

Vyhodnocování výsledků testování hypotéz na základě „ p -value“

1.1 Klasická metodologie vyhodnocování testování hypotéz

Uvědomme si na, jakém principu je založeno testování hypotéz: Za platnosti *nulové hypotézy* a splnění předpokladů pro použití testu tvrdíme, že bude *testové kritérium* – jakžto náhodná veličina – sledovat určité rozdělení. Testovým kritériem rozumíme hodnotu, vypočtenou na základě náhodného výběru z dat, jež testujeme. Proti nulové hypotéze (H_0) obvykle stojí *hypotéza alternativní* (H_A), která svým způsobem popírá hypotézu nulovou. Ty hodnoty testového kritéria, které jsou více méně málo pravděpodobné za platnosti nulové hypotézy, pak „obětujeme“ a nulovou hypotézu zamítneme, i když by mohla platit, a tvrdíme, že se nám podařilo prokázat hypotézu alternativní. *Chyba prvního druhu*, která je vlastně touto obětí (nulová hypotéza platí a my ji zamítáme) je kvantifikována¹ prostřednictvím *hladiny významnosti* α , obvykle volené 0,1, 0,05, 0,01 či 0,001. Obětované hodnoty testového kritéria vymezuje v *kritických oborech* pomocí *kritických hodnot*. Kritické hodnoty jsou vlastně kvantily rozdělení, jež testové kritérium sleduje, a odpovídají hladině významnosti. Kritické obory jsou navíc konstruovány tak, aby hodnoty v nich obsažené byly přijatelné pro alternativní hypotézu a pro nulovou málo pravděpodobné.

Pokud tedy hodnota testového kritéria spadne do kritického oboru, zamítáme nulovou hypotézu ve prospěch alternativní na hladině významnosti α . Čím nižší je hladina významnosti, tím obtížnější je zamítnutí nulové hypotézy (kritické obory jsou menší), ale test zároveň ztrácí svou *sílu*². Srovnání tedy probíhá v hodnotách náhodné veličiny a hladina významnosti je pro tento účel převáděna na kritické hodnoty. Konstrukce kritických oborů však není cílem tohoto textu.

Nešlo by tedy naopak modifikovat hodnotu testového kritéria a tu pak porovnávat přímo s hladinou významnosti? Srovnání by tedy probíhalo na úrovni pravděpodobnosti. Samozřejmě to jde. Modifikovaná hodnota testového kritéria bývá obvykle označována *p-value* a její možná interpretace a úskalí budou vyložena dále.

1.2 Interpretace významu hodnoty „ p -value“

Má-li být hodnota *p-value* jasně definována, potřebujeme vedle testu, který provádíme, znát hodnotu testového kritéria a alternativní hypotézu.

Podstata hodnoty *p-value* a hladiny významnosti α je v podstatě stejná (jedná se o pravděpodobnosti, vypovídající cosi o nulové hypotéze). Hladina významnosti je předpokládaná pravděpodobnost zamítnutí nulové hypotézy za předpokladu, že byla správná (pravděpodobnost chyby prvního druhu) a určujeme ji vždy před výpočtem testového kritéria (tj. před testem). *p-value* je taková nejnižší možná hladina významnosti – určená

¹Je vyjádřena její pravděpodobnost: $P(H_A|H_0) = \alpha$.

²Síla testu je pravděpodobnost, že je testem zamítnutá nulová hypotéza, když skutečně neplatí: $P(H_A|H_A) = 1 - \beta$.

na základě hodnoty testového kritéria (tj. na základě výsledku kvantifikace výběru) –, při které lze ještě zamítnout nulovou hypotézu.

p -value nám poskytuje obecněji více informací o výsledku statistického testování než pouhé zamítnutí nebo nezamítnutí nulové hypotézy. Předpokládejme, že p -value vyjde rovna 0,05. Z toho lze usoudit, že nulovou hypotézu lze zamítnout například na hladině významnosti $\alpha = 0,1$, ale již ne na hladině významnosti $\alpha = 0,01$ a $\alpha = 0,001$. Nejnižší možnou hladinu významnosti, na které ještě můžeme nulovou hypotézu zamítnout je právě $\alpha = 0,05$. Ještě jednou bude-li p -value rovna 0,03, potom můžeme nulovou hypotézu zamítnout na hladině významnosti $\alpha = 0,1$ nebo $\alpha = 0,05$. Na hladině významnosti nižší než $\alpha = 0,03$ nulovou hypotézu nezamítáme (tedy např. na hladině významnosti $\alpha = 0,01$). Čím nižší vyjde p -value, tím více jsme přesvědčeni, že nulová hypotéza není správná a je třeba jí zamítnout.

Nejjednodušší způsob, jak rozhodovat o výsledku testu spočívá v porovnání p -value (vypočte počítač – např. program STATISTICA) a hladiny významnosti α (určíme před testem sami). Platí následující pravidlo:

- a) Zamítni nulovou hypotézu H_0 , když $p\text{-value} \leq \alpha$.
- b) Nezamítej nulovou hypotézu H_0 , když $p\text{-value} > \alpha$.

Tento způsob testování je nejrychlejší a nejpohodlnější, pokud máme k dispozici počítač. Většina statistických programů počítá p -value, ať už je v softwarovém prostředí označováno symboly p , *Significance level* apod.

1.3 Není všechno zlato, co ...

Vše co bylo zmíněno výše platí, musíme však připustit, že i zde existuje **ALE**. Tím ALE je způsob výpočtu p -value pomocí statistického softwaru. Jak už jste jistě zjistili, existují z určitého hlediska dva typy alternativních hypotéz – *jednostranná* a *oboustranná* alternativní hypotéza. Některé testy hypotéz jsou používány jen pro jednostranné resp. oboustranné alternativní hypotézy. Je-li tomu tak, pak statistický software dává jako výsledek p -value, kterou můžeme interpretovat tak, jak bylo zmíněno výše.

Pro testy hypotéz, kde se běžně používají jak pro oboustranné tak jednostranné varianty alternativních hypotéz, vykalkuluje program hodnotu p -value pro oboustrannou hypotézu. To se týká především takzvaných t -testů. Software vypočte pravděpodobnost s jakou by náhodná veličina sledující teoretické rozdělení nabyla hodnotu větší, než je absolutní hodnota z hodnoty testového kritéria, a tuto hodnotu vynásobí dvěma. Vynásobená hodnota je následně uživateli předložena jako p -value. Proto je výše popsán způsob interpretace p -value (ve většině programech) bez modifikace možný jen pro takzvaný oboustranný test, tedy když dvojice hypotéz H_0 a H_A má např. následující matematické předpisy³:

$$H_0 : \mu = \mu_0, \quad H_A : \mu \neq \mu_0. \quad (1)$$

³Jedna se o jednovýběrový t -test.

Lze tedy tuto hodnotu p -value využít k rozhodnutí o výsledku testování i pro jednostranné testy? I v tomto případě lze naznačený problém vyřešit. Dokonce několikerým způsobem.

1.4 Modifikace oboustranného p -value

Vzhledem k tomu, že se tento text elegantně vyhýbá výpočetnímu pozadí testů, dokonce jsme v něm ani přesně nespecifikovali, jak se tvoří kritické obory, existuje jediná možnost, jak vyžrát na oboustranné p -value, aniž by jsme tento trend porušili. Dříve než tak učiníme, je nutnou dodat, že hodnotu p -value – příslušnou jakékoliv alternativní hypotéze – lze pomocí vhodného software vypočítat manuálně⁴, nebo na základě znalosti výpočtu testového kritéria a předpisu pro kritický obor odvodit pravidlo jak z oboustranného p -value získat správné „jednostranné“ p -value. Úmyslně na tomto místě není napsáno jednoduché pravidlo, protože – byť je triviální – musíme k výpočtu „jednostranného“ p -value přistupovat s rozmyslem. Zde toto pravidlo odvozovat nebudeme a omezíme se na popsání tohoto pravidla modifikace p -value pomocí známých pojmů.

Označme alternativní hypotézu $H_A : \theta_1 < \theta_2$ jako levostrannou a $H_A : \theta_1 > \theta_2$ jako pravostrannou. Přičemž θ_1 a θ_2 označují testované parametry (například střední hodnoty hodnoty u t -testů). Záludností pravidla, které popisujeme, je skutečnost, že modifikace p -value vypočteného softwarem pro oboustrannou alternativní hypotézu nezáleží jen na námi testované alternativní hypotéze, ale také na datech – tedy hodnotě testového kritéria. Pravidlo v sobě tedy kombinuje jak typ alternativní hypotézy tak hodnotu testového kritéria.

Pokud budeme uvažovat **levostrannou** alternativní hypotézu, pak oboustranné p -value **dělíme dvěma** v případě **záporné** hodnoty testového kritéria, respektive **dělíme dvěma a odečítáme od jedné** v případě **kladné** hodnoty testového kritéria.

Naopak v případě **pravostranné** alternativní hypotézy, oboustranné p -value **dělíme dvěma** v případě **kladné** hodnoty testového kritéria a respektive **dělíme dvěma a odečítáme od jedné** v případě **záporné** hodnoty testového kritéria.

Přehledně toto pravidlo shrňme v tabulce 1, v níž je naznačena i triviální situace pro oboustrannou alternativní hypotézu.

Tabulka 1: Modifikace oboustranného p -value. Modifikace oboustranného p -value, p -value je v tabulce pro přehlednost značenou pouze p

		test. kritérium < 0	test. kritérium > 0
$H_A : \theta_1 < \theta_2$	$t_{\text{hyp}} < 0$	$p/2$	$1 - p/2$
$H_A : \theta_1 > \theta_2$	$t_{\text{hyp}} > 0$	$1 - p/2$	$p/2$
$H_A : \theta_1 \neq \theta_2$	$t_{\text{hyp}} \neq 0$	p	p

⁴V programu STATISTICA lze hodnotu p -value dopočítat například pomocí pravděpodobnostního kalkulátoru rozdělení.