Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Martin Horáček

Výpočet velikosti výběru

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: Mgr. Michaela Šedová
Studijní program: Obecná matematika

2007

Chtěl bych poděkovat vedoucí této práce Mgr. Michaele Šedové za vynaložený čas, za pomoc při zajišťování literatury a za její užitečné rady nejen matematického charakteru.

Prohlašuji, že jsem svou bakalářskou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 2. srpna 2007 Martin Horáček

Obsah

Název práce: Výpočet velikosti výběru
Autor: Martin Horáček
Katedra (ústav): Katedra pravděpodobnosti a matematické statistiky
Vedoucí bakalářské práce: Mgr. Michaela Šedová
e-mail vedoucího: sedova@karlin.mff.cuni.cz

Abstrakt: V předložené práci se zabýváme způsoby výpočtu velikosti výběru, tj. počtu pozorování, které je potřeba provést při testování statistické hypotézy, má-li test splňovat dané požadavky. Zaměřili jsme se na metodu výpočtu zvanou analýza síly testu a na základní typy testů používané v lékařském výzkumu a v dalších oborech. Velikost výběru jsme odvodili pro některé běžně užívané testy o středních hodnotách normálně rozdělených náhodných veličin a pro přibližné a přesné testy parametru alternativního rozdělení. V poslední kapitole jsou nastíněny myšlenky skupinově-sekvenčních testů, které mohou díky větší flexibilitě velikost výběru lépe optimalizovat.

Klíčová slova: velikost výběru, síla testu, skupinově-sekvenční testy, testování hypotéz

Title: Calculating the sample size
Author: Martin Horáček
Department: Department of Probability and Mathematical Statistics
Supervisor: Mgr. Michaela Šedová
Supervisor's e-mail address: sedova@karlin.mff.cuni.cz

Abstract: In the present work we study techniques of calculating sample size that is required for succesful testing of a statistical hypothesis. We focus on a method called power analysis and we apply it on tests that are commonly used in clinical research and in other areas. The sample size is calculated for some regular tests, like tests that compare means of normally distributed random variables, large sample tests for proportions and exact tests for proportions. In the last chapter we present some basic ideas of group sequential methods that allow early termination of a test, which may lead to more optimized sample size.

Keywords: sample size, power analysis, group sequential tests, hypotheses testing

Kapitola 1

Úvod

Analýza velikosti výběru se zabývá problémy týkajícími se počtu pozorování nebo měření, která učiníme, testujeme-li při nějakém výzkumu určitou statistickou hypotézu. Řeší například, jak určit velikost výběru (potřebný počet pozorování) tak, aby výzkum měl dostatečnou vypovídající hodnotu a přitom se zbytečně neplýtvalo zdroji, jak tuto velikost upravit, pokud během výzkumu došlo k nějakému neočekávanému jevu, nebo jak ověřit, že velikost výběru při již proběhlém výzkumu byla správně zvolená. My se v této práci budeme soustředit na první z uvedených úloh, na určení velikosti výběru.

Snaha vhodně zvolit velikost výběru mívá více důvodů, nejzásadnější z nich ale bývají etické a ekonomické hledisko. Zbytečně nadsazená velikost výběru většinou značně zvedne náklady na výzkum, aniž bychom dostali odpovídající zlepšení výsledků, naopak nedostatečná velikost výběru může vést k nesprávným závěrům nebo k nutnosti výzkum opakovat. Jedná-li se o klinický výzkum, je také žádoucí co nejdříve rozhodnout o účinnosti zkoumaného léku, v případě pozitivního výsledku se lék může začít používat a výzkum se zaměří na další cíl. Navíc je neetické zbytečně podávat více pacientům (nebo testovaným zvířatům) přípravek, který je neefektivní a může mít vedlejší účinky. Testy používanými v lékařském výzkumu (ale nejen tam) se budeme převážně zabývat.

Testujeme-li proti sobě dvě hypotézy, nulovou hypotézu H_0 proti alternativní hypotéze H_a, vypovídající hodnota testu se většinou posuzuje podle pravděpodobností, že se nesprávně rozhodneme ve prospěch jedné z hypotéz. Můžeme udělat dva druhy chybného rozhodnutí:

Chyba I. druhu nastane, pokud zamítneme nulovou hypotézu a ta přitom platí. Pravděpodobnost tohoto jevu se většinou označuje řeckým písmenem α. Této pravděpodobnosti se také říká hladina testu.

Chyba II. druhu nastane, když nulová hypotéza neplatí a my ji nezamítneme. Tato pravděpodobnost se označuje písmenem β. Pravděpodobnosti doplňkového jevu, tj. že zamítneme nulovou hypotézu, když neplatí, se říká síla testu a je rovna $1-\beta$.

Velikost výběru můžeme zvolit například tak, aby intervalový odhad parametru měl danou šířku. Této metodě stanovení velikosti výběru se říká analýza přesnosti. Protože ale nebere v úvahu možnost chyby II. druhu, pravděpodobnost odhalení neplatnosti nulové hypotézy nemusí být velká.

Běžněji používaný přístup k problému se nazývá analýza síly testu (angl. power analysis). Touto metodou stanovíme velikost výběru tak, abychom za pevně dané pravděpodobnosti chyby I. druhu (nejčastěji 1 %, 5 % nebo 10 %) snížili pravděpodobnost chyby II. druhu na požadovanou hodnotu (většinou 10 % nebo 20 %). Někdy ani tento způsob není vhodný (například pokud chceme testovat hypotézu o nějakém velmi zřídka se vyskytujícím jevu), pak se používají jiné postupy. My se budeme zabývat téměř výhradně analýzou síly testu.

Častým cílem testů v lékařském výzkumu je ověřit, zda je nový lék účinnější nebo bezpečnější než lék původní, případně že má znatelné výsledky (je o poznání lepší než placebo). Nyní si nastíníme, jaké testy se za takovými účely nejčastěji používají a na jednom konkrétním příkladě si podrobněji ukážeme, jak se počítá velikost výběru pomocí analýzy síly a přesnosti. Blíže se analýzou síly uvedených testů budeme zabývat v kapitole 2 a 3. V kapitole 4 jsou vysvětleny základy skupinově-sekvenčních testů, které se týkají postupného vyhodnocování do dané doby shromážděných informací.

Test shodnosti (Test for Equality)

Test shodnosti bývá nazýván test hypotézy

$$H_0: \mu_1 = \mu_2 \quad \text{proti} \quad H_a: \mu_1 \neq \mu_2,$$

kde μ_1 a μ_2 jsou v případě dvouvýběrového testu střední hodnoty výsledků sledovaných náhodných veličin z prvního, resp. z druhého výběru. U jednovýběrového testu je μ_2 pevně zvolená hodnota (např. známá střední hodnota při užití placeba), s kterou porovnáváme výsledky sledované náhodné veličiny. Zamítnutí nulové hypotézy pak ukazuje na klinicky významný rozdíl mezi léky (resp. mezi lékem a placebem) a při vhodně zvolené velikosti výběru máme dostatečnou pravděpodobnost, že tento rozdíl (pokud existuje) bude při vlastním testu objeven.

Test inferiority (Test for Inferiority)

Pokud chceme v klinickém výzkumu ukázat, že nový lék je alespoň tak efektivní jako lék původní, volíme test hypotézy

$$H_0 : \mu - \mu_0 \leq -\delta \quad \text{proti} \quad H_a : \mu - \mu_0 > -\delta,$$

kde μ_0 je střední hodnota výsledků doposud používaného léku a $\delta \geq 0$ je taková hodnota, že rozdíl mezi středními hodnotami léků menší než δ je považován za klinicky nevýznamný. Zamítnutí nulové hypotézy pak ukazuje, že nový lék dává výsledky minimálně srovnatelné s původním, což je užitečné zejména, pokud má nový lék menší vedlejší účinky nebo je levnější na výrobu.

Test superiority (Test for Superiority)

$$H_0 : \mu - \mu_0 \leq \delta \quad \text{proti} \quad H_a : \mu - \mu_0 > \delta.$$

Test je podobný předchozímu testu inferiority, rozdíl spočívá ve znaménku u $\delta \geq 0$. Zamítnutí nulové hypotézy při testu superiority naznačuje, že nový lék je lepší než původní a že je mezi léky klinicky významný rozdíl. Poznamenejme, že tento test se také nazývá test klinické superiority. Pokud $\delta = 0$, říká se mu test statistické superiority.

Test ekvivalence (Test for Equivalence)

Někdy je cílem výzkumu ukázat, že rozdíl mezi dvěma přípravky je klinicky nevýznamný. Může tomu tak být tehdy, chceme-li mít k dispozici dva stejně efektivní léky a z nich vybírat až podle konkrétních potřeb pacienta. Pak použijeme test hypotézy

$$H_0 : |\mu_1 - \mu_2| \geq \delta \quad \text{proti} \quad H_a : |\mu_1 - \mu_2| < \delta.$$

V uvedených testech se objevují prvky, které nejsou přímo zadány a závisí na volbě statistika či lékaře, který připravuje test. Těmito nejistými faktory mohou být například hodnota klinicky nevýznamného rozdílu, rozptyl nebo to, zda a jak blízké normálnímu rozdělení bude rozdělení sledovaných hodnot (pokud budeme předpokládat normální rozdělení). Někdy je možné tyto prvky odhadnout například na základě dřívějších studií. Výpočty jsou ale i tak většinou jen přibližné a záleží na posouzení statistika, zda se jeho model příliš neodchýlil od skutečnosti.

1.1 Výpočet velikosti výběru

Analýza síly

Než začneme stanovovat velikost výběru, potřebujeme vědět, o jaké hypotézy v testu půjde, jaké je přibližně rozdělení náhodných veličin, o kterých se v testu jedná, na jaké hladině se bude zamítat nulová hypotéza a jakou sílu má test mít. Také je užitečné znát přibližně rozptyl, případně i další údaje o výzkumu, kterými se ale nebudeme blíže zabývat. Některé z těchto parametrů zvolí statistik či lékař, jiné známe jen přibližně nebo vůbec (např. rozptyl, někdy je možné ho odhadnout na základě předchozích výzkumů).

Odvodíme metodou analýzy síly velikost výběru pro konkrétní hypotézu. Řekněme, že plánujeme vyzkoušet účinnost nově vyvinutého léku a chceme za tímto účelem použít test hypotézy

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

kde μ je střední hodnota výsledků nového léku a μ_0 je známá střední hodnota výsledků u pacientů užívajících placebo.

Budeme předpokládat, že údaje získané od pacientů užívajících testovaný přípravek, tj. náhodný výběr $X_1 \dots X_n$, je z normálního rozdělení s (neznámou) střední hodnotou μ a se známým rozptylem σ^2. (Rozdělení značíme $N(\mu, \sigma^2)$.) Distribuční funkci standardního normálního rozdělení $N(0, 1)$ označíme Φ, $u_\alpha = -u_{1-\alpha} = \Phi^{-1}(1 - \alpha)$ je kritická hodnota tohoto rozdělení. Pro výběrový průměr $\bar{X}_n$ platí

$$\frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}_n \sim N(\mu, \tfrac{\sigma^2}{n}),$$

proto

$$\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1).$$

Dále zvolíme hladinu testu α a určíme sílu $1-\beta$, které chceme dosáhnou. Síla testu roste s velikostí výběru (viz příklad 1.3 na straně 11), najdeme proto nejmenší přirozené n takové, aby při velikosti výběru n byla síla alespoň $1-\beta$. Síla testu je rovna pravděpodobnosti

$$P\left(\text{zamítneme } H_0 | H_a \text{ platí}\right).$$

Potřebujeme zjistit, kdy zamítneme nulovou hypotézu (na hladině α). K tomu dospějeme například následující úvahou: Hypotézu H_0 zamítneme na hladině α, když

$$\begin{aligned} \alpha &= P(\text{zamítneme } H_0 | H_0 \text{ platí}) \\ &= P(\bar{X}_n > w | \mu \leq \mu_0), \end{aligned}$$

kde w je správně zvolená konstanta. Když bude $\mu < \mu_0$, je méně pravděpodobné, že zamítneme hypotézu H_0, stačí se zabývat případem $\mu = \mu_0$. Potom

$$\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} \sim N(0,1).$$

Nulovou hypotézu zamítneme, když

$$\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} > u_\alpha,$$

$w = \frac{u_\alpha}{\sqrt{n}}\sigma + \mu_0$. Další možností dávající stejné výsledky je vyjít z intervalových odhadů, což uděláme u testu ekvivalence. Hledaná velikost výběru je nejmenší přirozené číslo n takové, že platí:

$$\begin{aligned} 1-\beta &\leq P\left(\text{zamítneme } H_0 | H_a \text{ platí}\right) \\ &= P\left(\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} > u_\alpha \middle| \mu = \mu_0 + \epsilon\right), \end{aligned} \tag{1.1}$$

kde $\epsilon > 0$ je skutečný rozdíl mezi μ a μ_0.

Místo nerovnosti 1.1 položme rovnost. Výsledek, který dostaneme, stačí zaokrouhlit směrem nahoru na celé číslo.

$$\begin{aligned} 1-\beta &= P\left(\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} > u_\alpha \middle| \mu = \mu_0 + \epsilon\right) \\ &= P\left(\frac{\bar{X}_n - (\mu_0 + \epsilon)}{\sigma}\sqrt{n} > u_\alpha - \frac{\epsilon}{\sigma}\sqrt{n} \middle| \mu = \mu_0 + \epsilon\right). \end{aligned}$$

Náhodná veličina

$$Y = \frac{\bar{X}_n - (\mu_0 + \epsilon)}{\sigma}\sqrt{n}$$

má za podmínky $\mu = \mu_0 + \epsilon$ rozdělení $N(0,1)$. Proto

$$\begin{aligned}
1-\beta &= P\left(Y > u_\alpha - \frac{\epsilon}{\sigma}\sqrt{n}\middle|\mu = \mu_0 + \epsilon\right)\\
&= 1 - \Phi\left(u_\alpha - \frac{\epsilon}{\sigma}\sqrt{n}\right)\\
\Phi^{-1}(\beta) &= u_\alpha - \frac{\epsilon}{\sigma}\sqrt{n}\\
u_{1-\beta} &= u_\alpha - \frac{\epsilon}{\sigma}\sqrt{n}.
\end{aligned}$$

Hledaná velikost výběru je

$$n = \frac{(u_\alpha + u_\beta)^2\sigma^2}{\epsilon^2}. \tag{1.2}$$

K určení n zbývá zvolit hodnotu ϵ, případně ϵ/σ. To můžeme udělat například na základě požadavků na nový lék s přihlédnutím k pilotní studii nebo starším údajům. Pokud bude skutečná (relativní) odchylka rovna nebo větší než ϵ (ϵ/σ), test bude mít pro vypočítané n sílu nejméně $1-\beta$.

Příklad 1.1

Uvažujme následující problém: Máme sestavit test, jehož cílem bude ověřit, zda nový lék na krevní tlak je lepší než lék stávající. Střední hodnota průměrného snížení systolického tlaku za čtvrt roku u pacientů, kteří brali doposud užívaný přípravek, je 17,40 mmHg.

Po konzultaci s lékaři byla hladina testu stanovena na 5 %, vhodná síla testu 90 % a očekávaný rozdíl mezi léky určen jako 4 mmHg. Pilotní studie ukázala, že rozptyl výsledků nového léku je přibližně 181 mmHg. Pro test hypotézy

$$H_0 : \mu \leq 17{,}40 \quad \text{proti} \quad H_a : \mu > 17{,}40,$$

kde μ je střední hodnota snížení systolického tlaku u nového léku, vyjde velikost výběru

$$n = \frac{(u_{0{,}05} + u_{0{,}1})^2 181}{4^2} \doteq 96{,}88.$$

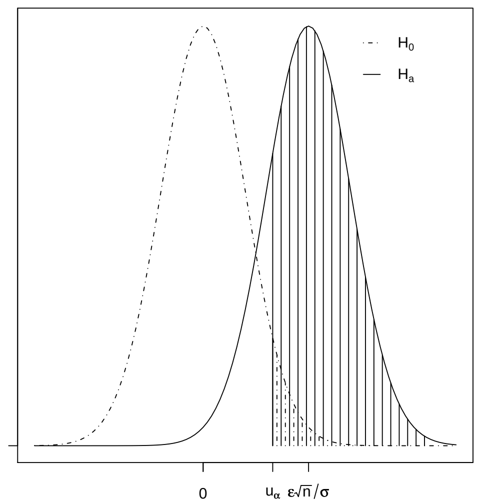
K realizaci testu budeme potřebovat 97 pacientů. Při vlastním testu zamítneme nulovou hypotézu, pokud bude $\bar{X}_{97} > 18{,}07$. Pravděpodobnost, že zamítneme nulovou hypotézu, když bude $\mu \leq 17{,}40$, je nejvýše 5 %. Pravděpodobnost, že zamítneme nulovou hypotézu, když bude $\mu > 21{,}40$, je nejméně 90 %.

Příklad 1.2

Na obrázku 1.1 je zakreslena hustota rozdělení náhodné veličiny

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

za platnosti nulové hypotézy $H_0 : \mu = \mu_0$ a za platnosti alternativní hypotézy $H_a : \mu = \mu_0 + \epsilon > \mu_0$.

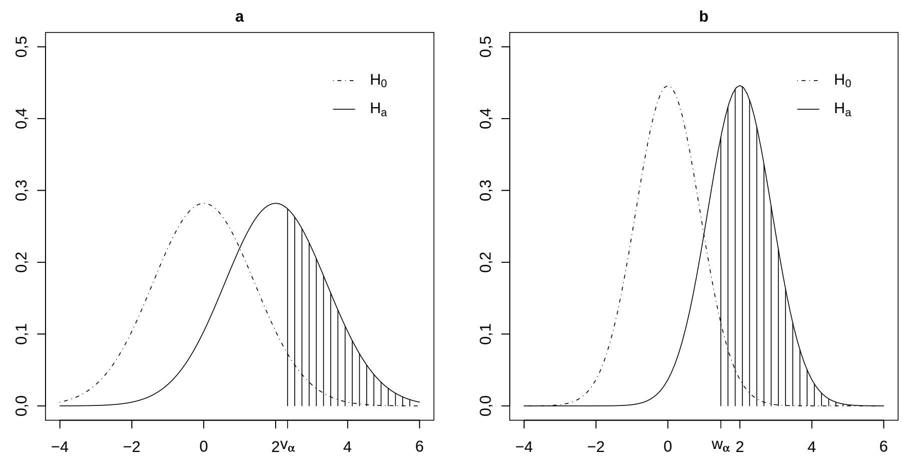


Obrázek 1.1: Hustota rozdělení náhodné veličiny $Z = \frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n}$ za platnosti nulové a alternativní hypotézy

Obsah plochy pod grafem na nějakém intervalu je roven pravděpodobnosti, se kterou náhodná veličina Z nabude hodnotu na na tomto intervalu, pokud platí nulová, resp. alternativní hypotéza. Obsah plochy vyšrafované přerušovanou čarou je tedy roven hladině testu, obsah nepřerušovaně vyšrafované plochy je roven síle testu, když $\mu = \mu_0 + \epsilon$ a velikost výběru je rovna n. Síla testu tedy roste s velikostí výběru, více v následujícím příkladu.

Příklad 1.3

Mějme náhodný výběr $X_1, X_2, \ldots$ z rozdělení $N(\mu, \sigma^2)$. Pokud je velikost výběru rovna n, výběrový průměr $\bar{X}_n$ má rozdělení $N(\mu, \frac{\sigma^2}{n})$. S rostoucím n rozptyl výběrového průměru klesá, rozdělení je více soustředěno kolem střední hodnoty a pravděpodobnost odhalení nesprávnosti nulové hypotézy, tj. síla testu, roste. Konkrétní případ je ilustrován na obrázku 1.2.



Obrázek 1.2: (a) Hustota rozdělení náhodné veličiny $\bar{X}_n$ a příslušná síla při platnosti nulové a alternativní hypotézy pro test hypotézy $H_0 : \mu = 0$ proti $H_a : \mu = \epsilon > 0$. Rozptyl je $\sigma^2 = 40$, $n = 20$ a $\epsilon = 2$. (b) Totéž pro $n = 50$. Body v_α a w_α označují kritické hodnoty odpovídajících rozdělení na hladině 5 %.

Analýza přesnosti

Metodou analýzy přesnosti se velikost výběru počítá na základě intervalového odhadu. Uvažujme test shodnosti pro normálně rozdělené veličiny. Je-li hladina testu α, šířka intervalu, který za platnosti nulové hypotézy s pravděpodobností $1 - \alpha$ překryje střední hodnotu μ, klesá s velikostí výběru. Velikost výběru vezmeme takovou, abychom zúžili intervalový odhad na požadovanou mez.

Požadovanou šířku intervalu spolehlivosti označme 2δ. Pokud $X_1 \dots X_n$ jsou nezávislé normálně rozdělené náhodné veličiny se střední hodnotou μ a rozptylem σ^2 a rozptyl známe, interval, který s pravděpodobností $1 - \alpha$ překryje μ, je

$$\left(\bar{X}_n - u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \qquad \text{a} \qquad \delta = u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Na základě toho můžeme velikost výběru zvolit jako

$$n = \frac{u_{\alpha/2}^2\sigma^2}{\delta^2}.$$

Kapitola 2

Velikost výběru u testů porovnávajících střední hodnoty

Nejen v lékařském výzkumu jsou běžné testy, jejichž účelem je porovnat střední hodnotu výsledků zkoumané skupiny subjektů s nějakou konstantou, či porovnat střední hodnoty výsledků získaných od dvou skupin. Na testu tohoto typu jsme si ukázali způsob odvození velikosti výběru analýzou síly. Touto metodou odvodíme velikost výběru u dalších testů při normálně rozdělených veličinách.

2.1 Jednovýběrové testy

Test shodnosti

$$H_0 : \mu = \mu_0 \quad \text{proti} \quad H_a : \mu \neq \mu_0.$$

Nejprve předpokládejme, že rozptyl je známý. Hypotézu H_0 zamítneme na hladině α, když

$$\left| \frac{\bar{X}_n - \mu_0}{\sigma} \right| \sqrt{n} > u_{\alpha/2}.$$

Když platí alternativní hypotéza, je konstanta $\epsilon = \mu - \mu_0$ různá od nuly. Síla testu je

$$1 - \beta \quad = \quad P\left(\left| \frac{\bar{X}_n - \mu_0}{\sigma} \right| \sqrt{n} > u_{\alpha/2} \middle| \mu = \mu_0 + \epsilon \right)$$

$$
\begin{aligned}
&= P\left(\frac{\bar{X}_n - (\mu_0 + \epsilon)}{\sigma}\sqrt{n} > u_{\alpha/2} - \frac{\epsilon}{\sigma}\sqrt{n}\Big|\mu = \mu_0 + \epsilon\right) \\
&\quad + P\left(\frac{-\bar{X}_n + \mu_0 + \epsilon}{\sigma}\sqrt{n} > u_{\alpha/2} + \frac{\epsilon}{\sigma}\sqrt{n}\Big|\mu = \mu_0 + \epsilon\right) \\
&= \Phi\left(-u_{\alpha/2} + \frac{\epsilon}{\sigma}\sqrt{n}\right) + \Phi\left(-u_{\alpha/2} - \frac{\epsilon}{\sigma}\sqrt{n}\right). \qquad (2.1)
\end{aligned}
$$

Jeden ze sčítanců v 2.1 je menší než $\alpha/2$ (který, závisí na znaménku ϵ). Vzhledem k malé hodnotě $\alpha/2$ si můžeme dovotil tento sčítanec zanedbat. Potom

$$
\begin{aligned}
1 - \beta &= \Phi\left(-u_{\alpha/2} + \frac{|\epsilon|}{\sigma}\sqrt{n}\right) \\
u_\beta &= -u_{\alpha/2} + \frac{|\epsilon|}{\sigma}\sqrt{n} \\
n &= \frac{(u_{\alpha/2} + u_\beta)^2\sigma^2}{\epsilon^2}. \qquad (2.2)
\end{aligned}
$$

Absolutní hodnotu ϵ můžeme zvolit jako v 1.2 nebo např. jako nejmenší klinicky významný rozdíl mezi μ a μ_0. Když bude skutečný rozdíl větší, bude větší i síla testu.

Pokud rozptyl neznáme, lze ho nahradit veličinou

$$
S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,
$$

která je pro $n > 1$ nestranný odhad σ^2 a je nezávislá na $\bar{X}_n$. Důkaz je možné najít v knize Anděl (2005), věty 2.10 a 4.21. Náhodná veličina

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

pak má za platnosti hypotézy H_0 Studentovo rozdělení (t-rozdělení) o $n-1$ stupních volnosti. Důkaz viz Anděl (2005), věta 4.23. Nulovou hypotézu zamítneme na hladině α, když

$$
\left|\frac{\bar{X}_n - \mu_0}{S_n}\right|\sqrt{n} > t_{\alpha,n-1},
$$

kde $t_{\alpha,n-1}$ je kritická hodnota t-rozdělení o $n-1$ stupních volnosti. (Jde o specifické značení u t-rozdělení, $P(|T| > t_{\alpha,n-1}) = \alpha$.) Síla testu je v tomto případě

$$\begin{aligned} 1-\beta &= P\left(\left|\frac{\bar{X}_n-\mu_0}{S_n}\right|\sqrt{n} > t_{\alpha,n-1}\middle|\mu=\mu_0+\epsilon\right) \\ &= P\left(\frac{\bar{X}_n-\mu_0}{S_n}\sqrt{n} > t_{\alpha,n-1}\middle|\mu=\mu_0+\epsilon\right) \\ &\quad +P\left(\frac{\bar{X}_n-\mu_0}{S_n}\sqrt{n} < -t_{\alpha,n-1}\middle|\mu=\mu_0+\epsilon\right). \end{aligned}$$

Když platí alternativní hypotéza, tj. $\mu = \mu_0 + \epsilon$, náhodná veličina

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

má necentrální t-rozdělení o $n-1$ stupních volnosti s parametrem necentrality $\frac{\mu-\mu_0}{\sigma}\sqrt{n}$. Když označíme distribuční funkci tohoto rozdělení jako $ncT_{n-1}(\cdot,\theta)$, kde θ je parametr necentrality, síla testu je

$$\begin{aligned} 1-\beta &= 1-ncT_{n-1}\left(t_{\alpha,n-1},\frac{\sqrt{n}\epsilon}{\sigma}\right)+ncT_{n-1}\left(-t_{\alpha,n-1},\frac{\sqrt{n}\epsilon}{\sigma}\right) \\ &= ncT_{n-1}\left(-t_{\alpha,n-1},-\frac{\sqrt{n}\epsilon}{\sigma}\right)+ncT_{n-1}\left(-t_{\alpha,n-1},\frac{\sqrt{n}\epsilon}{\sigma}\right). \end{aligned}$$

Zanedbáme-li sčítanec menší než α, je

$$1-\beta \doteq 1-ncT_{n-1}\left(t_{\alpha,n-1},\frac{\sqrt{n}\,|\epsilon|}{\sigma}\right).$$

Protože rozptyl není znám, zvolíme ϵ relativně vzhledem k směrodatné odchylce ($|\epsilon|/\sigma$). Pokud bude skutečný poměr větší, test bude mít při zvolené velikosti výběru sílu větší než $1-\beta$. Velikosti n pro některé hodnoty α, β a $\frac{|\epsilon|}{\sigma}$ jsou uvedeny v tabulce 2.1 (spočítáno v programu R).

Pokud je n velké, můžeme využít toho, že t-rozdělení se blíží normálnímu, $t_\alpha \approx u_\alpha$,

$$ncT_{n-1}\left(t_{\alpha,n-1},\frac{\sqrt{n}\epsilon}{\sigma}\right) \approx \Phi\left(u_{\alpha/2}-\frac{\epsilon}{\sigma}\sqrt{n}\right).$$

Pro velká n tedy lze použít vzorec 2.2.

U dalších testů je při neznámém rozptylu postup analogický a nebudeme ho rozepisovat.

Tabulka 2.1: Nejmenší n, pro které $ncT_{n-1}\left(t_{\alpha,n-1}, \sqrt{n}\frac{|\epsilon|}{\sigma}\right) \leq \beta$

	$\alpha = 0{,}01$		$\alpha = 0{,}05$	
	$1-\beta =$		$1-\beta =$	
$\lvert\epsilon\rvert/\sigma$	0,8	0,9	0,8	0,9
0,10	1172	1492	787	1053
0,15	532	665	350	469
0,20	296	376	199	265
0,25	191	242	128	171
0,30	134	169	90	119
0,40	77	97	52	68
0,50	51	63	34	44
0,60	36	45	24	32
0,80	22	27	15	19
1,00	16	19	10	13
1,50	9	11	6	7

Test non-inferiority a superiority

$$H_0 : \mu - \mu_0 \leq \delta \quad \text{proti} \quad H_a : \mu - \mu_0 > \delta.$$

Podle znaménka δ jde o test inferiority (záporné δ), resp. superiority. Nulovou hypotézu zamítneme, pokud

$$\frac{\bar{X}_n - \mu_0 - \delta}{\sigma}\sqrt{n} > u_\alpha.$$

Nechť platí alternativní hypotéza a $\mu - \mu_0 = \epsilon > \delta$. Síla testu je

$$\begin{aligned}
1 - \beta &= P\left(\frac{\bar{X}_n - \mu_0 - \delta}{\sigma}\sqrt{n} > u_\alpha \middle| \mu = \mu_0 + \epsilon\right) \\
&= P\left(\frac{\bar{X}_n - \mu_0 - \epsilon}{\sigma}\sqrt{n} > u_\alpha - \frac{\epsilon - \delta}{\sigma}\sqrt{n} \middle| \mu = \mu_0 + \epsilon\right) \\
&= \Phi\left(\frac{\epsilon - \delta}{\sigma}\sqrt{n} - u_\alpha\right) \\
u_\beta &= -u_\alpha + \frac{(\epsilon - \delta)}{\sigma}\sqrt{n}
\end{aligned}$$

$$n = \frac{(u_\alpha + u_\beta)^2\sigma^2}{(\epsilon - \delta)^2}.$$

Test ekvivalence

$$H_0 : |\mu - \mu_0| \geq \delta \quad \text{proti} \quad H_a : |\mu - \mu_0| < \delta.$$

Když je σ^2 známé, zamítneme nulovou hypotézu na hladině α, pokud

$$\frac{\bar{X}_n - \mu_0 - \delta}{\sigma}\sqrt{n} < -u_\alpha \quad \text{a} \quad \frac{\bar{X}_n - \mu_0 + \delta}{\sigma}\sqrt{n} > u_\alpha. \tag{2.3}$$

Odvození nerovností 2.3 je o něco složitější než v předchozích případech, proto si ho zde uvedeme. Další informace je možné najív v knize Wellek (2003) v kapitole 3. Označme

$$\mu_D = \bar{X}_n - \frac{\sigma}{\sqrt{n}}u_\alpha \qquad \text{a} \qquad \mu_H = \bar{X}_n + \frac{\sigma}{\sqrt{n}}u_\alpha,$$

tj. μ_D dolní a μ_H horní jednostranný intervalový odhad μ na hladině α na základě náhodného výběru $X_1 \ldots X_n$. Nulovou hypotézu zamítneme, právě když je interval (μ_D, μ_H) zcela obsažen v intervalu $(\mu_0 - \delta, \mu_0 + \delta)$. K alternativní hypotéze se proto přikloníme, pokud jsou splněny nerovnosti $\mu_D > \mu_0 - \delta$ a $\mu_H < \mu_0 + \delta$, z čehož plyne 2.3.

Ještě ověříme, že nulovou hypotézu zamítáme na hladině α: Nechť například $\mu < \mu_0 - \delta$. Nulovou hypotézu zamítneme, pokud $\mu_D > \mu_0 - \delta$ a $\mu_H < \mu_0 + \delta$. Chyba prvního druhu tak může nastat pouze v případě, kdy $\mu_D > \mu$. Dolní jednostranný odhad jsme ale zvolili tak, aby bylo $\mu_D > \mu$ s pravděpobností nejvýše α. Podobně postupujeme, když $\mu > \mu_0 + \delta$. Pokud $\mu \in (\mu_0 - \delta, \mu_0 + \delta)$, pak nulová hypotéza neplatí a její zamítnutí je správným rozhodnutím.

Nechť platí alternativní hypotéza a $\mu - \mu_0 = \epsilon$, kde $|\epsilon| < \delta$. Označíme-li

$$Y := \frac{\bar{X}_n - \mu_0 - \epsilon}{\sigma}\sqrt{n},$$

síla testu je

$$\begin{aligned}
1 - \beta &= P\left(Y < -u_\alpha + \frac{\delta - \epsilon}{\sigma}\sqrt{n} \wedge Y > u_\alpha - \frac{\delta + \epsilon}{\sigma}\sqrt{n}\right) \\
&= P\left(Y < -u_\alpha + \frac{\delta - \epsilon}{\sigma}\sqrt{n}\right) + P\left(Y > u_\alpha - \frac{\delta + \epsilon}{\sigma}\sqrt{n}\right) \\
&\quad -P\left(Y < -u_\alpha + \frac{\delta - \epsilon}{\sigma}\sqrt{n} \vee Y > u_\alpha - \frac{\delta + \epsilon}{\sigma}\sqrt{n}\right).
\end{aligned} \tag{2.4}$$

V 2.4 jsme použili pravidlo $P(A \cap B) = P(A) + P(B) - P(A \cup B)$. Povšimněme si, že vzhledem k charakteru jevů označených A a B platí, že pokud $P(A \cup B) < 1$, pak $P(A \cap B) = 0$ a síla testu je nulová. Bez újmy na obecnosti proto můžeme položit $P(A \cup B) = 1$. Pak

$$
\begin{aligned}
1 - \beta &= \Phi\left(-u_\alpha + \frac{\delta - \epsilon}{\sigma}\sqrt{n}\right) + \Phi\left(-u_\alpha + \frac{\delta + \epsilon}{\sigma}\sqrt{n}\right) - 1 \\
&\geq 2\Phi\left(-u_\alpha + \frac{\delta - |\epsilon|}{\sigma}\sqrt{n}\right) - 1 \qquad (2.5) \\
u_{\beta/2} &= -u_\alpha + \frac{\delta - |\epsilon|}{\sigma}\sqrt{n} \\
n &= \frac{(u_\alpha + u_{\beta/2})^2\sigma^2}{(\delta - |\epsilon|)^2}.
\end{aligned}
$$

Síla testu je větší, než výraz v 2.5. Položíme-li v 2.5 místo nerovnosti rovnost, dostaneme velikost výběru, pro kterou má test sílu alespoň $1 - \beta$.

2.2 Dvouvýběrové testy

Pokud je cílem výzkumu porovnat střední hodnoty normálně rozdělených výsledků dvou nezávislých skupin, můžeme užít některý z dvouvýběrových testů.

Označme X_j výsledek j-tého subjektu v první skupině a Y_i výsledek i-tého subjektu ve druhé skupině. Předpokládáme, že $X_j,\ j = 1 \dots n$ je náhodný výběr z $N(\mu_1, \sigma^2)$ a $Y_i,\ i = 1 \dots m$ je na něm nezávislý náhodný výběr z $N(\mu_2, \sigma^2)$. Čísla m a n mohou, ale nemusí být stejná, někdy je výhodné či potřebné mít jeden výběr větší než druhý. Velikost výběru pak ale při stejné síle a spolehlivosti vyjde větší, než když m a n jsou shodná. Poměr mezi velikostmi skupin označme $k = m/n$, dále $\epsilon = \mu_2 - \mu_1$,

$$
\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad \bar{Y}_m = \frac{1}{m}\sum_{i=1}^{m} Y_i,
$$

$$
s^2 = \frac{1}{n + m - 2}\Big(\sum_{j=1}^{n}(X_j - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_n)^2\Big).
$$

Výpočet velikosti výběru pro dvouvýběrové paralelní testy probíhá podobně jako u jednovýběrových testů. Pracujeme s náhodnou veličinou $Z = \bar{X}_n - \bar{Y}_m$, která má za našich předpokladů $N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n} + \frac{1}{m}))$ rozdělení. Postup je rozepsán pro test shodnosti, velikosti výběru pro test superiority, inferiority a ekvivalence se odvodí podobně. Stručný postup a výsledné vzorce lze nalézt v knize Chow, Shao, Wang (2003).

Test shodnosti

$$H_0 : \mu_2 - \mu_1 = 0 \quad \text{proti} \quad H_a : \mu_2 - \mu_1 \neq 0.$$

Pokud je rozptyl znám, zamítneme nulovou hypotézu na hladině α, když

$$\frac{|\bar{X}_n - \bar{Y}_m|}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} > u_{\alpha/2}.$$

Při platnosti alternativní hypotézy je

$$1 - \beta = \Phi\left(\frac{\epsilon}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} - u_{\alpha/2}\right) + \Phi\left(\frac{-\epsilon}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} - u_{\alpha/2}\right).$$

Zanedbáním hodnoty menší než $\alpha/2$ dostaneme, že síla testu je přibližně rovna

$$\Phi\left(\frac{|\epsilon|}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} - u_{\alpha/2}\right).$$

Po úpravě

$$n = \frac{(u_{\alpha/2} + u_\beta)^2\sigma^2(1 + 1/k)}{\epsilon^2} \qquad \text{a} \qquad k = \frac{m}{n}.$$

Pokud σ^2 neznáme, můžeme jej nahradit veličinou s^2. Potom

$$T = \frac{|\bar{X}_n - \bar{Y}_m|}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}.$$

Důkaz je uveden v knize Anděl (2005), věta 4.26. Náhodnou veličinu T lze upravit na tvar, pro který je věta dokázána. Nulovou hypotézu v tomto případě zamítneme na hladině α, když

$$\frac{|\bar{X}_n - \bar{Y}_m|}{s\sqrt{\frac{1}{n} + \frac{1}{m}}} > t_{\alpha,n+m-2}.$$

Síla testu je

$$
\begin{aligned}
1-\beta &= 1 - ncT_{n+m-2}\left(t_{\alpha,n+m-2}, \frac{\epsilon}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right) \\
&\quad + ncT_{n+m-2}\left(-t_{\alpha,n+m-2}, \frac{\epsilon}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right) \\
&\doteq 1 - ncT_{n+m-2}\left(t_{\alpha,n+m-2}, \frac{|\epsilon|}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}\right).
\end{aligned}
$$

Na základě tohoto výsledku můžeme pro zvolené $|\epsilon|/\sigma$ a k velikost n a m najít v tabulkách nebo spočítat na počítači.

Kapitola 3

Přibližné a přesné testy parametru alternativního rozdělení

Nyní se budeme zabývat velikostí výběru u testů prováděných na náhodných veličinách majících alternativní rozdělení. Veličina z alternativního rozdělení nabývá pouze dvou hodnot. Označme tyto hodnoty 0 a 1. Pokud veličina nabude hodnoty 1, budeme říkat, že výsledek testu byl pro tuto veličinu pozitivní. Pravděpodobnost tohoto jevu označme p.

Nechť je tedy $X_i, i = 1 \dots n$ náhodný výběr z alternativního rozdělení s neznámým parametrem $p = P(X_i = 1) \in (0, 1)$. Pokud budeme testovat dostatečně velký počet subjektů, má náhodná veličina

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

která je nejběžnějším (a v jistém smyslu nejlepším) odhadem parametru p, rozdělení blízké normálnímu. Na základě toho můžeme pro testy shodnosti, superiority, inferiority a ekvivalence přibližně vypočítat velikost výběru na dané hladině a při požadované síle podobným postupem, jako při výběru z normálního rozdělení. V některých případech (například testy zkoumající účinnost léčby rakoviny) ale bývají velikosti výběru malé a proto je předchozí přístup nevhodný. Pak se použije některý z přesných testů.

3.1 Přibližné jednovýběrové testy

Postup je uveden pro test shodnosti, u dalších testů ho můžeme odvodit na základě odpovídajících testů o středních hodnotách a uvedeného přibližného testu. Stručný postup a výslednou velikost výběru je také možné nalézt v knize Chow, Shao, Wang (2003).

Test shodnosti

$$H_0 : p = p_0 \quad \text{proti} \quad H_a : p \neq p_0.$$

Označme $\epsilon = p - p_0$ a $\hat{\epsilon} = \hat{p} - p_0$. Pro $n \to \infty$ se za platnosti nulové hypotézy rozdělení náhodné veličiny

$$\frac{\sqrt{n}\hat{\epsilon}}{\sqrt{\hat{p}(1-\hat{p})}}$$

blíží standardnímu normálnímu rozdělení. Tento poznatek vyplývá z centrálních limitních vět., viz např. Lachout (1998). Nulovou hypotézu proto zamítneme na hladině α, pokud

$$\left| \frac{\sqrt{n}\hat{\epsilon}}{\sqrt{\hat{p}(1-\hat{p})}} \right| > u_{\alpha/2}.$$

Když platí alternativní hypotéza, je $p = p_0 + \epsilon$, kde $\epsilon \neq 0$. Síla testu je

$$1 - \beta = P\left(\left| \frac{\sqrt{n}(\hat{p} - p_0 - \epsilon)}{\sqrt{\hat{p}(1-\hat{p})}} + \frac{\sqrt{n}\epsilon}{\sqrt{\hat{p}(1-\hat{p})}} \right| > u_{\alpha/2} \middle| p = p_0 + \epsilon \right),$$

což je přibližně rovno (zanedbáme hodnotu menší než $\alpha/2$ jako v 2.1)

$$\Phi\left(\frac{\sqrt{n}|\epsilon|}{\sqrt{\hat{p}(1-\hat{p})}} - u_{\alpha/2} \right).$$

Úpravou rovnice

$$u_\beta = \frac{\sqrt{n}|\epsilon|}{\sqrt{\hat{p}(1-\hat{p})}} - u_{\alpha/2}$$

dostaneme velikost výběru

$$n = \frac{(u_{\alpha/2} + u_\beta)^2 \hat{p}(1-\hat{p})}{\epsilon^2}.$$

K výpočtu n zbývá vhodně zvolit ϵ a $\hat{p}$, což můžeme provézt např. na základě pilotní studie. Poznamenejme, že nejkonzervativnější odhad vyjde pro $\hat{p} = 0{,}5$.

3.2 Přibližné dvouvýběrové testy

Podobně jako při testování středních hodnot můžeme porovnávat parametry při dvou výběrech ne nutně stejné velikosti. Nechť $X_i, i = 1 \dots m$ a $Y_j, j = 1 \dots n$ jsou na sobě nezávislé výběry z alternativního rozdělení s parametry $p_1 = P(X_i = 1) \in (0, 1)$ a $p_2 = P(Y_j = 1) \in (0, 1)$. Označme $\epsilon = p_1 - p_2$, $k = m/n$.

Test shodnosti

$$H_0 : p_1 = p_2 \quad \text{proti} \quad H_a : p_1 \neq p_2.$$

Analogicky jako u jednovýběrového testu jsou parametry p_1 a p_2 odhadnuty hodnotou

$$\hat{p}_1 = \frac{1}{m} \sum_{i=1}^{m} X_i \qquad \text{a} \qquad \hat{p}_2 = \frac{1}{n} \sum_{j=1}^{n} Y_j.$$

Za platnosti nulové hypotézy má náhodná veličina

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1(1 - \hat{p}_1)/m + \hat{p}_2(1 - \hat{p}_2)/n)}}$$

pro velká m a n přibližně standardní normální rozdělení. Postupujme-li analogicky jako u jednovýběrového testu, vyjde

$$\begin{aligned} n &= \frac{(u_{\alpha/2} + u_\beta)^2}{\epsilon^2} \left[\frac{\hat{p}_1(1 - \hat{p}_1)}{k} + \hat{p}_2(1 - \hat{p}_2) \right], \\ k &= m/n. \end{aligned}$$

3.3 Přesné testy

Jednostranný binomický test

Binomický test je jednovýběrový test, který může i při malém počtu pozorování umožnit rozhodnout, zda je parametr alternativního rozdělení p testovaného přípravku odlišný od stanovené hodnoty p_0. Jeho jednostrannou variantu lze zapsat jako test hypotézy

$$H_0 : p = p_0 \quad \text{proti} \quad H_a : p > p_0.$$

Nechť n je počet pozorování. Označme

$$m = \sum_{i=1}^{n} X_i,$$

tj. počet subjektů, u nichž je výsledek pozorování pozitivní (rovný jedné). V případě, že $p = p_0$, má m binomické rozdělení s parametry (n, p_0). Pravděpodobnost, že alespoň k ($0 \leq k \leq n$) subjektů má pozitivní výsledek, je rovna

$$P(m \geq k) = \sum_{i=k}^{n} \frac{n!}{i!(n-i)!} p_0^i (1-p_0)^{n-i}.$$

Pro danou hladinu α existuje přirozené číslo r_1 takové, že

$$P(m \geq r_1) \leq \alpha \quad \text{a} \quad P(m \geq r_1 - 1) > \alpha.$$

Nulovou hypotézu zamítneme na hladině α, když $m \geq r_1$. (Pozn. pro některá n a α může nastat situace, že $r_1 > n$, tj. nulovou hypotézu na dané hladině nikdy nezamítneme.) Abychom mohli spočítat velikost výběru, zbývá ještě určit, při jakém rozdílu mezi parametrem p a p_0 má síla testu být $1 - \beta$. Označme zvolený rozdíl δ a dále $p_1 = p_0 + \delta$. Velikost výběru spočteme pro případ, že $p = p_1 = p_0 + \delta$. Pokud bude skutečný rozdíl větší, bude mít test pro takto zvolenou velikost výběru sílu větší než $1 - \beta$. Za platnosti alternativní hypotézy je síla testu

$$\begin{aligned} 1 - \beta &= P(m \geq r_1 | p = p_1) \\ &= \sum_{i=r_1}^{n} \frac{n!}{i!(n-i)!} p_1^i (1-p_1)^{n-i}. \end{aligned}$$

Velikost výběru je nejmenší n takové, že

$$P(m \geq r_1 | p = p_1) \geq 1 - \beta. \tag{3.1}$$

Velikosti n a r_1 pro některé hodnoty parametrů jsou uvedeny v tabulce 3.1 (spočítáno v programu R, viz Dodatek).

Oboustranný binomický test

Při oboustranné variantě testujeme hypotézu

$$H_0 : p = p_0 \quad \text{proti} \quad H_a : p \neq p_0.$$

Tabulka 3.1: Velikost výběru n a kritická hodnota r_1 v jednostranném binomickém testu pro $\alpha = 0{,}05$

	$1-\beta = 0{,}8$		$1-\beta = 0{,}9$	
	$\delta = 0{,}15$	$\delta = 0{,}25$	$\delta = 0{,}15$	$\delta = 0{,}25$
p_0	$n(r_1)$	$n(r_1)$	$n(r_1)$	$n(r_1)$
0.05	27(4)	14(3)	38(5)	16(3)
0.10	40(8)	18(5)	55(10)	25(6)
0.15	48(12)	22(7)	64(15)	27(8)
0.20	56(17)	21(8)	77(22)	29(10)
0.25	62(22)	26(11)	83(28)	33(13)
0.30	67(27)	25(12)	93(36)	36(16)
0.35	68(31)	26(14)	96(42)	36(18)
0.40	71(36)	28(16)	94(46)	34(19)
0.45	70(39)	25(16)	98(53)	36(22)
0.50	69(42)	23(16)	93(55)	33(22)
0.55	70(46)	24(18)	92(59)	32(23)
0.60	62(44)	21(17)	85(59)	27(21)
0.65	55(42)	20(17)	75(56)	24(20)
0.70	49(40)	14(13)	69(55)	19(17)
0.75	45(39)	11(11)	55(47)	11(11)
0.80	30(28)	-	44(40)	-
0.85	19(19)	-	19(19)	-

Označme

$$m = \sum_{i=1}^{n} X_i,$$

$$D_k = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \quad \text{a} \quad H_k = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

Za platnosti nulové hypotézy existují pro každé n a p kritické hodnoty $r_1(n,p)$ a $r_2(n,p)$ takové, že $D_{r_1(n,p)+1} > \alpha/2$, $D_{r_1(n,p)} \leq \alpha/2$, $H_{r_2(n,p)-1} > \alpha/2$ a $H_{r_2(n,p)} \leq \alpha/2$. Nulovou hypotézu zamítneme, když $m \geq r_2(n,p_0)$ nebo $m \leq r_1(n,p_0)$, test přitom bude mít hladinu α. Velikost výběru při zvoleném δ získáme numerickým řešením nerovnice

$$\min_n : \{P(m \geq r_2(n,p_0) \vee m \leq r_1(n,p_0) | |p - p_0| = \delta) \geq 1 - \beta\}.$$

Zanedbáme-li hodnotu menší než $\alpha/2$, hledáme nejmenší n takové, že menší z výrazů

$$P(m \geq r_2(n,p_0)|p = p_0 + \delta) \quad \text{a} \quad P(m \leq r_1(n,p_0)|p = p_0 - \delta) \qquad (3.2)$$

je větší než $1 - \beta$. Protože platí

$$\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=n-k}^{n} \binom{n}{i} p^{n-i} (1-p)^i,$$

je $r_1(n,p) = n - r_2(n, 1-p)$,

$$\begin{aligned} P(m \leq r_1(n,p_0)|p = p_0 - \delta) &= \sum_{i=0}^{r_1(n,p_0)} \binom{n}{i} (p_0 - \delta)^i (1 - (p_0 - \delta))^{n-i} \\ &= \sum_{i=r_2(n,1-p_0)}^{n} \binom{n}{i} (p_0 - \delta)^{n-i} (1 - (p_0 - \delta))^i \\ &= P(m \geq r_2(n, 1-p_0)|p = 1 - p_0 + \delta). \end{aligned}$$

Díky tomu můžeme hledat nejmenší n takové, aby menší z výrazů

$$P(m \geq r_2(n,p_0)|p = p_0 + \delta) \quad \text{a} \quad P(m \geq r_2(n, 1-p_0)|p = 1 - p_0 + \delta)$$

byl větší než $1 - \beta$. Tento problém můžeme numericky řešit podobně jako nerovnici 3.1 u jednostranného testu.

Fisherův test

Dvouvýběrová varianta předchozích testů bývá nazývána Fisherův test. Tentokrát se zaměříme pouze na jednostranný test. Jeho hypotézy jsou

$$H_0 : p_1 = p_2 \quad \text{proti} \quad H_a : p_1 > p_2,$$

kde p_1 a p_2 jsou pravděpodobnosti pozitivního výsledku v prvním, resp. druhém výběru. Na rozdíl od binomického testu jsou i za platnosti nulové hypotézy tyto parametry neznámé. Rozhodnutí o případném zamítnutí alternativní hypotézy je založeno na následující úvaze: Nechť m_1 je počet pozitivních výsledků v prvním a m_2 v druhém výběru, n_1 a n_2 jsou velikosti výběrů. Pak celkový počet pozitivních výsledků je $m = m_1 + m_2$. Za platnosti nulové hypotézy $p_1 = p_2 = p$ je

$$\begin{aligned} P(m_1 = i | m_1 + m_2 = m) &= \frac{P(m_1 = i, m_1 + m_2 = m)}{P(m_1 + m_2 = m)} \\ &= \frac{P(m_1 + m_2 = m | m_1 = i) P(m_1 = i)}{P(m_1 + m_2 = m)} \\ &= \frac{P(m_2 = m - i) P(m_1 = i)}{P(m_1 + m_2 = m)} \\ &= \frac{\binom{n_1}{i} p^i (1-p)^{n_1 - i} \binom{n_2}{m-i} p^{m-i} (1-p)^{n_2 - m + i}}{\binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m}} \\ &= \frac{\begin{pmatrix} n_1 \\ i \end{pmatrix} \begin{pmatrix} n_2 \\ m-i \end{pmatrix}}{\begin{pmatrix} n_1 + n_2 \\ m \end{pmatrix}}. \end{aligned}$$

Pravděpodobnost, že za daných hodnot m, n_1 a n_2 bude počet pozitivních výsledků v první skupině větší nebo roven než nějaké k $(0 \leq k \leq m)$, je rovna

$$P_{k,m} = \sum_{i=k}^{m} \frac{\begin{pmatrix} n_1 \\ i \end{pmatrix} \begin{pmatrix} n_2 \\ m-i \end{pmatrix}}{\begin{pmatrix} n_1 + n_2 \\ m \end{pmatrix}}.$$

Pro daná m, n_1 a n_2 opět existuje kritická hodnota $r = r(m, n_1, n_2)$ taková, že $P_{r-1,m} > \alpha$, $P_{r,m} \leq \alpha$. Hypotézu H_0 zamítneme, jestliže $m_1 \geq r$. Přitom hladina testu bude α, protože pro každé m je pravděpodobnost chybného zamítnutí nejvýše rovna α. Zvolíme-li $\delta > 0$ a poměr mezi n_1 a n_2, hledáme $\min(n_1 + n_2)$ tak, aby

$$1 - \beta \quad \leq \quad \sum_{m=0}^{n_1+n_2} P(m_1 \geq r(m, n_1, n_2), m_1 + m_2 = m | p_1 = p_2 + \delta).$$

Kapitola 4

Skupinově-sekvenční testy

Ve výzkumu často nastává situace, kdy data jsou získávána postupně během delšího časového úseku. Tehdy může být výhodné použít některý ze skupinově-sekvenčních testů. Na rozdíl od nesekvenčního testu s pevnou velikostí výběru jsou při sekvenčním testu v určitých intervalech (např. vždy po získání jistého množství dat nebo v časových intervalech) vyhodnocována do dané doby získaná data a pokud tato vypovídají ve prospěch některé z hypotéz, je možné vyvodit příslušný závěr a test předčasně ukončit. Velikost výběru potřebná k provedení sekvenčního testu je náhodná veličina a závisí na tom, zda dojde k předčasnému ukončení. Pevně stanovena bývá jen její maximální hodnota, tj. velikost výběru v případě, že k předčasnému ukončení nedojde. Hlavním cílem takového postupu je díky šanci na předčasné ukončení snížit střední hodnotu potřebné velikosti výběru. Pokud je mezi testovanými přípravky dostatečný rozdíl, bude střední hodnota potřebné velikosti výběru menší než velikost výběru při nesekvenčním testu.

Nejprve budeme uvažovat pouze speciální případ sekvenčních testů, kdy je předem stanoven počet fází a množství dat, které přibude v každé fázi, je konstantní. Myšlenky sekvenčních testů si předvedeme na dvouvýběrovém testu shodnosti středdních hodnot normálně rozdělených náhodných veličin se stejným rozptylem. Při jiných testech je někdy postup podobný, někdy může být odvození složitější, popisovat ho zde ale již nebudeme.

Zvolíme-li počet fází jako K a velikost skupiny n, rozdělí se test na K částí a v každé z nich je získáno n nových údajů ($2n$ z obou výběrů dohromady). Označme množství dat z jednoho výběru, které vyhodnocujeme v k-té fázi ($k = 1 \dots K$), jako n_k. Protože vyhodnocujeme všechna do dané doby získaná data, je $n_k = nk$. V k-té fázi spočítáme testovou statistiku

$$Z_k = \frac{1}{\sqrt{2n_k\sigma^2}} \left(\sum_{j=1}^{n_k} X_i - \sum_{j=1}^{n_k} Y_j \right),$$

která má při platnosti nulové hypotézy standardní normální rozdělení. Nulovou hypotézu $H_0 : \mu_1 = \mu_2$ zamítneme a test ukončíme, pokud hodnota veličiny $|Z_k|$ překročí určitou kritickou hodnoutu c_k. Když ani v jedné fázi není kritická hodnota překročena, test končí a nulovou hypotézu nezamítáme.

Někdy je žádoucí mít možnost test předčasně ukončit ve prospěch nulové hypotézy. Pak bývají stanoveny (dolní) kritické hodnoty $b_1, \ldots, b_K$, a když je $|Z_k| < b_k$ pro $k \in (1, \ldots, K-1)$, můžeme test předčasně ukončit, aniž bychom zamítnuli nulovou hypotézu. Posloupnosti $c_1, \ldots, c_K$ a $b_1, \ldots, b_K$ kritických hodnot jsou přitom vybrány tak, aby celý test měl určenou hladinu. Způsobů, jak to udělat, se používá více. Nejprve si uvedeme několik variant, které uvažují možnost předčasného ukončení pouze při zamítnutí nulové hypotézy. Všechny členy posloupnosti $b_1, \ldots, b_K$ budou proto rovny nule. Pak se podíváme na testy obecnějšího charakteru, včetně testů, pro které není pevně stanoven počet fází a objem dat získaných v každé fázi.

4.1 Pocockův test

Při Pocockově testu rozhodujeme v každé fázi o zamítnutí nulové hypotézy na stejné hladině, tedy prvky posloupnosti $c_1, \ldots, c_K$ volíme shodné a tak, aby hladina celého testu byla α. Označme $C_P(K, \alpha) = c_1 = \ldots = c_K$. Schéma vyhodnocení jednotlivých fází je následující:

Fáze $k = 1 \ldots K-1$,

- když $|Z_k| > C_P(K, \alpha)$, zamítáme H_0 a test ukončíme,
- jinak pokračujeme fází $k+1$.

Fáze K,

- když $|Z_K| > C_P(K, \alpha)$, zamítáme H_0 a test ukončíme,
- jinak nezamítáme H_0 a test ukončíme.

Tabulka 4.1: Kritické hodnoty pro Pocockův a O'Brianův-Flemingův test

	$C_P(K,\alpha)$		$C_B(K,\alpha)$	
K	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$
1	2,576	1,960	2,576	1,960
2	2,772	2,178	2,580	1,977
3	2,873	2,289	2,595	2,004
4	2,939	2,361	2,609	2,024
5	2,986	2,413	2,621	2,040
6	3,023	2,453	2,631	2,053
10	3,117	2,555	2,660	2,087
15	3,182	2,626	2,681	2,110
20	3,225	2,672	2,695	2,126

Poznamenejme, že nestačí vzít $C_P(K,\alpha) = u_{\alpha/2}$, protože pak by kvůli opakovanému vyhodnocování pravděpodobnost chyby I. druhu hodnotu α překročila. Některé numericky spočítané hodnoty $C_P(K,\alpha)$ jsou uvedeny v tabulce 4.1. Postup vedoucí k jejich získání je možné nalézt v knize Jennison, Turnbull (2000), z této knihy jsou také převzaty v této kapitole uvedené tabulky.

Označme $\epsilon = \mu_1 - \mu_2$. Při platnosti alternativní hypotézy závisí maximální velikost výběru ($n_K = nK$) na α, β a K a je přímo úměrná σ^2/ϵ^2. Velikost výběru pro jednofázový test je také přímo úměrná σ^2/ϵ^2 a její hodnotu jsme odvodili v kapitole 2. Pro určení velikosti výběru n_K nutné k dosažení síly $1-\beta$ tedy stačí znát poměr $R_P(K,\alpha,\beta)$ mezi n_K a pevnou velikostí výběru jednofázového testu. Velikosti $R_P(K,\alpha,\beta)$ pro některé hodnoty proměnných jsou uvedeny v tabulce 4.2. Maximální velikost výběru pro K-fázový test je tak rovna

$$n_K = \frac{(u_{\alpha/2} + u_\beta)^2 2\sigma^2}{\epsilon^2} R_P(K,\alpha,\beta).$$

Po vydělení počtem fází a zaokrouhlení na celé číslo (většinou nahoru) dostaneme velikost skupiny v každé fázi.

Tabulka 4.2: Poměry $R_P(K, \alpha, \beta)$ a $R_B(K, \alpha, \beta)$

		$1-\beta = 0{,}8$		$1-\beta = 0{,}9\%$	
	K	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$
	1	1,000	1,000	1,000	1,000
	2	1,092	1,110	1,084	1,100
	3	1,137	1,166	1,125	1,151
	4	1,166	1,202	1,152	1,183
R_P	5	1,187	1,229	1,170	1,207
	6	1,203	1,249	1,185	1,225
	10	1,243	1,301	1,222	1,271
	15	1,272	1,338	1,248	1,305
	20	1,291	1,363	1,264	1,327
	1	1,000	1,000	1,000	1,000
	2	1,001	1,008	1,001	1,007
	3	1,007	1,017	1,006	1,016
	4	1,011	1,024	1,010	1,022
R_B	5	1,015	1,028	1,014	1,026
	6	1,017	1,032	1,016	1,030
	10	1,024	1,040	1,022	1,037
	15	1,028	1,045	1,026	1,042
	20	1,030	1,047	1,029	1,045

4.2 O'Brienův-Flemingův test

Při O'Brienově-Flemingově testu je posloupnost kritických hodnot zvolena jako

$$c_k = C_B(K, \alpha)\sqrt{K/k}$$

a je tedy klesající. Proto je narozdíl od Pocockova testu menší pravděpodobnost, že nulovou hypotézu zamítneme v počátečních fázích testu, ale větší pravděpodobnost, že ji zamítneme v pozdějších fázích. Přitom $C_B(K, \alpha)$ vezmeme opět tak, aby pravděpodobnost chyby I. druhu byla α. Některé hodnoty $C_B(K, \alpha)$ jsou uvedeny v tabulce 4.1. Schéma testu je podobné jako u Pocockova testu:

Fáze $k = 1 \dots K-1$,

– když $|Z_k| > C_B(K,\alpha)\sqrt{K/k}$, zamítáme H_0 a test ukončíme,

– jinak pokračujeme fází $k+1$.

Fáze K,

– když $|Z_K| > C_B(K,\alpha)$, zamítáme H_0 a test ukončíme,

– jinak nezamítáme H_0 a test ukončíme.

Pro určení velikosti výběru nám také stejně jako u Pocockova testu stačí znát poměr $R_B(K,\alpha,\beta)$ mezi maximální velikostí výběru vícefázového testu a pevnou velikostí výběru jednofázového testu. Velikosti tohoto poměru pro některé hodnoty proměnných jsou uvedeny v tabulce 4.2. Maximální velikost výběru O'Brienova-Flemingova testu je tedy rovna

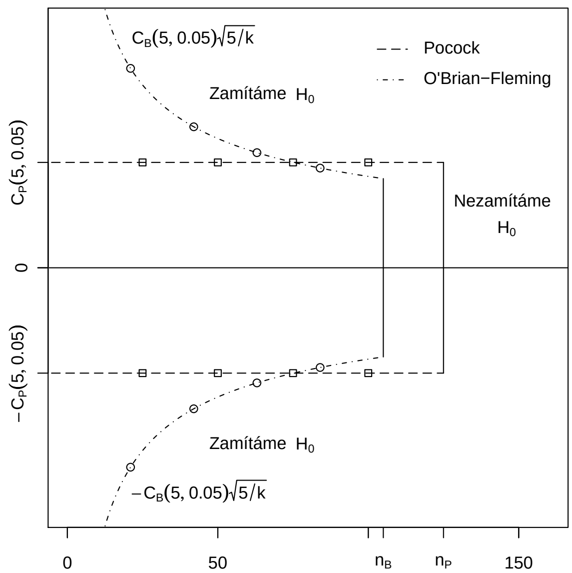
$$n_K = \frac{(u_{\alpha/2} + u_\beta)^2 2\sigma^2}{\epsilon^2} R_B(K,\alpha,\beta).$$

Srovnání

Sekvenční testy můžeme srovnávat z různých hledisek, nejdůležitější jsou ale většinou otázky týkající se střední hodnoty velikosti výběru, případně jeho maximální velikosti. Maximální velikost výběru zde uvedených skupinových sekvenčních testů je větší než u nesekvenčního testu při stejné hladině a síle. Je to způsobeno tím (volně řečeno), že v počátečních fázích sekvenčního testu použijeme část povolené hladiny testu, ale počet pozorování je zatím nízký a získaná síla proto malá.

Oba uvedené skupinově-sekvenční testy dávají možnost výzkum předčasně ukončit ve prospěch alternativní hypotézy, proto je-li zkoumaný lék dostatečně účinný, počet pacientů potřebných k provedení jednoho ze sekvenčních testů bude pravděpodobně menší, než u nesekvenčního testu. Protože kritické hodnoty Pocockova testu jsou v počátečních fázích menší a na konci naopak větší než u O'Brianova-Flemingova testu (Pocockův test využívá větší část povolené hladiny při menším počtu pozorování), je k dosažení požadované síly při Pocockově testu potřeba větší maximální velikost výběru než u O'Brianova-Flemingova testu.

Pocockův test je výhodnější, když je rozdíl mezi léky velmi výrazný a díky tomu je velká pravděpodobnost ukončení testu už v počátečních fázích.



Obrázek 4.1: Srovnání kritických hodnot Pocockova a O'Brianova--Flemingova testu pro pětifázový ($K = 5$) sekvenční test na hladině 5 % při velikosti výběru nesekvenčního testu $n = 100$. Maximální velikost výběru je $n_P = 125$ (Pocock), resp. $n_B = 105$, $C_P(5, 0{,}05) = 2{,}413$, $C_B(5, 0{,}05) = 2{,}040$. Pokud se hodnota testové statistiky v některé fázi více odchýlí od nuly a v příslušném bodě (dle prováděného testu) překročí přerušovanou čáru, zamítneme nulovou hypotézu.

V praxi ale spíše nastává situace, kdy rozdíl je malý, proto se častěji užívá O'Brianův-Flemingův test, jehož maximální velikost výběru je jen o málo větší, než u nesekvenčního testu. Přitom maximální velikost výběru budeme potřebovat s pravděpodobností alespoň $1 - \alpha$ vždy, když platí nulová hypotéza. Tabulka 4.3 srovnává konkrétní hodnoty na příkladě. Menší rozdíly v hladině testů jsou způsobeny zaokrouhlením velikostí skupin na celé číslo. Obrázek 4.1 nabízí grafické porovnání obou testů.

4.3 Wangův-Tsiatsiho test

Wangův-Tsiatsiho test je zobecněním Pocockova a O'Brianova-Flemingova testu. Kritická hodnota c_k se vezme jako $c_k = C_{WT}(K, \alpha, \beta)(k/K)^{\Delta - 1/2}$, Tabulka 4.3: Srovnání nesekvenčního testu, pětifázového Pocockova a pětifázového O'Brianova-Flemingova testu. Hladina testů je 5 %, síla je 90 %, když $|\mu_1 - \mu_2| = 1$. Rozptyl $\sigma^2 = 4$.

	Nesekvenční test	Pocockův test	O'Brianův-Flemingův test
	Pravděpodobnost zamítnutí H_0		
$\|\mu_1 - \mu_2\| = 0{,}0$	0,050	0,050	0,050
0,5	0,371	0,351	0,378
1,0	0,903	0,910	0,912
1,5	0,998	0,999	0,999
	Stř. hod. velikosti výběru		
$\|\mu_1 - \mu_2\| = 0{,}0$	170,0	204,8	178,7
0,5	170,0	182,3	167,9
1,0	170,0	116,9	129,8
1,5	170,0	70,1	94,4
	Rozptyl velikosti výběru		
$\|\mu_1 - \mu_2\| = 0{,}0$	0,0	26,1	8,6
0,5	0,0	50,8	24,7
1,0	0,0	57,9	35,5
1,5	0,0	34,1	25,7

kde Δ je parametr volený nejčastěji z intervalu $[0, \frac{1}{2}]$. Pro $\Delta = \frac{1}{2}$ dostaneme Pocockův test, při $\Delta = 0$ jde o O'Brianův-Flemingův test. Hodnoty funkce $C_{WT}(K, \alpha, \beta)(k/K)^{\Delta - 1/2}$ a hodnoty poměru mezi velikostí výběru sekvenčního a nesekvenčního testu pro různé velikosti proměnných jsou uvedeny v knize Jennison, Turnbull (2000), kapitola 2.

4.4 Test Inner-Wedge

Doposud uvedené sekvenční testy umožnovaly ukončit výzkum předčasně a přiklonit se k alternativní hypotéze, pokud data shromážděná do určitého období svědčí v její prospěch. Test Inner-Wedge dává možnost ukončit výzkum předčasně i v případě, že data nesvědčí o platnosti alternativní hypotézy a je značně nepravděpodobné, že by se situace ještě změnila. Schéma tohoto testu je následující:

Fáze $k = 1 \dots K-1$,

– když $|Z_k| \geq c_k$, zamítáme H_0 a test ukončíme,

– když $|Z_k| < b_k$, nezamítáme H_0 a test ukončíme,

– jinak pokračujeme fází $k+1$.

Fáze K,

– když $|Z_K| \geq c_K$, zamítáme H_0 a test ukončíme,

– když $|Z_K| < b_K$, nezamítáme H_0 a test ukončíme.

Kritické hodnoty $b_k < c_k$, $k = 1 \dots K-1$, $b_K = c_K$, jsou zvoleny tak, aby test měl hladinu α. Rovnost $b_K = c_K$ zaručuje, že test skončí nejpozději po K fázích.

Někdy není vhodné, aby bylo možné test ukončit ve prospěch nulové hypotézy již v počátečních fázích, první členy posloupnosti $(b_1, \dots, b_K)$ pak bývají položeny rovny nule. Běžná volba c_k a b_k je:

$$\begin{aligned} c_k &= C_1(K, \alpha, \beta, \Delta)\left(\tfrac{k}{K}\right)^{\Delta-1/2}, \\ b_k &= [C_1(K, \alpha, \beta, \Delta) + C_2(K, \alpha, \beta, \Delta)]\sqrt{\tfrac{k}{K}} - C_2(K, \alpha, \beta, \Delta)\left(\tfrac{k}{K}\right)^{\Delta-1/2}, \end{aligned}$$

kde Δ je volitelný parametr. Hodnoty $C_1(K, \alpha, \beta, \Delta)$, $C_2(K, \alpha, \beta, \Delta)$ a poměru $R_W(K, \alpha, \beta, \Delta)$ mezi velikostí výběru sekvenčního a nesekvenčního testu je možné najít v knize Jennison, Turnbull (2000), kap. 5.

4.5 Spotřebování hladiny testu

Nedostatkem zatím uvedených sekvenčních testů je fakt, že předpokládají dopředu stanovený počet fází a pevný počet subjektů ve skupině. V praxi se ale takový předpoklad může ukázat jako těžko splnitelný. Například vyhodnocování je někdy nutné provádět po určených časových intervalech, což ale nemůže zaručit shodné velikosti skupin a ani počet fází nutný k shromáždění určitého množství dat není možné zvolit pevně. Jedním z možných přístupů k tomuto problému je použití funkce spotřebování hladiny testu (angl. alpha spending function). Zde si představíme dvě metody využívající spotřebování hladiny testu. Bližší informace týkající se konkrétních velikostí kritických hodnot a velikosti výběru uvádí Jennison, Turnbull (2000), kap. 7.

Metoda při pevném počtu fází

První metoda přepokládá předem zvolený počet fází K, v každé ale může být shromážděno různé množtví dat. Celkovou pravděpodobnost chyby prvního druhu rozdělíme na K pravděpodobností $\alpha_1 \dots \alpha_K$ tak, aby $\sum_{k=1}^K \alpha_k = \alpha$. Kritické hodnoty $c_1 \dots c_k$ zvolíme tak, aby za platnosti nulové hypotézy

$$P\{|Z_1| < c_1, \dots, |Z_{k-1}| < c_{k-1}, |Z_k| \geq c_k\} = \alpha_k. \tag{4.1}$$

V každé fázi spotřebujeme část z povolené pravděpodobnosti, že nesprávně zamítneme nulovou hypotézu, část spotřebovaná v k-té fázi je díky rovnici 4.1 rovna α_k, a protože $\sum_{k=1}^K \alpha_k = \alpha$, celková hladina testu je α. Schéma vyhodnocení fází testu je analogické například jako u Pocockova testu, změní se pouze kritické hodnoty.

Nevýhodou tohoto postupu je nutnost dopředu zvolit celkový počet fází a hodnoty α_i, které budou spotřebovány v jednotlivých fázích, což může způsobit problém například při neočekávaných výkyvech v rychlosti získávání dat (když vyhodnocujeme v časových intervalech). Druhý přístup umožnuje počet fází a hodnoty α_i přizpůsobit tomu, jak probíhá získávání dat.

Metoda maximálních dat

Metoda maximálních dat je založena na použití funkce spotřebování hladiny testu $f(t)$. Hodnota funkce f v bodě t přitom značí, jaká čast z celkové hladiny testu má již být spotřebována v momentě, kdy množství shromážděných dat je rovno tn_{max}, kde n_{max} je předem stanovená horní mez množství dat (maximální velikost výběru). Funkce f je definována tak, aby na intervalu $[0, 1]$ byla neklesající, $f(0) = 0$ a $f(t) = \alpha$ pro $t \geq 1$.

Funkci spotřebování hladiny testu a horní mez potřebného množští dat je třeba zvolit před zahájením studie. Hodnoty $\alpha_1, \alpha_2, \dots$ jsou počítány až při vlastním testu jako

$$\begin{aligned}\alpha_1 &= f(n_1/n_{max}),\\ \alpha_k &= f(n_k/n_{max}) - f(n_{k-1}/n_{max}), \quad k = 2, 3, \dots,\end{aligned}$$

kde n_k je množství dat shromážděných do k-té fáze včetně. Kritické hodnoty c_k pak mohou být opět spočítány z rovnice 4.1. Příkladem funkcí spotřebovávajících hladinu jsou funkce

$$f(t) = \min(\alpha \ln[1 + (e - 1)t], \alpha), \tag{4.2}$$

$$f(t) = \min(2 - 2\Phi(z_{\alpha/2}/\sqrt{t}), \alpha), \tag{4.3}$$

jimž odpovídající kritické hodnoty jsou při stejně velkých skupinách blízké kritickým hodnotám Pocockova (4.2) a O'Brianova-Flemingova (4.3) testu, viz Lan, DeMets (1983).

Dodatek A

Výpočet velikosti výběru pro jednostranný binomický test v programu R

```
#Velikost vyberu jednostranneho binomickeho testu s parametry alfa, beta, p, delta.
a=0; b=0; c=0; d=0; g=0; h=0; m=0 #pomocne promenne

krit=function(n,p,alfa){
for(i in seq(from=1, to=n+1)) a[i]=(choose(n,i-1)*p^(i-1)*(1-p)^(n-i+1))
for(j in seq(from=1, to=n)) b[j+1]=sum(a[-(1:j)])
b[1]=1
c=b[b>alfa]
r=length(c)} #Funkce spocita pro dane n, p a alfa kritickou hodnotu r.

sila=function(n,r,px){
for(i in seq(from=1, to=n+1)) d[i]=(choose(n,i-1)*px^(i-1)*(1-px)^(n-i+1))
s=sum(d[-(1:r)])} #Pst, ze velicina s Bi(n,p) nabude hodnotu vetsi nebo rovnu r.

csila=function(n,p,alfa,delta){z=krit(n,p,alfa)
w=sila(n,z,p+delta)} #Hodnota fce sila pri z = krit. hodnota pro dane n.

vysl=function(p,alfa,beta,delta){j=1
g[1]=(do.call(csila,list(1,p,alfa,delta)))
while(g[j]<(1-beta)){j=j+1
g[j]=(do.call(csila,list(j,p,alfa,delta)))}
m=j} #Nejmensi n takove, ze fce csila pro toto n a jemu prislusne r prekroci hodnotu 1-beta.

alfa=0.05; beta=0.2; delta=0.15 # Zvolené hodnoty
```

```
do=17 #Zvolit podle hodnoty delta, aby do*0.05+delta≤1
u=0; h=0

#Vypsani vektoru vyslednych velikosti vyberu a kritickych hodnot pro ruzne
velikosti par. p.
for (j in seq(from=1, to=do)) u[j]=do.call(vysl,list(j*0.05,alfa,beta,delta))
u
for (i in seq(from=1, to=do)) h[i]=do.call(krit,list(u[i],i*0.05,alfa))
h
```

Literatura

[1] Anděl J. (2005): *Základy matematické statistiky*, Matfyzpress, Praha.

[2] Chow S. C., Shao J., Wang H. (2003): *Sample Size Calculations in Clinical Research*, Marcel Dekker, Inc., New York.

[3] Jennison C., Turnbull B. W. (2000): *Group Sequential Methods with Applications to Clinical Trials*, Chapman & Hall, Boca Raton.

[4] Lachout P. (1998): *Teorie pravděpodobnosti*, Karolinum, Praha, 74–78.

[5] Lan K. K. G., DeMets D. L. (1983): *Discrete sequential boundaries for clinical trials*, Biometrika **70**, 659–663.

[6] R Development Core Team (2006). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

[7] Wellek S. (2002): *Testing Statistical Hypotheses of Equivalence*, Chapman & Hall, Boca Raton, 29–30.