MI-SPOL-16

Testování statistických hypotéz. T-testy, testy nezávislosti, testy dobré shody.

Náhodný vektor $X = (X_1, ..., X_n)^T$, který má **nějaké rozdělení.**

Tvrzení o tomto rozdělení, jehož platnost je neznámá, je **hypotéza.**

Testování hypotéz: Mechanismus, jak na základě napozorovaných hodnot X ověrovat platnost hypotéz

Vždy se pracuje se dvěma hypotézami:

- **Nulová hypotéza** H_0**:** označuje tvrzení, o kterém chceme rozhodovat
- **Alternativní hypotéza** H_A**:** opačné tvrzení, které se v rozhodovacím procesu staví proti H_0

Předpoklad: ve skutečnosti platí buď H_0, nebo H_A

Test nulové hypotézy H_0 proti alternativní hypotéze H_A je rozhodovací proces založený na hodnotě X, na jehož základě **zamítneme** nebo **nezamítneme** H_0

Chyby při testování

Chyba prvního druhu: Zamítneme H_0, i když platí

Chyba druhého druhu: Nezamítneme H_0, i když neplatí

Nelze kontrolovat pravděpodobnost obou chyb $\Rightarrow$ H_0 je volena tak, aby chyba **1. druhu byla závažnější** než chyba 2. druhu

Hladina významnosti testu: Hodnota α určující maximální pravděpodobnost chyby 1. druhu (běžně $\alpha = 5\%$ nebo $\alpha = 1\%$)

Nejsilnější test: Test, který má mezi všemi testy na stejné hladině významnosti nejmenší pravděpodobnost chyby 2. druhu

Výsledky testování

Říkáme, že: **Testujeme hypotézu H_0 proti alternativě H_A na hladině významnosti α.**

Zamítnutí H_0 ve prospěch H_A je **silný výsledek**

- Pokud je H_0 zamítnuta, je tvrzení H_A **statisticky významné** (jinak je statisticky nevýznamné)

Hypotézu, kterou chceme dokázat, tedy volíme jako **alternativní** hypotézu H_A. Pokud je výsledkem testu zamítnutí H_0, víme, že H_A platí s pravděpodobností aspoň $1 - \alpha$

Matematická formulace

Test hypotézy H_0 proti alternativě H_A založený na pozorování náhodného vektoru X, jehož rozdělení je z nějaké množiny možných rozdělení $\mathcal{P} = \{P_\theta | \theta \in \Theta\}$, kde Θ je množina všech možných hodnot indexu θ.

Nulová a alternativná hypotéza jsou **podmnožiny** $\mathcal{P}$ tvořící jeho disjunktní rozklad: $H_0 \cup H_A = \mathcal{P}$

V indexové množine jim odpovídají podmnožiny $\Theta_0 = \{\theta | P_\theta \in H_0\}$ a $\Theta_A = \{\theta | P_\theta \in H_A\}$. Takže $\Theta_0 \cup \Theta_A = \Theta$

Nulová hypotéza H_0 platí, právě když X má rozdělení P_θ pro nějaké $\theta \in \Theta_0$

Kritický obor

Množina realizací X, pro které testování na hladině α skončí zamítnutím H_0

Značení: W_α

Pokud naměříme **data z W_α**, **zamítáme H_0** na hladině α:

- $X \in W_\alpha \Leftrightarrow$ Zamítáme H_0 na hladině α
- $X \notin W_\alpha \Leftrightarrow$ Nezamítáme H_0 na hladině α

P-hodnota

Minimální hladina významnosti $\hat{p}$, na které lze **zamítnout** hypotézu H_0

$$\hat{p} \equiv \hat{p}(X) = \inf\{\alpha | X \in W_\alpha\}$$

Význam: Pokud je p-hodnota menší než požadovaná hladina významnosti, zamítá se H_0
Velikost p-hodnoty říká, jak silné je zamítnutí H_0 (čím menší p, tím významnější zamítnutí)

$\hat{p}$ má rovnoměrné rozdělení na intervali $(0, 1), \hat{p} \sim Unif(0, 1)$

Typy hypotéz

Dělení podle toho, do jaké míry známě rozdělení X:

- **parametrické:** rozdělení X určeno parametrem $\theta \in \Theta \subset \mathbb{R}^d$. Tvrzení se týkají hodnot θ
- **neparametrické:** X má obecné rozdělení. Tvrzení se týkají různých vlastností rozdělení (medián, nezávislost...) nebo jeho tvaru (test dobré shody)

Dělení podle množství rozdělení uvedených v hypotézách:

- **jednoduchá hypotéza:** pouze jedno rozdělení
- **složená hypotéza:** více rozdělení

Intervaly spolehlivosti

Náhodný výběr $X = (X_1, ..., X_n)^T$ z rozdělení určeného parametrem $\theta \in \Theta \subset \mathbb{R}$

Chceme testovat jednoduchou parametrickou hypotézu $H_0 : \theta = \theta_0$ proti **oboustranné alternativě** $H_A : \theta \neq \theta_0$ pro konkrétní hodnotu θ_0

$(L(X), U(X))$ je **oboustranný** $100 \cdot (1 - \alpha)\%$ **interval spolehlivosti** pro parametr θ, sestavený na základě náhodného výběru X

Pro každé $\theta \in \Theta : P_\theta(\theta \in (L, U)) = 1 - \alpha$

Testování:

- **Zamítneme** H_0, pokud $\theta_0 \notin (L, U)$
- **Nezamítneme** H_0, pokud $\theta_0 \in (L, U)$

Kritický obor testu: $W_\alpha = \{x | \theta_0 \notin (L(x), U(x))\}$

Jednostranná alternativa: Testuje se $H_0 : \theta \le \theta_0$ proti $H_A : \theta > \theta_0$

- **Horní interval spolehlivosti:** Pro každé $\theta \in \Theta : P_\theta(\theta \in (L, +\infty) = P_\theta(\theta > L) = 1 - \alpha$
- **Zamítneme** H_0, pokud $\theta_0 \notin (L, +\infty)$
- **Kritický obor:** $W_\alpha = \{x | \theta_0 \le L(x)\}$

Testy o normálním rozdělení

V důsledku Centrální limitní věty.

Test $H_0 : \mu = \mu_0$ proti $H_A : \mu \ne \mu_0$ na hladině významnoti α:

- Při známém rozptylu σ^2 zamítneme H_0, pokud μ_0 neleží v intervalu

$$\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- Při neznámém rozptylu σ^2 zamítneme H_0, pokud μ_0 neleží v intervalu

$$\left(\bar{X}_n - t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}\right)$$

Testování hypotéz pomocí statistik

Testová statistika: $T \equiv T(X)$, funkce náhodného vektoru X, u které při platnosti nulové hypotézy známe její rozdělení

Kritický obor testové statistiky: Množina S_α -- podmnožina možných hodnot T, pro kterou platí $\sup_{\theta \in \Theta_0} P_\theta(T \in S_\alpha) \le \alpha$
(když platí H_0, má T hodnoty v S_α s pravděpodobností nejvýše α)

Provedení testu:

- **Zamítneme** H_0, pokud $T \in S_\alpha$
- **Nezamítneme** H_0, pokud $T \notin S_\alpha$

Kritický obor testu W_α implicitně $W_\alpha = \{x | T(x) \in S_\alpha\}$

T-testy: Testy o hodnotách parametrů normálního rozdělení

Testy o střední hodnotě normálního rozdělení (jednovýběrový t-test)

Známý rozptyl

Náhodný výběr $X_1, ..., X_n$ z normálního rozdělení $N(\mu, \sigma^2)$, rozptyl je **známý**

Testuje se hodnota μ a porovnává se s μ_0. **Statistika:**

$$T = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

Statistika má při platnosti $\mu = \mu_0$ rozdělení $T \sim N(0,1)$

Testování: $H_0 : \mu = \mu_0$ proti $H_A : \mu \neq \mu_0$ na hladině α

- $S_\alpha = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, +\infty)$
- $T \in S_\alpha \Leftrightarrow |T| \geq z_{\alpha/2}$

Podmínka zamítnutí je stejná jako by byla při testu založeném na konfidenčních intervalech.

Test $H_0 : \mu \leq \mu_0$ proti $H_A : \mu > \mu_0$ na hladině α:

- $S_\alpha = [z_\alpha, +\infty)$
- $T \in S_\alpha \Leftrightarrow T \geq z_\alpha$

Neznámý rozptyl

Statistika:

$$T = \frac{\bar{X}_n - \mu_0}{s_n/\sqrt{n}}$$

Kritický obor:

- Oboustranný: $|T| \geq t_{\alpha/2,n-1}$
- Jednostranný ($H_0 : \mu \leq \mu_0$): $T \geq t_{\alpha,n-1}$

Test o rozptylu normálního rozdělení

Statistika:

$$T = \frac{(n-1)s_n^2}{\sigma_0^2}$$

Kritický obor:

- Oboustranný: $T \leq \chi^2_{1-\alpha/2,n-1} \vee T \geq \chi^2_{\alpha/2,n-1}$
- Jednostranný ($\sigma^2 \leq \sigma_0^2$): $T \geq \chi^2_{\alpha,n-1}$

Párový t-test

Náhodný výběr $(X_1, Y_1)^T, ..., (X_n, Y_n)^T$ z dvourozměrného rozdělení s neznámým vektorem středních hodnot $(\mu_1, \mu_2)^T$

Testujeme hypotézu $H_0 : \mu_1 = \mu_2$ proti $H_A : \mu_1 \neq \mu_2$

Položme $Z_i = X_i - Y_i$. Veličiny $Z_1, ..., Z_n$ jsou i.i.d se střední hodnotou $\mu_\Delta = \mu_1 - \mu_2$.
Předpoklad: $Z_i \sim N(\mu_\Delta, \sigma^2)$, kde σ^2 neznáme

$\Rightarrow$ převedení testu na **jednovýběrový t-test** $H_0 : \mu_\Delta = 0$ proti $H_A : \mu_\Delta \neq 0$

Statistika:

$$T = \frac{\bar{Z}_n}{s_Z/\sqrt{n}}$$

s_Z -- odmocnina výběrového rozpylu veličiny Z

Kritický obor:

- Oboustranný: $|T| \geq t_{\alpha/2,n-1}$
- Jednostranný ($\mu_1 < \mu_2$): $T \geq t_{\alpha,n-1}$

Dvouvýběrový t-test

Výběr $X_1, ..., X_n$ z normálního rozdělení $N(\mu_1, \sigma_1^2)$ a nezávislý náhodný výběr $Y_1, ..., Y_m$ z normálního rozdělení $N(\mu_2, \sigma_2^2)$ (předpoklad, že σ neznáme)

Test: $H_0 : \mu_1 = \mu_2$ proti $H_A : \mu_1 \neq \mu_2$

Statistika (moc dlouhá) má při platnosti $\mu_1 = \mu_2$ studentovo rozdělení s určitým počtem stupňů volnosti (závisí na tom, zda $\sigma_1^2 = \sigma_2^2$, nebo ne)

Test dobré shody (χ^2)

Test shodnosti diskrétních rozdělení

Náhodný výběr $X = X_1, ..., X_n$ z diskrétního rozdělení p' *(vektor pravděpodobností $p_1, ..., p_k$, kterými nastávají jednotlivé výsledky náhodného pokusu o k možných výsledcích)*. Četnosti $N_1, ..., N_k$ jednotlivých hodnot X mají multinomické rozdělení $M(n, p')$

Test: H_0 : skutečné hodnoty pravděpodobností jsou $p_1, ..., p_k$

Statistika:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}$$

Kritický obor: $\chi^2 \geq \chi^2_{\alpha,k-1}$

Test je **asymptotický** -- nutné mít dostatečně velký rozsah výběru

Test dobré shody při **neznámých parameterch:** skutečné hodnoty $p_1, ..., p_k$ mohou záviset na m -rozměrném parametru $\theta = (\theta_1, ..., \theta_m)^T$, jehož hodnota se také odhaduje
Kritický obor: $\chi^2 \geq \chi^2_{\alpha,k-m-1}$

Test nezávislosti v kontingenčních tabulkách

Náhodný vektor $X = (Y, Z)^T$ s diskrétním rozdělením. Veličina Y nabývá hodnot $1, ..., r$, veličina Z nabývá $1, ..., c$.

Sdružené pravděpodobnosti: $p_{ij} = P(Y = i, Z = j)$
Marginální pravděpodobnosti: $p_{i\bullet} = \sum_j p_{ij}$, $p_\bullet j = \sum_i p_{ij}$

Náhodný výběr z rozdělení X o velikosti n.
N_{ij} -- počet výsledků, kdy nastala dvojice (i, j): $N_{ij} = |\{k|Y_k = i, Z_k = j\}|$
N_{ij} mají sdružené multinomické rozdělení s parametrem n a pravděpodobnostmi p_{ij}

Kontingenční tabulka: náhodná matice N rozměru $r \times c$ se složkami N_{ij}

Kontingenční tabulka

Y \ Z	1	$\dots$	c	Σ
1	N_{11}	$\dots$	N_{1c}	$N_{1\bullet}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
r	N_{r1}	$\dots$	N_{rc}	$N_{r\bullet}$
Σ	$N_{\bullet 1}$	$\dots$	$N_{\bullet c}$	n

Test: Nezávislost veličin Y a Z, $H_0 : p_{ij} = p_{i\bullet}p_{\bullet j}$ pro každé i, j

- Aplikuje se χ^2 test

Statistika:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(N_{ij} - \frac{N_{i\bullet}N_{\bullet j}}{n}\right)^2}{\frac{N_{i\bullet}N_{\bullet j}}{n}}$$

Kritický obor: $\chi^2 \geq \chi^2_{\alpha,(r-1)(c-1)}$

NIST: Runs above/below

Náhodné veličiny $X_1, ..., X_n$ se stejným rozdělením s $\mu = EX_i, P(X_i = \mu) = 0$ a μ je medián

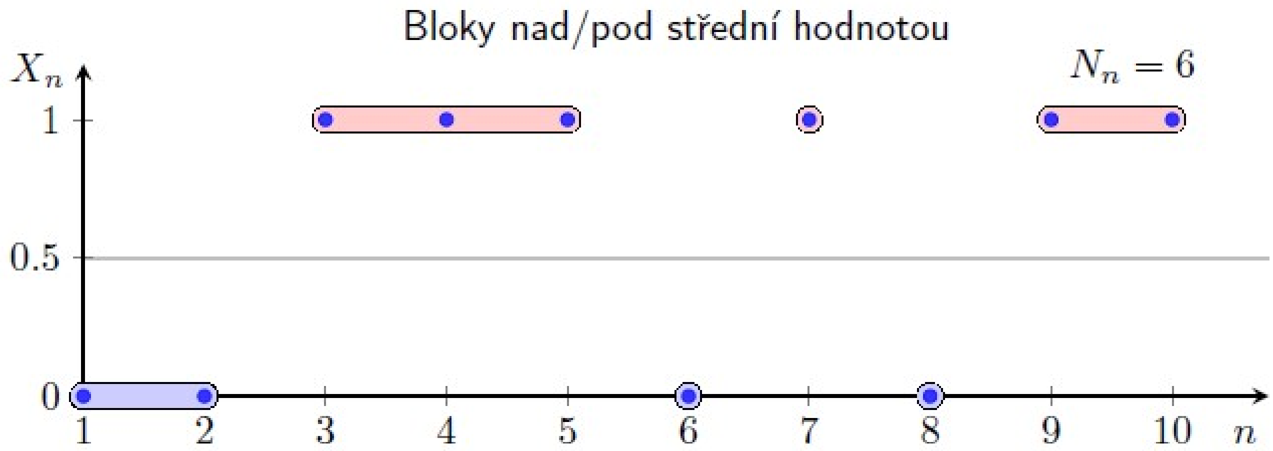
Test: $H_0 : X_i$ jsou nezávislé

$N_n = \#$ bloků hodnot nepřekračujících μ

Statistika:

$$T = \frac{2N_n - n - 1}{\sqrt{n-1}}$$

Kritický obor: $|T| \geq z_{\alpha/2}$



NIST: Runs up/down

$X_1, ..., X_n$ posloupnost náhodných veličin se stejným rozdělením, $P(X_i = X_i + 1) = 0$ pro každé $i < n$

Test: $H_0 : X_i$ jsou nezávislé

$N_n = \#blokůmonotónníchhodnot$

Statistika: $T = \frac{3N_n - 2n + 1}{\sqrt{1{,}6n - 2{,}9}}$

Kritický obor: $|T| \geq z_{\alpha/2}$