

Co potřebuje klinik vědět o (bio)statistice?

Testování hypotéz

Pravoslav Stránský

Ústav lékařské biofyziky a Oddělení výpočetní techniky, Univerzita Karlova v Praze, Lékařská fakulta v Hradci Králové, Česká republika / Department of Medical Biophysics and Division of Computer Science, Charles University in Prague, School of Medicine at Hradec Králové, Czech Republic

Stránský P. Co potřebuje klinik vědět o (bio)statistice? Testování hypotéz. *Folia Gastroenterol Hepatol* 2005; 3 (2): 74 – 76.

Souhrn. Je vysvětlen rozdíl mezi hypotézou výzkumnou a statistickou a vztah hladiny významnosti a pravděpodobnosti používané v testování hypotéz.

Klíčová slova: testování hypotéz, vědecká a statistická hypotéza, hladina významnosti, p -hodnota, chyba I. a II. druhu

Stránský P. What should a clinician know about (bio)statistics? Hypotheses testing. *Folia Gastroenterol Hepatol* 2005; 3 (2): 74 – 76.

Abstract. A difference between scientific and statistical hypothesis is explained. Meaning of significance level and probability value used in hypotheses testing are described.

Key words: hypotheses testing, scientific and statistical hypothesis, significance level, p -value, type I and II error

Hypotéza a její testování je ústředním pojmem induktivní, inferenční statistiky. Správné použití správného testu umožňuje kvantifikovat pravděpodobnost, s jakou můžeme výsledky zjištěné z pozorování výběru zobecnit na populaci. Každý výzkumný projekt by měl začít otázkou, na kterou bychom chtěli odpovědět s určitou jistotou tak, abychom byli přesvědčeni, že závěr, který předkládáme, je správný a v rozumných mezích nezpochybnitelný.

Touto první otázkou může například být: Jsou výsledky chirurgické léčby Parkinsonovy choroby jiné než výsledky léčby medikamentózní? Je nový lék účinnější než lék starý, má méně vedlejších účinků? Existuje vztah mezi dávkou léku a odezvou organismu? Je výsledek určitého vyšetření u nemocných s danou chorobou jiný než u zdravých jedinců? Všechny tyto otázky jsou vlastně formulací **vědecké hypotézy**, na kterou chceme svým výzkumem odpovědět. Statistika dává návod, jak při hledání odpovědi lze rámec rozumných pochybností vyjádřit pomocí pravděpodobnosti.

Určitým psychologickým problémem statistického

testování je skutečnost, že jeho základem je formulace hypotézy, která je popřením toho, co chceme prokázat. Při statistickém testování srovnáváme zjištěnou statistickou charakteristiku s nějakou hodnotou teoretickou (populační) nebo dvě či více charakteristik zjištěných v průběhu řešení výzkumného problému. Ve všech případech předpokládáme, že se srovnávané hodnoty od sebe neliší, tedy že jejich rozdíl je roven nule. To je důvodem, proč se tato základní statistická hypotéza nazývá **nulová hypotéza** (*null hypothesis*), H_0 .

Příslušný statistický test umožňuje říci, zda na předem zvolené hladině významnosti můžeme nulovou hypotézu zamítnout a přijmout **hypotézu alternativní** (*alternative hypothesis*), H_A , resp. H_1 . Alternativní hypotéza může být **oboustranná** (*two-sided*) a pak je formulována tak, že srovnávané hodnoty jsou rozdílné, nebo může být **jednostranná** (*one-sided*) a její formulace zní, že jedna srovnávaná charakteristika je menší (větší) než druhá. Někteří statistici mají vůči používání jednostranné alternativní hypotézy výhrady. Nicméně v medicíně za předpokladu, že ji formuluje-

me před začátkem výzkumného projektu, je její použití oprávněné. Budeme-li zkoumat účinek antihypertensiva, budeme a priori chtít, aby lék hodnotu krevního tlaku snižoval a ne pouze měnil.

V naprosté většině případů jsou vědecké hypotézy formulovány tak, že úspěšný experiment předpokládá, že nulovou hypotézu bude možno zamítnout a bude možné přijmout hypotézu alternativní. Existují však i výjimky. Budeme-li v okolí jaderné elektrárny provádět průzkum zdravotního stavu obyvatelstva, který bude financován vlastníkem elektrárny, cílem zřejmě bude prokázat, že se parametry charakterizující tento stav neliší od stejných parametrů obyvatelstva v jiné lokalitě.

Tím se dostáváme ke kvantifikaci rozumných pochybností, se kterými můžeme svoje tvrzení považovat za správná. Nejprve zvolíme **hladinu významnosti** (*significance level*), na které jsme ochotni nulovou hypotézu zamítnout. Ta se nazývá *p* hodnota (*p-value*) a v klinickém výzkumu to je obvykle 0,05 (5 %) nebo 0,01 (1 %). Znamená to, že provedeme-li 100 srovnání statistických charakteristik, které jsou shodné, např. dvou výběrových průměrů pomocí nepárového *t*-testu, pak v pěti případech bude výsledek testu pro zamítnutí nulové hypotézy, i když je ve skutečnosti platná. Takovýto výsledek se nazývá **chyba I. druhu** (*type I error*) a *p* je pravděpodobnost, že se chyby I. druhu dopustíme, a označuje se α .

Zamítneme-li nulovou hypotézu, přijímáme hypotézu alternativní. Takovýto výsledek testování označujeme jako **statisticky významný** rozdíl srovnávaných charakteristik, v uvedeném příkladu statisticky významný rozdíl mezi výběrovými průměry. Častá chyba v interpretaci přijetí alternativní hypotézy spočívá v tom, že se tvrdí, že $1-\alpha$ je pravděpodobnost toho, že alternativní hypotéza je správná. Ve skutečnosti je $1-\alpha$ pravděpodobnost toho, že v $(1-\alpha)\cdot 100$ procentech případů bude pozorovaný rozdíl charakteristik menší než námi zjištěný (a v $\alpha\cdot 100$ procentech stejný nebo větší).

Opakem zamítnutí platné nulové hypotézy je její přijetí, i když je neplatná. Takovýto výsledek testu se nazývá **chyba II. druhu** (*type II error*). Čtyři možné výsledky testování hypotézy jsou v této tabulce:

Výsledek testu	Skutečný stav	
	H_0 je platná	H_0 je neplatná
Přijmi H_0	správně	chyba II. druhu
		$p = \beta$
Zamítni H_0	chyba I. druhu	správně
	$p = \alpha$	

Mezi oběma chybami existuje spojitost spočívající v tom, že při snižování α (což je žádoucí) vzrůstá β , (což je nežádoucí). Zmenšením hodnoty α snižujeme pravděpodobnost chyby I. druhu a toho, že prohlásíme výsledek za statisticky významný, i když tomu tak ve skutečnosti není. Nicméně však vzrůstá pravděpodobnost, že neodhalíme významný rozdíl v případě nesprávné nulové hypotézy. Jinými slovy zmenšením α vzrůstá hodnota β , a je třeba najít rozumný kompromis. $1-\beta$, je tzv. **síla testu** (*power*) a je mírou jeho robustnosti, což kromě jiného znamená, jak je test citlivý na splnění předpokladů (data nejsou spojitá, nemají Gaussovo rozložení apod.). O síle testu blíže v (1).

Pokud jde o vlastní rozhodnutí v příslušném testu, tak základem je srovnání tzv. testového kritéria s kritickou hodnotou teoretického rozdělení nějaké náhodné veličiny. Množina hodnot teoretického rozdělení, které jsou menší než hodnota kritická, se nazývá obor přijetí (nulové hypotézy). Množina hodnot ležících mimo tento obor se označuje jako obor zamítnutí. Parametrické testy předpokládají, že v populaci, ze které jsou subjekty v projektu vybrány, má zkoumaná proměnná Gaussovo rozdělení. Nejčastěji používané teoretické rozdělení, jehož kritická hodnota se k testování charakteristiky v této skupině testů používá, je *t*-rozdělení nebo též Studentovo rozdělení. Je odvozeno z rozdělení výběrových průměrů a je platykurtické (3). Svým tvarem (a hodnotami kvantilů) se blíží rozdělení Gaussovu, vzrůstá-li počet stupňů volnosti pro daný test (pro počet stupňů volnosti blízký se nekonečnu přechází Studentovo rozdělení v rozdělení Gaussovo). Počet stupňů volnosti je obecně počet nezávislých pozorování potřebných k výpočtu dané charakteristiky a jeho výpočet nemusí vždy mít intuitivní charakter a v řadě případů je třeba použít k jeho určení odpovídající statistický program.

Jako příklad k vysvětlení právě uvedených pojmů můžeme zvolit zřejmě nejčastěji používaný dvouvýběrový (nepárový) *t*-test. Jeho testové kritérium se vypočte jako podíl absolutní hodnoty rozdílu srovnávaných charakteristik a standardní chyby odhadu tohoto rozdílu. Jestliže hodnota testového kritéria je menší než odpovídající kritická hodnota *t*-rozdělení, přijímáme nulovou hypotézu. Jestliže je hodnota testového kritéria rovna hodnotě kritické nebo je větší, nulovou hypotézu zamítáme a přijímáme hypotézu alternativní. Volba kvantilu pravděpodobnosti *t*-rozdělení závisí na tom, o jakou alternativní hypotézu jde.

V případě oboustranné je to $1-\alpha/2$, pro jednostrannou alternativní hypotézu to je $1-\alpha$. Takže pro hladinu významnosti 0,05 jsou kritické hodnoty dány 0,975 a 0,95 kvantilem (97,5 a 95 percentilem).

Uvědomíme-li si, že standardní chyba je nepřímo úměrná druhé odmocnině z počtu pozorování, dospějeme snadno k závěru, že jejich zvyšováním můžeme dosáhnout toho, že hodnota testového kritéria dosáhne hodnoty kritické. Výsledek bude statisticky významný při jinak nezměněných podmínkách projektu. Z toho také vyplývá, že statisticky významný rozdíl nemusí být významný klinicky či vědecky. Ve skutečnosti tento rozdíl nemusí být důležitý nebo zajímavý. Zjistíme-li, že na předem zvolené hladině významnosti je hodnota rozdílu krevních tlaků ve dvou skupinách léčených dvěma různými léky rovna 1 mmHg, tak jistě nebudeme takovýto výsledek považovat za důležitý při rozhodování o tom, který lék použijeme.

V souvislosti se statistickou významností výsledku testů je vcelku přirozené si myslet, že odpovídá-li hodnota pravděpodobnosti vypočteného testového kritéria $p = 0,02$, je výsledek pozorování významnější než jestliže $p = 0,05$. Z hlediska statistického rozhodování to však není správné. Hodnotu α musíme zvolit na samém začátku experimentu, a pak už je výsledek významný (můžeme přijmout alternativní hypotézu), nebo významný není (nemůžeme zamítnout hypotézu nulovou).

Testování hypotéz můžeme srovnat se soudním procesem (2). Na začátku výzkumného projektu zformulujeme vědeckou hypotézu. K jejímu potvrzení vybereme vhodný statistický test a nulovou hypotézu. Této hypotéze odpovídá presumpce nevinu obžalovaného. V dalším kroku experimentu shromažďujeme data potřebná k průkazu vědecké hypotézy. V soudním řízení tomu odpovídá shromažďování důkazů svědčících pro vinu obžalovaného. Předtím než přistoupíme k vlastnímu testu, ověříme, zda jsou splněny všechny předpoklady pro jeho použití, čemuž odpovídá posouzení, zda věříme předloženým důkazům. Pokud jsou předpoklady splněny, provedeme výpočet testového kritéria a v případě soudu se rozhodneme o tom, zda presumpce nevinu není v rozpo-

ru s předloženými důkazy. Jestliže je hodnota testového kritéria stejná nebo větší než hodnota kritická, zamítneme nulovou hypotézu a přijmeme hypotézu alternativní. Jsou-li předložené důkazy inkonzistentní s presumpcí nevinu, prohlásíme, že obžalovaný je vinen. Přehledně jsou jednotlivé kroky v obou procesech uvedeny v tabulce:

test	soud
Platí H_0	Presumpce nevinu
Posud' validitu shromážděných dat	Vyslechni přeložené důkazy a svědectví
Jaká je hodnota testového kritéria?	Jsou důkazy ve shodě s presumpcí nevinu?
Je-li jeho hodnota rovna nebo větší než hodnota kritická, zamítni H_0	Není-li tomu tak, zamítni presumpci nevinu
Přijmi H_A	Oznam obžalovanému, že je vinen

Chyba I. druhu odpovídá odsouzení nevinného, chyba II. druhu osvobození viníka. Ve většině společenských systémů se dává přednost tomu, aby pravděpodobnost chyby I. druhu byla velmi malá. Raději osvobodíme viníka, než bychom odsoudili nevinného. V medicíně situace není tak jednoznačná. Volba hladiny významnosti závisí na druhu výzkumného projektu. Chceme-li prokázat účinnost nového léku při léčbě onemocnění, pro které již jsou dobré léky známé, zvolíme α malé, např. 0,001, zatímco v případě, že dobrý lék neexistuje, může být α rovno 0,1 nebo i 0,2, abychom nezavrhlí potenciálně účinnou terapii.

LITERATURA

1. Knižek J, Stránský P. Příspěvek k nápravě opomíjení významu síly testu ve studiích z experimentální medicíny. Čas lék čes 2005; 144: 56 – 58.
2. Motolsky H. Intuitive Biostatistics. Oxford: Oxford University Press, 1995: 108 – 110.
3. Stránský P. Co potřebuje klinik vědět o (bio)statistice. Jak opsat data. Folia Gastroenterol Hepatol 2005; 3: 42 – 46.

Adresa pro korespondenci / correspondence to:

Prof. MUDr. Pravoslav Stránský, Ústav lékařské biofyziky a Oddělení výpočetní techniky, Univerzita Karlova v Praze, Lékařská fakulta v Hradci Králové, Šimkova 870, P.O. Box 38, 500 38 Hradec Králové, Česká republika / Czech Republic.
E-mail: str@lfhk.cuni.cz

