Testování hypotéz - základní principy

Předpokládejme test hypotézy H_0: $\theta = \theta_0$ proti alternativě H_1: $\theta \neq \theta_0$.

θ je neznámý parametr pravděpodobnostního modelu pro zkoumaná data $x_1,\ldots,x_n$,
např. střední hodnota rozdělení, ze kterého data pocházejí,
θ_0 je známé číslo.

Test H_0 proti H_1 je rozhodnutí, zda H_0 máme zamítnout, či ne.

Při rozhodování může nastat **chyba 1. druhu** (zam. H_0, když platí)
a **chyba 2. druhu** (nezam. H_0, když neplatí).

Hladina testu je předepsaná pravděpodobnost chyby 1. druhu α.
Zpravidla volíme $\alpha = 0.05$.
Při menší hodnotě α by narůstala pravděpodobnost chyby 2. druhu β a klesala **síla testu** $1-\beta$ = P(zam. H_0, když neplatí).

Provedení testu se realizuje na základě práce s daty. Existují 3 ekvivalentní způsoby.

(I) Pomocí testové statistiky T_n a kritického oboru C

Testová statistika:
Teoreticky je $T_n(X_1,\ldots,X_n)$ funkce náhodného výběru, tedy náhodná veličina.
Prakticky je $T_n(x_1,\ldots,x_n)$ funkce dat, tedy číslo.

Zpravidla je T_n odvozena od bodového odhadu parametru θ.

Rozhodovací pravidlo:
T_n leží v C => zam. H_0 ve prospěch H_1 na hladině α,
T_n neleží v C => H_0 nelze zam. na základě našich dat.

Konstrukce kritického oboru: potřebujeme znát rozdělení, ze kterého pocházejí data.

Příklad: **Test hypotézy o střední hodnotě.**
Nechť jsou data pozorovanou realizací náhodného výběru $X_1,\dots,X_n$ z rozdělení N(μ, σ^2).
Testujeme H_0: $\mu = \mu_0$ proti H_1: $\mu \neq \mu_0$.

1)
Bodový odhad střední hodnoty μ je průměr $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$.
$|\bar{X} - \mu_0|$ velké svědčí proti H_0 => potřebujeme najít dolní mez D pro $|\bar{X} - \mu_0|$ takovou,
aby P(zam. H_0, když platí) = P($|\bar{X} - \mu_0| > D$) = α .

2)
Platí: $X_1,\dots,X_n$ mají rozdělení N(μ, σ^2) => $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ má rozdělení N(μ, σ^2/n)
=> **centrovaný a normovaný průměr** $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ **má rozdělení N(0,1).**

3)
Označme α – kvantil rozdělení N(0,1) symbolem u(α) a distribuční funkci N(0,1) jako $\Phi(z)$.

Pro veličinu Z s rozdělením N(0,1) platí: P[Z < u(α)] = Φ[u(α)] = α .

Odtud a ze symetrie hustoty N(0,1) kolem 0 máme u(α) = - u(1-α).

Pomocí distribuční funkce lze počítat P($a < Z < b$) = $\Phi(b) - \Phi(a)$.

Proto P[|Z| < u(1-α/2)] = P[- u(1-α/2) < Z < u(1-α/2)] = P[u(α/2) < Z < u(1-α/2)] =

Φ[u(1-α/2)] - Φ[u(α/2)] = 1- α/2 - α/2 = 1 - α.

Tedy za platnosti H_0 pro $Z_0 = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$ a pro α = 0.05 platí P[$|Z_0|$ < u(0.975)] = 0.95,
P[$|Z_0| \geq$ u(0.975)] = 0.05.

4)
Testová statistika je $T_n = |Z_0|$,
kritický obor je $\dfrac{|\bar{X} - \mu_0|}{\sigma}\sqrt{n} \geq$ u(0.975) = 1.96 => zam. H_0 na hladině 5%.

(II) Pomocí intervalového odhadu parametru θ

Intervalový odhad ***I*** má teoreticky náhodné meze a
pokryje neznámý odhadovaný parametr $\boldsymbol{\theta}$
s předepsaně velkou pravděpodobností (např. 95 %),
prakticky má **číselné meze spočítané z dat**.

Rozhodovací pravidlo:
θ_0 neleží v I => zam. H_0 ve prospěch H_1 na hladině α,
θ_0 leží v I => H_0 nelze zam. na základě našich dat.

Konstrukce intervalového odhadu: potřebujeme znát rozdělení, ze kterého pocházejí data.

Příklad: Test hypotézy o střední hodnotě.
Nechť jsou data pozorovanou realizací náhodného výběru $X_1,\dots,X_n$ z rozdělení N(μ, σ^2).
Testujeme H_0: $\mu = \mu_0$ proti H_1: $\mu \neq \mu_0$.

1)

Platí: $X_1,\dots,X_n$ mají rozdělení N(μ, σ^2) => $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ **má rozdělení N(0,1).**

2)

Pro $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ a pro $\alpha = 0.05$ platí P[|Z| < u(0.975)] = 0.95,

odtud po úpravě dostaneme $P(\bar{X} - 1.96\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\dfrac{\sigma}{\sqrt{n}}) = 0.95$,

tedy **95 %-ní interval spolehlivosti pro střední hodnotu** $\boldsymbol{\mu}$ je

$$I = (\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}; \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}).$$

3)
Z konstrukce kritického oboru víme

$$0.05 = \alpha = \text{P(zam. } H_0\text{, když platí)} = P\left(\frac{|\bar{X}-\mu_0|}{\sigma}\sqrt{n} \geq u(0.975) = 1.96\right)$$

$$= 1 - P\left(\frac{|\bar{X}-\mu_0|}{\sigma}\sqrt{n} < 1.96\right) = 1 - P(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu_0 < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}})$$

$$= 1 - P(\mu_0 \in I).$$

Tedy jev $\dfrac{|\bar{X}-\mu_0|}{\sigma}\sqrt{n} \geq u(0.975) = 1.96$ je ekvivalentní jevu μ_0 neleží v I, a proto

μ_0 neleží v $(\bar{X} - 1.96\dfrac{\sigma}{\sqrt{n}}; \bar{X} + 1.96\dfrac{\sigma}{\sqrt{n}})$ => zam. H_0 na hladině 5 %.

(III) Pomocí p-hodnoty

p-hodnota je číslo *p* - nejmenší hladina, na které zamítáme H_0.

Ke zpracovávaným datům a vybranému testu zkoumané hypotézy ji zkonstruuje statistický software.

Rozhodovací pravidlo:
$p \leq \alpha$ => zam. H_0 ve prospěch H_1 na hladině α,
$p > \alpha$ => H_0 nelze zam. na základě našich dat.

Výhodou p-hodnoty je její univerzální použitelnost. Testovaná hypotéza se nemusí týkat jen neznámého parametru a jeho domnělé číselné hodnoty, ale můžeme například testovat, zda data pocházejí z normálního rozdělení aj. Rozhodovací pravidlo má stále stejný tvar, porovnáváme p-hodnotu se zvolenou hladinou testu.

Příklad: Práce s p-hodnotou.

a) Statistický software poskytl p-hodnotu ***p* = 0.02**.
Nejmenší hladina, na které zamítáme H_0, je 0.02, zamítáme tedy H_0 na hladinách od 2 % výše
=> zam. H_0 na hladině 5%.

b) Statistický software poskytl p-hodnotu ***p* = 0.08**.
Nejmenší hladina, na které zamítáme H_0, je 0.08, zamítáme tedy H_0 na hladinách od 8 % výše
=> na hladině 5 % H_0 ještě nezamítáme.

Poznámky k testování hypotéz

A)
Mezi hypotézou H_0 a alternativou H_1 je nesymetrie obsažená v rozhodovacím pravidle:
říkáme zam. H_0 ve prospěch H_1 na hladině α,
ale opačně H_0 nelze zam. na základě našich dat,
neměli bychom říkat, že H_0 přijímáme.

Proč? Abychom vyvrátili platnost H_0, stačí mít jediná data, která ji vyvracejí (naše data). Ale abychom H_0 s obecnou platností přijali, musela by se její platnost potvrdit na všech možných sadách dat, což není možné prokázat.

B)
Rozhodovací pravidlo závisí na zvolené alternativě. Proto bychom, není-li to z kontextu jasné, měli říkat zam. H_0 ve prospěch H_1 na hladině α.

Příklad: Test hypotézy o střední hodnotě.
Nechť jsou data pozorovanou realizací náhodného výběru $X_1,\ldots,X_n$ z rozdělení N(μ, σ^2).

1)
Testujeme H_0: $\mu = \mu_0$ proti H_1: $\mu \neq \mu_0$ …**oboustranná alternativa**.

Idea: $|\bar{X} - \mu_0|$ velké svědčí proti H_0 => potřebujeme najít dolní mez D pro $|\bar{X} - \mu_0|$ takovou, aby P(zam. H_0, když platí) = $P(|\bar{X} - \mu_0| > D) = \alpha$.

Kritický obor: $\frac{|\bar{X} - \mu_0|}{\sigma}\sqrt{n} > u(1\text{-}\alpha/2) = u(0.975) = 1.96$ => zam. H_0 na hladině 5%.

2)
Testujeme H_0: $\mu = \mu_0$ proti H_1: $\mu > \mu_0$ … **jednostranná alternativa**.

Idea: $\bar{X} - \mu_0$ velké svědčí proti H_0 => potřebujeme najít dolní mez D pro $\bar{X} - \mu_0$ takovou, aby P(zam. H_0, když platí) = $P(\bar{X} - \mu_0 > D) = \alpha$.

Kritický obor: $\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} > u(1\text{-}\alpha) = u(0.95) = 1.64$ => zam. H_0 na hladině 5%.

Kritické obory testu stejné nulové hypotézy proti oboustranné a jednostranné alternativě jsou různé!