náhodných premenných s rozdelením  $N(\mu, \sigma^2)$  pre nejaké  $\mu$  a  $\sigma^2$ . Čiže  $\{f(\cdot, \theta) \mid \theta \in \Theta\}$  je množina všetkých hustôt rozdelení  $N(\mu, \sigma^2)$  pre nejaké  $\mu$  a  $\sigma^2$ , parameter  $\theta = (\mu, \sigma^2)$  patrí do  $\Theta = R \times (0, \infty)$ .

- *Nulová a alternatívna hypotéza:*

$$\Theta_0 = \{\mu_0\} \times (0, \infty), \Theta_1 = \Theta \setminus \Theta_0$$

- *Test:*

Nech  $x_1, \dots, x_n$  sú dáta, nech  $\bar{x}$  je ich aritmetický priemer a

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

je výberový rozptyl. Nech

$$T = \sqrt{n} \frac{\bar{x} - \mu_0}{S_x}$$

Nech  $\alpha \in (0,1)$  je stanovené číslo.

„Tradičný“ postup testovania:

Nech  $t_{\alpha/2}(n-1)$  je  $100(1 - \frac{\alpha}{2})$  percentný kvantil

Studentovho t-rozdelenia s  $n-1$  stupňami voľnosti. Ak  $T$  padne do „kritickej oblasti“ (*critical region*)

$$W = \left(-\infty, -t_{\alpha/2}(n-1)\right) \cup \left(t_{\alpha/2}(n-1), \infty\right),$$

tak výsledok testu je „zamietam“. Ak  $T$  nepadne do  $W$ , tak výsledok testu je „nezamietam“.

„Moderný“ postup testovania:

Nech  $G_{n-1}$  je kvantilová funkcia Studentovho t-rozdelenia s  $n - 1$  stupňami voľnosti. Ak je „p-hodnota“ (*p-value*)

$$p = 2(1 - G_{n-1}(|T|))$$

menšia ako  $\alpha$ , tak výsledok testu je „zamietam“, v opačnom prípade je výsledok testu „nezamietam“.

Oba prístupy vedú na ten istý test, no „moderný“ prístup ešte poskytne p-hodnotu, ktorá je sama osebe informatívna.

### *Chyba prvého druhu (type I error)*

„Chybou prvého druhu“ je pravdepodobnosť, že výsledkom ŠT bude „zamietam“, v situácii, že  $H_0$  platí, čiže ak  $\theta^* \in \Theta_0$ .

Vo všeobecnosti by pravdepodobnosť výsledku "zamietam" mohla závisieť od toho, aké presne je  $\theta^*$  (ktoré nepoznáme).

Štatistické testy pre konkrétne pravdepodobnostné modely a konkrétne hypotézy sú však vymyslené tak, že *chyba 1. druhu nezávisí od neznámeho  $\theta^*$  a je rovná stanovenému  $\alpha$ .*

Taký test nazývame testom hypotézy  $H_0$  na hladine významnosti  $\alpha$ .

## Test z príkladu vyššie je presne tak vymyslený!

### *Korektné testovanie štatistických hypotéz*

1) Na základe „pilotných“ experimentov, „exploratívnej“ (*exploratory*) analýzy dát a/alebo iných znalostí si stanovíme:

1a) Pravdepodobnostný model na pozorované dáta,

1b) Nulovú a alternatívnu štatistickú hypotézu (v rámci 1a),

1c) Štatistický test, ktorý použijeme (na testovanie 1b),

1d) Návrh a protokol „potvrdzujúceho“ (*confirmatory*) experimentu;

2) Vykonáme potvrdzujúci experiment v súlade s 1d) a vykonáme štatistický test hypotézy 1b) pomocou testu 1c).

Poznámka: Každý z bodov v časti 1) si v praxi vyžaduje ďalšie znalosti. Špeciálne, v bode 1d) je dôležité stanoviť dostatočný rozsah experimentu, aby mal požadovanú „silu“ (*test power*).

### *Nekorektné testovanie štatistických hypotéz*

Spomením na prednáške na konci semestra týkajúcej sa "etiky dátovej vedy".