

UNIVERZITA PALACKÉHO V OLOMOUCI  
PŘÍRODOVĚDECKÁ FAKULTA

## BAKALÁŘSKÁ PRÁCE

Klinické studie a statistika



**Katedra matematické analýzy a aplikací matematiky**

Vedoucí práce: **doc. Mgr. Ondřej Vencálek, Ph.D.**

Vypracoval(a): **Jáchym Jelínek**

Studijní program: Aplikovaná matematika

Specializace: Matematika v ekonomické praxi

Forma studia: prezenční

Rok odevzdání: 2024

# BIBLIOGRAFICKÁ IDENTIFIKACE

**Autor:** Jáchym Jelínek

**Název práce:** Klinické studie a statistika

**Typ práce:** Bakalářská práce

**Pracoviště:** Katedra matematické analýzy a aplikací matematiky

**Vedoucí práce:** doc. Mgr. Ondřej Vencálek, Ph.D.

**Rok obhajoby práce:** 2024

**Abstrakt:** Pomocí klinických studií se vyvíjejí a testují léčiva. V této práci představíme, kde v klinických studiích hraje roli matematika a statistika. Nejpodrobněji se zaměříme na velikost testovací skupiny subjektů. Odvodíme její výpočet při použití jednovýběrového z-testu, dvouvýběrového t-testu, jednofaktorové analýzy rozptylu, Chí-kvadrát testu nezávislosti a McNemarova testu efektu ošetření. Poté provedeme stejný výpočet pomocí analytických programů SAS Studio a RStudio. Výsledky našich výpočtů a výsledky použitého softwaru porovnáme.

**Klíčová slova:** klinické studie, statistika, velikost testovací skupiny

**Počet stran:** 56

**Počet příloh:** 1

**Jazyk:** český

## BIBLIOGRAPHICAL IDENTIFICATION

**Author:** Jáchym Jelínek

**Title:** Clinical studies and statistics

**Type of thesis:** Bachelor's

**Department:** Department of Mathematical Analysis and Application of Mathematics

**Supervisor:** doc. Mgr. Ondřej Vencálek, Ph.D.

**The year of presentation:** 2024

**Abstract:** Clinical studies are used for development and testing of new medications. We will look at statistic side of this process. We will focus on required sample size of patients for the study. We will derive the formula for one-sample z-test, two-sample t-test, one-factor ANOVA, Chi-square test of independence and McNemar's test. Then, we will make the same calculations in programs SAS Studio and RStudio. We will compare our calculations and calculations by the software.

**Key words:** clinical studies, statistics, sample size

**Number of pages:** 56

**Number of appendices:** 1

**Language:** czech

### **Prohlášení**

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením pana doc. Mgr. Ondřeje Vencálka, Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci, dne .....

.....  
podpis



## **Poděkování**

Rád bych tímto poděkoval mému vedoucímu práce, panu doc. Mgr. Ondřeji Vencáčkovi, Ph.D., za velkou podporu, ochotu a čas, který mi při tvorbě této práce věnoval.

# Obsah

<b>1</b>	<b>Úvod</b>	<b>7</b>
<b>2</b>	<b>Klinické studie</b>	<b>8</b>
2.1	Průběh studie . . . . .	8
2.2	Cíl studie . . . . .	9
2.3	Výběr testovací skupiny a uspořádání studie . . . . .	11
<b>3</b>	<b>Velikost testovací skupiny</b>	<b>12</b>
3.1	Používané statistické testy . . . . .	12
3.2	Chyba I. a II. druhu . . . . .	13
3.3	Velikost účinku . . . . .	14
3.4	Drop-out rate . . . . .	15
<b>4</b>	<b>Velikost testovací skupiny podle zvolené metody</b>	<b>18</b>
4.1	Jednovýběrový z-test . . . . .	18
4.2	Dvouvýběrový t-test . . . . .	21
4.3	Jednofaktorová ANOVA . . . . .	25
4.4	Chí-kvadrát test nezávislosti . . . . .	29
4.5	McNemarův test efektu ošetření . . . . .	31
<b>5</b>	<b>Použití softwaru pro výpočet velikosti testovací skupiny</b>	<b>38</b>
5.1	Jednovýběrový z-test . . . . .	38
5.2	Dvouvýběrový t-test . . . . .	41
5.3	Jednofaktorová ANOVA . . . . .	45
5.4	Chí-kvadrát test nezávislosti . . . . .	48
5.5	McNemarův test efektu ošetření . . . . .	50
<b>6</b>	<b>Závěr</b>	<b>53</b>

# 1 Úvod

Každý z nás již někdy užíval léky, dostal očkování nebo podstoupil lékařské vyšetření. Těžko jsme přitom ale přemýšleli nad tím, kolik práce, času a peněz stojí jen za takovou pilulkou na alergii. Součástí vývoje každého léčiva nebo lékařské metody je tzv. klinická studie. Cílem této práce bude klinické studie představit a ukázat, jak moc s nimi souvisí matematika a statistika. Nejpodrobněji se zaměříme na to, jak velká by měla být testovací skupina subjektů pro konkrétní klinickou studii. Chtěli bychom ukázat, jak tuto velikost vypočítat, a poté jak tento výpočet provést pomocí používaných analytických programů.

V první kapitole si představíme klinické studie jako takové. Vysvětlíme, proč jsou důležité, jak probíhají a jak se mohou lišit cíle jednotlivých studií. Nakonec se v této kapitole podíváme na výběr testovací skupiny, protože právě testovací skupině se budeme v dalších kapitolách věnovat nejvíce.

Ve druhé kapitole se zaměříme na faktory, které mají při plánování studie vliv na rozhodnutí o velikosti testovací skupiny. Podíváme se na nejčastěji používané statistické testy, vysvětlíme statistickou chybu I. a II. druhu, velikost účinku a tzv. drop-out rate studie.

Ve třetí kapitole se vrátíme k nejčastěji používaným statistickým testům v klinických studiích. Spojením znalostí o těchto testech a zavedených pojmů z předchozích kapitol pro každý test odvodíme vzorec pro výpočet velikosti testovací skupiny.

V poslední kapitole porovnáme náš výpočet velikosti testovací skupiny pomocí odvozených vzorců z předchozí kapitoly a výpočet pomocí softwaru. Ukážeme, jak úlohu řešit v analytických programech SAS Studio a RStudio.

## 2 Klinické studie

Informace o klinických studiích v této kapitole jsou čerpány z knihy Klinické studie v praxi (viz [2]).

Klinická studie představuje klíčový nástroj při vývoji nových léčiv a léčebných metod. Jejím hlavním cílem je pečlivé testování účinnosti a bezpečnosti léčivých přípravků na vybrané skupině pacientů nebo dobrovolníků. Pomocí klinické studie jsou získávána data a důkazy, které jsou nezbytné pro rozhodnutí o dalším použití daného léčiva či metody.

Rostoucí důraz na účinnost a bezpečnost léčiv s sebou nese i rostoucí finanční a časovou náročnost jejich vývoje. Průměrně z 50 000 testovaných molekul postoupí jen 500 molekul do další fáze studie. Na těchto 500 molekulách potom probíhá série laboratorních a tzv. preklinických testů na zvířatech, aby se zajistila maximální bezpečnost před prvním testováním na lidech. Do samotné klinické studie postoupí jen 5 molekul, ze kterých se obvykle jen u 1 molekuly podaří prokázat dostatečnou účinnost a bezpečnost pro použití v praxi. Proces vývoje, testování, registrace a uvedení na trh této 1 molekuly trvá 10 – 15 let a stojí 800 mil. – 1 mld. USD.

### 2.1 Průběh studie

Klinická studie probíhá ve čtyřech fázích, které na sebe postupně navazují.

#### **Fáze I**

V této fázi je testované léčivo poprvé podáno člověku. Probíhá na velmi malém vzorku subjektů a jsou během ní shromažďovány např. informace o bezpečnosti, toleranci a nežádoucích účincích.

## **Fáze II**

Ve II. fázi probíhá tzv. pilotní studie na malé skupině subjektů (nejčastěji pár desítek lidí), kdy se kromě bezpečnosti léčiva poprvé testuje i jeho účinnost. Podle pilotní studie se také upravuje připravovaná hlavní studie v další fázi.

## **Fáze III**

Ve III. fázi se léčivo testuje na již mnohem větších skupinách subjektů. Zjišťují se podrobné účinky léčiva, bezpečnost, nebo např. výskyt a četnost nežádoucích účinků. V této fázi také bývá léčivo porovnáváno s jiným již zavedeným léčivem na trhu.

## **Fáze IV**

Poslední fáze klinické studie probíhá až po registraci léčiva. Cílem této fáze je sbírání informací o účinku a bezpečnosti léčiva na mnohem rozsáhlejší a méně selektované populaci, než tomu bylo v předchozích fázích.

V posledních letech se však stále častěji ustupuje od tohoto dělení studie na 4 fáze. Místo toho si doktoři a výzkumníci volí takový postup, který se nejvíce hodí pro jejich primární cíl studie, případně různě kombinují jednotlivé fáze dohromady.

## **2.2 Cíl studie**

Existují různé primární cíle studie. Může to být např. zavedení nového léčiva na trh, dokázání stejného účinku více srovnávaných léčiv, nebo třeba prokázání účinnosti nové léčebné metody. Třemi základními hypotézami v klinických studiích, kterými můžeme tyto cíle matematicky popsat, jsou ekvivalence, superiorita a non-inferiorita. Při testování těchto hypotéz většinou

zkoumané léčivo porovnáváme s jiným kontrolním léčivem, případně s placebem.

### **Ekvivalence**

Pokud je primárním cílem studie prokázat, že hodnocené a kontrolní léčivo jsou stejně nebo velmi podobně účinné, pak budeme testovat hypotézu o ekvivalenci. Při testování ekvivalence musíme určit hranici, či maximální přípustný rozdíl, v rámci které budeme léčiva ještě považovat za stejně účinná. S tímto cílem studie se setkáme např. při vývoji nových léčiv, ke kterým již existují alternativy, se kterými je můžeme srovnávat.

### **Superiorita**

Je-li primárním cílem studie dokázat, že testované léčivo je lepší než kontrolní, pak budeme testovat hypotézu o superioritě. U tohoto cíle studie je nutné zvolit tzv. klinicky významný rozdíl, který hodnotí, zda je testované léčivo o tolik lepší, aby se vyplatilo jím nahradit kontrolní léčivo (např. prodloužení života nemocného o rok navíc).

### **Non-inferiorita**

Pokud je primárním cílem studie dokázat, že testované léčivo není horší než kontrolní léčivo, pak budeme testovat hypotézu o non-inferioritě. Při této hypotéze je opět nutné stanovit tzv. klinicky významný rozdíl, který nyní bude určovat, o kolik horší by muselo být testované léčivo, aby se skutečně nedalo považovat za stejně účinné jako kontrolní léčivo.

## 2.3 Výběr testovací skupiny a uspořádání studie

Dalším důležitým aspektem přípravy studie je vybrat testovací skupinu subjektů tak, aby se dala co nejlépe zobecnit na cílovou populaci. Proto musí být kritéria pro zařazení do studie dostatečně volná, ale zároveň je potřeba dát pozor na nechtěné korelace v testovací skupině, jako mohou být např. dvojčata. Uspořádáním se testované subjekty rozdělí do více menších skupin. Nejpoužívanějšími jsou paralelní uspořádání subjektů a cross-over uspořádání subjektů. Obě tato uspořádání si blíže vysvětlíme níže. V kombinaci s těmito uspořádáními se často používá tzv. dvojitě zaslepení studie. Testovaný subjekt ani lékař potom neví, zda subjekt užívá testované léčivo, nebo kontrolní léčivo či placebo.

### Paralelní uspořádání

Při paralelním uspořádání jsou subjekty náhodně rozděleny do více (většinou stejně velkých) skupin. Každé skupině je potom podáváno jiné léčivo či placebo. Každá z těchto skupin se označuje jako rameno studie. Většinou se vytvoří tolik ramen studie, kolik je srovnávaných léčiv.

### Cross-over uspořádání

Při cross-over uspořádání jsou subjekty studie náhodně rozděleny do sekvencí. Sekvence, do které daný subjekt spadá, určí, v jakém pořadí budou subjektu podávány srovnávané přípravky. Například při srovnávání léčiv  $A$  a  $B$  budou subjekty rozděleny do dvou sekvencí. Subjekty v první sekvenci budou nejdříve dostávat léčivo  $A$  a poté léčivo  $B$ , a subjekty ve druhé sekvenci dostanou léčiva přesně naopak. Při srovnávání  $N$  různých léčiv bude maximální počet různých sekvencí roven  $N!$ .

### 3 Velikost testovací skupiny

Po zvolení primárního cíle studie a zvolení uspořádání testovaných subjektů můžeme začít řešit, kolik subjektů vlastně do studie potřebujeme. Anglicky se tento proces označuje jako „Sample size calculation“. Odhad počtu subjektů, které budeme ke studii potřebovat, je klíčový pro výpočet nákladů a naplánování celé studie. Podcenění počtu testovaných subjektů totiž může znamenat neprůkazné výsledky celé studie a tedy zbytečně vynaložené finance a čas. Zároveň nesmí být testovaných subjektů zbytečně moc, a to z důvodů finančních i etických. Optimální počet testovaných subjektů je takový, aby šly výsledky studie zobecnit na reálné prostředí v akceptovatelné míře pravděpodobnosti.

Existuje mnoho faktorů, které mají vliv na požadovanou velikost testovací skupiny. Hlavními jsou cíl studie (viz 2.2), uspořádání subjektů ve studii (viz 2.3), typ primární sledované veličiny (diskrétní, spojitý), zvolené statistické metody (viz 3.1), hladina významnosti a síla testu (viz 3.2), klinicky významný rozdíl – velikost účinku (viz 3.3), a tzv. drop-out rate (viz 3.4).

Čerpáno z knihy *Klinické studie v praxi* (viz [2], Kapitola 4.4.4).

#### 3.1 Používané statistické testy

Volba metody (statistického testu) pro test hypotézy o primárním parametru hraje velkou roli v tom, kolik bude ve studii potřeba subjektů. Obecně platí, že u metod pro testování diskrétního parametru bývá potřeba více subjektů než u metod pro spojitý parametr (viz [2], str. 52). Oboustranné testy se používají pro prokázání ekvivalence a jednostranné testy se používají pro prokázání non-inferiority a superiority (viz Kapitola 2.2).



### **Spojité primární sledovaná veličina**

Při práci se spojitou proměnnou či proměnnými se v klinických studiích používají např. tyto metody – jednovýběrový z-test, jednovýběrový a dvouvýběrový t-test, ANOVA, dvouvýběrový Wilcoxonův test a F-test pro porovnání rozptylů.

### **Diskrétní primární sledovaná veličina**

U diskrétních proměnných se v klinických studiích používají např. tyto metody – chí-kvadrát test nezávislosti, McNemarův test efektu ošetření a Fisherův test nezávislosti.

Čerpáno z knihy *Klinické studie v praxi* (viz [2], Kapitola 4.4.9) a z knihy *Sample Size Calculations in Clinical Research* (viz [1]).

## **3.2 Chyba I. a II. druhu**

Chybou I. druhu je případ, kdy chybně zamítneme pravdivou nulovou hypotézu. Pravděpodobnost chyby I. druhu se označuje jako  $\alpha$ . Tuto pravděpodobnost také nazýváme hladinou významnosti testu. U klinických studií je akceptována hranice  $\alpha = 0.05$ , výjimečně se potom používá  $\alpha = 0.01$ .

Chybou II. druhu označujeme případ, kdy nezamítneme nulovou hypotézu, která ve skutečnosti není pravdivá. Pravděpodobnost chyby II. druhu se označuje jako  $\beta$  a úzce souvisí s termínem síla testu. Síla testu je rovna  $1 - \beta$ , tedy pravděpodobnosti zamítnutí neplatné nulové hypotézy. U klinických studií se nejčastěji používá  $\beta = 0.2$ , občas potom  $\beta = 0.1$ .

Čerpáno z knihy *Základy počtu pravděpodobnosti a metod statistiky* (viz [3], str. 175) a z knihy *Klinické studie v praxi* (viz [2], Kapitola 4.4.4).

### 3.3 Velikost účinku

Velikost účinku je míra síly jevu, která je nezávislá na jednotkách měření. Je jí libovolná statistika, která má tyto tři vlastnosti:

- odráží míru porušení nulové hypotézy (např. velikost rozdílu dvou středních hodnot),
- její střední hodnota není ovlivněna počtem subjektů,
- není ovlivněna jednotkami měření primárního parametru.

Čerpáno z knihy *Statistické metody v psychologii* (viz [6], str. 104).

Jako příklad takového ukazatele uvedeme Cohenovo  $d$ :

$$d = \frac{\mu_1 - \mu_2}{S},$$

kde parametry  $\mu_1$  a  $\mu_2$  jsou střední hodnoty náhodných veličin  $X$  a  $Y$  a parametr  $S$  je výběrová směrodatná odchylka veličin  $X$  a  $Y$ . Tento tvar velikosti účinku se vyskytuje např. ve vzorci pro určení minimální velikosti testované skupiny u dvouvýběrového t-testu (viz Kapitola 4.2). V tomto případě je u finálního vzorce pro minimální velikost testované skupiny dobře vidět, jak volba konkrétní hodnoty Cohenova  $d$  ovlivní velikost skupiny – s většími hodnotami Cohenova  $d$  připouštíme větší standardizovaný rozdíl srovnávaných středních hodnot z nulové hypotézy, a proto nám postačí menší velikost testovací skupiny, a naopak.

### 3.4 Drop-out rate

Drop-out rate je podíl subjektů z testovací skupiny, který při vyhodnocení studie nebude možné zahrnout do statistické analýzy. S ubývajícím počtem subjektů klesá průkaznost celé studie, proto je třeba se proti této situaci pojistit.

Odhadovaný drop-out rate  $p$  bereme z intervalu  $(0, 1)$ , kdy např.  $p = 0.3$  značí, že očekávaných 30 % z  $n$  subjektů studii nedokončí. Abychom na konci studie stále pracovali s  $n$  subjekty, navýšíme jejich počet před zahájením studie na  $N$  subjektů podle vzorce

$$N = n(1 + p).$$

S tímto opatřením jsme v průběhu studie připraveni na ztrátu  $pn$  subjektů se zachováním požadované hladiny významnosti  $\alpha$  a síly testu  $1 - \beta$ .

Čerpáno z knihy Sample Size Calculations in Clinical Research (viz [1], str. 39 a 40).

## Věty a definice

Uvedeme zde pár složitějších vět, které využijeme v další kapitole.

**Věta 1.** (Centrální limitní) *Nechť  $\{X_n\}_{n=1}^{\infty}$  je posloupnost nezávislých, stejně rozdělených náhodných veličin s konečnou střední hodnotou  $\mu$  a s konečným nenulovým rozptylem  $\sigma^2$ . Potom posloupnost distribučních funkcí normovaných součtů*

$$Z_n = \frac{\sum_1^n X_k - E(\sum_1^n X_k)}{\sqrt{\text{var}(\sum_1^n X_k)}} = \frac{\sum_1^n (X_k - \mu)}{\sigma\sqrt{n}}$$

*konverguje na  $R^1$  k distribuční funkci normovaného normálního rozdělení  $N(0, 1)$ , tj.*

$$\lim_{n \rightarrow \infty} P\left(\frac{\sum_1^n (X_k - \mu)}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x), \forall x \in R^1.$$

Čerpáno z knihy Základy počtu pravděpodobnosti a matematické statistiky (viz [3], str. 149).

**Věta 2.** *Nechť  $X_1, \dots, X_n$  jsou nezávislé náhodné veličiny, pro které platí  $X_i \sim N(\mu_i, 1), i = 1, \dots, n$ . Nechť  $\lambda = \sum_{i=1}^n \mu_i^2 \neq 0$ . Pak rozdělení náhodné veličiny  $Y = \sum_{i=1}^n X_i^2$  závisí jen na  $n$  a  $\lambda$  a nazývá se necentrální  $\chi^2$  rozdělení s  $n$  stupni volnosti a s parametrem necentrality  $\lambda$ . Toto rozdělení značíme  $\chi_{n,\lambda}^2$ .*

*Hustota necentrálního  $\chi^2$  rozdělení s  $n$  stupni volnosti a s parametrem necentrality  $\lambda$  je*

$$h_{n,\lambda}(z) = \frac{e^{-(z+\lambda)/2} \cdot z^{(n-2)/2}}{2^{n/2} \Gamma(\frac{n-1}{2}) \Gamma(\frac{1}{2})} \sum_{r=0}^{\infty} \frac{\lambda^r z^r}{(2r)!} B\left(\frac{n-1}{2}, \frac{1}{2} + r\right).$$

Čerpáno z knihy Základy matematické statistiky (viz [4], str. 69).

**Věta 3.** (Cramérova-Sluckého) *Nechť  $X_1, X_2, \dots$  je posloupnost náhodných veličin s distribučními funkcemi  $F_1, F_2, \dots$ . Nechť  $F$  je distribuční funkce a  $c$  konstanta. Nechť  $F_n$  konvergují slabě k  $F$ . Nechť  $Y_1, Y_2, \dots$  je taková posloupnost náhodných veličin, že  $Y_n$  konverguje v pravděpodobnosti k  $c$ . Definujme*

$$R_n = X_n + Y_n, \quad S_n = X_n Y_n, \quad T_n = X_n / Y_n.$$

*Nechť  $F_n^R, F_n^S$  a  $F_n^T$  jsou po řadě distribuční funkce veličin  $R_n, S_n$  a  $T_n$ . Pak  $F_n^R(x)$  konvergují slabě k  $F(x - c)$ . Je-li  $c > 0$ , pak  $F_n^S(x)$  konvergují slabě k  $F(x/c)$  a  $F_n^T(x)$  konvergují slabě k  $F(cx)$ .*

Čerpáno z knihy Základy matematické statistiky (viz [4], str. 333).

## 4 Velikost testovací skupiny podle zvolené metody

V této kapitole se zaměříme na to, jak vypadají vzorce pro výpočet minimální velikosti testovací skupiny pro nejpoužívanější statistické testy, a na to, jak jsou tyto vzorce odvozeny.

Vzhledem k rozsahu práce budeme pracovat jen s paralelním uspořádáním subjektů (viz 2.3). Nulové hypotézy v jednotlivých testech uvádíme ve tvaru, kdy je cílem studie dokázání ekvivalence (viz 2.2), proto i odvozené vzorce minimální velikosti platí pro testování hypotézy ekvivalence.

Obrázky v této kapitole jsou vytvořeny pomocí RStudia (viz [9]).

### 4.1 Jednovýběrový z-test

Informace o jednovýběrovém z-testu čerpány z knihy Základy počtu pravděpodobnosti a metod matematické statistiky (viz [3], str. 193 a 194).

Máme náhodný výběr  $X_1, X_2, \dots, X_n$  z normálního rozdělení se střední hodnotou  $\mu$  a rozptylem  $\sigma^2$ . Známe hodnotu rozptylu  $\sigma^2$  a chceme testovat nulovou hypotézu  $H_0 : \mu = \mu_0$ , kde  $\mu_0$  je pevně dané reálné číslo (hypotetická střední hodnota), na dané hladině významnosti  $\alpha$  a s pravděpodobností chyby II. druhu  $\beta$ . Nestranným odhadem střední hodnoty je veličina

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

a testovou statistikou veličina

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

kteřá se za platnosti nulové hypotézy  $H_0 : \mu = \mu_0$  řídí normálním normovaným rozdělením. Její kritický obor je

$$W = (-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; \infty).$$

Za platnosti alternativní hypotézy  $H_A : \mu \neq \mu_0$  se veličina  $Z$  řídí normálním rozdělením se střední hodnotou  $\gamma\sqrt{n}$ , kde  $\gamma = \frac{\mu - \mu_0}{\sigma}$  a rozptylem 1. Bez ztráty na obecnosti dále uvažujme  $\mu < \mu_0$  a tedy  $\gamma \in \mathbb{R}^-$ .

Za platnosti alternativní hypotézy platí pro veličinu  $Z$

$$P_A(Z \in W) = 1 - \beta,$$

tedy

$$P_A(Z < -u_{1-\frac{\alpha}{2}} \cup u_{1-\frac{\alpha}{2}} < Z) = 1 - \beta$$

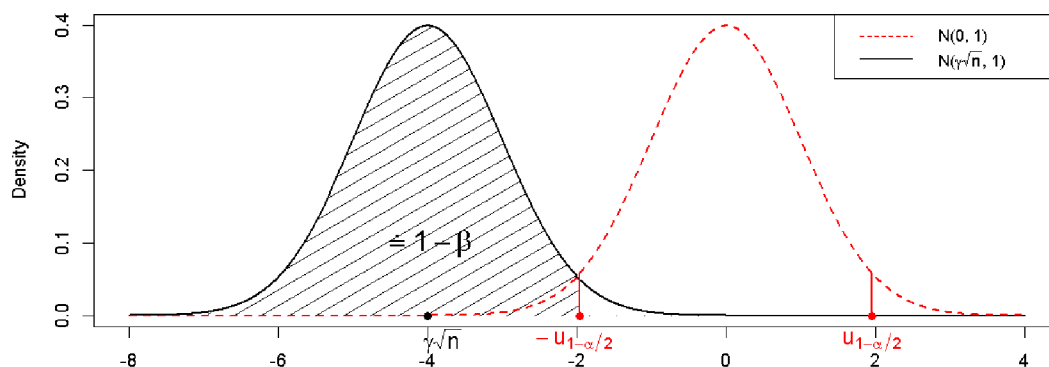
a díky disjunktnosti intervalů kritického oboru můžeme psát

$$P_A(Z < -u_{1-\frac{\alpha}{2}}) + P_A(u_{1-\frac{\alpha}{2}} < Z) = 1 - \beta. \quad (1)$$

Dle předpokladu je  $\gamma$  menší než 0, můžeme proto předpokládat, že pro dostatečně velké hodnoty  $|\gamma|$  platí

$$P_A(Z < -u_{1-\frac{\alpha}{2}}) \gg P_A(u_{1-\frac{\alpha}{2}} < Z), \quad (2)$$

jak můžeme vidět na následujícím obrázku 1.



Obrázek 1: Grafické znázornění nerovnosti (2)

Zanedbáme tedy menší člen z rovnosti (1) a uvažujeme přibližnou rovnost

$$P_A\left(Z < -u_{1-\frac{\alpha}{2}}\right) \doteq 1 - \beta. \quad (3)$$

Na obrázku 1 je dobře vidět, že čím menší je rozdíl  $|\mu - \mu_0|$ , tím menší je  $|\gamma| = \frac{|\mu - \mu_0|}{\sigma}$ , a tím bližší jsou si střední hodnoty rozdělení při platnosti  $H_0$  a  $H_A$ , což povede k větší nepřesnosti aproximace (3).

Rovnost (3) můžeme odečtením střední hodnoty veličiny  $Z$  upravit na tvar

$$P_A\left(Z - \gamma\sqrt{n} < -u_{1-\frac{\alpha}{2}} - \gamma\sqrt{n}\right) \doteq 1 - \beta, \quad (4)$$

který se nám bude hodit za chvíli. Veličina

$$Z_1 = Z - \gamma\sqrt{n} = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} - \frac{\mu - \mu_0}{\sigma}\sqrt{n} = \frac{\bar{X} - \mu}{\sigma}\sqrt{n}$$

se za platnosti alternativní hypotézy řídí normálním normovaným rozdělením. Proto podle definice kvantilů (viz [3], str. 71) platí

$$P_A\left(Z_1 < u_{1-\beta}\right) = 1 - \beta. \quad (5)$$

Z rovností (4) a (5) plyne rovnost

$$-u_{1-\frac{\alpha}{2}} - \gamma\sqrt{n} = u_{1-\beta}$$

a vyjádřením  $n$  získáme vzorec pro minimální velikost testovací skupiny na hladině významnosti  $\alpha$ , s pravděpodobností chyby II. druhu  $\beta$  a s velikostí účinku  $\gamma$

$$n = \frac{(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\gamma^2}, \text{ kde } \gamma = \frac{\mu - \mu_0}{\sigma}. \quad (6)$$

V situaci, kdy hodnotu parametru  $\sigma^2$  neznáme, místo něj můžeme použít jeho odhad

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



a místo kvantilů normálního rozdělení ve výsledném vzorci potom použijeme kvantily  $t$  rozdělení o  $n - 1$  stupních volnosti (viz [1], str. 51). Odvozovali bychom tedy vzorec pro minimální velikost testovací skupiny pro použití jednovýběrového  $t$ -testu. Výsledný vzorec by byl

$$n = \frac{(t_{n-1, 1-\frac{\alpha}{2}} + t_{n-1, 1-\beta})^2}{\gamma^2}, \text{ kde } \gamma = \frac{\mu - \mu_0}{S}.$$

V klinické praxi se ale setkáme s použitím vzorce (6) i při neznámé přesné hodnotě parametru  $\sigma^2$ , protože  $t$  rozdělení o  $n - 1$  stupních volnosti pro dostatečně velká  $n$  (což je v případě klinických studií téměř vždy splněno) velmi dobře aproximuje normální normované rozdělení. Jako hodnotu parametru  $\sigma^2$  potom použijeme jeho odhad.

Stejný vztah mezi  $t$  rozdělením a normálním rozdělením bude platit i u dalších testů, a proto i u nich budeme vzorec pro  $n$  odvozovat tak, jako bychom hodnotu parametru  $\sigma^2$  znali.

Je zajímavé, že velikost  $n$  s použitím kvantilů  $t$  rozdělení bude vždy větší nebo rovna velikosti  $n$  s použitím kvantilů normálního rozdělení. Vzorec s kvantily  $t$  rozdělení tedy můžeme použít pro „přísnější“ výpočet velikosti  $n$ .

Požadovaná velikost účinku  $\gamma$  je při použití vzorce (6) v praxi většinou určena podle předchozích podobných studií.

Postup a výsledný vzorec ověřen podle knihy *Sample Size Calculations in Clinical Research* (viz [1], str. 50–52).

## 4.2 Dvouvýběrový $t$ -test

Informace o dvouvýběrovém  $t$ -testu čerpány z knihy *Základy počtu pravděpodobnosti a metod matematické statistiky* (viz [3], str. 259).

Máme náhodný výběr  $X_1, X_2, \dots, X_n$  z normálního rozdělení s parametry  $\mu_1$  a  $\sigma^2$  a náhodný výběr  $Y_1, Y_2, \dots, Y_n$  z normálního rozdělení s parametry  $\mu_2$  a  $\sigma^2$ . U obou výběrů uvažujeme stejný rozsah  $n$ , protože při klinickém testování chceme porovnávat stejně velké skupiny subjektů. Náhodné výběry jsou nezávislé. Budeme chtít testovat nulovou hypotézu  $H_0 : \mu_1 = \mu_2$ . Hodnotu parametru  $\sigma^2$  budeme považovat za známou ze stejného důvodu, jako je vysvětleno na konci předchozí kapitoly o z-testu.

Umíme sestavit výběrové průměry

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{a} \quad \bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}.$$

Veličina  $\bar{X}$  se řídí normálním rozdělením se střední hodnotou  $\mu_1$  a rozptylem  $\frac{\sigma^2}{n}$ , veličina  $\bar{Y}$  se řídí normálním rozdělením se střední hodnotou  $\mu_2$  a stejným rozptylem  $\frac{\sigma^2}{n}$ .

Normalizací (viz [3], str. 87) dostaneme náhodné veličiny

$$\frac{\bar{X} - \mu_1}{\sigma} \sqrt{n} \quad \text{a} \quad \frac{\bar{Y} - \mu_2}{\sigma} \sqrt{n},$$

které si zachovávají vzájemnou nezávislost a řídí se normálním normovaným rozdělením. Rozdílem těchto veličin vytvoříme novou náhodnou veličinu

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma} \sqrt{n}$$

s normálním rozdělením, střední hodnotou 0 a rozptylem 2. Veličinu podělíme její směrodatnou odchylkou  $\sqrt{2}$  a dostaneme opět veličinu s normálním normovaným rozdělením

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{2}\sigma} \sqrt{n}. \tag{7}$$

Nulovou hypotézu  $\mu_1 = \mu_2$  lze upravit na  $\mu_1 - \mu_2 = 0$ . Za platnosti nulové hypotézy tedy můžeme veličinu (7) upravit na veličinu

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{2}\sigma} \sqrt{n}, \quad (8)$$

která se řídí normálním normovaným rozdělením. Testová statistika  $Z$  má kritický obor

$$W = (-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty).$$

Při platnosti alternativní hypotézy, tedy  $\mu_1 \neq \mu_2$ , bude existovat nenulový člen  $\gamma = \frac{\mu_1 - \mu_2}{\sigma}$ , bez ztráty na obecnosti uvažujme  $\gamma \in \mathbb{R}^-$ . Veličinu (7) můžeme napsat jako veličinu

$$\frac{\bar{X} - \bar{Y} - \gamma\sigma}{\sqrt{2}\sigma} \sqrt{n}, \quad (9)$$

která se řídí normálním normovaným rozdělením. Dále přičtením výrazu  $\frac{\gamma}{\sqrt{2}}\sqrt{n}$  dostaneme stejnou veličinu jako (8), tedy

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{2}\sigma} \sqrt{n},$$

nyň ale s normálním rozdělením, se střední hodnotou  $\frac{\gamma}{\sqrt{2}}\sqrt{n}$  a rozptylem 1. Pro tuto veličinu platí

$$P_A(Z \in W) = 1 - \beta,$$

neboli

$$P_A(Z < -u_{1-\frac{\alpha}{2}} \cup u_{1-\frac{\alpha}{2}} < Z) = 1 - \beta,$$

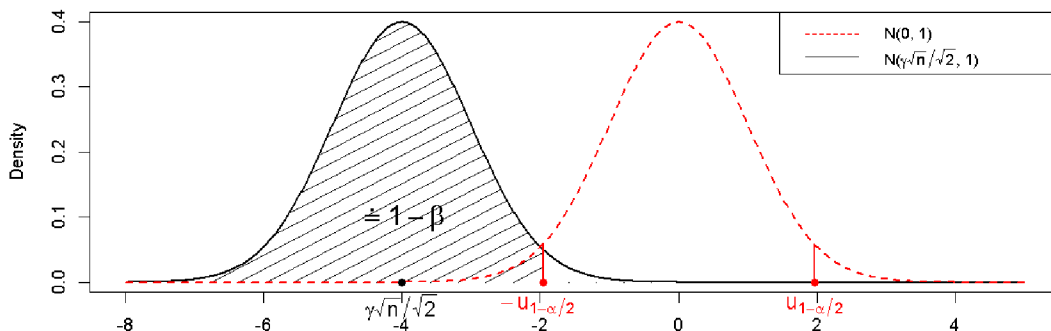
což můžeme díky disjunktnosti intervalů kritického oboru napsat jako

$$P_A(Z < -u_{1-\frac{\alpha}{2}}) + P_A(u_{1-\frac{\alpha}{2}} < Z) = 1 - \beta. \quad (10)$$

Protože dle předpokladu platí  $\gamma < 0$ , můžeme předpokládat, že pro dostatečně velké hodnoty  $|\gamma|$  platí

$$P_A(Z < -u_{1-\frac{\alpha}{2}}) \gg P_A(u_{1-\frac{\alpha}{2}} < Z), \quad (11)$$

což je dobře vidět na obrázku 2.



Obrázek 2: Grafické znázornění nerovnosti (11)

Proto rovnost (10) aproximujeme jako

$$P_A\left(Z < -u_{1-\frac{\alpha}{2}}\right) \doteq 1 - \beta. \quad (12)$$

Na obrázku 2 je dobře vidět, že čím menší je  $|\mu_1 - \mu_2|$ , tím menší je  $|\gamma| = \frac{|\mu_1 - \mu_2|}{\sigma}$ , a tím bližší jsou si střední hodnoty rozdělení při platnosti  $H_0$  a  $H_A$ , což povede k větší nepřesnosti aproximace (12).

Rovnost (12) nyní odečtením střední hodnoty veličiny  $Z$  upravíme do tvaru, který se nám bude za chvíli hodit

$$P_A\left(Z - \frac{\gamma}{\sqrt{2}}\sqrt{n} < -u_{1-\frac{\alpha}{2}} - \frac{\gamma}{\sqrt{2}}\sqrt{n}\right) \doteq 1 - \beta. \quad (13)$$

Podle veličiny (9) víme, že totožná veličina

$$Z_1 = Z - \frac{\gamma}{\sqrt{2}}\sqrt{n} = \frac{\bar{X} - \bar{Y}}{\sqrt{2}\sigma}\sqrt{n} - \frac{\gamma}{\sqrt{2}}\sqrt{n} = \frac{\bar{X} - \bar{Y} - \gamma\sigma}{\sqrt{2}\sigma}\sqrt{n}$$

se za platnosti alternativní hypotézy řídí normálním normovaným rozdělením. Proto dle definice kvantilů (viz [3], str. 71) platí

$$P_A\left(Z_1 < u_{1-\beta}\right) = 1 - \beta. \quad (14)$$

Z rovností (13) a (14) plyne

$$-u_{1-\frac{\alpha}{2}} - \frac{\gamma}{\sqrt{2}}\sqrt{n} = u_{1-\beta}.$$

Vyjádřením  $n$  získáme hledaný vzorec pro výpočet minimální velikosti testovací skupiny na hladině významnosti  $\alpha$ , s pravděpodobností chyby II. druhu  $\beta$  a velikostí účinku  $\gamma$

$$n = \frac{2(u_{1-\frac{\alpha}{2}} + u_{1-\beta})^2}{\gamma^2}, \text{ kde } \gamma = \frac{\mu_1 - \mu_2}{\sigma}. \quad (15)$$

Postup a výsledný vzorec ověřen podle knihy Sample Size Calculations in Clinical Research (viz [1], str. 57-59).

### 4.3 Jednofaktorová ANOVA

Informace o jednofaktorové analýze rozptylu (známé též jako ANOVA) čerpány z knihy Základy počtu pravděpodobnosti a metod matematické statistiky (viz [3], str. 266–271) a z knihy Základy matematické statistiky (viz [4], Kapitola analýza rozptylu).

Všechny testované subjekty jsme rozdělili do  $k$  skupin o velikosti  $n$ . Jednotlivými subjekty se od teď budeme odkazovat na hodnotu primárního testovaného parametru u daného subjektu. Chceme testovat nulovou hypotézu, že střední hodnoty v jednotlivých skupinách jsou stejné, tedy

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k,$$

oproti alternativní hypotéze, že alespoň jedna dvojice středních hodnot se liší

$$H_A : \exists p \in \{1, \dots, k\} \exists q \in \{1, \dots, k\} \setminus \{p\} : \mu_p \neq \mu_q \quad .$$

Nechť  $x_{ij}$  je  $j$ -tý subjekt z  $i$ -té skupiny,  $i = 1, \dots, k$  a  $j = 1, \dots, n$ . Model jednofaktorové analýzy rozptylu můžeme popsat jako

$$x_{ij} = \mu_i + \epsilon_{ij},$$

kde  $\mu_i$  je střední hodnota v  $i$ -té skupině a  $\epsilon_{ij}$  je náhodná odchylka při pozorování vybraného subjektu. Předpokládáme, že všechny veličiny  $\epsilon_{ij}$  se řídí normálním rozdělením se střední hodnotou 0 a rozptylem  $\sigma^2$  a jsou vzájemně nezávislé.

Výběrový průměr v jednotlivých skupinách je roven

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}, \quad \text{pro } i = 1, \dots, k.$$

Reziduální součet čtverců je roven

$$SSE = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2.$$

Veličina  $\frac{SSE}{\sigma^2}$  se za platnosti nulové hypotézy řídí  $\chi^2$  rozdělením o  $k(n-1)$  stupních volnosti. Skupinový součet čtverců je roven

$$SSA = \sum_{i=1}^k n(\bar{x}_i - \bar{x}_{..})^2, \quad \text{kde } \bar{x}_{..} = \frac{1}{kn} \sum_{i=1}^k \sum_{j=1}^n x_{ij}.$$

Veličina  $\frac{SSA}{\sigma^2}$  se za platnosti nulové hypotézy řídí  $\chi^2$  rozdělením o  $k-1$  stupních volnosti.

Dále můžeme odhadnout rozptyl náhodné složky  $\epsilon$  jako

$$\hat{\sigma}^2 = \frac{SSE}{k(n-1)}. \quad (16)$$

Testovou statistikou je veličina

$$F_1 = \frac{\frac{SSA}{k-1}}{\frac{SSE}{k(n-1)}}.$$

Z výše uvedených vlastností  $SSE$  a  $SSA$  a z jejich nezávislosti plyne, že se testová statistika  $F_1$  za platnosti nulové hypotézy řídí  $F$  rozdělením o  $k-1$  a  $k(n-1)$  stupních volnosti.

Nulovou hypotézu zamítneme na hladině významnosti  $\alpha$ , pokud

$$F_1 > F_{k-1, k(n-1)}(1 - \alpha),$$

kde  $F_{k-1, k(n-1)}(1 - \alpha)$  je  $1 - \alpha$  kvantil  $F$  rozdělení o  $k - 1$  a  $k(n - 1)$  stupních volnosti. Kritický obor testu je tedy

$$W = \left( F_{k-1, k(n-1)}(1 - \alpha); \infty \right).$$

Pro sílu testu  $1 - \beta$  platí

$$P_A(F_1 \in W) = 1 - \beta,$$

tedy

$$P_A \left( \frac{\frac{SSA}{k-1}}{\frac{SSE}{k(n-1)}} > F_{k-1, k(n-1)}(1 - \alpha) \right) = 1 - \beta.$$

Použitím odhadu (16) upravíme výraz na

$$P_A \left( \frac{\frac{SSA}{k-1}}{\hat{\sigma}^2} > F_{k-1, k(n-1)}(1 - \alpha) \right) \doteq 1 - \beta$$

a dále použitím asymptotického vztahu

$$\chi_{k-1}^2(1 - \alpha) \approx (k - 1)F_{k-1, k(n-1)}(1 - \alpha)$$

a nahrazením odhadu  $\hat{\sigma}^2$  jeho limitní hodnotou  $\sigma^2$  dostaneme

$$P_A \left( \frac{SSA}{\sigma^2} > \chi_{k-1}^2(1 - \alpha) \right) \doteq 1 - \beta. \quad (17)$$

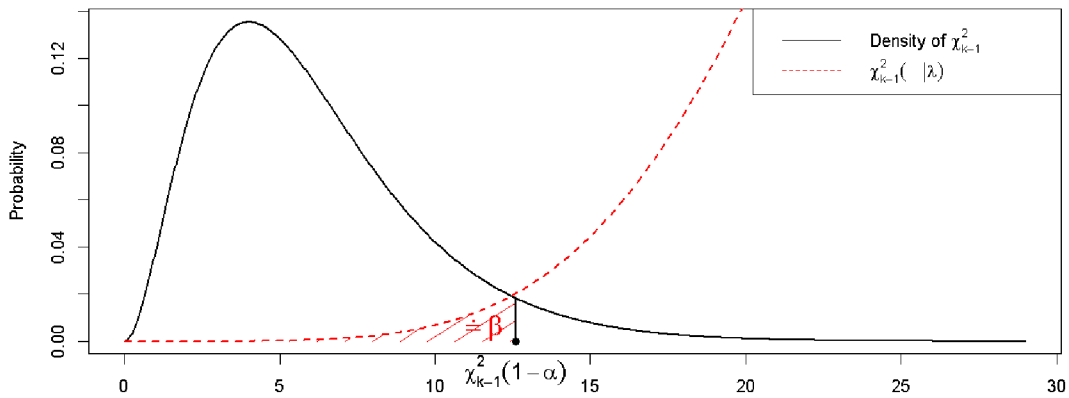
Za platnosti alternativní hypotézy se veličina  $\frac{SSA}{\sigma^2}$  řídí necentrálním  $\chi^2$  rozdělením (viz Věta (2)) o  $k - 1$  stupních volnosti s parametrem necentrality

$$\lambda = \frac{n}{\sigma^2} \sum_{i=1}^k (\mu_i - \mu)^2, \text{ kde } \mu = \frac{1}{k} \sum_{j=1}^k \mu_j \quad . \quad (18)$$

Z přibližné rovnosti (17) a z rozdělení veličiny  $\frac{SSA}{\sigma^2}$  plyne

$$\chi_{k-1}^2(\chi_{k-1}^2(1-\alpha)|\lambda) \doteq \beta, \quad (19)$$

kde  $\chi_{k-1}^2(\cdot|\lambda)$  je distribuční funkce necentrálního  $\chi^2$  rozdělení o  $k-1$  stupních volnosti s parametrem necentrality  $\lambda$ .



Obrázek 3: Grafické znázornění rovnosti (19)

Z rovnosti (19) získáme hodnotu parametru  $\lambda$  na dané hladině významnosti  $\alpha$  a s danou pravděpodobností chyby II. druhu  $\beta$ , označme ho  $\lambda_{\alpha,\beta}$ .

Nyní už z definice parametru  $\lambda$  (viz (18)) vyjádříme  $n$  a dostaneme minimální velikost jednotlivých testovaných skupin na hladině významnosti  $\alpha$  a s pravděpodobností chyby II. druhu  $\beta$

$$n = \frac{\lambda_{\alpha,\beta}\sigma^2}{\sum_{i=1}^k (\mu_i - \mu)^2}, \text{ kde } \mu = \frac{1}{k} \sum_{j=1}^k \mu_j.$$

Výraz  $\frac{\sum_{i=1}^k (\mu_i - \mu)^2}{\sigma^2}$  reprezentuje velikost účinku, označme ho  $\gamma$ . Výsledný vzorec tedy můžeme napsat také jako

$$n = \frac{\lambda_{\alpha,\beta}}{\gamma}. \quad (20)$$



Postup a výsledný vzorec ověřen podle knihy *Sample Size Calculations in Clinical Research* (viz [1], str. 70–72).

#### 4.4 Chí-kvadrát test nezávislosti

Informace o Chí-kvadrát testu nezávislosti čerpány z knihy *Základy počtu pravděpodobnosti a metod matematické statistiky* (viz [3], str. 223).

Máme dvě diskrétní náhodné veličiny  $X$  a  $Y$  a z nich vytvořenou kontingenční tabulku

X/Y	$y_1$	$y_2$	$\dots$	$y_s$	$\sum$
$x_1$	$n_{11}$	$n_{12}$	$\dots$	$n_{1s}$	$n_{1\cdot}$
$x_2$	$n_{21}$	$n_{22}$	$\dots$	$n_{2s}$	$n_{2\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$x_r$	$n_{r1}$	$n_{r2}$	$\dots$	$n_{rs}$	$n_{r\cdot}$
$\sum$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\dots$	$n_{\cdot s}$	$n$

Tabulka 1: Kontingenční tabulka pro Chí-kvadrát test nezávislosti

Chceme testovat nulovou hypotézu, že jsou veličiny  $X$  a  $Y$  nezávislé. Testovou statistikou je veličina

$$T = \sum_{i=1}^r \sum_{j=1}^s \frac{\left( n_{ij} - \frac{n_{i\cdot} n_{\cdot j}}{n} \right)^2}{\frac{n_{i\cdot} n_{\cdot j}}{n}}. \quad (21)$$

Definujme teoretické pravděpodobnosti

$$p_{ij} = P(X = x_i, Y = y_j), \quad p_{i\cdot} = P(X = x_i), \quad p_{\cdot j} = P(Y = y_j)$$

a jejich odhady

$$\hat{p}_{ij} = n_{ij}/n, \quad \hat{p}_{i\cdot} = n_{i\cdot}/n, \quad \hat{p}_{\cdot j} = n_{\cdot j}/n \quad . \quad (22)$$

Veličinu (21) nyní můžeme napsat jako

$$T = \sum_{i=1}^r \sum_{j=1}^s \frac{n(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2}{\hat{p}_{i.}\hat{p}_{.j}}. \quad (23)$$

Za platnosti nulové hypotézy se veličina (23) asymptoticky řídí chí-kvadrát rozdělením o  $(r-1)(s-1)$  stupních volnosti. Za platnosti alternativní hypotézy, tedy při závislosti veličin X a Y, a při platnosti

$$\lim_{n \rightarrow \infty} \sum_{i=1}^r \sum_{j=1}^s \frac{n(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2}{\hat{p}_{i.}\hat{p}_{.j}} = \lambda, \quad (24)$$

se veličina (23) asymptoticky řídí necentrálním chí-kvadrát rozdělením (viz Věta (2)) o  $(r-1)(s-1)$  stupních volnosti s parametrem necentrality  $\lambda$ .

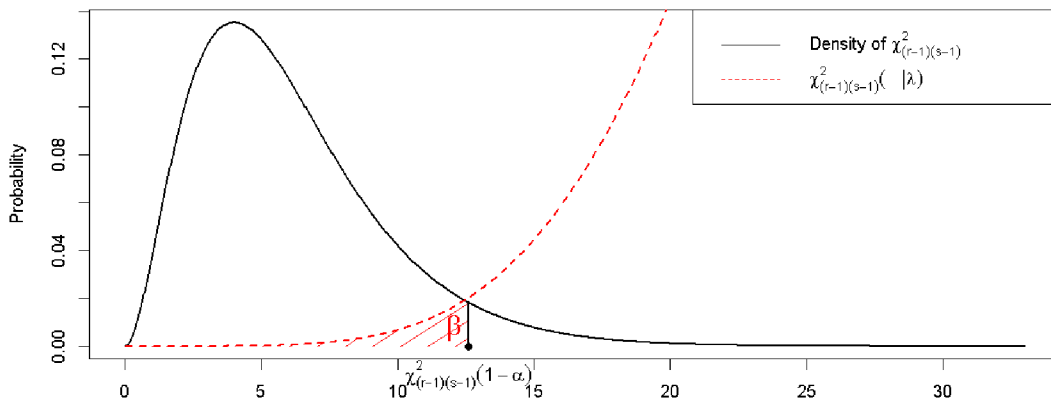
Nulovou hypotézu zamítáme na hladině významnosti  $\alpha$  pro hodnoty testové statistiky spadající do kritického oboru  $W$

$$W = \left( \chi_{(r-1)(s-1)}^2(1-\alpha); \infty \right).$$

Za platnosti alternativní hypotézy, s pravděpodobností chyby II. druhu  $\beta$  platí

$$\chi_{(r-1)(s-1)}^2(\chi_{(r-1)(s-1)}^2(1-\alpha)|\lambda) = \beta, \quad (25)$$

kde výraz  $\chi_{(r-1)(s-1)}^2(\quad|\lambda)$  je distribuční funkce necentrálního  $\chi^2$  rozdělení o  $(r-1)(s-1)$  stupních volnosti s parametrem necentrality  $\lambda$ .



Obrázek 4: Grafické znázornění rovnosti (25)

Hodnotu parametru  $\lambda$  na zvolené hladině významnosti  $\alpha$  a s pravděpodobností chyby II. druhu  $\beta$  vypočítáme z rovnosti (25) a řešení označíme  $\lambda_{\alpha,\beta}$ .

Z asymptotického vztahu (24) získáme vyjádřením  $n$  vzorec pro minimální velikost testovací skupiny

$$n = \lambda_{\alpha,\beta} \left( \sum_{i=1}^r \sum_{j=1}^s \frac{(\hat{p}_{ij} - \hat{p}_{i\cdot} \hat{p}_{\cdot j})^2}{\hat{p}_{i\cdot} \hat{p}_{\cdot j}} \right)^{-1}. \quad (26)$$

Postup a výsledný vzorec ověřen podle knihy Sample Size Calculations in Clinical Research (viz [1], str. 148 a 149).

## 4.5 McNemarův test efektu ošetření

Informace o McNemarově testu efektu ošetření čerpány z knihy Základy počtu pravděpodobnosti a metod matematické statistiky (viz [3], str. 229).

Chceme zjistit, zda určitý zásah (ošetření) ovlivňuje pravděpodobnost výskytu jisté vlastnosti (má efekt). Náhodně vybereme  $n$  subjektů a zjistíme počet těch, u nichž se sledovaná vlastnost vyskytla. Poté provedeme na všech

$n$  subjektech určitý zásah, který výskyt sledované vlastnosti může ovlivnit. Následně opět zjistíme počet statistických jednotek s výskytem dané vlastnosti. Na základě zjištěných změn výskytu sledované vlastnosti pak chceme rozhodnout, zda provedený zásah pravděpodobnost výskytu této vlastnosti ovlivnil.

Nechť  $x_{ij}$  je binární proměnná ( $x_{ij} = 1$ , pokud se vlastnost vyskytla a  $x_{ij} = 0$ , pokud se vlastnost nevyskytla). Index  $i \in \{1, \dots, n\}$  značí vybraný subjekt z  $n$  subjektů, a  $j \in \{1, 2\}$  značí, zda je  $i$ -tý subjekt měřen před daným ošetřením ( $j = 1$ ) nebo po ošetření ( $j = 2$ ). Výsledky měření můžeme zapsat do kontingenční tabulky

před/po	Bez vlastnosti	Výskyt vlastnosti	$\sum$
Bez vlastnosti	$n_{00}$	$n_{01}$	$n_{0.}$
Výskyt vlastnosti	$n_{10}$	$n_{11}$	$n_{1.}$
$\sum$	$n_{.0}$	$n_{.1}$	$n$

Tabulka 2: Kontingenční tabulka pro McNemarův test efektu ošetření

Hodnoty v této tabulce jsou definovány jako

$$n_{00} = \sum_{i=1}^n (1 - x_{i1})(1 - x_{i2}), \quad n_{01} = \sum_{i=1}^n (1 - x_{i1})x_{i2},$$

$$n_{10} = \sum_{i=1}^n x_{i1}(1 - x_{i2}), \quad n_{11} = \sum_{i=1}^n x_{i1}x_{i2}.$$

Definujme ještě teoretické pravděpodobnosti

$$p_{00} = P(x_{i1} = 0, x_{i2} = 0), \quad p_{01} = P(x_{i1} = 0, x_{i2} = 1),$$

$$p_{10} = P(x_{i1} = 1, x_{i2} = 0), \quad p_{11} = P(x_{i1} = 1, x_{i2} = 1),$$

$$p_{1+} = P(x_{i1} = 1) = p_{10} + p_{11}, \quad p_{+1} = P(x_{i2} = 1) = p_{01} + p_{11}$$

a jejich odhady

$$\hat{p}_{00} = n_{00}/n, \quad \hat{p}_{01} = n_{01}/n, \quad \hat{p}_{10} = n_{10}/n, \quad \hat{p}_{11} = n_{11}/n .$$

Jako nulovou a alternativní hypotézu položíme

$$H_0 : p_{1+} = p_{+1} \quad \text{a} \quad H_A : p_{1+} \neq p_{+1},$$

což je z definice  $p_{1+}$  a  $p_{+1}$  ekvivalentní hypotézám

$$H_0 : p_{10} = p_{01} \quad \text{a} \quad H_A : p_{10} \neq p_{01}.$$

Nulová hypotéza tedy říká, že pravděpodobnost změny výskytu vlastnosti při ošetření je stejná pro nový výskyt vlastnosti u subjektu, který vlastnost před ošetřením neměl, a pro vymizení vlastnosti u subjektu, který vlastnost před ošetřením měl.

Testovou statistikou pro McNemarův test bude veličina

$$T = \frac{n_{10} - n_{01}}{\sqrt{n_{10} + n_{01}}}, \quad (27)$$

která se za platnosti nulové hypotézy asymptoticky řídí normálním normovaným rozdělením. Kritický obor testu na hladině významnosti  $\alpha$  je tedy

$$W = (-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; \infty).$$

Za platnosti alternativní hypotézy  $H_A : p_{10} \neq p_{01}$  můžeme testovou statistiku  $T$  upravit následujícím způsobem

$$\begin{aligned} T &= \frac{n_{10} - n_{01}}{\sqrt{n_{10} + n_{01}}} = \frac{\sqrt{n}}{\sqrt{n}} \frac{n_{10} - n_{01}}{\sqrt{n_{10} + n_{01}}} = \sqrt{\frac{n}{n_{10} + n_{01}}} \frac{1}{\sqrt{n}} (n_{10} - n_{01}) = \\ &= \sqrt{\frac{1}{\frac{n_{10}}{n} + \frac{n_{01}}{n}}} \frac{1}{\sqrt{n}} \left[ \sum_{i=1}^n [x_{i1}(1 - x_{i2})] - \sum_{i=1}^n [(1 - x_{i1})x_{i2}] \right] = \\ &= \frac{1}{\sqrt{\hat{p}_{10} + \hat{p}_{01}}} \frac{1}{\sqrt{n}} \sum_{i=1}^n (x_{i1} - x_{i2}) = \frac{1}{\sqrt{\hat{p}_{10} + \hat{p}_{01}}} \frac{1}{\sqrt{n}} \sum_{i=1}^n d_i , \end{aligned}$$

kde  $d_i = x_{i1} - x_{i2}$ . Z počátečního výběru  $n$  subjektů plyne, že jsou veličiny  $d_i$  nezávislé a stejně rozdělené. Nyní odvodíme jejich střední hodnotu a rozptyl

$$\begin{aligned} E(d_i) &= E(x_{i1} - x_{i2}) = E(x_{i1}) - E(x_{i2}) = \\ &= 0 \cdot P(x_{i1} = 0) + 1 \cdot P(x_{i1} = 1) - 0 \cdot P(x_{i2} = 0) - 1 \cdot P(x_{i2} = 1) = \\ &= p_{10} + p_{11} - p_{01} - p_{11} = p_{10} - p_{01} \end{aligned}$$

$$\begin{aligned} \text{Var}(d_i) &= \text{Var}(x_{i1} - x_{i2}) = E(x_{i1} - x_{i2})^2 - [E(x_{i1} - x_{i2})]^2 = \\ &= E(x_{i1}^2 - 2x_{i1}x_{i2} + x_{i2}^2) - (p_{10} - p_{01})^2 = \\ &= E(x_{i1}^2) - 2 \cdot E(x_{i1}x_{i2}) + E(x_{i2}^2) - (p_{10} - p_{01})^2 = \\ &= 0 \cdot P(x_{i1}^2 = 0) + 1 \cdot P(x_{i1}^2 = 1) - 2[0 \cdot 0 \cdot P(x_{i1} = 0, x_{i2} = 0) + \\ &+ 1 \cdot 0 \cdot P(x_{i1} = 1, x_{i2} = 0) + 0 \cdot 1 \cdot P(x_{i1} = 0, x_{i2} = 1) + \\ &+ 1 \cdot 1 \cdot P(x_{i1} = 1, x_{i2} = 1)] + 0 \cdot P(x_{i2}^2 = 0) + 1 \cdot P(x_{i2}^2 = 1) - \\ &- (p_{10} - p_{01})^2 = p_{10} + p_{11} - 2 \cdot p_{11} + p_{01} + p_{11} - (p_{10} - p_{01})^2 = \\ &= p_{10} + p_{01} - (p_{10} - p_{01})^2. \end{aligned}$$

Vraťme se nyní k testové statistice  $T$  ve tvaru

$$\frac{1}{\sqrt{\hat{p}_{10} + \hat{p}_{01}}} \frac{1}{\sqrt{n}} \sum_{i=1}^n d_i .$$

Dle zákona velkých čísel (viz [3], Kapitola 4) veličina  $\frac{1}{\sqrt{\hat{p}_{10} + \hat{p}_{01}}}$  konverguje v pravděpodobnosti ke konstantě  $c = \frac{1}{\sqrt{p_{10} + p_{01}}}$ . Proto dle Cramérový-Sluckého věty (viz Věta (3)) můžeme v úvahách o asymptotickém chování přistoupit k této veličině jako ke konstantě  $c$ . Testovou statistiku  $T$  nahlíženou s touto asymptotickou úpravou označíme  $T_{as}$

$$T_{as} = \frac{c}{\sqrt{n}} \sum_{i=1}^n d_i .$$

Protože veličiny  $d_i$  jsou navzájem nezávislé, mají stejnou střední hodnotu i rozptyl, a sumu těchto veličin násobíme v testové statistice  $T_{as}$  pouze konstantou  $\frac{c}{\sqrt{n}}$ , platí podle centrální limitní věty (viz Věta (1)), že se veličina

$$\frac{T_{as} - E(T_{as})}{\sqrt{\text{Var}(T_{as})}}$$

asymptoticky řídí normálním normovaným rozdělením. Střední hodnota a rozptyl veličiny  $T_{as}$  jsou

$$\begin{aligned} E(T_{as}) &= E\left(\frac{c}{\sqrt{n}} \sum_{i=1}^n d_i\right) = \frac{\sum_{i=1}^n E(d_i)}{\sqrt{p_{10} + p_{01}}\sqrt{n}} = \\ &= \frac{n(p_{10} - p_{01})}{\sqrt{p_{10} + p_{01}}\sqrt{n}} = \frac{\sqrt{n}(p_{10} - p_{01})}{\sqrt{p_{10} + p_{01}}} \end{aligned}$$

$$\begin{aligned} \text{Var}(T_{as}) &= \text{Var}\left(\frac{c}{\sqrt{n}} \sum_{i=1}^n d_i\right) = \frac{c^2}{n} \sum_{i=1}^n \text{Var}(d_i) = \\ &= \frac{1}{(p_{10} + p_{01})n} n[p_{10} + p_{01} - (p_{10} - p_{01})^2] = \\ &= \frac{p_{10} + p_{01} - (p_{10} - p_{01})^2}{p_{10} + p_{01}} \end{aligned}$$

Podle definice kvantilů (viz [3], str. 71) pro tuto normovanou veličinu platí asymptotická rovnost

$$P\left(\frac{T_{as} - E(T_{as})}{\sqrt{\text{Var}(T_{as})}} < u_{1-\beta}\right) \approx 1 - \beta, \quad (28)$$

kterou použijeme za chvíli.

Bez ztráty na obecnosti nyní uvažujme porušení nulové hypotézy ve tvaru  $p_{10} < p_{01}$ . Pro střední hodnotu testové statistiky  $T_{as}$  za platnosti alternativní hypotézy potom bude platit

$$E(T_{as}) = \frac{\sqrt{n}(p_{10} - p_{01})}{\sqrt{p_{10} + p_{01}}} < 0. \quad (29)$$

Za platnosti alternativní hypotézy platí pro sílu testu  $1 - \beta$

$$P_A(T_{as} \in W) = 1 - \beta,$$

tedy

$$P_A(T_{as} < -u_{1-\frac{\alpha}{2}} \cup u_{1-\frac{\alpha}{2}} < T_{as}) = 1 - \beta$$

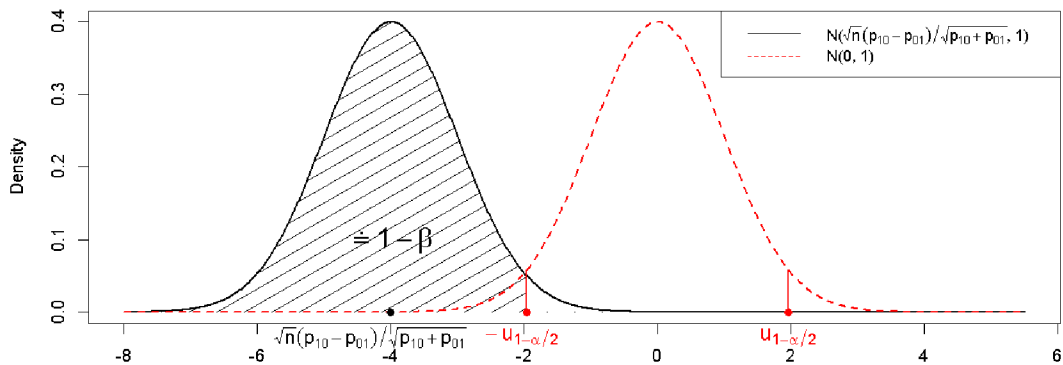
a kvůli disjunktním intervalům kritického oboru můžeme psát

$$P_A(T_{as} < -u_{1-\frac{\alpha}{2}}) + P_A(u_{1-\frac{\alpha}{2}} < T_{as}) = 1 - \beta. \quad (30)$$

Ze záporné střední hodnoty (viz (29)) veličiny  $T_{as}$  za platnosti alternativní hypotézy můžeme předpokládat, že pro dostatečně velké hodnoty  $|E(T_{as})|$  platí

$$P_A(T_{as} < -u_{1-\frac{\alpha}{2}}) \gg P_A(u_{1-\frac{\alpha}{2}} < T_{as}), \quad (31)$$

což je dobře vidět na obrázku 5.



Obrázek 5: Grafické znázornění nerovnosti (31)

Proto rovnost (30) aproximujeme přibližnou rovností

$$P_A(T_{as} < -u_{1-\frac{\alpha}{2}}) \doteq 1 - \beta.$$



Tuto přibližnou rovnost můžeme též napsat jako

$$P_A \left( \frac{T_{as} - E(T_{as})}{\sqrt{\text{Var}(T_{as})}} < \frac{-u_{1-\frac{\alpha}{2}} - E(T_{as})}{\sqrt{\text{Var}(T_{as})}} \right) \doteq 1 - \beta. \quad (32)$$

Z rovností (28) a (32) plyne

$$u_{1-\beta} = \frac{-u_{1-\frac{\alpha}{2}} - E(T_{as})}{\sqrt{\text{Var}(T_{as})}},$$

tedy

$$u_{1-\beta} = \frac{-u_{1-\frac{\alpha}{2}} - \frac{\sqrt{n}(p_{10} - p_{01})}{\sqrt{p_{10} + p_{01}}}}{\sqrt{\frac{p_{10} + p_{01} - (p_{10} - p_{01})^2}{p_{10} + p_{01}}}}.$$

Vyjádřením  $n$  získáme hledaný vzorec pro minimální velikost testovací skupiny na hladině významnosti  $\alpha$  a při pravděpodobnosti chyby II. druhu  $\beta$  pro McNemarův test efektu ošetření

$$n = \frac{[u_{1-\frac{\alpha}{2}}\sqrt{p_{01} + p_{10}} + u_{1-\beta}\sqrt{p_{01} + p_{10} - (p_{01} - p_{10})^2}]^2}{(p_{10} - p_{01})^2}. \quad (33)$$

V praxi neznáme přesné hodnoty teoretických pravděpodobností  $p_{01}$  a  $p_{10}$ , a proto místo nich použijeme jejich odhady  $\hat{p}_{01}$  a  $\hat{p}_{10}$ .

Postup a výsledný vzorec ověřen podle knihy *Sample Size Calculations in Clinical Research* (viz [1], str. 153-155).

## 5 Použití softwaru pro výpočet velikosti testovací skupiny

V této kapitole u testů z předchozí kapitoly porovnáme svůj výpočet požadovaného počtu subjektů a výpočet pomocí softwaru. Podíváme se, jaké možnosti nabízí analytické programy SAS Studio (viz [7]) a RStudio (viz [9]). V SAS Studiu budeme používat knihovnu POWER (viz [7], Kapitola 89) a v RStudiu knihovnu pwr (viz [10]).

RStudio tvoří přívětivější uživatelské rozhraní pro R software (viz [8]), který používáme ve verzi 4.3.2. RStudio používáme ve verzi 2023.12.1. R software lze bezplatně stáhnout na stránkách „[www.r-project.org](http://www.r-project.org)“ a RStudio na stránkách „[www.rstudio.com](http://www.rstudio.com)“. SAS Studio je placený nástroj, ke kterému jsme získali přístup přes prohlížeč registrací v portálu SAS OnDemand for Academics na stránkách „[welcome.oda.sas.com](http://welcome.oda.sas.com)“, který vývojáři SAS nabízí. Používáme jej ve verzi 3.81.

V příloze práce je k dispozici script z RStudia i SAS Studia s použitými kódy.

Veškeré výpočty budeme provádět s přesností na dvě desetinná místa. Použitím odvozených vzorců zpravidla dostaneme neceločíselné výsledky, které budeme zaokrouhlovat na nejbližší celé číslo nahoru.

### 5.1 Jednovýběrový z-test

Nejdříve vypočítáme minimální velikost testovací skupiny podle vzorce (6) v kapitole 4.1. Zvolíme hladinu významnosti  $\alpha = 0.01$ , pravděpodobnost chyby II. druhu  $\beta = 0.2$  a velikost účinku  $\gamma = \frac{\mu - \mu_0}{\sigma} = \frac{14 - 10}{8} = 0.5$ .

Dosazením do odvozeného vzorce (6) dostaneme

$$n = \frac{(u_{0.995} + u_{0.8})^2}{0.5^2} \doteq 46.79.$$

Po zaokrouhlení výsledku nahoru tedy pro studii potřebujeme 47 subjektů.

Nyní budeme se stejným zadáním počítat v RStudios. Použijeme kód

```
pwr.norm.test(n = ,  
              d = 0.5,  
              sig.level = 0.01,  
              power = 0.8,  
              alternative = "two.sided")
```

kde `n` je parametr pro velikost testovací skupiny, `d` je parametr pro velikost účinku (my jsme ho při odvození označili  $\gamma$ ), `sig.level` je zvolená hladina významnosti  $\alpha$ , `power` je síla testu  $1 - \beta$  a `alternative` značí oboustrannou alternativní hypotézu. Samotný znak „`=`“ u parametru `n` značí, že jeho hodnotu chceme vypočítat. Další informace k parametrům nalezneme v dokumentaci knihovny `pwr` (viz [10]).

Po spuštění kódu dostaneme výstup

```
Mean power calculation for normal distribution with known variance  
d = 0.5  
n = 46.71587  
sig.level = 0.01  
power = 0.8  
alternative = two.sided
```

Po nahlédnutí do definice funkce `pwr.norm.test` v knihovně `pwr` zjistíme, že hodnotu  $n$  počítá numericky. Získaný výsledek se proto od našeho výpočtu podle vzorce (6) liší s přesností na setiny. Po zaokrouhlení nahoru je požadovaná velikost testované skupiny opět  $n = 47$ .

Nyní se podíváme, jak bychom stejnou úlohu řešili v SAS Studiu. Program neumí pracovat s jednovýběrovým z-testem pro test střední hodnoty, ale umí pracovat s jeho dobrou alternativou – jednovýběrovým t-testem. Zkusíme tedy se stejným zadáním spočítat  $n$  pro jednovýběrový t-test. Použijeme kód

```
PROC POWER;  
    ONESAMPLEMEANS  
    TEST = T  
    MEAN = 14  
    NULLMEAN = 10  
    STDDEV = 8  
    POWER = 0.8  
    ALPHA = 0.01  
    NTOTAL = .  
;  
RUN;
```

kde parametrem `ONESAMPLEMEANS` zvolíme jednovýběrový test pro výpočet střední hodnoty, parametrem `TEST` zvolíme t-test, parametrem `NULLMEAN` nulovou střední hodnotu  $\mu_0$ , parametrem `MEAN` zvolíme hodnotu parametru  $\mu$  ze vzorce pro velikost účinku. Pro velikost rozdílu  $\mu - \mu_0$  bude mít test požadovanou sílu  $1 - \beta$ . Parametrem `STDDEV` výběrovou směrodatnou odchylku  $S$ , parametrem `POWER` sílu testu  $1 - \beta$ , parametrem `ALPHA` hladinu významnosti  $\alpha$  a parametrem `NTOTAL` velikost testovací skupiny. U parametru `NTOTAL` je znak „.“, protože jeho hodnotu chceme vypočítat. Podrobnější vysvětlení použitých parametrů či dalších možných parametrů najdeme v Uživatelské příručce knihovny `POWER` (viz [7], Kapitola `ONESAMPLEMEANS`).

Po spuštění kódu dostaneme jako výstup tabulku

Computed N Total	
Actual Power	N Total
0.809	51

Tabulka 3: Výstup v SAS Studiu pro jednovýběrový z-test

Počet požadovaných subjektů pro jednovýběrový t-test pomocí SAS Studia je tedy  $n = 51$ . Výsledek je podobný jako výsledky při předchozích výpočtech pro jednovýběrový z-test ( $n = 47$ ). Nejsou si ale rovny kvůli poměrně malému výslednému  $n$ . Jak jsme totiž uvedli na konci kapitoly 4.1, pro výpočet se používají kvantily t-rozdělení, které pro malé hodnoty  $n$  hůře aproximují kvantily normálního rozdělení.

## 5.2 Dvouvýběrový t-test

Opět začneme naším výpočtem podle vzorce (15). Mějme např. požadovanou hladinu významnosti  $\alpha = 0.05$ , sílu testu  $1 - \beta = 0.9$  a velikost účinku  $\frac{\mu_1 - \mu_2}{\sigma} = \frac{36 - 24}{30}$ . Dosazením do vzorce (15) dostaneme

$$n = \frac{2(u_{0.975} + u_{0.9})^2}{\left(\frac{36-24}{30}\right)^2} \doteq 131.22 \quad ,$$

po zaokrouhlení nahoru tedy  $n = 132$ . Protože jsou testované skupiny dvě, bude celkový požadovaný počet subjektů ve studii 264.

Nyní stejnou úlohu vyřešíme v SAS Studiu. Použijeme kód

```
PROC POWER;
  TWOSAMPLEMEANS
  TEST = DIFF
  MEANDIFF = 12
```

```

STDDEV = 30
POWER = 0.9
ALPHA = 0.05
NPERGROUP = .
;
RUN;

```

kde parametrem TWOSAMPLEMEANS zvolíme test týkající se středních hodnot ve dvou výběrech, parametrem TEST zvolíme t-test, parametrem MEANDIFF zvolíme velikost rozdílu středních hodnot  $\mu_1 - \mu_2$ , která když skutečně nastane, má test požadovanou sílu  $1 - \beta$ , parametrem STDDEV výběrovou směrodatnou odchylku  $S$ , parametrem POWER sílu testu  $1 - \beta$ , parametrem ALPHA hladinu významnosti  $\alpha$  a parametrem NPERGROUP velikost jedné testované skupiny. U parametru NPERGROUP je znak „.“, protože jeho hodnotu chceme vypočítat. Podrobnější vysvětlení použitých parametrů či dalších možných parametrů najdeme v Uživatelské příručce knihovny POWER (viz [7], Kapitola TWOSAMPLEMEANS).

Jako výstup dostaneme tabulku

Computed N per Group	
Actual Power	N per Group
0.901	133

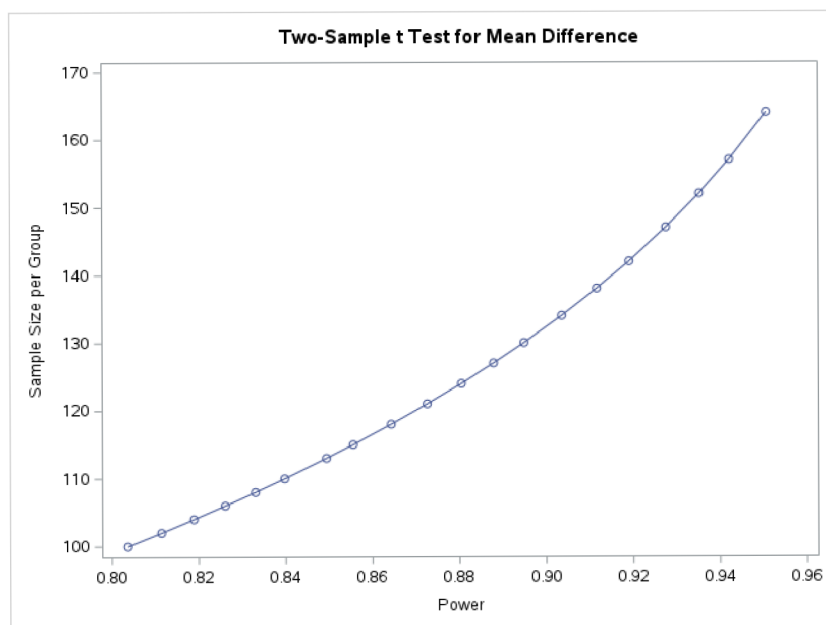
Tabulka 4: Výstup v SAS Studiu pro dvouvýběrový t-test

Požadovaný počet subjektů v každé skupině je tedy 133 a celkový počet subjektů ve studii potom 266. Díky vyššímu výslednému  $n$  a tedy i lepší aproximaci kvantilů normálního rozdělení kvantily t-rozdělení jsme dostali podobnější výsledek našemu výpočtu, než u jednovýběrového z-testu. V rozdílné přesnosti taky hraje roli různá volba hladiny významnosti u obou testů.

V SAS Studiu dále můžeme použitím kódu

```
PROC POWER;  
    TWOSAMPLEMEANS  
    TEST = DIFF  
    MEANDIFF = 12  
    STDDEV = 30  
    POWER = 0.9  
    ALPHA = 0.05  
    NPERGROUP = .  
;  
PLOT X=POWER MIN=0.8 MAX=0.95  
;  
RUN;
```

nechat vytvořit graf



Obrázek 6: Graf závislosti síly testu a počtu subjektů při pevně dané velikosti účinku  $\gamma = 12/30$  a hladině významnosti  $\alpha = 0.05$

Z grafu na obrázku 5.2 vidíme, jak se v závislosti na zvolené síle testu (osa x) mění hledaná velikost testované skupiny (osa y), a to při pevně dané velikosti účinku  $\gamma = 12/30$  a hladině významnosti  $\alpha = 0.05$ . Podobný graf závislosti bychom stejným způsobem vytvořili i u ostatních testů, nebo mezi hledaným  $n$  a jiným vybraným vstupním parametrem.

Stejnou úlohu vyřešíme v RStudiu. Použijeme kód

```
pwr.t.test(n = ,  
           d = (36-24)/30,  
           sig.level = 0.05,  
           power = 0.9,  
           type = "two.sample")
```

kde  $n$  je hledaný počet subjektů v testované skupině,  $d$  je velikost účinku ve tvaru  $d = \frac{\mu_1 - \mu_2}{\sigma}$ , `sig.level` je hladina významnosti  $\alpha$ , a `power` je síla testu  $1 - \beta$ . Samotný znak „=“ u parametru  $n$  znamená, že jeho hodnotu má algoritmus vypočítat. Další informace k parametrům nalezneme v dokumentaci knihovny `pwr` (viz [10]).

Po spuštění kódu dostaneme jako výstup

```
Two-sample t test power calculation  
n = 132.3105  
d = 0.4  
sig.level = 0.05  
power = 0.9  
alternative = two.sided  
  
NOTE: n is number in *each* group
```

Počet subjektů v jedné testovací skupině je po zaokrouhlení nahoru opět 133 a celkový počet subjektů potom 266. Získali jsme tedy stejný výsledek



jako při použití SAS Studia a téměř stejný výsledek jako při použití našeho výpočtu, opět kvůli použití kvantilů t-rozdělení místo kvantilů normálního rozdělení.

### 5.3 Jednofaktorová ANOVA

Začneme naším výpočtem. Mějme tři testované skupiny ( $k = 3$ ) o velikosti  $n$ . Velikost skupin přitom chceme stanovit tak, aby test byl schopen s danou silou  $1 - \beta$  odhalit následující rozdíly mezi středními hodnotami v jednotlivých skupinách:  $\mu_1 = 61$ ,  $\mu_2 = 56$ ,  $\mu_3 = 48$ , a to při směrodatné odchylce  $\sigma = 20$ . Známe tedy požadovanou velikost účinku (viz závěr Kapitoly 4.3)

$$\gamma = \frac{\sum_{i=1}^3 (\mu_i - \frac{\sum_{j=1}^3 \mu_j}{3})^2}{\sigma^2} = \frac{(61 - 55)^2 + (56 - 55)^2 + (48 - 55)^2}{20^2} \doteq 0.21 ,$$

dále hladinu významnosti  $\alpha = 0.05$  a sílu testu  $1 - \beta = 0.8$ .

Dosazením zadaných veličin  $\alpha, \beta$  a  $k$  do rovnosti (19) získáme rovnici

$$\chi_2^2(\chi_2^2(0.95)|\lambda) \doteq 0.2 ,$$

kterou vyřešíme pro  $\lambda$  a získáme  $\lambda_{\alpha, \beta} \doteq 9.63$ .

Dosazením vypočítaných hodnot parametrů  $\gamma$  a  $\lambda_{\alpha, \beta}$  do vzorce (20) dostaneme

$$n = \frac{9.63}{0.21} \doteq 45.86.$$

Po zaokrouhlení nahoru tedy budeme pro jednu testovací skupinu potřebovat 46 subjektů, pro všechny tři skupiny dohromady potom 138 subjektů.

V SAS Studiu budeme stejnou úlohu řešit pomocí následujícího kódu

```
PROC POWER;  
    ONEWAYANOVA  
    TEST = OVERALL  
    GROUPMEANS = 61 | 56 | 48  
    STDDEV = 20  
    POWER = 0.8  
    ALPHA = 0.05  
    NPERGROUP = .  
;  
RUN;
```

kde parametrem `ONEWAYANOVA` zvolíme výpočet pro jednofaktorovou analýzu rozptylu, parametrem `GROUPMEANS` zvolíme hodnoty středních hodnot  $\mu_1$ ,  $\mu_2$ , a  $\mu_3$  ze vzorce pro velikost účinku, parametrem `STDDEV` zvolíme výběrovou směrodatnou odchylku  $S$ , parametrem `POWER` zvolíme sílu testu  $1 - \beta$ , parametrem `ALPHA` hladinu významnosti  $\alpha$ , a parametrem `NPERGROUP` velikost jedné testovací skupiny. Za parametr `NPERGROUP` dosadíme „.“, protože jej chceme spočítat. Podrobnější vysvětlení použitých parametrů či dalších možných parametrů najdeme v Uživatelské příručce knihovny `POWER` (viz [7], Kapitola `ONEWAYANOVA`).

Jako výstup dostaneme tabulku

Computed N per Group	
Actual Power	N per Group
0.802	46

Tabulka 5: Výstup v SAS Studiu pro jednofaktorovou analýzu rozptylu

Podle SAS Studia jsme tedy opět dospěli ke skupině o velikosti 46 subjektů. Celkem potom potřebujeme 138 subjektů.

Nakonec úlohu vyřešíme v RStudiu. Budeme ji řešit pomocí kódu

```
pwr.anova.test(n = ,  
               k = 3,  
               f = 0.27,  
               sig.level = 0.05,  
               power = 0.8)
```

kde parametr  $n$  je hledaný počet subjektů, parametr  $k$  je počet porovnávaných skupin, parametr  $f$  je velikost účinku, `sig.level` je hladina významnosti a `power` je síla testu. U parametru  $n$  je samotný znak „=“, protože jeho hodnotu chceme vypočítat. Pozorný čtenář si může všimnout, že počítáme s velikostí účinku  $f = 0.27$  místo spočítané velikosti účinku  $\gamma = 0.21$  z části s ručním výpočtem. Kdybychom použili velikost účinku  $f = 0.21$  pro výpočet v RStudiu, dostali bychom výslednou velikost skupiny  $n = 74$ . Vysvětlení tohoto rozdílu najdeme v knize Statistical Power Analysis od Jacoba Cohena (viz [5], str. 275), ze které knihovna `pwr` čerpá (viz dokumentace knihovny [10]). V ní zjistíme, že je zde velikost účinku počítána podle vzorce

$$f = \frac{\sqrt{\frac{\sum_{i=1}^k (\mu_i - \mu)^2}{k}}}{\sigma}, \text{ kde } \mu = \frac{\sum_{j=1}^k \mu_j}{k}.$$

Dosazením do tohoto vzorce a po zaokrouhlení získáme  $f = 0.27$ , které jsme v kódu použili. Jako výstup po spuštění kódu dostaneme

```
Balanced one-way analysis of variance power calculation  
k = 3  
n = 45.06812  
f = 0.27
```

sig.level = 0.05

power = 0.8

NOTE: n is number in \*each\* group

Po zaokrouhlení nahoru jsme opět zjistili, že potřebujeme 46 subjektů pro každou skupinu. Celkový požadovaný počet subjektů pro studii je 138.

## 5.4 Chí-kvadrát test nezávislosti

Znovu si vše vyzkoušíme u chí-kvadrát testu nezávislosti. Začneme naším výpočtem. Mějme diskrétní veličiny  $X$  a  $Y$ , které mohou nabývat hodnot  $x_1, x_2, x_3$ , respektive  $y_1, y_2, y_3$ . Dále známe odhady  $\hat{p}_{ij}$  teoretických pravděpodobností  $p_{ij}$  pro  $i \in \{1, 2\}$  a  $j \in \{1, 2, 3\}$ , které můžeme zapsat do následující tabulky

X/Y	$y_1$	$y_2$	$y_3$	$\Sigma$
$x_1$	$\hat{p}_{11} = 0.1$	$\hat{p}_{12} = 0.2$	$\hat{p}_{13} = 0.15$	$\hat{p}_{1.} = 0.45$
$x_2$	$\hat{p}_{21} = 0.2$	$\hat{p}_{22} = 0.25$	$\hat{p}_{23} = 0.1$	$\hat{p}_{2.} = 0.55$
$\Sigma$	$\hat{p}_{.1} = 0.3$	$\hat{p}_{.2} = 0.45$	$\hat{p}_{.3} = 0.25$	$\hat{p} = 1$

Tabulka 6: Odhady  $\hat{p}_{ij}$  pro Chí-kvadrát test nezávislosti

Zvolili jsme požadovanou hladinu významnosti  $\alpha = 0.01$  a pravděpodobnost chyby II. druhu  $\beta = 0.2$ .

Pro získání hodnoty parametru  $\lambda_{\alpha,\beta}$  budeme řešit rovnici

$$\chi_{1.2}^2(\chi_{1.2}^2 0.99 | \lambda_{\alpha,\beta}) = 0.2,$$

viz (25). Výsledkem je  $\lambda_{\alpha,\beta} \doteq 13.88$ . Nyní už známe hodnoty všech parametrů pro dosazení do vzorce (26) pro výpočet  $n$

$$n = 13.88 \left( \frac{(0.1 - 0.3 \cdot 0.45)^2}{0.3 \cdot 0.45} + \dots + \frac{(0.1 - 0.25 \cdot 0.55)^2}{0.25 \cdot 0.55} \right)^{-1} \doteq 353.38.$$

Výsledek zaokrouhlíme nahoru. Zjistili jsme, že pro splnění požadavků daných vstupními parametry budeme pro studii potřebovat 354 subjektů.

Nyní budeme stejnou úlohu řešit v RStudiu. Použijeme kód

```
pwr.chisq.test(N = ,  
              df = 1*2,  
              w = 0.1982,  
              sig.level = 0.01,  
              power = 0.8)
```

Parametr `N` je požadovaný počet subjektů, který chceme vypočítat. Proto je u něj jen znak „`=`“. Parametr `df` je počet stupňů volnosti, `sig.level` je hladina významnosti, `power` je síla testu a parametr `w` je velikost účinku podle knihy *Statistical Power Analysis* od Jacoba Cohena (viz [5], str. 221) ve tvaru

$$w = \sqrt{\sum_{i=1}^r \sum_{j=1}^s \frac{(\hat{p}_{ij} - \hat{p}_{i.}\hat{p}_{.j})^2}{\hat{p}_{i.}\hat{p}_{.j}}}$$

Další informace k parametrům funkce `pwr.chisq.test` nalezneme v dokumentaci knihovny `pwr` (viz [10]). Spuštěním kódu dostaneme výstup

```
Chi squared power calculation  
w = 0.1982  
N = 353.3492  
df = 2  
sig.level = 0.01  
power = 0.8  
NOTE: N is the number of observations
```

Po zaokrouhlení nahoru potřebujeme pro studii podle RStudia 354 subjektů. Dostali jsme stejný výsledek jako při našem výpočtu.

V SAS Studiu se nám ani po důkladném pročtení Uživatelské příručky knihovny POWER (viz [7]) nepodařilo najít způsob, jak pracovat s Chí-kvadrát testem nezávislosti pro kontingenční tabulku větší než 2x2. Pro tabulku splňující tuto velikost, tedy pro testování nezávislosti mezi dvěma binárními proměnnými, což ale nyní není náš případ, můžeme použít parametr TWOSAMPLEFREQ (viz [7], Kapitola TWOSAMPLEFREQ).

## 5.5 McNemarův test efektu ošetření

Nejprve vypočítáme potřebný počet subjektů  $n$  ručně. Známe odhady  $\hat{p}_{ij}$  teoretických pravděpodobností  $p_{ij}$  pro  $i \in \{1, 2\}$  a  $j \in \{1, 2\}$ , které jsme zapsali do následující kontingenční tabulky

před/po	Bez vlastnosti	Výskyt vlastnosti
Bez vlastnosti	$\hat{p}_{00} = 0.2$	$\hat{p}_{01} = 0.4$
Výskyt vlastnosti	$\hat{p}_{10} = 0.25$	$\hat{p}_{11} = 0.15$

Tabulka 7: Odhady  $\hat{p}_{ij}$  pro McNemarův test efektu ošetření

Dále víme, že chceme testovat na hladině významnosti  $\alpha = 0.05$  a s pravděpodobností chyby II. druhu  $\beta = 0.2$ . Vše dosadíme do vzorce (33) a dostaneme

$$n = \frac{[u_{0.975}\sqrt{0.4 + 0.25} + u_{0.8}\sqrt{0.4 + 0.25 - (0.4 - 0.25)^2}]^2}{(0.25 - 0.4)^2} \doteq 224.12.$$

Po zaokrouhlení výsledku nahoru jsme zjistili, že budeme pro studii potřebovat 225 subjektů.

V SAS Studiu bychom úlohu řešili použitím kódu

```
PROC POWER;
    PAIREDFREQ
    DIST = NORMAL
```

```

METHOD = CONNOR
ALPHA = 0.05
POWER = 0.8
DISCPRORPORTIONS = 0.25 | 0.4
NPAIRS = .
;
RUN;

```

kde parametrem `PAIREDFREQ` volíme výpočet pro McNemarův test, parametr `ALPHA` je hladina významnosti  $\alpha$ , parametr `POWER` je síla testu  $1 - \beta$ , v parametru `DISCPRORPORTIONS` jsou hodnoty  $\hat{p}_{10} | \hat{p}_{01}$  a parametr `NPAIRS` je počet subjektů v testu. U parametru `NPAIRS` bude znak „.“, protože jde o parametr, který chceme vypočítat. Podrobnější vysvětlení použitých parametrů či dalších možných parametrů najdeme v Uživatelské příručce knihovny `POWER` (viz [7], Kapitola `PAIREDFREQ`).

Jako výstup po spuštění kódu dostaneme tabulku

Computed N Pairs	
Actual Power	N Pairs
0.801	225

Tabulka 8: Výstup v SAS Studiu pro McNemarův test efektu ošetření

Vidíme, že také podle SAS Studia potřebujeme 225 subjektů.

Knihovna `pwr` v RStudiu výpočet  $n$  pro McNemarův test nenabízí. Po hledání alternativy, kterou bychom mohli v RStudiu použít, jsme našli pouze funkci `sampleSizeMcNemar` z knihovny `biostatUZH` (viz [11]). Podle této funkce vyjde požadovaný počet subjektů 210 (kód je pro zajímavost k dispozici v příloze práce). Tento výsledek se značně liší od našeho výpočtu a od výpočtu

pomocí SAS Studia. Proto výpočet velikosti testovací skupiny pro McNemarův test v RStudiosu nedoporučujeme.



## 6 Závěr

Po představení problematiky klinických studií bylo cílem této bakalářské práce čtenáři ukázat, jak se klinické studie prolínají s matematikou a statistikou. Nejpodrobněji jsme se zaměřili na výpočet velikosti testovací skupiny podle předpokladů dané studie. Povedlo se nám ukázat, jak tento výpočet provést u všech pěti statistických testů, které jsme vybrali – konkrétně u jednovýběrového z-testu, dvouvýběrového t-testu, jednofaktorové analýzy rozptylu, chí-kvadrát testu nezávislosti a McNemarova testu efektu ošetření. Dále jsme úspěšně našli způsob, jak pro tyto testy stejnou úlohu řešit pomocí analytických programů SAS Studio a RStudio. Měli jsme možnost porovnat, zda se výpočty se stejnými vstupními parametry pomocí odvozených vzorců a softwaru lišily, nebo byly stejné. Zjistili jsme, že pokud software výpočet pro daný test nabízel, což platilo ve většině případů, byly výsledky stejné nebo velmi podobné. U výsledků kde nenastala rovnost, jsme odhalili, proč k tomu došlo. Scripty s použitými kódy jsme dali k dispozici v příloze práce.

## Seznam obrázků

1	Grafické znázornění nerovnosti (2) . . . . .	19
2	Grafické znázornění nerovnosti (11) . . . . .	24
3	Grafické znázornění rovnosti (19) . . . . .	28
4	Grafické znázornění rovnosti (25) . . . . .	31
5	Grafické znázornění nerovnosti (31) . . . . .	36
6	Graf závislosti síly testu a počtu subjektů při pevně dané velikosti účinku $\gamma = 12/30$ a hladině významnosti $\alpha = 0.05$ . . .	43

## Seznam tabulek

1	Kontingenční tabulka pro Chí-kvadrát test nezávislosti . . . .	29
2	Kontingenční tabulka pro McNemarův test efektu ošetření . .	32
3	Výstup v SAS Studiu pro jednovýběrový z-test . . . . .	41
4	Výstup v SAS Studiu pro dvouvýběrový t-test . . . . .	42
5	Výstup v SAS Studiu pro jednofaktorovou analýzu rozptylu .	46
6	Odhady $\hat{p}_{ij}$ pro Chí-kvadrát test nezávislosti . . . . .	48
7	Odhady $\hat{p}_{ij}$ pro McNemarův test efektu ošetření . . . . .	50
8	Výstup v SAS Studiu pro McNemarův test efektu ošetření . .	51

## Reference

- [1] CHOW Shein-Chung, SHAO Jun, WANG Hansheng *Sample Size Calculations in Clinical Research* Chapman and Hall/CRC , 2007. ISBN: 978-1-58488-982-3.
- [2] SVOBODNÍK, Adam; DEMLOVÁ, Regina a PECEN, Ladislav. *Klinické studie v praxi*. Brno: Facta Medica, 2014. ISBN 978-80-904731-8-8.
- [3] HRON, Karel; KUNDEROVÁ, Pavla a VENCÁLEK, Ondřej. *Základy počtu pravděpodobnosti a metod matematické statistiky*. 4. doplněné vydání. Olomouc: Univerzita Palackého v Olomouci, Přírodovědecká fakulta, 2021. ISBN 978-80-244-5990-5.
- [4] ANDĚL, Jiří. *Základy matematické statistiky*. Matfyzpress 2007. ISBN: 80-7378-001-1.
- [5] COHEN, Jacob. *Statistical power analysis for the behavioral sciences*. Routledge, 2013. ISBN: 0-8058-0283-5.
- [6] DOSTÁL, Daniel. *Statistické metody v psychologii*. Olomouc. Filozofická fakulta, 2016.
- [7] SAS Institute Inc. *SAS/STAT® 14.1 User's Guide*. Cary, NC: SAS Institute Inc. 2015.
- [8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. <https://www.r-project.org/>.
- [9] RStudio Team (2020). *RStudio: Integrated Development for R*. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.

- [10] CHAMPELY S. *pwr: Basic Functions for Power Analysis*. 2020. R package version 1.3-0. <https://CRAN.R-project.org/package=pwr>
- [11] HAILE SR, HELD L, MEYER S, RUEEGER S, RUFIBACH K, SCHWAB S. *biostatUZH: Misc Tools of the Department of Biostatistics*, EBPI, University of Zurich, 2020. R package version 1.8.0/r103. <https://R-Forge.R-project.org/projects/ebuzh/>