

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Myšlení rychlé a pomalé



Katedra matematické analýzy a aplikací matematiky
Vedoucí bakalářské práce: **Mgr. Ondřej Vencálek, Ph.D.**
Vypracovala: **Věra Kumová**
Studijní program: B1103 Aplikovaná matematika
Studijní obor Aplikovaná statistika
Forma studia: prezenční
Rok odevzdání: 2016

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Věra Kumová

Název práce: Myšlení rychlé a pomalé

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

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

Rok obhajoby práce: 2016

Abstrakt: Práce je inspirována stejnojmennou knihou Daniela Kahnemana. Jejím cílem je zamyslet se nad statistickým vyhodnocováním experimentů, které jsou v knize popsány, a některý z nich zopakovat. Prostor je věnován vhodně stanovené velikosti výběru při testování hypotéz a otázce, zda je velikost výběru podhodnocována. Dále se práce zabývá testováním vlivu takzvaného ukotvení. K tomu je použita analýza rozptylu.

Klíčová slova: Velikost výběru, síla testu, míra účinku, analýza rozptylu

Počet stran: 41

Počet příloh: 0

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Věra Kumová

Title: Thinking, fast and slow

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: Mgr. Ondřej Vencálek, Ph.D.

The year of presentation: 2016

Abstract: The thesis is inspired by the book of the same name written by Daniel Kahneman. The aim of the thesis is to deal with a statistical evaluation of the experiments, which are described in the book, and to repeat one of them. Main focus is on an appropriately determined sample size and an issue, whether the sample size is underestimated. The thesis also deals with the testing of an effect of so called anchoring, to which analysis of variance is used.

Key words: Sample size, power of a test, effect size, analysis of variance

Number of pages: 41

Number of appendices: 0

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracovala samostatně pod vedením Mgr. Ondřeje Vencálka, Ph.D. a všechny použité zdroje jsem uvedla v seznamu literatury.

V Olomouci dne 19. dubna 2016

.....
podpis

Poděkování

Ráda bych poděkovala svému vedoucímu Mgr. Ondřeji Vencálkovi, Ph.D. za odborné rady, cenné připomínky a především za čas, který mi věnoval.

Obsah

Úvod	7
1 Zákon malých čísel	8
1.1 Úvod do problému	8
1.1.1 Jednovýběrový t–test	9
1.2 Výpočet velikosti vzorku	10
1.2.1 Odvození vzorce pro rozsah výběru	13
1.3 Průběh skutečného testování	15
1.4 Vlastní návrh testu hypotézy o podhodnocování velikosti výběrového vzorku	20
2 Ukotvení	26
2.1 Úvod do problému	26
2.2 Efekt ukotvení: vlastní experiment	27
2.2.1 Volba dat a následný sběr	29
2.3 Analýza dat - ANOVA	31
2.4 Závěry experimentu	35
2.4.1 Sběr dat	36
2.4.2 Analýza dat	37
2.4.3 Výsledky experimentu	37
Závěr	39
Literatura	41

Úvod

V roce 2011 vyšla ve Spojených státech kniha s názvem *Thinking, Fast and Slow*. Jejím autorem je Daniel Kahneman, psycholog a nositel Nobelovy ceny za ekonomii. Ve své knize popisuje výsledky experimentů, které prováděl často ve spolupráci s Amosem Tverským. Experimenty těchto dvou psychologů se týkaly lidského myšlení, mimo jiné například toho, jak lidem chybí intuice pro statistiku. Kniha se stala inspirací pro téma této bakalářské práce.

Cílem mé práce je zamyslet se nad tím, jak se podobné experimenty statisticky vyhodnocují a některý z experimentů zopakovat. Práce je rozčleněna do dvou kapitol, přičemž v každé z nich se věnuji jednomu samostatnému tématu.

První kapitola vychází z kapitoly výše uvedené knihy *Zákon malých čísel* a nese stejný název. V této kapitole se zabývám správně zvolenou velikostí výběrového vzorku při provádění statistických testů. Součástí kapitoly je také vlastní návrh testu hypotézy o podhodnocování velikosti výběrového vzorku.

Druhá kapitola se zabývá fenoménem *ukotvení*, což je jev, kdy konkrétní hodnota neznámého množství ovlivňuje následný odhad. V této kapitole provedu i vlastní experiment, přičemž data k tomuto experimentu sesbírám pomocí sociální sítě a následně provedu vhodný statistický test.

Veškeré výpočty či analýzy dat provedu buď ručně, nebo v softwaru R.

Kapitola 1

Zákon malých čísel

V první kapitole se budeme zabývat tématem volby velikosti výběrového vzorku při statistickém testování hypotéz. Teorie k jednotlivým testům byla čerpána z [3]. Dalšími důležitými zdroji této kapitoly byly články [2] a [7].

1.1. Úvod do problému

V knize [6] v kapitole *Zákon malých čísel* píše autor o článku Jacoba Cohena, ve kterém byl demonstrován omyl vědeckých pracovníků týkající se velikosti výběrového vzorku. Dle článku vybírají psychologové tak malé vzorky, že se vystavují 50% riziku nemožnosti potvrdit svoji vlastní pravdivou hypotézu. Tento problém je zapříčiněn tím, že vědečtí pracovníci nerozhodují o velikosti vzorku na základě výpočtu, ale pouze intuitivně.

Kahneman se svým kolegou následně vytvořili svůj vlastní dotazník, pomocí kterého chtěli otestovat, jestli se „vědečtí pracovníci zběhlí v matematice dopouštějí stejných omylů“. V dotazníku byly popsány realistické výzkumné situace a respondenti měli mimo jiné zvolit pro jednotlivé případy velikost vzorku. Výsledkem experimentu bylo, že ani jejich respondenti, odborníci na matematiku, nevěnovali dostatečnou pozornost velikosti vzorků.

1.1.1. Jednovýběrový t–test

Na test hypotézy, že vědečtí pracovníci podhodnocují velikost vzorků, by byl vhodný jednovýběrový t–test. Zkoumaná náhodná veličina X by zde byla velikost vzorku uvedená respondenty. Předpokládejme, že má normální rozdělení o parametrech μ a σ^2 , přičemž tyto parametry neznáme. O parametru μ se však domníváme, že by mohl být roven hodnotě, která odpovídá vhodně zvolené velikosti vzorku. Tu pro daný případ určíme výpočtem. Tuto hodnotu označíme jako μ_0 . Následně tedy testujeme nulovou hypotézu $H_0: \mu = \mu_0$ proti jednostranné alternativě $H_A: \mu < \mu_0$. Pro výpočet použijeme výběrovou funkci

$$T(\mathbf{X}) = \frac{\bar{X}_n - \mu}{S_n} \sqrt{n} \sim t_{n-1}.$$

Za předpokladu, že H_0 je správná, dosazujeme do T ve všech případech $\mu = \mu_0$, přičemž ve prospěch alternativy budou svědčit „malé“ hodnoty testového kritéria. Kritický obor bude proto vypadat takto

$$W = (-\infty, -t_{n-1, 1-\alpha}).$$

Slabá stránka t–testu ovšem spočívá v tom, že předpokládá normální rozdělení, které zde není zaručeno. Tvar rozdělení bude totiž vycházet ze správné velikosti vzorku, která zároveň určí, o kolik je možné velikost vzorku podhodnotit. Pokud například velikost vzorku bude 1 000, máme celkem dost prostoru na podhodnocení a rozdělení by mohlo být symetrické. Pokud však bude stačit 10 respondentů, můžeme podhodnotit vzorek maximálně o devět, přičemž tedy předpokládáme nejmenší rozsah výběru roven jednomu respondentovi. Oproti tomu však máme mnohem více možností, jak vzorek nadhodnotit. Je tedy vhodné předpokládat, že rozdělení bude spíše zešíkmené.

Porušení předpokladu normality nás tedy vede k použití neparametrického testu. Než však přistoupíme k jeho popisu, měli bychom se podrobněji podívat na to, jak se vlastně velikost vzorku počítá.

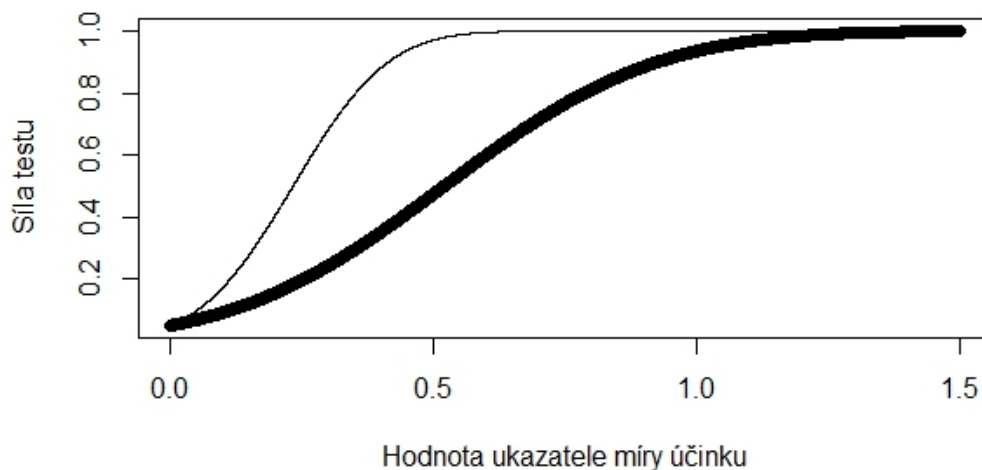
1.2. Výpočet velikosti vzorku

Velikost vzorku počítáme na základě zvolené pravděpodobnosti chyby 1. druhu α , síly testu a takzvané očekávané *míry účinku*. Je na místě jednotlivé aspekty krátce popsat.

Chyba 1. druhu znamená, že jsme zamítli správnou hypotézu. Test konstruueme tak, aby pravděpodobnost, že zamítneme správnou hypotézu, byla nejvýše α , což je vhodně zvolené malé číslo, které bývá nejčastěji 0,05 a říká se mu *hladina testu*.

Síla testu souvisí s chybou 2. druhu, tedy s nezamítnutím chybné nulové hypotézy. Pravděpodobnost této chyby se značí β . Síla testu je potom $1 - \beta$ a představuje pravděpodobnost, že zamítneme nulovou hypotézu, pokud je chybná.

Konečně očekávaná míra účinku, neboli ES (z anglického *effect size*), je „rozdíl“ mezi hodnotou v nulové a alternativní hypotéze. Zde je třeba upozornit na to, že alternativní hypotéza musí být specifická a nestačí nám pouhá negace hypotézy nulové. Velikost výběru se potom počítá pomocí *ukazatelů míry účinku* neboli *ES indexů*.



Obrázek 1.1: Graf silofunkce pro jednovýběrový t-test. Tenká čára vykresluje silofunkci pro $n = 50$, spodní tlustá čára pro $n = 10$.

Na obrázku 1.1 je patrné, proč je nutné stanovit specifickou alternativní hypotézu. Tenká čára v grafu vykresluje silofunkci pro rozsah výběru $n = 50$, spodní tlustá je pro rozsah výběru $n = 10$. Obor hodnot silofunkce se pohybuje v rozmezí $\langle \alpha, 1 \rangle$, přičemž tento graf je vykreslen pro $\alpha = 0,05$. Z grafu je zřejmé, že abychom dokázali spočítat sílu testu pro dané n , která je na ose y , musíme znát přesnou hodnotu ukazatele míry účinku, která je na ose x .

Každý statistický test má svůj vlastní způsob výpočtu tohoto ukazatele. Uvedme si tři příklady ukazatelů tak, jak je zavedl Cohen.

- Pro dvouvýběrový t -test Cohen zavádí ukazatel míry účinku d , který se počítá jako rozdíl středních hodnot vydělený směrodatnou odchylkou, tedy

$$d = \frac{|\mu_1 - \mu_2|}{\sigma},$$

kde μ_1 je střední hodnota prvního výběru, μ_2 je střední hodnota druhého výběru a σ je směrodatná odchylka v jednotlivých výběrech, přičemž předpokládáme, že je pro oba výběry stejná.

- Pro testování toho, zda se velikost proporce liší od jedné poloviny, zavádí ukazatel míry účinku g . Pokud bychom chtěli například otestovat, zda je na vysokých školách více dívek než chlapců, zvolíme nulovou hypotézu, že populační proporce p je rovna jedné polovině, tedy $H_0: p = p_0 = 0,5$. Alternativní hypotéza musí být specifická, zvolme například $H_A: p = p_A = 0,6$ ve prospěch děvčat. Potom

$$g = p_A - p_0 = 0,6 - 0,5 = 0,1.$$

- Pro analýzu rozptylu zavádí ukazatel míry účinku f , který se vypočítá jako směrodatná odchylka jednotlivých středních hodnot vydělaná směrodatnou odchylkou v jednotlivých skupinách, která je ovšem pro všechny skupiny stejná:

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

kde k je počet skupin, μ je společná střední hodnota, μ_k je střední hodnota v k -té skupině a σ je směrodatná odchylka v jednotlivých výběrech.

Nicméně vzhledem k faktu, že občas bývá náročné určit konkrétní alternativní hypotézu, pro usnadnění zavádí Cohen „malé“, „střední“ a „velké“ hodnoty ukazatelů míry účinku. Například pro ukazatel d je stanovil na 0,2, 0,5 a 0,8. Pro ukazatel f určil 0,1, 0,25 a 0,4.

Mezi velikostí očekávaného efektu a velikostí vzorku platí při pevně stanovené hladině a síle testu nepřímá úměra, stejně tak mezi velikostí vzorku a hladinou testu při pevně daném efektu a pevně dané síle. Naopak čím silnější chceme na dané hladině test, tím větší musí být náš vzorek k prokázání daného efektu.

Cohen uvádí příklad, kdy pro hladinu testu $\alpha = 0,01$, sílu testu 0,99 a rozdíl středních hodnot se středně velkým efektem ($d = 0,5$) potřebujeme dvě skupiny o 194 pozorováních. Pokud však zvolíme hladinu testu $\alpha = 0,05$, stačí nám v každé skupině 148 pozorování. Tyto dva příklady si teď vypočítáme, přičemž budeme vycházet ze vzorců uvedených v [5].

Prvotně je třeba si uvědomit, co znamená $d = 0,5$. Dle vzorce výše dosadíme:

$$\frac{1}{2} = \frac{|\mu_1 - \mu_2|}{\sigma}$$

$$|\mu_1 - \mu_2| = \frac{1}{2}\sigma$$

Střední hodnota v první skupině je tedy o polovinu směrodatné odchylky menší či větší než střední hodnota ve druhé skupině.

Velikost vzorku n vypočítáme následovně:

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(u_{\frac{\alpha}{2}} + u_{\beta})^2}{(\mu_1 - \mu_2)^2}.$$

Pro $\sigma_1^2 = \sigma_2^2 = \sigma^2$ můžeme upravit na

$$n = \frac{2\sigma^2(u_{\frac{\alpha}{2}} + u_{\beta})^2}{(\mu_1 - \mu_2)^2} \quad (1.1)$$

a pro $\mu_1 - \mu_2 = c\sigma$ dostáváme

$$n = \frac{2(u_{\frac{\alpha}{2}} + u_{\beta})^2}{c^2}.$$

Hodnoty $u_{\frac{\alpha}{2}}$ a u_{β} jsou kvantily normálního normovaného rozdělení pro zadanou hladinu testu α a sílu testu $1 - \beta$, $c \in \mathbb{R} \setminus \{0\}$.

Velikosti vzorků v uvedených příkladech nyní jednoduše vypočítáme dosazením do posledního vzorce:

$$n_1 = \frac{2(u_{0,005} + u_{0,01})^2}{0,5^2} \doteq 192,25$$

$$n_2 = \frac{2(u_{0,025} + u_{0,01})^2}{0,5^2} \doteq 146,98.$$

Velikosti vzorku jsou tedy při hladině $\alpha = 0,01$ 193 pozorování v obou skupinách, pro hladinu testu $\alpha = 0,05$ 147 pozorování v obou skupinách. Velikosti vzorku jsou v obou případech o jedno menší, což by však mohlo být způsobeno zaokrouhlováním kvantilů. Naše výpočty byly provedeny pomocí softwaru R, takže k výraznému zaokrouhlení došlo až u konečného výsledku.

1.2.1. Odvození vzorce pro rozsah výběru

V této kapitole si odvodíme vzorec použitý pro výpočet velikosti vzorku u dvouvýběrového t-testu, který je převzatý z [5]. Vzhledem k tomu, že dvouvýběrový t-test předpokládá u obou skupin stejné rozptyly, budeme odvozovat vzorec pro $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Dále předpokládáme, že počet probandů je v obou skupinách stejný, tedy $n_1 = n_2 = n$.

Testová statistika dvouvýběrového t-testu pro $n_1 = n_2 = n$ vypadá takto:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n-1)(S_X^2 + S_Y^2)}} \sqrt{\frac{n(2n-2)}{2}} \sim t_{2n-2},$$

kde \bar{X} a S_X^2 je výběrový průměr a rozptyl v první skupině a \bar{Y} a S_Y^2 je výběrový průměr a rozptyl ve druhé skupině.

Při odvozování vzorce však nebudeme pracovat s výběrovými rozptyly ani se Studentovým rozdělením. Využijeme toho, že s rostoucím n se výběrový rozptyl blíží skutečnému rozptylu a s rostoucími stupni volnosti se Studentovo rozdělení blíží normálnímu rozdělení. Budeme tedy vycházet z testové statistiky Z :

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{2}{n}\sigma^2}} \sim N(\mu_1 - \mu_2, 1),$$

kde σ^2 je rozptyl v jednotlivých výběrech. Předpokládáme, že tento rozptyl je v obou výběrech stejný.

Nulová hypotéza u dvouvýběrového z -testu zní, že se střední hodnoty rovnají, tedy $H_0: \mu_1 = \mu_2$. Za platnosti nulové hypotézy má tedy testová statistika Z normované normální rozdělení. Při oboustranné alternativě nulovou hypotézu zamítneme, pokud

$$|\bar{X} - \bar{Y}| > u_{1-\frac{\alpha}{2}} \frac{\sqrt{2}\sigma}{\sqrt{n}}.$$

Jak již bylo uvedeno, pro výpočet velikosti vzorku musíme znát konkrétní alternativní hypotézu, abychom mohli zajistit požadovanou sílu testu $1 - \beta$. Zvolme proto specifickou alternativu, kdy rozdíl $|\mu_1 - \mu_2|$ je nějaké konkrétní nenulové číslo. Za platnosti alternativy můžeme psát, že

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{2\sigma^2}{n}\right),$$

kde $\mu_1 - \mu_2$ je nám známý rozdíl středních hodnot výběrů. Bez újmy na obecnosti předpokládejme, že $\mu_1 > \mu_2$. Pro chybu 2. druhu musí platit

$$P_A\left(|\bar{X} - \bar{Y}| < u_{1-\frac{\alpha}{2}} \frac{\sqrt{2}\sigma}{\sqrt{n}}\right) \leq \beta,$$

tedy pokud platí alternativa, pak pravděpodobnost nezamítnutí nulové hypotézy musí být menší nebo rovna β . Pokud rozdíl mezi středními hodnotami rozdělení za alternativní a nulové hypotézy není příliš malý, můžeme psát

$$P_A\left(|\bar{X} - \bar{Y}| < u_{1-\frac{\alpha}{2}} \frac{\sqrt{2}\sigma}{\sqrt{n}}\right) \doteq P_A\left(\bar{X} - \bar{Y} < u_{1-\frac{\alpha}{2}} \frac{\sqrt{2}\sigma}{\sqrt{n}}\right),$$

protože $P_A \left(\bar{X} - \bar{Y} < u_{\frac{\alpha}{2}} \frac{\sqrt{2\sigma}}{\sqrt{n}} \right)$ je zanedbatelně malá.

Výraz v závorce si nyní upravíme tak, aby na levé straně byla náhodná veličina s normovaným normálním rozdělením a na pravé straně kvantil normovaného normálního rozdělení:

$$P_A \left(\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{2\sigma}} \sqrt{n} < \frac{u_{1-\frac{\alpha}{2}} \frac{\sqrt{2\sigma}}{\sqrt{n}} - (\mu_1 - \mu_2)}{\sqrt{2\sigma}} \sqrt{n} \right) \leq \beta.$$

Platí, že $\frac{u_{1-\frac{\alpha}{2}} \frac{\sqrt{2\sigma}}{\sqrt{n}} - (\mu_1 - \mu_2)}{\sqrt{2\sigma}} \sqrt{n} = u_\beta$ a $u_{1-\frac{\alpha}{2}} = -u_{\frac{\alpha}{2}}$. Po úpravách dostaneme:

$$-\frac{\mu_1 - \mu_2}{\sqrt{2\sigma}} \sqrt{n} = u_\beta + u_{\frac{\alpha}{2}}$$

$$-\sqrt{n} = \frac{(u_\beta + u_{\frac{\alpha}{2}}) \sqrt{2\sigma}}{\mu_1 - \mu_2}$$

$$n = \frac{2\sigma^2 (u_{\frac{\alpha}{2}} + u_\beta)^2}{(\mu_1 - \mu_2)^2},$$

což je vzorec 1.1. Obdobně lze odvodit vzorec pro výpočet velikosti vzorku u jednovýběrového t-testu, případně pro jednostranné alternativy.

V dalším textu už nebudeme velikosti vzorku počítat ručně, ale budeme pracovat se softwarem R a s balíčkem `pwr` od Stéphane Champelyho [4].

1.3. Průběh skutečného testování

V této podkapitole se podíváme, jak k testování přistoupili Kahneman s Tverským. Jejich postup je částečně popsán v [7].

Autoři vytvořili dotazník, který předložili dvěma menším skupinám respondentů. Jednu skupinu tvořili účastníci konference o matematické psychologii, druhou pak účastníci srazu Americké psychologické asociace. Následně prezentovali mediány odpovědí na jednotlivé otázky, které si nyní představíme. První otázka zní následovně.

- Představte si, že jste provedli experiment na dvaceti subjektech a získali jste signifikantní výsledek, který potvrzuje vaši teorii ($z = 2,23$, $p < 0,05$, oboustranný test). Nyní provedete stejný experiment na deseti subjektech. Jaká je podle vás pravděpodobnost, že výsledek bude signifikantní, pokud použijete jednostranný test?

Medián uváděných pravděpodobností byl 0,85. Pouze 9 z 84 respondentů dalo odpověď v rozsahu 0,4 – 0,6, přičemž správná odpověď by se měla pohybovat kolem 0,48. Nyní si spočítáme, jak se k této pravděpodobnosti dopracujeme.

Postup je takový, že hodnotu získanou v prvním testu použijeme pro stanovení alternativní hypotézy druhého testu, tedy $H_A: \mu = \mu^* = \bar{x}_1$, kde \bar{x}_1 je průměr získaný v prvním experimentu. Pravděpodobnost, že výsledek druhého experimentu bude signifikantní, je potom síla druhého testu proti alternativní hypotéze určené výsledkem prvního testu. Využijeme tedy tento postup a pravděpodobnost spočítáme.

Autoři uvedli výsledek $z = 2,23$, použili tedy z -test, který vypadá takto:

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

kde μ_0 je střední hodnota za platnosti nulové hypotézy, přičemž běžně se stanovuje $\mu_0 = 0$, a σ je známá směrodatná odchylka. Když do testové statistiky dosadíme $n = 20$, což je velikost vzorku v prvním experimentu, a $\mu_0 = 0$, získáme

$$\frac{\bar{x}_1}{\sigma} \doteq 0,4986.$$

Sílu testu obdržíme, když spočítáme pravděpodobnost, že za platnosti alternativní hypotézy zamítneme hypotézu nulovou. Tedy

$$1 - \beta = P_A(\bar{X} - \mu_0 > u_{1-\alpha} \frac{\sigma}{\sqrt{n}}) = 1 - P_A(\bar{X} - \mu_0 < u_{1-\alpha} \frac{\sigma}{\sqrt{n}}).$$

Počítejme s $\alpha = 0,05$. V tabulkách najdeme nebo pomocí softwaru spočítáme, že $u_{0,95} = 1,645$. Protože za platnosti H_A má $\bar{X} \sim N(\bar{x}_1, \frac{\sigma^2}{n})$, tak pro druhý test o

rozsahu výběru $n = 10$ vypočítáme:

$$1 - \beta = 1 - \Phi\left(u_{1-\alpha} - \frac{\bar{x}_1}{\sigma} \sqrt{n}\right)$$

$$1 - \beta = 1 - \Phi(1,645 - 0,4986\sqrt{10})$$

$$1 - \beta \doteq 0,473$$

Výsledek jsme spočítali v softwaru R pomocí kódu

```
1-pnorm(qnorm(0.95)-0.4986*sqrt(10)).
```

Podívejme se na další otázku, kterou autoři položili svým respondentům.

- *Jeden z vašich doktorandů dokončil obtížný a časově náročný experiment provedený na 40 zvířatech. Vyhodnotil a zanalyzoval spoustu proměnných. Jeho výsledky jsou převážně neprůkazné, ale u jednoho srovnání před a po mu vyšla vysoce signifikantní hodnota $t = 2,7$, což je překvapivé a mohlo by to mít zásadní význam. Když zvážíte význam výsledku, jeho překvapivou hodnotu a počet analýz, které váš student provedl, doporučili byste mu studii před publikováním zopakovat? Pokud ano, na kolika zvířatech?*

Autoři uvádí, že 66 z 75 respondentů by doporučilo studii zopakovat, přičemž medián odpovědí byl, aby student experiment zopakoval na 20 subjektech. Pokud by však student poslechl tuto radu, tak při stejném průměru a výběrovém rozptylu jako u prvního experimentu by získal hodnotu testové statistiky $t_2 = 1,88$ a jeho šance na získání signifikantního výsledku by při jednostranném testu byla lehce nad polovinou. Pojdme si to spočítat.

Student získal testovou statistiku $t_1 = 2,7$, přičemž dělal srovnání před a po. Použil tedy párový t -test, který se počítá jako jednovýběrový t -test aplikovaný na rozdíly průměrů. Spočítáme testovou statistiku t_2 pro výběr o 20 subjektech, přičemž za μ_0 dosadíme nulu:

$$t_1 = 2,7 = \frac{\bar{X} - \mu_0}{S} \sqrt{n_1 - 1} = \frac{\bar{X}}{S} \sqrt{39}$$

Zjištěnou hodnotu podílu $\frac{\bar{X}}{S}$ pak dosadíme do vzorce pro testovou statistiku t_2 :

$$t_2 = \frac{\bar{X}}{S} \sqrt{n_2 - 1} = \frac{2,7}{\sqrt{39}} \sqrt{19} \doteq 1,88$$

Pro tuto testovou statistiku ještě v R spočítáme p-hodnotu. Vyjde nám 0,038. Výsledek je tedy pořád významný, avšak p-hodnota je výrazně vyšší než pro výsledek $t_1 = 2,7$, kdy vychází 0,007.

Pravděpodobnost, že student získá v druhém experimentu s poloviční velikostí výběru signifikantní výsledek spočítáme stejně jako v předchozím příkladě:

$$1 - \beta \doteq 1 - \Phi\left(u_{1-\alpha} - \frac{\bar{x}_1 - \mu_0}{\sigma} \sqrt{n_2}\right)$$

$$1 - \beta \doteq 1 - \Phi(1,645 - 0,4323\sqrt{20})$$

$$1 - \beta \doteq 0,613$$

V tomto výpočtu jsme využili asymptotiky a počítali jsme s kvantily normálního normovaného rozdělení. Pravděpodobnost, že doktorand získá signifikantní výsledek, nám vyšla přibližně 0,61. Nyní zkusme sílu testu spočítat pomocí kvantilů Studentova rozdělení. Kód v R bude vypadat takto:

```
1-pt(qt(0.95,19)-0.4323*sqrt(19),19).
```

V tomto případě nám síla testu vyšla 0,56.

Dle autorů článku, pokud by doktorand předpokládal, že jeho výsledek je platný a zároveň by byl ochotný přijmout riziko jen 0,1, že při zopakování experimentu získá testovou statistiku t_2 nižší než 1,7 (což je 95% kvantil Studentova rozdělení konkrétně pro 28 stupňů volnosti), měl by druhou studii provést se zhruba 50 zvířaty. Pokud by hodnota testové statistiky v prvním experimentu vyšla nižší ($t_1 = 2,2$), velikost vzorku u druhé studie se stejnou mírou účinku by se zvýšila až na zhruba 75. Uvedené velikosti vzorku si opět spočítáme, tentokrát pomocí softwaru R a funkce `pwr.t.test`, kterou najdeme ve výše zmíněném balíčku `pwr`.

Funkce `pwr.t.test` vyžaduje tři z následujících čtyř parametrů: velikost vzorku, míra účinku, hladina významnosti a síla testu. Když tři z nich zadáme, čtvrtý nám dopočítá. Pokud u této funkce nezadáme typ testu, bude automaticky počítat hodnotu pro dvouvýběrový t-test. My chceme spočítat velikost vzorku pro jednostrannou alternativu. Síla testu je 0,9, hladina významnosti 0,05. Míra účinku je v tomto případě rovna $\frac{2,7}{\sqrt{39}}$. Kód v R bude tedy vypadat takto:

```
pwr.t.test(d=2.7/sqrt(39),sig.level=0.05,power=0.9,
           type="one.sample",alternative="greater")
```

Vyšlo nám $n = 47,2$, po zaokrouhlení tedy $n = 50$. Pro výpočet velikosti vzorku za předpokladu, že by první hodnota testové statistiky vyšla nižší, upravíme pouze velikost míry účinku:

```
pwr.t.test(d=2.2/sqrt(39),sig.level=0.05,power=0.9,
           type="one.sample",alternative="greater")
```

V tomto případě nám vyšlo $n = 70,4$ a při hrubém zaokrouhlení dostáváme $n = 75$.

Nyní se podívejme na poslední příklad, který autoři uvádějí ve svém článku.

- *Provedli jste korelační studii, ve které jste zkoumali 20 proměnných na 100 subjektech. 27 ze 190 korelačních koeficientů¹ bylo významných na hladině $\alpha = 0,05$, 9 z nich na hladině $\alpha < 0,01$. Průměr z absolutních hodnot významných korelačních koeficientů byl 0,31 a výsledek je dobře teoreticky odůvodnitelný. Kolik z těchto 27 významných korelačních koeficientů by podle vás bylo významných i v případě, že by byla ta samá studie provedena na výběru o velikosti 40?*

Nejdříve spočítáme sílu testu při daných parametrech, tedy $n = 40$, $r = 0,31$ a $\alpha = 0,05$. Tentokrát použijeme ze stejného balíčku funkci `pwr.r.test` a kód v R bude vypadat takto:

¹Počet korelačních koeficientů z 20 proměnných zjistíme tak, že spočítáme počet všech možných dvojic z 20 prvků, tedy $\binom{20}{2} = 190$.

`pwr.r.test(n=40,r=0.31,sig.level = 0.05)`

Vyšla nám síla testu 0,5, tedy počet významných korelačních koeficientů by měl klesnout na polovinu, tj. z původních 27 nenulových koeficientů by jich mělo být významných 13 až 14². Korelace v druhé studii se však budou od těch původních lišit a dle autorů dojde kvůli regresi k dalšímu snížení počtu významných korelačních koeficientů na 8 až 10. Přesto byl medián odhadů respondentů 18.

Autoři na základě svého dotazníku usuzují na obecnou víru v zákon malých čísel, který říká, že zákon velkých čísel platí i pro malá čísla. Jejich typický respondent tomuto zákonu věřil bez ohledu na to, jestli pocházel ze skupiny účastníků konference o matematické psychologii nebo to byl účastník srazu Americké psychologické asociace, i když samozřejmě není zaručeno, že jejich výběry byly reprezentativní.

V další podkapitole se budeme zabývat vlastním návrhem testování hypotézy, zda vědečtí pracovníci podhodnocují velikost testovacího vzorku.

1.4. Vlastní návrh testu hypotézy o podhodnocování velikosti výběrového vzorku

Nyní navrhne postup, jakým by se dala otestovat hypotéza, zda vědečtí pracovníci obecně podhodnocují velikosti výběrových vzorků. Výzkum skutečně provádět nebudeme, jedná se o fiktivní studii, ve které se také zamyslíme nad možnými úskalími výzkumu.

Autoři provedli dotazníkové šetření, přičemž jejich respondenty byli účastníci konferencí, kteří byli považováni za odborníky zblhlé v matematice. Při návrhu vlastního testování také využijeme formu dotazníku, jelikož je časově poměrně nenáročná, dokáže zaručit anonymitu a také není moc finančně nákladná. Pro získání potřebných hodnot by bylo možné využít i formu rozhovoru, avšak ta nezaručuje výše zmíněná pozitiva.

²Máme předpoklad, že v první studii vyšlo 27 nenulových koeficientů, zbývajících 163 je nulových.

Jelikož testujeme znalosti *vědeckých pracovníků zběhlých v matematice*, měli bychom blíže specifikovat naše respondenty. Za vědecké pracovníky bychom mohli považovat zaměstnance univerzit, po kterých jejich povolání může vyžadovat testování různých hypotéz. Může se jednat o odborníky z oblastí medicíny, chemie, fyziky, ale také právě psychologie či sociologie. Všichni tito lidé by za sebou měli mít nějaký základní kurz statistiky, který by jim měl poskytnout znalosti potřebné pro testování hypotéz v jejich oboru. Mezi naše respondenty bychom mohli samozřejmě zařadit také vědecké pracovníky zabývající se přímo matematikou a statistikou, u kterých bychom nejspíš i předpokládali nejpřesnější odpovědi. Dotazníky bychom jim mohli rozeslat mailem, případně poštou. Pokud bychom však posílali poštou, museli bychom počítat s nižší návratností, případně s vyššími náklady spojenými s ofrankovanou obálkou pro každého respondenta.

Co se týče právě návratnosti, můžeme provést určitá opatření, která by ji mohla zvýšit. Například můžeme vytvořit stručnou webovou stránku, kde bude výzkum podrobněji popsán. Také můžeme v respondentech vhodným způsobem vzbudit zvědavost a slíbit jim, že se s nimi podělíme o svá výzkumná zjištění. I přes tato opatření bychom měli raději počítat s návratností zhruba 50 %.

V otázce respondentů nestačí jen určit, koho se budeme ptát, ale také kolika lidí se budeme ptát. K tomu je potřeba promyslet hlavně to, jaký očekáváme efekt neboli míru účinku. Pro výpočet velikosti vzorku použijeme R a už zmíněnou funkci `pwr.t.test`. Hladinu významnosti zvolíme standardně $\alpha = 0,05$, sílu testu zvolme 0,9. Co se týče míry účinku, zkusme využít Cohenem zavedené zjednodušení a vypočítejme velikosti vzorku pro malou, střední a velkou míru účinku. Mějme na paměti, že naše alternativní hypotéza je jednostranná. Do R tedy zadáme kódy:

```
pwr.t.test(d=0.2,sig.level=0.05,power=0.9,type="one.sample",
alternative="greater")
pwr.t.test(d=0.5,sig.level=0.05,power=0.9,type="one.sample",
alternative="greater")
pwr.t.test(d=0.8,sig.level=0.05,power=0.9,type="one.sample",
```

alternative="greater")

Velikosti výběrů nám vyšly postupně 216, 36 a 15. Řekněme, že výběr o velikosti 36 respondentů je pro nás reálný. Potom se však musíme spokojit s tím, že s touto velikostí výběru můžeme při zadané síle testu „bezpečně“ prokázat jen středně velkou či velkou míru účinku, pravděpodobnost prokázání „malých“ efektů bude nižší. Pokud by mezi průměrnou velikostí výběru udávanou respondenty a správně zvolenou velikostí výběru byl rozdíl menší než polovina směrodatné odchylky, museli bychom pro udržení požadované síly testu zvolit vyšší velikost výběrového vzorku. Navíc, vzhledem k předpokládané 50% návratnosti dotazníků budeme muset pro získání vzorku 36 respondentů oslovit zhruba 72 respondentů.

Jak již bylo uvedeno, výběr statistického testu bude záviset na očekávaném tvaru rozdělení zjišťované veličiny (udaný rozsah výběru). Tento tvar si musíme stanovit dopředu a závisí na správné velikosti vzorku. Vždy však budeme chtít testovat hypotézu o poloze příslušného rozdělení, u parametrického testu to bude střední hodnota, u neparametrických testů medián. Budeme mít tedy na výběr mezi parametrickým t–testem, neparametrickým znaménkovým testem nebo neparametrickým jednovýběrovým Wilcoxonovým testem. Než tedy určíme konkrétní testy, zvolme nejprve otázky, které zařadíme do dotazníku. Aby naše výsledky měly co největší vypovídající hodnotu, měly by být otázky pro všechny respondenty stejné. Situaci jsme si poněkud zkomplikovali tím, že naši respondenti budou odborníci z různých oborů. Otázky by tedy měly být spíše obecné, případně k nim budeme muset doplnit potřebné informace.

Chceme testovat, zda odborníci velikost výběru podhodnocují, avšak platnost této hypotézy by obecně neměla záviset na správné velikosti vzorku. Pokud totiž obecně tvrdíme, že odborníci velikost vzorku podhodnocují, mělo by to platit jak pro vzorky velké, tak pro vzorky menší. A to i přes to, že u malých vzorků je na podhodnocení menší prostor. Jeví se tedy jako rozumné zařadit i otázky, pro které je vhodně zvolená velikost vzorku relativně malá. Díky tomu zjistíme, zda odborníci skutečně podhodnocují i malé vzorky, nebo jestli třeba u malých vzorků nedochází naopak spíše k nadhodnocení. Zároveň však otázek nesmí být

hodně, jinak by se také mohlo stát, že by respondent poctivě vyplnil prvních 5 až 10 otázek (číslo by se odvíjelo nejspíš na základě odhodlání a dobré vůle) a zbývající odpovědi by mohly být *špatné*, čili hodně by se lišily od správné velikosti vzorku, ne kvůli neznalosti, ale „otrávenosti“ respondenta.

Řekněme, že na začátku dotazníku respondentům sdělíme, aby vždy počítali s hladinou významnosti $\alpha = 0,05$. U každé otázky však sami musíme spočítat správnou velikost vzorku, která pak bude sloužit jako nulová hypotéza pro naše vlastní testování, a k tomu musíme stanovit požadovanou sílu testu. Tuto hodnotu můžeme respondentům také sdělit³. Stanovme tedy sílu testu 0,9. První otázka by mohla vypadat následovně:

- **Kolik byste potřebovali studentů k prokázání hypotézy, že na filozofických fakultách v České republice není zastoupení mužů a žen stejné, ale že na nich studuje o 40 % více žen než mužů?**

Jedná se o otázku na velikost proporce, přičemž jsme v ní uvedli vše potřebné, a to tvar nulové hypotézy i tvar specifické alternativní hypotézy. Tedy

$$H_0: p = p_0 = 0,5$$

$$H_A: p = p_A = 0,7$$

Míra účinku je $p_A - p_0 = 0,2$. V R použijeme funkci `pwr.p.test`, která slouží pro výpočty při testování velikosti proporce. Alternativa je jednostranná. Oboustranná alternativa by nám říkala, že mohou nastat dvě situace: buď bude 70 % studentek, nebo 70 % studentů. Nám jde však jen o první variantu. Po zadání kódu

```
pwr.p.test(h=0.2, sig.level = 0.05, power = 0.9,  
alternative = "greater")
```

získáme velikost vzorku $n = 215$. Tato hodnota není zrovna nízká, proto by k výraznému zešikmení rozdělení dojít nemuselo. Pokud budeme předpokládat, že

³Mohli bychom předpokládat, že respondenti si sami chtějí být dost jistí, že se jim podaří prokázat platnou hypotézu. Přestože o síle testu jako takové neuvažují, mají o ní nějaké povědomí a očekávají aspoň 90% šanci, že zamítnou neplatnou hypotézu.

rozdělení sice není normální, ale je symetrické, použijeme jednovýběrový Wilcoxonův test. Budeme testovat hypotézu $H_0: \tilde{x} = 215$ proti alternativě $H_A: \tilde{x} < 215$, kde \tilde{x} je medián rozdělení. Test provedeme tak, jak je popsáno v [3].

Vzhledem k celkem vysokému $n = 215$ bychom si však mohli dovolit předpokládat i normalitu rozdělení. Pak bychom použili jednovýběrový t -test, který je popsán na začátku této kapitoly. Pro ověření normality můžeme použít například Lillieforsův test, který je speciálním případem Kolmogorova–Smirnova testu, nebo Shapiro–Wilkův test. Symetričnost pak můžeme ověřit z grafu získaných hodnot sledované veličiny. Data můžeme také zlogaritmovat, díky čemuž se rozdělení může stát symetričtější. Test pak provádíme na zlogaritmovaných datech.

Uveďme si další otázku:

- **Chceme zjistit, zda muži vydělávají v určitém odvětví v průměru o 10 000 Kč více než ženy, přičemž směrodatná odchylka platů u mužů i žen je 5 000 Kč. Kolik respondentů potřebujeme v jedné skupině (skupina mužů a skupina žen), abychom mohli rozhodnout?**

Jedná se o příklad na dvouvýběrový test. Přestože výše platů není náhodná veličina s normálním rozdělením, využijme asymptotiky a spočítejme si velikost vzorku pomocí funkce `pwr.t.test`. V tomto případě míra účinku $d = \frac{10000}{5000} = 2$, alternativa je opět jednostranná. V R zjistíme, že potřebujeme 6 respondentů v každé skupině. Tohle číslo je docela malé, takže může dojít k zešikmení rozdělení. Použijeme proto znaménkový test. Naše hypotézy budou obdobně jako v předchozím příkladě $H_0: \tilde{x} = 6$ proti alternativě $H_A: \tilde{x} < 6$. Při výpočtu opět postupujeme tak, jak je popsáno v [3]. Vzhledem k tomu, že jsme na začátku zvolili $n = 36$, můžeme využít tvrzení klasické limitní věty a spočítat velikost testové statistiky

$$U = \frac{2Y - n}{\sqrt{n}},$$

kde Y je počet rozdílů $X_1 - 6, \dots, X_{36} - 6$ s kladným znaménkem. Ta má za platnosti nulové hypotézy asymptoticky rozdělení $N(0, 1)$. Hodnotu testové statistiky porovnáváme s kvantilem normálního normovaného rozdělení.

Třetí otázka bude následující:

- **Domníváte se, že mezi výší platu a výší IQ existuje silná kladná lineární závislost (například v určitém oboru). Jak velký byste museli mít výběr, abyste byli schopni o této hypotéze rozhodnout? Řekněte, že silnou lineární závislost značí korelační koeficient roven dvěma třetinám.**

Zde máme příklad na korelovanost dvou náhodných veličin, kterými jsou výše IQ a výše platu. Alternativa je jednostranná, protože jsme stanovili, že předpokládaná závislost je kladná. Velikost výběru spočítáme pomocí kódu

```
pwr.r.test(r=(2/3), sig.level = 0.05, power = 0.9,  
alternative = "greater").
```

Vyjde nám $n = 16$. Předpokládejme, že rozdělení nebude úplně normální, ale bude aspoň symetrické. Pro test hypotézy použijeme proto jednovýběrový Wilcoxonův test.

Řekněme, že tyto tři otázky nám pro naše účely postačí. Jak již bylo zmíněno, kdyby otázek bylo příliš, buď by to mohlo značně zredukovat počet vrácených dotazníků nebo by respondenti nevěnovali velkému množství otázek dostatečnou pozornost. Při velkém množství otázek, kde by se v podstatě jen měnila velikost efektu, by navíc mohlo dojít k tomu, že by se dotazník sám o sobě stal návodným. To by znamenalo, že by si respondent až po přečtení dotazníku uvědomil, že velikost míry účinku má významný vliv (tedy pokud tomu nevěnoval pozornost sám od sebe). Odpovědi by se pak lišily od případu, kdy by musel dané problémy řešit jednotlivě.

Kapitola 2

Ukotvení

V druhé kapitole budeme zkoumat takzvaný efekt ukotvení. Informace o ukotvení byly čerpány z [6], teorie ke statistickým testům byla čerpána z [3] a [1]. Data k experimentu byla získána pomocí sociální sítě.

2.1. Úvod do problému

Takzvaný *efekt ukotvení* je jev, ke kterému dochází tehdy, „když lidé uvažují o konkrétní hodnotě neznámého množství předtím, než tuto hodnotu odhadují“. Následné odhady se pak pohybují kolem čísla, o kterém lidé uvažovali. Tomuto číslu se říká „kotva“. Kahneman tento efekt popisuje v [6] na stranách 130 až 140. Líčí zde experimenty, ve kterých se efekt ukotvení projevil v menší či větší míře.

Jedním z nich byl například experiment na účastnících studie v Exploratoriu¹, které informovali o jisté ekologické katastrofě a zeptali se jich, kolik by byli ochotni ročně přispívat na nápravu škod způsobených danou katastrofou. Některým účastníkům byla však nejdříve položena kotvící otázka, zda by byli ochotni přispět 5 nebo 400 \$ a až poté byli dotázáni na výši příspěvku. Když otázce kotva nepředcházela, byl průměrný příspěvek účastníků 64 \$ za rok. Účastníci, kteří měli v kotvící otázce částku 5 \$, by přispěli průměrně 20 \$ a účastníci s kotvou 400 \$ by byli ochotni zaplatit průměrně 143 \$ ročně. Efekt ukotvení se udává v procentech a počítá se jako podíl rozdílu průměrných částek v daných

¹Muzeum v San Franciscu.

skupinách a rozdílu příslušných kotev. V tomto případě je rozdíl průměrných příspěvků 123 a rozdíl kotev 395. Efekt ukotvení je tedy $\frac{123}{395} \doteq 0,31$, tj. 31 %.

Další zajímavý experiment byl proveden na realitních agentech. Ti měli ohodnotit dům, který si prošli a ke kterému dostali leták s informacemi. V letáku byla mimo jiné požadovaná cena za dům. Jedna skupina agentů měla v letáku podstatně vyšší cenu, než byla skutečná cena domu, druhá skupina měla v letáku cenu podstatně nižší. Přestože pak agenti mezi faktory, které ovlivnily jejich úsudek o ceně, neuvedli cenu v letáku, efekt ukotvení v tomto případě činil 41 %.

V následující podkapitole provedeme vlastní experiment a otestujeme vliv kotvy na výši příspěvku na charitu.

2.2. Efekt ukotvení: vlastní experiment

Když testujeme, zda má nějaká kotva efekt, testujeme vlastně, zda existuje rozdíl mezi střední hodnotou (například výše příspěvku, ceny domu) udanou bez kotvy a střední hodnotou udanou s kotvou. Pokud bychom náhodné veličiny měli jen dvě, mohli bychom statistickou významnost rozdílu otestovat pomocí dvouvýběrového t -testu, případně pokud bychom měli podezření na porušení předpokladů pro t -test, použili bychom test neparametrický, například dvouvýběrový Wilcoxonův test.

Jestliže chceme otestovat vliv několika kotev najednou, musíme zvolit analýzu rozptylu jednoduchého třídění. U ní předpokládáme, že máme $k > 2$ nezávislých výběrů z normálních rozdělání se stejným rozptylem. Na hladině testu α testujeme nulovou hypotézu $H_0: \mu_1 = \mu_2 = \dots = \mu_k$, tedy že se všechny střední hodnoty rovnají. Alternativní hypotéza tvrdí, že se aspoň jedna dvojice středních hodnot liší. Testová statistika pro analýzu rozptylu vypadá takto:

$$F_A = \frac{\frac{S_A}{k-1}}{\frac{S_e}{n-k}},$$

kde S_A je skupinový součet čtverců, S_e je reziduální součet čtverců, n je celkový počet pozorování a k je počet skupin. F_A má za platnosti nulové hypotézy

rozdělení $F_{k-1, n-k}$. Ve prospěch alternativy svědčí velké hodnoty testové statistiky.

Pokud nulovou hypotézu zamítneme, budeme chtít přirozeně zjistit, které střední hodnoty se liší, čili které kotvy mají vliv. K tomu použijeme metodu pro mnohonásobná porovnání, například Scheffého metodu nebo Tukeyho metodu.

Obecně, pokud pomocí testové statistiky F_A zamítneme nulovou hypotézu a současně nenajdeme významný rozdíl u žádné dvojice středních hodnot, platí, že se nějaká složitější kombinace středních hodnot neboli kontrast významně liší oproti zbývajícím středním hodnotám nebo jejich kombinaci. Otázkou však je, zda je testování kontrastů konkrétně pro situaci testování vlivu kotvy přínosné, respektive co můžeme testováním kontrastů zjistit. Obecný tvar nulové hypotézy o kontrastu je

$$H_0: \sum_{i=1}^k c_i \mu_i = 0$$

za podmíněk $\sum_{i=1}^k c_i = 0$, $c_i \in \mathbb{R}$, $i = 1, \dots, k$. Můžeme tedy například zjišťovat, zda je součet středních hodnot několika určitých skupin odlišný od součtu středních hodnot stejného počtu jiných skupin, případně zda by mělo význam kotvy nakombinovat například procentuálně. Pokud bychom třeba předpokládali, že nízká kotva snižuje odpověď respondentů oproti žádné kotvě a vysoká kotva odpověď naopak zvýší, mohli bychom tuto hypotézu otestovat právě pomocí kontrastu. Vektor koeficientů c_i by měl tři nenulové složky, tedy $\mathbf{c} = (-\frac{1}{2}, -\frac{1}{2}, 1)$. Hodnoty $c_i = -\frac{1}{2}$ bychom přiřadili skupinám s vysokou a nízkou kotvou, hodnotu $c_i = 1$ by měla skupina bez kotvy. Oproti oboustranné alternativní hypotéze bychom testovali, zda se součet polovin středních hodnot pro skupiny s kotvami rovná střední hodnotě pro skupinu bez kotvy.

Postup pro analýzu rozptylu však můžeme použít tehdy, jsou-li splněny předpoklady nezávislosti náhodných veličin, normálního rozdělení veličin a shodných rozptylů ve všech skupinách. Jestliže máme podezření, že by náhodné veličiny nemusely pocházet z normálního rozdělení, použijeme neparametrickou alternativu, kterou je Kruskalův–Wallisův test. Ten je zobecněním dvouvýběrového Wilcox-

onova testu pro $k > 2$ skupin. Nulová hypotéza zní, že všechny výběry pocházejí ze stejného rozdělení, alternativní hypotéza tvrdí, že aspoň dva z výběrů pocházejí z různých rozdělení. Postup Kruskalova–Wallisova testu je popsán v [3] na straně 287.

2.2.1. Volba dat a následný sběr

Vzhledem k tomu, že chceme efekt ukotvení skutečně testovat a ukázat tak výše uvedený postup na konkrétním příkladě, musíme se rozhodnout, jaká data k tomu použijeme, případně jakým způsobem provedeme jejich sběr.

V [6] Kahneman tvrdí, že nejvíce patrný bývá vliv kotvy v situacích, ve kterých se jedná o určení finanční částky. Budeme se tedy inspirovat experimentem, který byl proveden na účastnících studie v Exploratoriu. Našich respondentů se zeptáme, kolik by byli ochotni přispět na charitu dle svého výběru. Zvolíme si celkem tři typy kotev. Jedna bude velmi nízká, druhá relativně vysoká a třetí řekněme střední. Jejich konkrétní výše jsme volili 20 Kč, 200 Kč a 500 Kč. Přesné znění otázek bude následující:

- **Představte si, že Vás osloví zástupci nějaké charity (v tomto kontextu nezáleží na tom jaké, můžete si zvolit tu, která je Vám nejvíce sympatická). Kolik byste jim byli ochotni jednorázově přispět?**
- **Představte si, že Vás osloví zástupci nějaké charity (v tomto kontextu nezáleží na tom jaké, můžete si zvolit tu, která je Vám nejvíce sympatická). Byli byste ochotni přispět jim jednorázově 20 Kč (200 Kč, 500 Kč)? Kolik bys jim byli ochotni přispět, kdybyste mohli přispět jakoukoliv částku?**

Vzhledem k dobré dostupnosti využijeme ke sběru dat sociální síť. Respondenty budou autorčiny kontakty na sociální síti Facebook. Protože je věková struktura těchto kontaktů převážně homogenní, věkem se zabývat nebudeme. Zamyšleme se však nad tím, co by mohlo představovat tzv. *confounder*, neboli rušivý

faktor. Mohlo by se jednat například o pohlaví nebo dosažené vzdělání. Abychom zamezili vlivu těchto faktorů a zajistili, aby výše příspěvku byla ovlivněna hlavně výší kotvy, uděláme stratifikovaný výběr. To znamená, že se v každé skupině respondentů dotážeme stejného počtu žen a mužů a stejného počtu vysokoškoláků a středoškoláků. Ke každému respondentovi si pak kromě výše příspěvku poznamenejeme i pohlaví a vzdělání.

Předchozí kapitola byla věnována velikosti výběrového vzorku. Na správnou velikost výběrového vzorku nesmíme zapomínat ani při tomto experimentu. Hladinu významnosti zvolme opět $\alpha = 0,05$, sílu testu 0,9. Opět využijeme Cohenem zavedené zjednodušení a v R spočítáme velikosti výběru pro malou, střední a velkou míru účinku, přičemž tentokrát použijeme funkci `pwr.anova.test`. Kromě hladiny testu, síly testu a velikosti efektu musíme nyní zadat i počet skupin, v našem případě čtyři. Kódy budou vypadat takto:

```
pwr.anova.test(k=4,f=0.1,sig.level=0.05,power=0.9)
```

```
pwr.anova.test(k=4,f=0.25,sig.level=0.05,power=0.9)
```

```
pwr.anova.test(k=4,f=0.4,sig.level=0.05,power=0.9)
```

Postupně nám vyšlo, že v každé skupině potřebujeme 356, 58 a 24 respondentů. Vzhledem k omezenému počtu autorčinných kontaktů na sociální síti se musíme spokojit s tím, že při zadané síle testu dokážeme „bezpečně“ prokázat jen velkou míru účinku. Celkem totiž bude potřeba 96 respondentů, přičemž nemůžeme počítat se 100% návratností. Návratnost by však mohla být vyšší než u anonymních dotazníků, jelikož bychom se mohli spoléhat na ochotu respondentů pomoci autorce s výzkumem. Přesto budeme muset oslovit zhruba 120 respondentů, protože někteří nemusí stihnout odpovědět dostatečně rychle.

Nezávislost výběrů bude zaručena tím, že z případných párů oslovíme jen jednoho partnera, ze společné domácnosti oslovíme jen jednoho člena a dále zařadíme jen odpovědi, které dostaneme v řádu několika dnů, aby se respondenti nemohli „radit“.

Vzhledem k tomu, že se jedná o výši příspěvku, může snadno dojít k porušení předpokladu normality. Někteří respondenti mohou přispět velmi vysokou částku

nebo, a to je pravděpodobnější, nebudou chtít přispět nic. K testování proto použijeme neparametrický přístup. Neparametrický test však snižuje sílu testu, takže si musíme být vědomi toho, že výše uvedené počty respondentů jsou spíše orientační. Veškeré analýzy provedeme v softwaru R.

2.3. Analýza dat - ANOVA

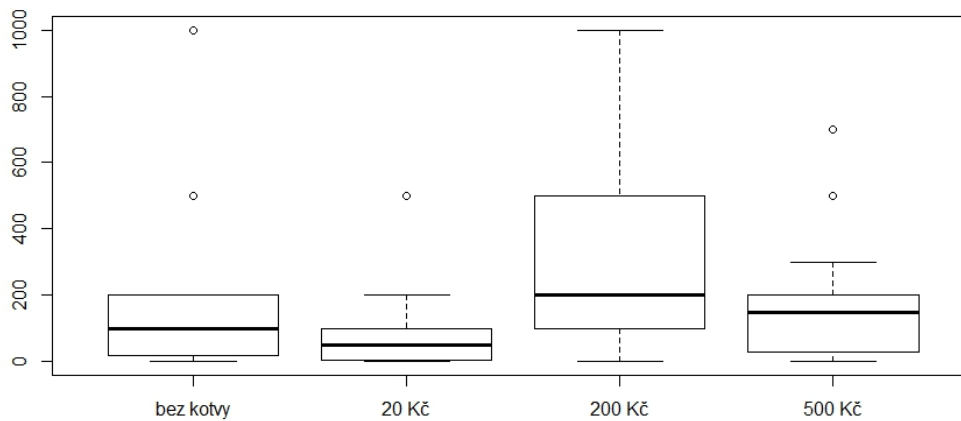
Data byla sesbírána v průběhu února. Než přistoupíme k samotné analýze, nejprve se s nimi seznámíme.

Celkem odpovědělo 102 respondentů. Protože jsme neměli 100% návratnost, došlo k nevyváženému třídění. Zastoupení respondentů v jednotlivých skupinách ukážeme v tabulce 2.1:

Skupina	Počet respondentů	Pohlaví	Vzdělání
bez kotvy	26	Muži 11	SŠ 14
		Ženy 15	VŠ 12
kotva 20 Kč	24	Muži 12	SŠ 12
		Ženy 12	VŠ 12
kotva 200 Kč	25	Muži 12	SŠ 13
		Ženy 13	VŠ 12
kotva 500 Kč	27	Muži 12	SŠ 13
		Ženy 15	VŠ 14

Tabulka 2.1: Zastoupení respondentů v jednotlivých skupinách.

Z tabulky 2.1 je patrné, že zastoupení mužů a žen, stejně tak vysokoškoláků a středoškoláků, je ve všech skupinách skoro vyrovnané. Ke confoundingu ze strany pohlaví či vzdělání by tedy dojít nemělo. Minimální počet respondentů v jedné skupině stanovený na 24 je dodržen. Abychom získali ucelenou představu o odpovědích respondentů v jednotlivých skupinách, vykreslíme si graf. V souboru je však jedno odlehlé pozorování s hodnotou 50 000 Kč ze skupiny s kotvou 20 Kč. Toto odlehlé pozorování není v grafu zohledněno, stejně tak druhé nejvyšší odlehlé pozorování s hodnotou 2 000 Kč ze skupiny s kotvou 500 Kč.



Obrázek 2.1: Odpovědi respondentů v závislosti na kotvící otázce bez dvou odlehlých pozorování.

Z obrázku 2.1 na první pohled vidíme, že mediány ve skupinách nejsou stejné a že se nejvýrazněji odlišuje třetí skupina s kotvou 200 Kč.

Než začneme s testováním, musíme ověřit předpoklad shody rozptylů. K tomu použijeme Leveneův test s tím, že jako střed nastavíme medián, protože naše data nepochází z normálního rozdělení. Medián jako střed nám poskytne robustnější test, který provedeme na datech s odlehlými pozorováními i bez nich. Hodnoty testové statistiky i p–hodnoty vypíšeme do tabulky 2.2:

Data	testová statistika F	p–hodnota
celý soubor	0.962	0,414
bez 1 odlehlého pozorování	1.318	0,273
bez 2 odlehlých pozorování	1.983	0,122

Tabulka 2.2: Výsledky Leveneova testu shody rozptylů.

Ve všech případech homoskedasticitu nezamítáme. Vidíme však, že odstraněním jen dvou pozorování se p–hodnota celkem výrazně zmenšila, což je zapříčiněno malým rozsahem výběrů.

Po ověření platnosti předpokladu můžeme přistoupit k neparametrické analýze rozptylu. V R použijeme funkci `kruskal.test` a test provedeme opět na datech

s odlehlými pozorováními i bez nich. P–hodnoty pro Kruskalův–Wallisův test jsou v tabulce 2.3:

Data	p–hodnota
celý soubor	0,018
bez 1 odlehlého pozorování	0,007
bez 2 odlehlých pozorování	0,007

Tabulka 2.3: P–hodnoty pro Kruskal–Wallisův test.

Zamítáme tedy nulovou hypotézu, která tvrdí, že se mediány ve skupinách rovnají. Vidíme, že p–hodnota se výrazněji snížila po odstranění prvního odlehlého pozorování a po odstranění druhého odlehlého pozorování se změnila jen minimálně (změna není zjevná kvůli zaokrouhlení). To proto, že první odlehlé pozorování bylo ve druhé skupině s kotvou 20 Kč, ve které jsou jinak hodnoty nízké a přítomnost pozorování s nejvyšším pořadím značně zvýšila celkový součet pořadí ve skupině. Druhou nejvyšší částku udal respondent ze skupiny s nejvyšší kotvou, kde je však pořadí částek celkově vyšší. Odstranění jediné hodnoty z této skupiny proto nemělo takový vliv.

Přirozeně nás nyní zajímá, které rozdíly dvojic jsou statisticky významné. Vypíšeme nejprve hodnoty mediánů v jednotlivých skupinách do tabulky 2.4. \tilde{x}_1 jsou mediány pro neočištěná data, \tilde{x}_2 jsou mediány pro soubor bez jednoho odlehlého pozorování a \tilde{x}_3 jsou mediány pro soubor bez dvou odlehlých pozorování.

Skupina	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3
bez kotvy	100	100	100
kotva 20 Kč	50	50	50
kotva 200 Kč	200	200	200
kotva 500 Kč	200	200	150

Tabulka 2.4: Hodnoty mediánů pro neočištěná data (první sloupec), pro data bez jednoho odlehlého pozorování (druhý sloupec) a pro data bez dvou odlehlých pozorování (třetí sloupec).

Z tabulky 2.4 je zřejmé, že odstranění odlehlé hodnoty z druhé skupiny s kotvou 20 Kč nemělo na hodnotu mediánu vliv. Naopak odstraněním druhé odlehlé

hodnoty z poslední skupiny klesl medián z 200 Kč na 150 Kč. Na první pohled se zdá, že existují významné rozdíly mezi prvními třemi skupinami. Mezi kotvou 200 Kč a 500 Kč však rozdíl patrný není.

Pro zjištění toho, které dvojice výběrů se od sebe významně liší, provedeme obecnou metodu mnohonásobného porovnávání, při které budeme pracovat se vzorcem z [1]

$$|t_i - t_j| > \sqrt{\frac{1}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) N(N+1) h_{I-1}(\alpha)}, \quad (2.1)$$

kde t_i je průměrné pořadí v i -té skupině, N je celkový počet pozorování, n_i je počet pozorování v i -té skupině, I je počet skupin a $h_{I-1}(\alpha)$ je kritická hodnota Kruskalova–Wallisova testu na hladině α , přičemž při větších rozsazích výběru lze použít aproximace $h_{I-1}(\alpha) \doteq \chi_{I-1}^2(1 - \alpha)$. Řekneme, že se dvojice výběrů významně liší, pokud pro ně platí nerovnost 2.1.

Pomocí R spočítáme pro všechny dvojice skupin absolutní rozdíly průměrných pořadí a také hodnoty odmocniny, která je ve vzorci 2.1 na pravé straně. Využijeme aproximace a místo kritické hodnoty Kruskalova–Wallisova testu použijeme $1 - \alpha$ kvantil χ^2 rozdělení o $I - 1$ stupních volnosti. Výsledky porovnáme a následně určíme, která dvojice se statisticky významně liší. Tento postup provedeme pro data s odlehlými pozorováními i bez nich. Ve všech třech případech zjistíme, že se statisticky významně liší druhá a třetí skupina, tedy skupina s kotvou 20 Kč a skupina s kotvou 200 Kč. Pro představu si uveďme výsledky pro neočištěná data. V tabulce 2.5 jsou hodnoty absolutních rozdílů, v tabulce 2.6 jsou hodnoty odmocniny. Srovnáním tabulek zjistíme, že pouze pro dvojici s kotvami 20 a 200 Kč je hodnota v první tabulce větší než hodnota ve druhé. Pro ostatní dvojice jsme tímto způsobem statisticky významný rozdíl neprokázali.

Na začátku kapitoly jsme psali o efektu ukotvení. Spočítejme ho tedy i pro dvojici skupin, jejíž rozdíl jsme prokázali jako statisticky významný. Když budeme počítat s průměry tak, jak popisuje autor v [6] a nezbavíme se odlehlého

	Kotva 20 Kč	Kotva 200 Kč	Kotva 500 Kč
Skupina bez kotvy	9,02	16,76	5,65
Kotva 20 Kč		25,78	14,67
Kotva 200 Kč			11,11

Tabulka 2.5: Hodnoty absolutních rozdílů pro dvojice průměrných pořadí skupin pro neočištěná data.

	Kotva 20 Kč	Kotva 200 Kč	Kotva 500 Kč
Skupina bez kotvy	23,41	23,17	22,73
Kotva 20 Kč		23,64	23,21
Kotva 200 Kč			22,96

Tabulka 2.6: Hodnoty odmocniny ze vzorce 2.1 pro dvojice skupin pro neočištěná data.

pozorování ze skupiny s kotvou 20 Kč, odhad efektu vyjde záporný, konkrétně

$$\frac{312 - 2173}{200 - 20} = \frac{-1861}{180} \doteq -10,34,$$

tedy - 1034 %. To se obecně může stát, avšak v tomto případě to nedává smysl. Po odstranění zmíněného odlehlého pozorování bude odhad efektu kotvy smysluplnější:

$$\frac{312 - 94}{200 - 20} = \frac{218}{180} \doteq 1,21$$

Efekt počítaný z průměrů bez odlehlého pozorování je odhadnut na 121 %. Ještě spočítejme odhad efektu kotvy pro rozdíly mediánů, na které, jak jsme ukázali, nemá přítomnost odlehlého pozorování vliv.

$$\frac{200 - 50}{200 - 20} = \frac{150}{180} \doteq 0,83$$

V tomto případě je efekt kotvy odhadnut na 83 %.

2.4. Závěry experimentu

Nyní shrneme průběh experimentu a zamyslíme se nad jeho výsledky, alternativami řešení, případně nad možnostmi zlepšení.

2.4.1. Sběr dat

Během sběru dat pro experiment bylo zjištěno, že výše uvedená otázka nebyla položena úplně nešťastněji. Řada respondentů ji pochopila tak, že má odpovídat na hypotetickou situaci „co by, kdyby“. Následně bylo třeba otázku zpřesňovat a doptávat se. Vhodnější znění otázky by bylo následující:

- **Představte si, že Vás osloví zástupci nějaké charity (v tomto kontextu nezáleží na tom jaké, můžete si zvolit tu, která je Vám nejvíce sympatická). Kolik byste jim ve své aktuální situaci byli ochotni jednorázově přispět?**

Obdobně by zněla otázka i pro situaci s kotvou.

Dále, jak jsme předpokládali, panoval mezi respondenty odlišný postoj k přispívání na charitu. Je zřejmé, že postoj k přispívání na charitu pravděpodobně ovlivní výši příspěvku stejně, ne-li více než výše kotvy. Tento fakt mohl být ošetřen doplňující otázkou na postoj respondentů k přispívání na charitu, kdy by respondenti hodnotili sami sebe například škálou od jedné do tří, kdy jednička by znamenala, že je respondent odpůrcem přispívání, trojka by naopak znamenala, že respondent přispívá na charitu rád. Problém by však mohl nastat v okamžiku, kdy respondent sice přispívá rád, nikoliv však penězi, ale například oblečením. Pak je jeho vztah k přispívání na charitu kladný, zároveň by jeho odpověď byla nula.

Kromě postoje k přispívání na charitu může samozřejmě výši příspěvku ovlivnit i výše příjmů. Na ten se však dotazuje poměrně těžko a zkušenost je spíš taková, že lidé výši svého příjmu neradi sdělují.

V těchto případech nám tedy nezbyvá, než se zkusit spolehnout na náhodnost a předpokládat, že zastoupení lidí s různými postoji k charitě, případně s různými příjmy, bude ve všech skupinách zhruba stejné. Informativně si do tabulky 2.7 vypíšeme, kolik lidí v jednotlivých skupinách odmítlo přispět.

Skupina	Počet nul
Skupina bez kotvy	6
Kotva 20 Kč	6
Kotva 200 Kč	2
Kotva 500 Kč	6

Tabulka 2.7: Počet respondentů, kteří odmítli přispět na charitu jakoukoliv částku.

V tabulce 2.7 vidíme, že celkem odmítlo přispět 20 respondentů. Jejich počet je kromě třetí skupiny rozdělen rovnoměrně, jen ve skupině s kotvou 200 Kč je oproti ostatním skupinám třetinový.

2.4.2. Analýza dat

Před samotným sběrem dat byla provedena úvaha, zda by bylo vhodné použít k analýze dat lineární model. Podnětem k této úvaze byla domněnka, že čím vyšší kotva, tím vyšší bude průměrný příspěvek respondentů. Ukázalo se však, že mezi výší kotvy a výší příspěvku nelze očekávat lineární, dokonce ani monotónní vztah.

Z dat je patrné, že nízká kotva snižuje příspěvek oproti žádné kotvě. Pokud žádné kotvě přiřadíme hodnotu nula, pak je v tomto případě vztah klesající. Vyšší kotva naopak výši příspěvku vskutku zvyšuje, ale jen do určité doby. Pokud bychom chtěli výsledky experimentu aproximovat nějakou křivkou, bylo by zapotřebí znát účinky více kotev. Takto by aproximace byla velmi hrubá.

2.4.3. Výsledky experimentu

K analýze jsme použili robustní testy, protože data nepocházela z normálního rozdělení. Díky tomu jsme také došli ke stejným závěrům jak pro data s odlehlými pozorováními, tak bez nich.

Na první pohled se z dat jeví, že existují významnější rozdíly mezi více dvojicemi skupin, avšak metodou mnohonásobného porovnávání jsme prokázali statisticky významný rozdíl pouze mezi dvěma skupinami. Důvodem je malý rozsah výběru. Před sběrem dat jsme stanovili, že při počtu 24 respondentů v každé skupině dokážeme „bezpečně“ prokázat jen velkou míru účinku. Pokud bychom

měli rozsah výběru větší, je možné, že bychom prokázali i menší efekty a našli tak statisticky významný rozdíl mezi více skupinami.

Naše výsledky se však shodují s výsledky Kahnemanova experimentu, který je popsán na začátku této kapitoly. Pro připomenutí, Kahnemanovi respondenti uvedli na otázku bez kotvy průměrný příspěvek 64 \$ za rok, účastníci s kotvicí otázkou s částkou 5 \$ by přispěli průměrně 20 \$ a účastníci s kotvou 400 \$ by byli ochotni zaplatit průměrně 143 \$ ročně.

Když spočítáme efekty ukotvení pro Kahnemanův experiment, přičemž pro případ žádné kotvy dosadíme velikost kotvy nulovou, zjistíme následující:

$$\frac{64 - 20}{0 - 5} = \frac{44}{-5} = -8,8$$

$$\frac{143 - 64}{400 - 0} = \frac{79}{400} \doteq 0,198$$

$$\frac{143 - 20}{400 - 5} = \frac{123}{395} \doteq 0,311$$

Nyní provedeme totéž pro naše data. Vezmeme první tři skupiny a z druhé skupiny vyloučíme odlehlé pozorování.

$$\frac{216 - 94}{0 - 20} = \frac{122}{-20} = -6,1$$

$$\frac{312 - 216}{200 - 0} = \frac{96}{200} = 0,48$$

$$\frac{312 - 94}{200 - 20} = \frac{218}{180} \doteq 1,21$$

Vidíme, že nízká kotva měla v obou případech záporný efekt. Další dva efekty jsou již kladné a v těchto případech jsou v našem experimentu dokonce vyšší. Nutno ovšem podotknout, že Kahneman ve své knize neuvedl, kolik respondentů během svého experimentu v Exploratoriu oslovil.

Přestože byl efekt kotvy prokázán, nemáme ambice na zobecňování výsledků. Náš vzorek byl poměrně specifický a cílem experimentu bylo pouze názorně předvést postup statistického testování.

Závěr

Na tomto místě shrnu poznatky, ke kterým jsem během psaní své práce dospěla.

Nejdříve bych ráda upozornila na to, že v knize *Myšlení: rychlé a pomalé* je celá řada dalších zajímavých experimentů, které se týkají lidského myšlení. Z kapacitních důvodů jsem se však nemohla věnovat všem. Tímto chci tedy čtenáře pobídnout k tomu, aby si v případě zájmu přečetl i zmíněnou knihu.

V první kapitole práce jsem se zabývala vhodně zvolenou velikostí vzorku, kterou dle Kahnemana i Cohena řada vědeckých pracovníků podhodnocuje, a to dokonce i vědeckých pracovníků zběhlých v matematice. Odvodila jsem i vzorec pro výpočet správné velikosti vzorku, avšak tuto velikost můžeme také velmi snadno zjistit pomocí softwaru. V závěru kapitoly jsem navrhla, jak by se mohla testovat hypotéza o podhodnocování velikosti vzorku. Test jsem navrhla ještě předtím, než jsem si přečetla vlastní experiment Kahnemana a Tverskyho. Otázky se samozřejmě liší a řekla bych, že mé nejsou tak komplexní. Myslím si však, že svůj účel by splnily.

Druhá kapitola popisuje efekt *ukotvení*, což je jev, kdy konkrétní hodnota neznámého množství ovlivňuje následný odhad. Jejím ústředním bodem je můj vlastní experiment, na kterém jsem ukázala, jak můžeme rozhodnout o tom, zda je efekt kotvy statisticky významný. Dle popisu v knize se jeví, že Kahneman s Tverským svůj experiment nevyhodnocovali stejným způsobem jako já, a pokud ano, tak se o tom nezmiňují. Počítali však efekty kotvy, které jsem spočítala i pro svůj experiment a došla jsem ke stejnému závěru jako uvedení autoři. Výsledky svého experimentu však nechci nijak zobecňovat, protože můj vzorek nebyl dostatečně

rozmanitý, abych z něj mohla vyvozovat obecně platné závěry.

Myslím, že jsem cíl práce naplnila a osvětlila tak čtenáři některé z postupů statistického testování experimentů, v tomto případě experimentů z oblasti psychologie a lidského myšlení.

Literatura

- [1] Anděl, J.: *Základy matematické statistiky*. MatfyzPress, Praha, 2011.
- [2] Cohen, J.: *Statistical Power Analysis* [online]. 1992, [cit. 2016-02-19]. dostupné z: [http://courses.csusm.edu/psyc393md/2011-Fall-41892/readings/StatisticalPower/Cohen\(1992\)StatisticalPower.pdf](http://courses.csusm.edu/psyc393md/2011-Fall-41892/readings/StatisticalPower/Cohen(1992)StatisticalPower.pdf).
- [3] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky*. Univerzita Palackého v Olomouci, Olomouc, 2013.
- [4] Champely, S.: *pwr: Basic Functions for Power Analysis*. R package version 1.1-3. 2015, [cit. 2016-03-03]. dostupné z: <https://CRAN.R-project.org/package=pwr>.
- [5] Chow, S., Shao, J., Wang, H.: *Sample Size Calculations In Clinical Research (Second Edition)*. Chapman & Hall/CRC, Boca Raton, 2008.
- [6] Kahneman, D.: *Myšlení: rychlé a pomalé*. Jan Melvil, Brno, 2012.
- [7] Kahneman, D., Tversky, A.: *Belief in the law of small numbers* [online]. 1971, [cit. 2016-02-19]. dostupné z: <http://isites.harvard.edu/...pdf>.