Testování hypotéz na základě jednoho a dvou výběrů [1]

[1] Tyto materiály byly vytvořeny za pomoci grantu FRVŠ číslo 1145/2004.

Testování hypotéz

Pokud nás zajímá zda platí, či neplatí tvrzení o určitém parametru, např. o parametru Θ, pak takovéto tvrzení lze nazvat hypotézou, resp. statistickou hypotézou.

Například, chceme testovat hypotézu, že střední hodnota je nějaké konkrétní číslo, nebo patří do určitého intervalu. Pak θ je právě střední hodnota a θ_0 právě naše konkrétní hodnota. Statistickou hypotézu lze pak zapsat například ve tvaru

$$H_0 : \theta = \theta_0 \tag{1}$$

Takto formulovanou hypotézu nazveme testovanou hypotézou. Někdy se lze setkat s pojmenováním nulová hypotéza.

Proti testované hypotéze formulujeme hypotézu alternativní. Značí se symbolem H_A nebo H_1. Existují různé alternativní hypotézy. V podstatě však při testování hypotéz rozeznáváme pouze tři typy alternativních hypotéz. Ty lze koncipovat následujícím způsobem:

Pravostranná hypotéza	$H_A :$	$\Theta > \Theta_0$
Levostranná hypotéza	$H_A :$	$\Theta < \Theta_0$
Oboustranná hypotéza	$H_A :$	$\Theta \neq \Theta_0$.

Na tomto místě je nutné upozornit čtenáře na to, že je velmi důležité, jak budeme tyto hypotézy specifikovat. Je nutné se správně rozhodnout mezi těmito variantami:

$$H_0 : \Theta = \Theta_0 \quad \text{vs.} \quad H_A : \Theta \neq \Theta_0,$$

nebo

$$H_0 : \Theta \leq \Theta_0 \quad \text{vs.} \quad H_A : \Theta > \Theta_0$$

nebo

$$H_0 : \Theta \geq \Theta_0 \quad \text{vs.} \quad H_A : \Theta < \Theta_0 \ .$$

Testové kritérium

Pro rozhodnutí o tom, která z výše formulovaných hypotéz je pravdivá, tj. zda bude platit H_0 nebo naopak H_A, rozhodujeme za pomoci tzv. testové statistiky T. To je jistá námi zvolená funkce, která závisí na našem výběru. Příkladem může být testová statistika pro odhad střední hodnoty - aritmetický průměr. Tedy funkce závisející na našem pozorování. Testovací statistiku pak zpravidla nazýváme testovým kritériem.

Statistika T je náhodnou veličinou nabývající určitého oboru hodnot, resp. hodnot z určité podmnožiny množiny reálných čísel.

Na definovaném oboru hodnot testové statistiky T lze vymezit jistým způsobem dvě podmnožiny. První z nich, označíme ji symbolem K, nazveme kritickým oborem. Druhou podmnožinu, označíme ji H, nazveme oborem přijetí.

Heuristickým vysvětlením těchto dvou množin může být následující představa. Chceme rozhodnout, zda se naše rozdělení chová podle naší hypotézy H_0, nebo už spíše dle alternativy H_A.

Například máme šestistěnou kostku a hodláme rozhodnout, zda je poctivá, tj. všechna čísla padají se stejnou pravděpodobností, nebo zda je falešná a některá čísla padají častěji. Test provedeme tak, že si řekneme, že hodíme stokrát kostkou a spočteme si, kolikrát které číslo padlo.

Víme, že se může stát, že i když je kostka poctivá, mohou nám padnout například jen samé pětky a šestky. Ale pravděpodobnost, že se tak stane, je u poctivé kostky mizivá. V takovém případě je daleko pravděpodobnější, že je kostka falešná.

Naším cílem je právě dopředu určit množinu všech výsledků, které už jsou "daleko" od naší testované hypotézy. Pravděpodobnost této množiny má být právě naše α. Pravděpodobnost (za platnosti H_0), že nastanou tyto nebo všechny "horší" výsledky je velmi malá. Jak malá, to je právě naše hladina významnosti testu.

A pak už jen házíme a pokud nastal výsledek, který je v této naší množině, pak zamítneme hypotézu.

Pokud jde o samotné testování hypotézy, pak to spočívá v aplikaci jednoduchého rozhodovacího pravidla:

Leží-li hodnota testového kritéria T v kritickém oboru tj., $T \in K$, zamítáme nulovou hypotézu ve prospěch hypotézy alternativní.

Naopak, neleží-li hodnota testového kritéria v kritickém oboru, pak testovanou hypotézu nezamítáme a tvrdíme, že se nepodařilo zamítnout nulovou hypotézu na předem zvolené hladině významnosti α.

Chyby spojené s testováním hypotéz

Při testování hypotéz se můžeme dopustit v podstatě chyb dvojího typu. Ve statistické teorii je nazýváme chybami prvního a druhého druhu. Pojednejme o nich blíže.

Chyby prvního druhu se dopustíme tehdy, zamítneme-li testovanou hypotézu, přestože platí. Pokud bychom tedy chtěli určit pravděpodobnost vzniku chyby I. druhu, platilo by následující:

$$P(\text{chyby I.}) = P(\text{přijmu } H_A|H_0) = P(T \in K|\text{platí } H_0). \tag{2}$$

Pravděpodobnost vzniku chyby I. typu chceme zpravidla jistým způsobem omezit. Ve většině případů požadujeme, aby tato pravděpodobnost nepřekročila určitou, předem danou hodnotu α.

Hodnotu α nazýváme hladinou významnosti. Nejčastější volbou hodnoty α pro testování hypotéz je $\alpha = 0,05$ či $\alpha = 0,01$. V takovém případě připouštíme existenci vzniku chyby I. druhu s pravděpodobností 0,05 resp. 0,01. Kritický obor je pak konstruován tak, že platí:

$$P(\text{chyby I.}) = P(T \in K|\text{platí } H_0) = \alpha \ .$$

Chyby druhého druhu se dopustíme tehdy, nezamítneme-li hypotézu H_0, přestože tato hypotéza ve skutečnosti neplatí a my bychom se měli přiklonit k alternativní hypotéze. Pravděpodobnost toho, že se dopustíme chyby II. druhu lze vyjádřit následujícím způsobem:

$$P(\text{chyby II.}) = P(\text{nezamítnu } H_0|H_A) = P(T \notin K|H_A) = \beta. \tag{3}$$

Většinou se však zajímáme spíše o doplněk k této pravděpodobnosti. Jinými slovy, zajímáme se o pravděpodobnost toho, že skutečně zamítneme nulovou hypotézu, pokud skutečně neplatí. Určujeme tedy pravděpodobnost toho, že se této chyby nedopustíme. Symbolicky lze hledanou pravděpodobnost definovat následovně:

$$P(\text{přijmu } H_A|H_A) = P(T \in K|H_A) = 1 - \beta. \tag{4}$$

Tento doplněk k pravděpodobnosti chyby II. typu, tj. hodnotu $1 - \beta$, zpravidla nazýváme silou testu.

Ve statistické teorii se lze setkat s celou řadou různých testů, které slouží k vyslovení závěrů o hodnotách odhadovaných parametrů či o tvarech studovaných rozdělení. Tato problematika však přesahuje rámec našeho kurzu, proto se omezíme pouze na některé z nich. Další testy lze nalézt v odborné statistické literatuře. Jejich prostudování přenecháváme laskavému čtenáři.

Druhy testů

Z hlediska toho, jaké předpoklady činíme o rozdělení sledovaného statistického znaku, lze rozlišit dvě třídy testů:

Parametrické testy: Jsou testy založené na znalosti charakteru rozdělení sledovaného statistického znaku, tj. předpokládáme např. že víme, že naše pozorování pochází například z normálního, binomického, apod. rozdělení, ale neznáme hodnoty jednoho či více parametrů. Parametrickými testy se pak testujeme předpoklady o těchto neznámých parametrech (může jít například o střední hodnotu či rozptyl). V převážné většině jde o početně náročnější, ale silné testy.

Neparametrické testy: Jsou takové testy, které nevyžadují znalost předpokladů o charakteru rozdělení náhodných veličin. Neparametrické, se nazývají proto, že se netýkají parametrů rozdělení. Tyto testy mají obecně menší sílu ve srovnání s parametrickými testy. Jejich výhodu je však jejich univerzálnost, neboť lze tyto testy použít jak pro kvantitativní, tak i kvalitativní znaky. Po výpočetní stránce jsou jednoduché.

Testy hypotéz o parametru μ při neznámém σ^2

Předpokládejme, že máme k dispozici výběr $\mathbf{x} = [x_1, x_2, \cdots, x_n]'$, pocházející z normálního rozdělení, jehož parametry μ a σ^2 neznáme. Tato situace bývá v biologické a zejména v ekonomické praxi nejčastější. Dále předpokládejme, že se budeme snažit učinit závěr o velikosti střední hodnoty μ v základním výběru.

Kritický obor je v případě testu hypotéz o parametru μ při neznámém σ^2 založen na testovém kritériu

$$t = \frac{\bar{x} - \mu_0}{s}\sqrt{n}\ . \tag{5}$$

Uvědomte si, že t je náhodnou veličinou. Tato náhodná veličina sleduje za platnosti nulové hypotézy Studentovo rozdělení s $n-1$ stupni volnosti. Symbolicky lze tedy psát, že $t \sim t(n-1)$. Testujeme-li hypotézu

$$H_0 : \mu \leq \mu_0 \ \text{ proti } \ H_A : \mu > \mu_0\ ,$$

je kritickým oborem K následující množina:

$$\{t; t \geq t_{1-\alpha}(n-1)\}\ . \tag{6}$$

Protože, pokud by platila alternativa, pak by složky vektoru pozorování x_i nabývaly spíše vyšších hodnot, a tedy i $\bar{x}$ určený na základě těchto hodnot by byl vyšší. My tedy hledáme tak vysokou hodnotu, aby, pokud platí nulová hypotéza, $\bar{x}$ nabýval této nebo vyšší hodnoty jen s pravděpodobností menší než α. Pokud se tedy stane, že $\bar{x}$ dosáhne této nebo vyšší hodnoty, pak hypotézu zamítneme na hladině α.

Rozmyslete si, že právě toto nám říká nerovnost

$$t \geq t_{1-\alpha}(n-1) \quad ! \tag{7}$$

Naopak testujeme-li hypotézu

$$H_0 : \mu \geq \mu_0 \ \text{ proti } \ H_A : \mu < \mu_0\ ,$$

je kritickým oborem K množina

$$\{t; t \leq -t_{1-\alpha}(n-1)\}\ . \tag{8}$$

S ohledem na symetričnost rozdělení je tento kritický obor totožný s množinou $\{t; t \leq t_\alpha(n-1)\}$. V případě, že nás zajímá pouze oboustranná alternativa, tj. předpokládáme následující hypotézy:

$$H_0 : \mu = \mu_0 \ \text{ proti } \ H_A : \mu \neq \mu_0\ ,$$

pak je kritickým oborem množina

$$\{t; |t| \geq t_{1-\frac{\alpha}{2}}(n-1)\}\ . \tag{9}$$

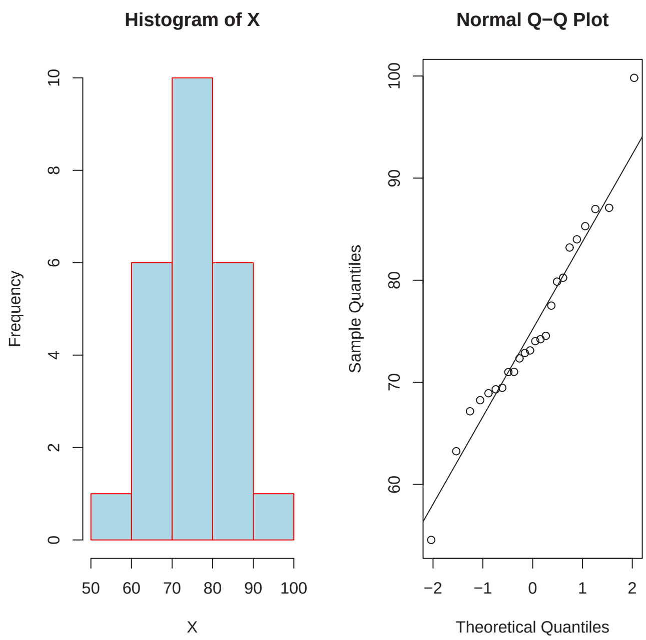
Celý postup demonstrujme na následujícím příkladě. V populárním časopise bylo uvedeno, že česká populace žen vlivem změn životního stylu a špatných stravovacích návyků tloustne. Zatímco před pěti lety byla průměrná hmotnost českých žen 72,5 Kg dnes jsou ženy podstatně obéznější. Má autor tohoto článku pravdu nebo ženám křivdí?

Přistupme k tomuto problému statisticky. Předpokládejme, že jsme pořídili náhodný výběr z populace českých žen o rozsahu $n = 24$.

```
[1] 67.16 86.97 84.00 85.29 74.55 68.93 74.22 72.34 87.09 77.51
[11] 63.25 71.02 99.82 54.56 73.12 74.03 71.00 69.46 79.86 69.31
[21] 68.25 72.87 80.24 83.20
```

Pro posouzení normality dat využijeme histogram a tzv. Q-Q plot[1] viz obrázek 1. Z těchto grafů lze vyvodit předběžný závěr, že testovaný soubor může pocházet z normálního rozdělení. Závěr učiněný na základě těchto grafů je nutné brát pouze jako informativní závěr.

Obrázek 1: Q-Q graf pro vizuální posouzení normality dat



Předepišme tedy testovanou a alternativní hypotézu:

$$H_0 : \mu \leq 72,5 \qquad H_A : \mu > 72,5$$

Průměrná hodnota hmotnosti žen ve výběrovém souboru je

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 75,33542\ .$$

[1] Velmi hrubě řečeno, jsou-li body v tomto grafu přibližně na přímce lze se domnívat, že pozorovaná data pocházejí z normálního rozdělení.

Směrodatná odchylka pak

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})} = \sqrt{88,07907} = 9,385045\ .$$

Vyjádřeme hodnotu testové statistiky t.

$$t = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} = \frac{75,33542 - 72,5}{9,385045} \cdot \sqrt{(24)} = 1,480083$$

Dále předpokládejme, že budeme chtít interpretovat naše výsledky s 95% spolehlivostí. Hladinu α specifikujeme tedy hodnotu $0,05$. Zbývá určit kritický obor K, ten je definován jako množina hodnot t pro které platí:

$$t \geq t_{1-\alpha}(n-1) = t_{0,95}(23) = 1,713872\ .$$

Je tedy zřejmé, že testová statistika t neleží v kritickém oboru K. Z tohoto důvodu nelze zamítnout testovanou hypotézu $H_0 : \mu \leq 72,5$. Závěrem lze tedy říci, že se nám s 95% spolehlivostí na základě pozorovaných dat nepodařilo prokázat, že hmotnost žen vzrostla v porovnání s průměrnou hmotností žen před pěti lety. Autor článku ženám zřejmě křivdí.

P-value, aneb dosažená hladina významnosti

Pokud budeme pracovat se statistickým software, tak se téměř s jistotou setkáme s hodnotou nazývanou p-value. Ta udává pravděpodobnost s jakou by náhodná veličina T, za platnosti nulové hypotézy, nabyla alespoň takové hodnoty, jakou má testová statistika t, vypočtená na základě pozorovaných dat. Symbolicky pak pro spojité náhodné veličiny:

$$P(T \geq t) = p\text{-value} \tag{10}$$

Při testování hypotéz lze uplatnit následující pravidlo:

Pokud je p-value menší než α, zamítáme nulovou hypotézu ve prospěch hypotézy alternativní.

Pro předešlý příklad činí hodnota p-value $0,07620986$. Je tedy větší než námi specifikovaná hladina $\alpha = 0.05$, a proto nelze zamítnout testovanou hypotézu $H_0 : \mu \leq 72,5$.

Pokud bychom chtěli určit dosaženou hladinu významnosti pro levostrannou alternativu, tj. pro $H_A : \mu < 72,5$, museli bychom postupovat následujícím způsobem:

$$1 - P(T \geq t) = 1 - p\text{-value} = 1 - 0,07620986 = 0,9237901$$

V případě, že bychom se zajímali o dosaženou hladinu významnosti pro oboustranný test, tj. $H_A : \mu \neq 72,5$, stačí vynásobit hodnotu p-value získanou pro pravostrannou alternativu dvěma. Pokud postupujeme přes p-value získané pomocí levostranné alternativy je nutno před vynásobením odečíst od p-value jedničku.

Pozn.: U některých statistických balíků si můžeme zvolit pro jaký typ testu (jakou alternativu) chceme p-value spočítat. Komerční statistický balík STATISTICA však tuto vlastnost nemá. Hodnota p-value je počítána vždy pro oboustranný test. V případě jednostranných testů musíme skutečné p-value dopočítat.

Testy hypotéz o parametru σ^2 normálního rozdělení

Testy hypotézy o rozptylu normálního rozdělení lze provádět za podmínky, že střední hodnotu μ známe, což je dosti řídký případ, nebo za podmínky že μ neznáme. Všimněme si častějšího případu. V takovém případě jsou kritické obory pro jednotlivé hypotézy založeny na testovém kritériu

$$\chi^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} \ , \tag{11}$$

které má, za předpokladu platnosti nulové hypotézy χ^2 rozdělení s $n-1$ stupni volnosti.

Při testu hypotézy $H_0 : \sigma^2 \leq \sigma_0^2$ proti pravostranné alternativě $H_A : \sigma^2 > \sigma^2$ využíváme kritický obor K, který lze vymezit jako

$$\{\chi^2; \chi^2 \geq \chi^2_{1-\alpha}(n-1)\} \ . \tag{12}$$

Naopak při testu hypotézy $H_0 : \sigma^2 \geq \sigma_0^2$ proti levostranné alternativě $H_A : \sigma^2 < \sigma^2$ využíváme kritický obor K, který je specifikován jako

$$\{\chi^2; \chi^2 \leq \chi^2_{\alpha}(n-1)\} \ . \tag{13}$$

Konečně, při testu hypotézy $H_0 : \sigma^2 = \sigma_0^2$ proti oboustranné alternativě $H_A : \sigma^2 \neq \sigma^2$, využíváme kritický obor K, který je specifikován jako sjednocení dvou podmnožin, tedy:

$$\{\chi^2; \chi^2 \leq \chi^2_{\frac{\alpha}{2}}(n-1) \cup \chi^2 \geq \chi^2_{1-\frac{\alpha}{2}}(n-1)\} \ . \tag{14}$$

Testy hypotéz o shodě středních hodnot dvou normálních rozdělení při nezávislých výběrech a neznámých rozptylech

Testujeme-li hypotézu o shodě středních hodnot dvou normálních rozdělení, jejichž rozptyly neznáme, je způsob testu závislý na tom, zda můžeme předpokládat, že rozptyly jsou stejné, tj. je nutno nejprve provést test hypotézy $H_0 : \sigma_1^2 = \sigma_2^2$.

Testovacím kritériem je statistika

$$F = \frac{s_1^2}{s_2^2} \ , \tag{15}$$

která má za platnosti nulové hypotézy rozdělení $F(m-1, n-1)$. Jednotlivé kritické obory jsou pak vymezeny následovně:

Tabulka 1: Testy hypotézy o shodě dvou rozptylů

H_A	K
$\sigma_1^2 > \sigma_2^2$	$\{F : F \geq F_{1-\alpha}(m-1, n-1)\}$
$\sigma_1^2 < \sigma_2^2$	$\{F : F \leq F_{\alpha}(m-1, n-1)\}$
$\sigma_1^2 \neq \sigma_2^2$	$\{F : F \leq F_{\frac{\alpha}{2}}(m-1, n-1) \ \cup F \geq F_{1-\frac{\alpha}{2}}(m-1, n-1)\}$

Varianta a) shodné rozptyly - homoskedasticita

Pokud tedy předpokládáme, na základě předchozího testu, že jsou oba rozptyly shodné, lze k testování nulové hypotézy o shodě středních hodnot využít testovací statistiky

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^*}\sqrt{\frac{mn}{m+n}}\ , \tag{16}$$

kde

$$s^* = \sqrt{\frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}}\ . \tag{17}$$

Dle formulace alternativní hypotézy pak používáme kritické obory uvedené v tabulce 2:

Tabulka 2: Testy hypotézy o shodě středních hodnot

H_A	K
$\mu_1 > \mu_2$	$\{t : t \geq t_{1-\alpha}(m+n-2)\}$
$\mu_1 < \mu_2$	$\{t : t \leq -t_{1-\alpha}(m+n-2)\}$
$\mu_1 \neq \mu_2$	$\{t : \lvert t\rvert \geq t_{1-\frac{\alpha}{2}}(m+n-2)\}$

Varianta b) neshodné rozptyly - heteroskedasticita

V praxi se setkáváme i s případem, kdy rozptyly neznáme a navíc nemůžeme předpokládat, že jsou shodné. Neboli testem na shodu rozptylů jsme zamítli nulovou hypotézu $\sigma_1^2 = \sigma_2^2$. V takovém případě jsme nuceni použít poněkud jiné testové statistiky:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} \tag{18}$$

Tato statistika sleduje, za předpokladu platnosti nulové hypotézy, rozdělení $t(f)$, kde

$$f = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{1}{m-1}\left(\frac{s_1^2}{m}\right)^2 + \frac{1}{n-1}\left(\frac{s_2^2}{n}\right)^2} \ . \tag{19}$$

Dle alternativní hypotézy pak rozeznáváme následující kritické obory viz tabulka 3.

Tabulka 3: Testy hypotéz o shodě středních hodnot

H_A	K
$\mu_1 > \mu_2$	$\{t : t \geq t_{1-\alpha}(f)\}$
$\mu_1 < \mu_2$	$\{t : t \leq -t_{1-\alpha}(f)\}$
$\mu_1 \neq \mu_2$	$\{t : \|t\| \geq t_{1-\frac{\alpha}{2}}(f)\}$

Testy hypotéz o shodě středních hodnot při závislých výběrech

Jde o takzvaný Studentův párový t-test. V tomto případě lze nulovou hypotézu tvrdící, že náhodné výběry $\mathbf{x}_1$ a $\mathbf{x}_2$ pocházejí z rozdělení se stejnými středními hodnotami, lze převést na hypotézu H_0, tvrdící, že náhodný výběr $\mathbf{d} = [d_1, d_2, \cdots, d_n]$, kde pro d_i platí $d_i = d_{1i} - d_{2i}$, pochází z rozdělení jehož střední hodnota se rovná nule. Testujeme hypotézu $H_A : \mu = \mu_1 - \mu_2 = 0$. Testovým kritériem je statistika

$$t = \frac{\bar{d}}{s_d}\sqrt{n} \ . \tag{20}$$

Kritické obory pro jednotlivé alternativní hypotézy udává tabulka 4.

Tabulka 4: Test hypotézy o shodě středních hodnot – závislé soubory

H_A	K
$\mu = \mu_1 - \mu_2 > 0$	$\{t : t \geq t_{1-\alpha}(n-1)\}$
$\mu = \mu_1 - \mu_2 < 0$	$\{t : t \leq -t_{1-\alpha}(n-1)\}$
$\mu = \mu_1 - \mu_2 \neq 0$	$\{t : \|t\| \geq t_{1-\frac{\alpha}{2}}(n-1)\}$