MATEMATICKO-FYZIKÁLNÍ FAKULTA
Univerzita Karlova

BAKALÁŘSKÁ PRÁCE

Anna Benešová

Využití základních statistických metod v psychologickém výzkumu

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. Ing. Marek Omelka, Ph.D.
Studijní program: Matematika
Studijní obor: Finanční matematika

Praha 2023

Prohlašuji, že jsem tuto bakalářskou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů. Tato práce nebyla využita k získání jiného nebo stejného titulu.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V ..... dne ....

Podpis autora

Tímto bych ráda poděkovala svému vedoucímu Ing. Marku Omelkovi, Ph.D za ochotu, neocenitelné rady a trpělivost. Dále děkuji profesorovi Nathanovi Novemskému za poskytnutí dat, rodině za psychickou podporu a přátelům jakožto čtenářům prvotních verzí.

Název práce: Využití základních statistických metod v psychologickém výzkumu

Autor: Anna Benešová

Department: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: doc. Ing. Marek Omelka, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Tato práce se zabývá metodami testování pro dvouvýběrový problém a následnou aplikací na data z oblasti behaviorální ekonomie. Hlavní otázkou analýzy je zjistit existenci averze ke ztrátě a majetnického efektu. V práci jsou uvedeny testy vhodné pro analýzu daných dat. Konkrétně se věnuje testování podílu a rozdílu středních hodnot a Mann-Whitneyho formulaci pro Wilcoxonův test. V další části je představen podkladový článek a zkoumané téma. Na konci jsou aplikovány metody na poskytnutý datový soubor. Existence majetnického efektu je potvrzena alespoň v polovině případů.

Klíčová slova: testování hypotéz dvouvýběrové testy t-test Wilcoxonův test averze ke ztrátě

Title: Elementary statistical methods in psychological research

Author: Anna Benešová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Ing. Marek Omelka, Ph.D., Department of Probability and Mathematical Statistics

Abstract: This work deals with testing methods for a two-sample problem and their subsequent application to data from the field of behavioural economics. The main question of the analysis is to determine the existence of loss aversion and the endowment effect. This thesis presents tests suitable for analyzing the given data. Specifically, it focuses on testing the ratio and difference of means and the Mann-Whitney formulation for the Wilcoxon test. In the next section, the underlying article and the research topic are introduced. In the end, the methods are applied to the provided data set. The existence of the endowment effect is confirmed in at least half of the cases.

Keywords: hypothesis testing two-sample tests t-test Wilcoxon test loss aversion

Obsah

Úvod **2**

1 Metodologie testování **4**
1.1 Charakteristika dat 4
1.2 Inference rozdílu středních hodnot 6
1.2.1 Welchův test 6
1.2.2 Odvození konfidenčního intervalu 7
1.3 Inference podílu středních hodnot 7
1.3.1 I. způsob odvození 7
1.3.2 II. způsob odvození 9
1.4 Neparametrický přístup 11
1.4.1 I. formulace 11
1.4.2 II. formulace 14

2 Averze ke ztrátě a majetnický efekt **16**
2.1 Prospektová teorie 16
2.1.1 Averze ke ztrátě 16

3 Seznámení s podkladovou studií **18**
3.1 Formulace zadání 18
3.2 Průběh experimentu 18

4 Diskuze nad výsledky **21**
4.1 Výsledky testování 21
4.2 Rozdíly oproti původnímu článku 23

Závěr **26**

Seznam použité literatury **27**

Úvod

V posledních desetiletích byla zaznamenána významná změna v uvažovaní nad komplexními vědami typu ekonomie či finance. Běžným postulátem ekonomických a finančních modelů bývá předpoklad plné racionality člověka, který se s přibývajícím výzkumem ukazuje jako nepravdivý, ne-li přímo zavádějící. Pro řešení jsme se obrátili na obor, který se zabývá chováním člověka - psychologii. Následkem korekce zmíněného postulátu o poznatky lidského chování vznikly behaviorální vědy.

Cílem této práce je představit statistické metody vhodné pro získaná data, demonstrovat jejich použití a porovnat je s původním výzkumem. Text by měl být srozumitelný čtenáři, který absolvoval základní kurz statistiky a je obeznámen s pojmy jako náhodný výběr, testová statistika či p-hodnota.

Celá práce je rozdělena do čtyř kapitol. První část se věnuje statistickým metodám pro dvouvýběrové problémy. Popíšeme si teorii statistických testů na základě zvolených parametrů a vhodně je použijeme na vzorku dat. V druhé kapitole si uvedeme faktický kontext pro podkladové studie a ilustrujeme vybraný jev na příkladech z finančního světa. V třetí kapitole si představíme podkladový článek, ze kterého jsme získali data. Vysvětlíme si použité značení a design experimentů. V poslední části si shrneme naše poznatky a porovnáme je s těmi, které vyvodil sám autor.

Výzkum

Zmíněným článkem je *Boundary of Loss Aversion* od Novemsky (2005)[1], který se věnuje averzi ke ztrátě a majetnickému efektu, což je jev, při kterém si lidé více váží věcí, které vlastní, než věcí, které nevlastní. Jinak řečeno, v článku se zabývají otázkou, zda existuje asymetrie ceny mezi kupcem a prodejcem.

Asymetrie cen se vztahuje k situaci, kdy jeden ze subjektů obchodu uvádí cenovou nabídku, která není založena na objektivní tržní ceně, ale spíše na vnitřních subjektivních faktorech. Tato situace nastává, když subjekt uplatňuje své osobní preference, vnímání hodnoty, očekávání nebo jiné kognitivní zkreslení při stanovování ceny zboží nebo služby. Pro účely první kapitoly si představme první studii článku.

Jednalo se o simulaci obchodu na trhu, kde každému účastníkovi byla přidělena jedna ze tří rolí:

1. *prodejce*, který obdržel zboží a měl uvést cenu, za kterou by ho prodal,
2. *volitel*, který vyslovil částku, kdy preferuje peníze před zbožím,
3. *kupec*, který měl říci cenu, za kterou by zboží koupil.

Role *volitele* byla zavedena jako kontrolní skupina a neměla by vykazovat majetnický efekt, kdežto *prodejce*, který zboží vlastnil, ano. Označme si veličinami X, Y, Z ceny v dolarech, které jsme zaznamenali u rolí *prodejce*, *volitele*, *kupce* respektive. Formálně se jedná o náhodné výběry $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m), \mathbf{Z} = (Z_1, Z_2, \ldots, Z_l)$, kde například X_i je zvolená cena zboží i-tým účastníkem v roli *prodejce*. Každý subjekt experimentu obdržel právě jednu roli, a proto lze očekávat sdruženou nezávislost výběru, jakožto i nezávislost po dvou. Přirozená čísla n, m, l odpovídají počtům účastníků pro související role.

Kapitola 1

Metodologie testování

V této kapitole si rozebereme vybrané testy v závislosti na zvoleném parametru a poskytneme ucelený pohled na celou problematiku testování hypotéz. Odvodíme si příslušné testové statistiky a konfidenční intervaly, rovněž prodiskutujeme vhodnost pro naše zadání. Vše demonstrujeme na první studii uvedené v úvodní části.

Naším úkolem bude vyšetřit potenciální účinek majetnického efektu mezi *prodejcem* a *volitelem*. V řeči statistiky, zda parametry rozdělení X_i, Y_j jsou rozdílné či nikoliv.

1.1 Charakteristika dat

Připomeňme si, že se budeme zabývat dvěma nezávislými náhodnými výběry $\mathbf{X}, \mathbf{Y}$ zmíněnými v úvodu. Díky povaze experimentu data nemusí pocházet z normálního rozdělení a navíc výskyt extrémních hodnot nebude ojedinělý. Proto se zaměříme na asymptotické a pořadové testy či na jejich varianty. Další obtíž nám způsobí diskrétní povaha veličin. V důsledku toho budeme muset upravit některé testy vyžadující předpoklad spojitosti.

Označme si výběrové průměry $\overline{X}_n, \overline{Y}_m$ a výběrové rozptyly S_X^2, S_Y^2 dané vzorci

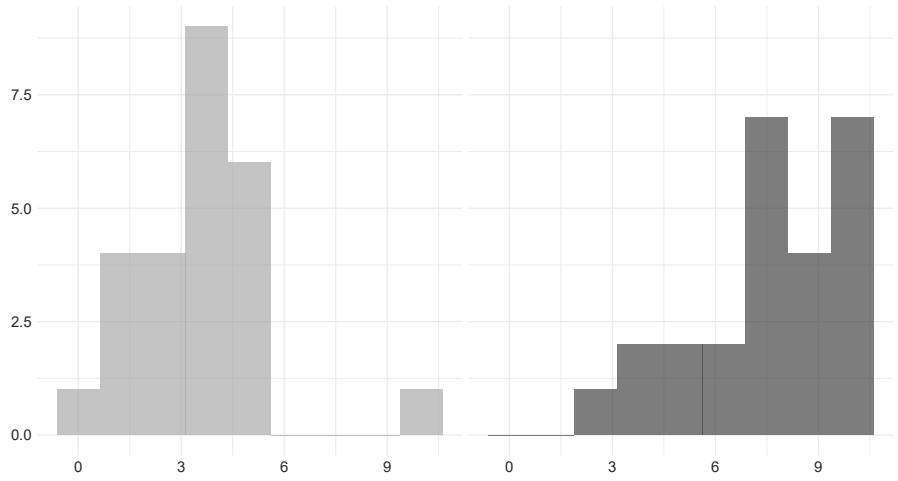
$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2,$$

které hojně využijeme v následujících řádcích. Vypočtené hodnoty nalezneme v tabulce 1.1.

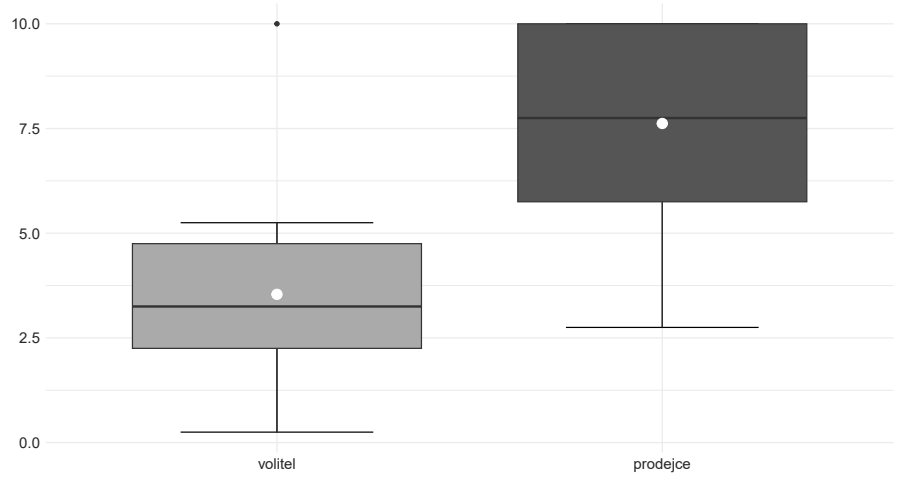
	X	**Y**
Rozsah výběru	$n = 32$	$m = 25$
Výběrový průměr	$\overline{X}_n = 7.62$	$\overline{Y}_m = 3.54$
Výběrový rozptyl	$S_X^2 = 5.19$	$S_Y^2 = 4.28$

Tabulka 1.1: Základní charakteristiky rozdělení X_i, Y_j. Již na první pohled je znatelný rozdíl mezi $\overline{X}_n$ a $\overline{Y}_m$. To ovšem je potřebné ověřit pomocí testování.

Nyní se podívejme na grafické znázornění realizací náhodných vektorů $\mathbf{X}, \mathbf{Y}$. Pro dvouvýběrový problém diskrétních rozdělení jich existuje řada, pro naše účely postačí histogram a boxplot. Histogram nám poskytne prvotní pohled na charakter našich dat; zda mohou pocházet ze symetrického rozdělení či jestli obsahují extrémní hodnoty. Zaměříme-li se na související histogram 1.1, zjistíme, že výběr s vysokou pravděpodobností není z rodiny normálních rozdělení, dokonce není symetrický kolem své střední hodnoty. Boxplot je užitečný v porovnání mezi dvěma či více výběry. Někdy již jeho vykreslení dává představu, jestli budeme nulovou hypotézu zamítat či nikoliv. Nezapomínejme však na to, že bychom měli naše rozhodnutí vždy zakládat na výsledcích testování. Podívejme se na příklad boxplotu (viz obrázek 1.2). Zde vidíme jasný rozdíl mezi mediány i středními hodnotami veličin X_i, Y_j indikující existenci majetnického efektu.



Obrázek 1.1: Histogramy náhodných výběrů $\mathbf{X}$ (vpravo) a $\mathbf{Y}$ (vlevo). Na x-ové ose se nachází cena zboží (v dolarech), na y-ové je počet subjektů, který zvolil danou cenu.



Obrázek 1.2: Boxplot náhodných výběrů $\mathbf{X}$ (vpravo) a $\mathbf{Y}$ (vlevo). Na x-ové ose máme role: Na y-ové ose cenu zboží. Bílé tečky značí výběrové průměry.

1.2 Inference rozdílu středních hodnot

Označme si střední hodnoty μ_X a μ_Y respektive. Testujeme parametr $\theta_1 \equiv \mu_X - \mu_Y$ s odhadem $\hat{\theta}_1 \equiv \overline{X}_n - \overline{Y}_m$. V našem případě θ_1 popisuje rozdíl středních hodnot cen, který *prodejce* a *volitel* přijme, přičemž *prodejce* v průměru přijme o θ_1 dolarů více než *volitel*.

Model $\mathcal{F}$ vyžaduje pouze konečné druhé momenty distribučních funkcí F_X, F_Y. Nulovou a alternativní hypotézu formulujeme následovně

$$H_0 : \theta_1 \equiv \mu_X - \mu_Y = \delta, \quad H_1 : \theta_1 \equiv \mu_X - \mu_Y \neq \delta,$$

kde δ je konstanta (v našem případě volíme $\delta = 0$). V celé práci budeme symbolem $\xrightarrow{D}$ značit konvergenci v distribuci.

Věta 1. *Mějme nezávislé náhodné výběry $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_m$ z rozdělení se středními hodnotami μ_X a μ_Y respektive a s konečnými rozptyly. Pak pro $\min(n, m) \to \infty$ a $m/n \to p \in (0, \infty)$ platí*

$$\frac{\hat{\theta}_1 - \theta_1}{\sqrt{S_X^2/n + S_Y^2/m}} \xrightarrow{D} \mathcal{N}(0, 1). \tag{1.1}$$

Důkaz. Důkaz je obdobný důkazu věty 3 s tím rozdílem, že vynecháme logaritmické transformace marginálních výběrů (tedy první aplikaci delta věty). □

Z (1.1) odvodíme testovou statistiku. Pro $\delta = 0$ dostaneme

$$Z_r \equiv \frac{\hat{\theta}_1}{\sqrt{S_X^2/n + S_Y^2/m}}. \tag{1.2}$$

V praxi nazýváme tento test **dvouvýběrový z-test**.[1] Dále s použitím věty 1 snadno určíme kritický obor testu $\mathbf{C} = (-\infty, -u_{1-\alpha/2}) \cup (u_{1-\alpha/2}, \infty)$, kde u_x je x-tý kvantil normovaného normálního rozdělení. Nulovou hypotézu H_0 zamítáme právě tehdy, když $|Z_r| \geq u_{1-\alpha/2}$, kde Z_r je pozorovaná hodnota testové statistiky. Prakticky se pro testování hypotéz používá p-hodnota. Asymptotickou variantu lze vypočítat pomocí distribuční funkce normálního rozdělení Φ jako $p = 2\left(1 - \Phi\left(|Z_r|\right)\right)$.

1.2.1 Welchův test

Bez shodnosti rozptylů jsme byli odkázáni na použití asymptotické varianty t-testu. Avšak existuje několik aproximací, které zlepšují přesnost aproximace rozdělení testové statistiky. Jedna z nejpoužívanějších je **Welchův test**.

Welch (1938)[2] nalezl užitečnou aproximaci skutečného rozdělení veličiny Z_r (1.2) za platnosti nulové hypotézy pomocí t-rozdělení o f stupních volnosti dané následujícím vzorcem

$$f = \frac{\left(\dfrac{S_X^2}{n} + \dfrac{S_Y^2}{m}\right)^2}{\dfrac{(S_X^2)^2}{n^2(n-1)} + \dfrac{(S_Y^2)^2}{m^2(m-1)}}.$$

[1] alternativně **asymptotický dvouvýběrový t-test**

P-hodnotu spočteme jako $p = 2(1 - G_f|Z_r|)$, kde Z_r je pozorovaná hodnota testové statistiky a G_f je distribuční funkce t-rozdělení o f stupních volnosti. Pro tuto aproximaci jsme napočítali $p = 2.83 \times 10^{-8}$ s testovou statistkou $Z_r = 6.63$. Jelikož je p významně menší než $\alpha = 0.05$, můžeme s rezervou zamítnout H_0 ve prospěch H_1. To znamená, že rozdíl mezi střední hodnotami cen *prodejce* a *volitele* na hladině α není nulový, což indikuje asymetrii cen a existenci majetnického efektu.

1.2.2 Odvození konfidenčního intervalu

Označme si $M \equiv \sqrt{S_X^2/n + S_Y^2/m}$ a využijeme větu 1 a aproximované kvantily t-rozdělení t_f pro odvození intervalu spolehlivosti pro θ_1.
Pro $\min(m, n) \to \infty$ a $m/n \to p \in (0, \infty)$ platí:

$$P\left[t_f(\alpha/2) \leq \frac{\hat{\theta}_1 - \theta_1}{M} \leq t_f(1 - \alpha/2)\right] \to 1 - \alpha,$$

$$P\left[\hat{\theta}_1 - t_f(1 - \alpha/2)M \leq \theta_1 \leq \hat{\theta}_1 - t_f(\alpha/2)M\right] \to 1 - \alpha.$$

Po dosazení dostaneme, že rozdíl středních hodnot cen θ_1 mezi zmíněnými subjekty je mezi 2.84 a 5.32 dolary s 0.95 pravděpodobností pokrytí.

Poznámka. Jednou z nevýhod testu je jeho citlivost na extrémní hodnoty, které mohou mít výrazný vliv na střední hodnotu a tudíž i na výsledky testu. Vzhledem k tomu, že naše podkladová data nepocházejí ze symetrického rozdělení a obsahují větší množství extrémních hodnot, tak nemusí být nejvhodnější založit testovou statistiku na rozdílu středních hodnot.

1.3 Inference podílu středních hodnot

Zvolme si parametr $\theta_2 \equiv \frac{\mu_X}{\mu_Y}$ s odhadem $\hat{\theta}_2 = \frac{\overline{X}_n}{\overline{Y}_m}$. Parametr θ_2 lze interpretovat jako poměr střední hodnoty ceny *prodejce* a *volitele*, tudíž kolikrát je střední hodnota ceny zboží *prodejce* větší než střední hodnota ceny zboží *volitele*. Všimněme si, že na rozdíl od parametru θ_1, je tato volba relativní a nemá žádné jednotky. Díky tomu můžeme porovnávat výsledky různých studií, které mohou používat odlišné míry nebo odlišné druhy zboží, což poskytuje obsáhlejší a srovnatelnější pohled na zkoumaný majetnický efekt.

Opět pracujeme s modelem $\mathcal{F}$, který vyžaduje pouze konečné druhé momenty rozdělení X_i, Y_j. Testovaný parametr lze zavést dvěma způsoby: podíl středních hodnot $\frac{\mu_X}{\mu_Y}$ anebo rozdíl logaritmů středních hodnot $\log \mu_X - \log \mu_Y$.

1.3.1 I. způsob odvození

Začněme s „klasickou“ variantou, kde formulujeme nulovou a alternativní hypotézu pro θ_2 jako

$$H_0 : \theta_2 \equiv \frac{\mu_X}{\mu_Y} = \delta, \quad H_1 : \theta_1 \equiv \frac{\mu_X}{\mu_Y} \neq \delta,$$

kde δ je konstanta (v našem případě volíme $\delta = 1$).

Věta 2. *Mějme nezávislé náhodné výběry $X_1, \dots, X_n$ a $Y_1, \dots, Y_m$ z rozdělení se středními hodnotami μ_X a μ_Y respektive a konečnými rozptyly, kde $\mu_Y > 0$, poté pro $\min(m,n) \to \infty$ a $m/n \to p \in (0,\infty)$ dostáváme*

$$\frac{\hat{\theta}_2 - \theta_2}{\sqrt{\dfrac{S_X^2}{\overline{Y}_m^2 n} + \dfrac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 m}}} \xrightarrow{D} \mathcal{N}(0,1). \tag{1.3}$$

Důkaz. Tvrzení dokážeme pomocí centrální limitní věty (CLV) a delta věty. Nechť existuje $p > 0$ takové, že $m/n \to p$. Nejdříve aplikujeme CLV na marginální konvergence a provedeme jednoduché úpravy

$$\begin{aligned}
&\sqrt{n}(\overline{X}_n - \mu_X) \xrightarrow{D} \mathcal{N}(0, \sigma_X^2),\\
&\sqrt{m}(\overline{Y}_m - \mu_Y) \xrightarrow{D} \mathcal{N}(0, \sigma_Y^2),\\
&\implies \sqrt{np}(\overline{Y}_m - \mu_Y) \xrightarrow{D} \mathcal{N}(0, \sigma_Y^2),\\
&\implies \sqrt{n}(\overline{Y}_m - \mu_Y) \xrightarrow{D} \mathcal{N}(0, \frac{\sigma_Y^2}{p}).
\end{aligned}$$

Díky nezávislosti výběrových průměrů $\overline{X}_n, \overline{Y}_m$, která vyplývá z nezávislosti náhodných výběrů $\mathbf{X}, \mathbf{Y}$, lze marginální konvergence sdružit

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu_X \\ \overline{Y}_m - \mu_Y \end{pmatrix} \xrightarrow{D} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \dfrac{\sigma_Y^2}{p} \end{pmatrix}\right) \text{ pro } n \to \infty. \tag{1.4}$$

Nulové kovariance jsou zajištěné opět nezávislostí výběrů $\mathbf{X}, \mathbf{Y}$. Na výslednou matici aplikujeme delta větu s funkcí $g(x,y) = x/y$, kde $\nabla g(x,y) = (1/y, -x/y^2)^T$ je gradient. Označme si výsledný asymptotický rozptyl V_r a počítejme

$$V_r \equiv \begin{pmatrix} \dfrac{1}{\mu_Y} & -\dfrac{\mu_X}{\mu_Y^2} \end{pmatrix}^T \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \dfrac{\sigma_Y^2}{p} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\mu_Y} & -\dfrac{\mu_X}{\mu_Y^2} \end{pmatrix} = \frac{\sigma_X^2}{\mu_Y^2} + \frac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 p}. \tag{1.5}$$

Pro další postup využijeme základních vlastností výběrového průměru a rozptylu. Víme, že $S_X^2 \xrightarrow{P} \sigma_X^2, \overline{X}_n \xrightarrow{P} \mu_X$, kde symbolem $\xrightarrow{P}$ myslíme konvergenci v pravděpodobnosti pro $n \to \infty$.

Dále známe fakt, že funkce umocnění, součinu, odmocniny a podílu jsou spojité pro kladná reálná čísla. Můžeme proto využít větu o spojitém zobrazení, díky které máme pro $m/n \to p, \min(m,n) \to \infty$ zajištěnou konvergenci

$$\frac{S_X^2}{\overline{Y}_m^2} + \frac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 p} \xrightarrow{P} \frac{\sigma_X^2}{\mu_Y^2} + \frac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 p}. \tag{1.6}$$

Nakonec vezmeme výraz (1.1) a rozšíříme ho o napočítaný asymptotický rozptyl V_r (1.5)

$$\sqrt{n}\frac{\hat{\theta}_2 - \theta_2}{\sqrt{\dfrac{S_X^2}{\overline{Y}_m^2} + \dfrac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 p}}} \cdot \frac{\sqrt{\dfrac{\sigma_X^2}{\mu_Y^2} + \dfrac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 p}}}{\sqrt{\dfrac{\sigma_X^2}{\mu_Y^2} + \dfrac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 p}}}. \tag{1.7}$$

Provedeme takové algebraické úpravy, abychom mohli na vzniklý zlomek aplikovat Cramer-Sluckého větu. Pro $\min(n,m) \to \infty, m/n \to p$ máme

$$\frac{\hat{\theta}_2 - \theta_2}{\sqrt{\frac{\sigma_X^2}{\mu_Y^2 n} + \frac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 m}}} \cdot \underbrace{\frac{\sqrt{\frac{\sigma_X^2}{\mu_Y^2 n} + \frac{\mu_X^2 \sigma_Y^2}{\mu_Y^4 m}}}{\sqrt{\frac{S_X^2}{\overline{Y}_m^2 n} + \frac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 m}}}}_{\xrightarrow{P} 1} \xrightarrow{D} \mathcal{N}(0,1). \tag{1.8}$$

Z rovnice (1.6) plyne, že jmenovatel konverguje v pravděpodobnosti k téže hodnotě jako čitatel, což znamená, že celý zlomek konverguje v pravděpodobnosti k jednotce. Dokázali jsme, že výraz (1.3) konverguje k normálnímu rozdělení, a tím je důkaz hotov. □

Testová statistika pro $\delta = 1$ plyne z (1.3)

$$Z_q \equiv \frac{\hat{\theta}_2 - 1}{\sqrt{\frac{S_X^2}{\overline{Y}_m^2 n} + \frac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 m}}}. \tag{1.9}$$

Kritický obor i p-hodnota jsou obdobné jako v sekci 1.2. Asymptotický interval spolehlivosti pro θ_2 odpovídá

$$\left(\hat{\theta}_2 - u_{1-\alpha/2}\sqrt{\frac{S_X^2}{\overline{Y}_m^2 n} + \frac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 m}}, \hat{\theta}_1 + u_{1-\alpha/2}\sqrt{\frac{S_X^2}{\overline{Y}_m^2 n} + \frac{\overline{X}_n^2 S_Y^2}{\overline{Y}_m^4 m}}\right).$$

Pro naše data vyšla statistika $Z_q = 4.08$, konfidenční interval $(1.60, 2.71)$, tedy cena, kterou přijal *prodejce* je 1.6 až 2.7 krát větší než u *volitele*. Navíc nám vyšla p-hodnota $p = 4.53 \times 10^{-5} \ll \alpha = 0.05$, a proto H_0 s rezervou zamítáme ve prospěch H_1, čímž jsme prokázali přítomnost majetnického efektu.

1.3.2 II. způsob odvození

Nyní se zaměříme na transformaci pomocí logaritmu, která oproti svému protějšku má zásadní výhodu. Konfidenční intervaly pro θ_2 odvozené z konfidenčního intervalu pro $\log\theta_2$ má kladné meze, což je přirozený parametrický prostor pro daný parametr. Zlogaritmujme-li parametr $\theta_2 = \frac{\mu_X}{\mu_Y}$ na $\log\theta_2 = \log\frac{\mu_X}{\mu_Y} = \log\mu_X - \log\mu_Y$ s odhadem $\log\hat{\theta}_2 = \log\frac{\overline{X}_n}{\overline{Y}_m}$. Pro $\log\theta_2$ testujeme

$$H_0 : \log\theta_2 = \delta \text{ oproti } H_1 : \log\theta_2 \neq \delta,$$

kde δ je dané číslo (v našem případě volíme $\delta = 0$).

Věta 3. *Mějme nezávislé náhodné výběry $X_1, \ldots, X_n$ a $Y_1, \ldots, Y_m$ z rozdělení s kladnými středními hodnotami μ_X a μ_Y respektive a konečnými rozptyly. Pro $\min(m,n) \to \infty$ a $m/n \to p \in (0,\infty)$ dostaneme*

$$\frac{\log\hat{\theta}_2 - \log\theta_2}{\sqrt{\frac{S_X^2}{\overline{X}_n^2 n} + \frac{S_Y^2}{\overline{Y}_m^2 m}}} \xrightarrow{D} \mathcal{N}(0,1). \tag{1.10}$$

Důkaz. Uvažujme funkci $g(x) = \log x$ s derivací $g'(x) = \frac{1}{x}$. Využijeme CLV a delta větu a odvodíme marginální konvergence pro $n \to \infty$, $m \to \infty$ respektive

$$\sqrt{n}(\log(\overline{X}_n) - \log \mu_X) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\mu_X^2}\sigma_X^2),$$
$$\sqrt{m}(\log(\overline{Y}_m) - \log \mu_Y) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\mu_Y^2}\sigma_Y^2).$$

Díky nezávislosti výběrových průměrů $\overline{X}_n, \overline{Y}_m$ lze konvergence vyjádřit sdruženě

$$\sqrt{n}\begin{pmatrix}\log(\overline{X}_n) - \log \mu_X \\ \log(\overline{Y}_m) - \log \mu_Y\end{pmatrix} \xrightarrow{D} \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\frac{\sigma_X^2}{\mu_X^2} & 0 \\ 0 & \frac{\sigma_Y^2}{\mu_Y^2 p}\end{pmatrix}\right).$$

Opět použijeme delta větu, nyní ale pro sdruženou konvergenci. Mějme funkci $g(x, y) = x - y$ a její gradient $\nabla g(x, y) = (1, 1)^T$. Zbytek důkazu povedeme v podobném duchu jako u věty 1. Označme si výsledný asymptotický rozptyl $V_{\log}$

$$V_{\log} \equiv \begin{pmatrix}1 & 1\end{pmatrix}^T \begin{pmatrix}\frac{\sigma_X^2}{\mu_X^2} & 0 \\ 0 & \frac{\sigma_Y^2}{\mu_Y^2 p}\end{pmatrix} \begin{pmatrix}1 & 1\end{pmatrix} = \frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2 p}. \tag{1.11}$$

Dle věty o spojitém zobrazení získáme pro $m/n \to p, \min(m, n) \to \infty$

$$\frac{S_X^2}{\overline{X}_n^2} + \frac{S_Y^2}{\overline{Y}_m^2 p} \xrightarrow{P} \frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2 p}. \tag{1.12}$$

Označme si $\log \theta_2 = \log \mu_X - \log \mu_Y$ a $\log \hat{\theta}_2 = \log(\overline{X}_n) - \log(\overline{Y}_m)$. Analogicky s důkazem věty 2, výraz upravíme

$$\sqrt{n}\frac{\log \hat{\theta}_2 - \log \theta_2}{\sqrt{\frac{S_X^2}{\overline{X}_n^2} + \frac{S_Y^2}{\overline{Y}_m^2 p}}} \cdot \frac{\sqrt{\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2 p}}}{\sqrt{\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2 p}}}, \tag{1.13}$$

a pro $m/n \to p, \min(m, n) \to \infty$ aplikujeme Cramer-Sluckého větu

$$\frac{\log \hat{\theta}_2 - \log \theta_2}{\sqrt{\frac{\sigma_X^2}{\mu_X^2 n} + \frac{\sigma_Y^2}{\mu_Y^2 m}}} \cdot \underbrace{\frac{\sqrt{\frac{\sigma_X^2}{\mu_X^2 n} + \frac{\sigma_Y^2}{\mu_Y^2 m}}}{\sqrt{\frac{S_X^2}{\overline{X}_n^2 n} + \frac{S_Y^2}{\overline{Y}_m^2 m}}}}_{\xrightarrow{P} 1} \xrightarrow{D} \mathcal{N}(0, 1). \tag{1.14}$$

□

Z (1.10) plyne testová statistika. Pro $\delta = 0$ máme

$$Z_{log} \equiv \frac{\log \hat{\theta}_2}{\sqrt{\frac{S_X^2}{\overline{X}_n^2 n} + \frac{S_Y^2}{\overline{Y}_m^2 m}}}.$$

Kritický obor i p-hodnota se počítají shodně s prvním způsobem. Označme si $M_2 = \sqrt{\frac{S_X^2}{\overline{X}_n^2 n} + \frac{S_Y^2}{\overline{Y}_m^2 m}}$. Asymptotický interval spolehlivosti parametru θ_2 pro $\min(m,n) \to \infty$ a $m/n \to p$ odvodíme

$$P\left[-u_{1-\alpha/2} \leq \frac{\log\hat{\theta}_2 - \log\theta_2}{M_2} \leq u_{1-\alpha/2}\right] \to 1-\alpha,$$

$$P\left[e^{\log\hat{\theta}_2 - u_{1-\alpha/2}M_2} \leq \theta_2 \leq e^{\log\hat{\theta}_2 + u_{1-\alpha/2}M_2}\right] \to 1-\alpha.$$

Na celý výraz jsme aplikovali exponenciální funkci a vyjádřili θ_2:

$$\theta_2 \in \left(\hat{\theta}_2 e^{-u_{1-\alpha/2}M_2}, \hat{\theta}_2 e^{u_{1-\alpha/2}M_2}\right).$$

Pro naše data jsme napočítali $Z_{\log} = 5.84$, p-hodnotu rovnou 5.22×10^{-9} a konfidenční interval $(1.66, 2.78)$. Jelikož je $p \ll \alpha = 0.05$, nulovou hypotézu zamítáme ve prospěch alternativy. Prokázali jsme, že logaritmus podílu středních hod je různý od nuly, tedy střední hodnoty jsou různé. To implikuje asymetrii ceny mezi subjekty a s tím související majetnický efekt.

1.4 Neparametrický přístup

Doposud jsme se zabývali testováním středních hodnot, což nemusí být vždy optimální volba. Existují však alternativní přístupy, které mají neparametrický model. Jednou z nejvyužívanějších je **dvouvýběrový Wilcoxonův test**, který podrobněji prozkoumáme. Mějme model

$$\mathcal{F} = \{F_X, F_Y \text{ distribuční funkce, kde } F_X(x) = F_Y(x - \delta_{XY}) \text{ pro } \delta_{XY} \in \mathbb{R}\},$$

kde F_X je spojitá distribuční funkce.

Někteří autoři tento test označují také jako **Mann-Whitney test**. Ve skutečnosti se jedná o alternativní formulaci Wilcoxonova testu. Na tuto formulaci se podíváme v sekci 1.4.2, kde si ukážeme, jaký je mezi nimi vztah a jak lze tuto alternativu využít.

1.4.1 I. formulace

Klasický Wilcoxonův test je vhodnou alternativou v případě, kdy výběry mají odlehlé hodnoty, poněvadž Wilcoxonovu statistiku zajímá pouze pořadí hodnot. Je-li ale porušen předpoklad spojitosti, je nutno úprav rozebraných v sekci 1.4.1.

Testovaným parametrem je posunutí $\theta_3 = \delta_{XY}$[2] s hypotézami

$$H_0 : \delta_{XY} = 0 \text{ oproti } H_1 : \delta_{XY} \neq 0.$$

Definice 1 (Testová statistika $W_{n,m}$)**.** Mějme spojený náhodný výběr $(G_1, \ldots, G_{n+m}) \equiv (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$. Definujme

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

[2]Platí-li model $\mathcal{F}$ a hypotéza H_0, pak jsou X, Y totožné, platí rovnost mediánů $m_X = m_Y$ a středních hodnot $\mu_X = \mu_Y$.

kde $R_1, R_2, \ldots, R_n$ jsou pořadí náhodných veličin X_i ve uspořádaném sdruženém náhodném výběru $\mathbf{G}^*$.

Testová statistika $W_{n,m}$ je založena na pořadích pozorovaných hodnot, což nám poskytuje určitou flexibilitu ohledně rozdělení výběru. Co ovšem potřebujeme, je možnost výběr jednoznačné uspořádat.

Věta 4 (Střední hodnota a rozptyl $W_{n,m}$ statistiky)**.** *Nechť $N = n + m$. Platí-li model $\mathcal{F}$ a hypotéza H_0, pak*

1. $U \equiv \dfrac{W_{n,m} - \dfrac{n(m+n+1)}{2}}{\sqrt{\dfrac{mn(m+n+1)}{12}}}$

2. *Pokud* $\min(n, m) \to \infty$, $n/N \to q \in (0, 1)$, *pak* $U \xrightarrow{D} \mathcal{N}(0, 1)$.

Důkaz. První část nalezneme například v knize Anděl (2011)[3], kapitola 8, strana 102; druhou v Kulla (2020)[4], kapitola 1, věta 3. □

Věta nám dává nástroj, jak spočítat testovou statistiku a jak provést asymptotický test.

Poznámka. Důkaz první části lze jednoduše odvodit, známe-li součty základních konečných řad. Důkaz druhé části je složitější než se na první pohled zdá. Všimněme si, pořadí R_i nejsou nezávislé, tudíž nelze použít běžnou verzi centrální limitní věty.

Porušení předpokladů

Vlivem diskretizace [3] či zaokrouhlení veličin se během zpracování dat mohou vyskytnout shodné hodnoty, což porušuje předpoklad spojitosti distribuční funkce. Testovou statistiku $W^*_{n,m}$ vypočteme pomocí tzv. průměrného pořadí, což znamená, že v případě, kdy se ve uspořádaném náhodném výběru vyskytnou shodné hodnoty, vypočítáme aritmetický průměr z původních pořadí těchto hodnot. Například, pokud se hodnota 1.25 vyskytuje ve sdruženém výběru $(\mathbf{X}, \mathbf{Y})$ dvakrát s původními pořadími 4 a 5, použijeme metodu průměrného pořadí a vypočítáme nové průměrné pořadí jako $\frac{4+5}{2} = 4.5$. Tento přístup demonstrujeme na uspořádaném sdruženém výběru $(\mathbf{X}^*, \mathbf{Y}^*)$ v tabulce 1.2.

Následující věta nám formuluje statistiku, kterou můžeme v této situaci využít. Všimněme si stejné střední hodnoty jako u neupravené statistiky. Tato skutečnost vychází přirozeně z definice $W^*_{n,m}$.

[3] Převod spojitých hodnot na diskrétní, aby bylo možné s nimi snáze pracovat.

Hodnota	Původ	Pořadí	Průměrné pořadí
0.25	Y^*	1	1.0
0.75	Y^*	2	2.5
0.75	Y^*	3	2.5
1.25	Y^*	4	4.5
1.25	Y^*	5	4.5
...			
5.25	X^*	25	26.5
5.25	Y^*	26	26.5
5.25	Y^*	27	26.5
5.25	Y^*	28	26.5
5.25	Y^*	29	26.5
...			
10.00	X^*	43	46.5
10.00	X^*	44	46.5
10.00	X^*	45	46.5
10.00	X^*	46	46.5
10.00	X^*	47	46.5
10.00	X^*	48	46.5
10.00	X^*	49	46.5
10.00	Y^*	50	46.5

Tabulka 1.2: Ukázka techniky průměrného pořadí pro uspořádané náhodné výběry $\mathbf{X}^*, \mathbf{Y}^*$.

Věta 5. *Nechť $N = n + m$. Za platnosti hypotézy H_0 pro $\min(n, m) \to \infty$ a $n/N \to q \in (0, 1)$ dostaneme*

$$U_T \equiv \frac{W^*_{n,m} - \mathbb{E}\, W^*_{n,m}}{\sqrt{\mathcal{V}}} \xrightarrow{D} \mathcal{N}(0, 1),$$

kde

$$\mathbb{E}\, W^*_{n,m} = \frac{n(m + n + 1)}{2},$$

$$\mathcal{V} = \frac{mn}{12}\left[(n + m + 1) - \frac{\sum_q t_q^3 - t_q}{(n + m)(n + m - 1)}\right],$$

t_q je počet pozorování s q-tou shodnou hodnotou a suma $\sum_q$ značí součet přes všechny různé hodnoty sdruženého výběru $(\mathbf{X}, \mathbf{Y})$.

Důkaz. Důkaz je k nahlédnutí v práci Kulla (2020)[4], kapitola 2, věta 8. □

Pro naše data vychází pořadová statistika $W^*_{n,m} = 884$ a $U_T = 4.80$. Jelikož se jedná o asymptotický test, tak p-hodnota se počítá obdobně jako v předchozích sekcích, takže $p = 2(1 - \Phi(|U_T|))$, konkrétně $p = 1.55 \times 10^{-6}$. Jelikož je p řádově menší než $\alpha = 0.05$, hypotézu H_0 zamítáme. Prokázali jsme, že distribuční funkce F_X a F_Y jsou různé.

Poznámka. Protože Wilcoxonův test se zaměřuje na shodnost rozdělení výběrů, může poskytnout užitečné informace o tom, zda jsou distribuční funkce výběrů odlišné, a to i v případě, kdy výběry mají stejnou střední hodnotu. Nicméně pro vyšetřování majetnického efektu nás zajímají spíše charakteristiky typu střední hodnota než-li samotné distribuční funkce. Ve zkratce je pro náš problém první formulace těžko interpretovatelná a nedává nám dodatečný náhled do problematiky.

1.4.2 II. formulace

Konečně se dostáváme ke slibované Mann-Whitneyho formulaci. Na rozdíl od první formulace model $\mathcal{F}$ vyžaduje pouze spojitost distribučních funkcí F_X, F_Y, což ale naše data hrubě porušují. Naštěstí lze požadavek opět obejít úpravou testové statistiky. Zajímá nás parametr $\theta_4 \equiv P[X_i < Y_j]$, tedy pravděpodobnost, že náhodná veličina X_i je menší než Y_j. Formulujeme hypotézu

$$H_0 : \theta_4 = \frac{1}{2} \text{ oproti } \quad H_1 : \theta_4 \neq \frac{1}{2}.$$

Definice 2 (Upravená Mann-Whitneyho statistika)**.** Mějme náhodné výběry $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$. Uvažujme všechny dvojice (X_i, Y_j) pro $i = 1, \ldots, n$, $j = 1, \ldots, m$, poté

$$W_{MW} \equiv \sum_{i=1}^{n} \sum_{j=1}^{m} (\mathbb{I}_{X_i < Y_j} + \tfrac{1}{2}\mathbb{I}_{X_i = Y_j}).$$

Uvedli jsme obecnou verzi, která v případě shodných hodnot připočte $\frac{1}{2}$ namísto 1. Pokud by byly hodnoty unikátní, tak bychom indikátor $\mathbb{I}_{X_i = Y_j}$ mohli vynechat, jelikož by se rovnal 0. V následujícím tvrzení si ukážeme souvislost $W^*_{n,m}$ a W_{MW}, díky které snadno určíme momenty W_{MW} a odvodíme příslušnou testovou statistiku.

Věta 6 (Vztah $W^*_{n,m}$ a W_{MW} statistiky)**.** *Mějme $W^*_{n,m}$ a W_{MW} dle definice 1 a definice 2 respektive. Platí*

1. $W^*_{n,m} + W_{MW} = mn + \frac{n(n+1)}{2}$,
2. *Pokud* $\min(m, n) \to \infty$, *pak* $\frac{W_{MW}}{nm} \xrightarrow{P} P[X_i < Y_j] + \frac{1}{2}P[X_i = Y_j]$.

Důkaz. První část nalezneme v práci Kulla (2020)[4], kapitola 2, tvrzení 4. V druhé části je důkaz konzistence k nahlédnutí v knize van der Vaart (2000)[5], kapitola 12, věta 6. □

Testy založené na Wilcoxonově a Mann-Whitneyho statistice jsou ekvivalentní. Mann-Whitneyho statistika lépe ilustruje, co vlastně Wilcoxonův test testuje i pro obecný model $\mathcal{F}$.

Nyní máme všechny potřebné nástroje k formulování upravené testové statistiky pro Mann-Whitneyho variantu. Použijeme vztah z prvního bodu věty 6 a formuli z věty 5. Po dosazení dostaneme

$$U_{MW} \equiv \frac{\frac{mn}{2} - W_{MW}}{\sqrt{\mathcal{V}}}. \tag{1.15}$$

Pomocí vzorce (1.15) vypočítáme hodnotu $U_{MW} = 4.80$ s $p = 1.55 \times 10^{-6}$. Tím můžeme zamítnout H_0 na hladině α a získáváme důkaz, že mezi cenami *prodejce* a *volitele* existuje statisticky významný rozdíl.

Pokud by platila nulová hypotéza, tedy že neexistuje rozdíl mezi rozděleními X_i a Y_j, pak by pravděpodobnost, že cena *prodejce* je nižší než cena *volitele*, byla $\frac{1}{2}$. Zamítnutím nulové hypotézy získáváme ovšem evidenci, že takový rozdíl existuje. Například, pokud by skutečný parametr θ_4 byl $\frac{1}{4}$, pak bychom měli pravděpodobnost $\frac{3}{4}$, že cena *prodejce* X_i je vyšší než cena *volitele* Y_j, což naznačuje asymetrii cen a majetnický efekt.

Testování ekvivalencí

V úvodu jsme zmínili náhodný výběr **Z**, čili *kupce*. V podkladovém článku Novemsky (2005)[1] je testován vztah mezi *volitelem* a *kupcem*, kde se snaží potvrdit rovnost zvolených charakteristik, tedy že *kupec* nevykazuje majetnický efekt. Poznamenejme, že není možné použít metody rozebrané v této kapitole pro testování rovností. Avšak problém není beznadějný a odpověď můžeme nalézt v teorii o testování ekvivalencí, která je však nad rámec této práce. Této problematice se věnuje například Wellek (2010)[6].

Kapitola 2

Averze ke ztrátě a majetnický efekt

V předchozí kapitole byl několikrát zmíněn pojem *majetnický efekt* a *asymetrie ceny*. V úvodu jsme si již nastínili, co testujeme a na co se snažíme najít odpověď, ale nevysvětlili jsme podstatu problému do podrobností. V této kapitole si proto problematiku řádně vyložíme, uvedeme do kontextu a ilustrujeme na příkladech z finančního světa.

2.1 Prospektová teorie

Prospektová teorie byla představena v roce 1979 psychologem Danielem Kahnemanem a ekonomem Amosem Tverskym a stala se klíčovým pojmem v oblasti behaviorální ekonomie. Jejich pilotní práce *Prospect Theory: An Analysis of Decision under Risk*[7] se zabývala rozhodováním člověka vystavenému možné ztrátě či zisku. Podle Kahnemana a Tverského se lidé při rozhodování v rizikových situací neřídí pouze pravděpodobnostmi, ale také subjektivními hodnotami a emocemi spojenými s možnými výsledky. Tato teorie vysvětluje skutečnost, že člověk více vnímá změnu hodnoty vůči referenčnímu bodu než samotnou absolutní velikost. **Referenčním bodem** (angl. Reference Point) je myšlena hodnota, kterou má jedinec nastavenou jako výchozí pro porovnání výsledků v dané situaci. Často vychází ze zkušeností, očekávání nebo společenských norem. Příklad referenčního bodu je ilustrován v Clarkovi (2008)[8].

Ve studii uvedené v úvodu a použité v kapitole 1 jsme ukázali, že *prodejci* měli tendenci prodávat zboží za vyšší cenu než kontrolní skupina, *volitelé*. Tento fakt je názornou ukázkou iracionálního myšlení subjektu nad daným obchodem. *Prodejci* vnímali výměnu jako ztrátu, kdežto *volitelé* jako zisk.

2.1.1 Averze ke ztrátě

Averze ke ztrátě (angl. Loss Aversion) představuje fenomén, kdy jednotlivci přikládají větší váhu ztrátám než ziskům. Novemsky ve svém článku *Boundary of Loss Aversion*[1] zdůrazňuje, že *„ztráta je vnímána jako větší a psychologicky významnější než zisk stejné výše, což má silný vliv na jednání lidí v oblasti ekonomiky“*. Například ztráta 100 dolarů je jedincem vnímána hůře než zisk 100 dolarů.

V našem případě *prodejce* byl při prodeji vystaven averzi ke ztrátě, a proto vyžadoval jakési prémium v podobě zvýšené prodejní ceny.

V kapitole 1 byl několikrát zmíněn termín **majetnický efekt** (angl. Endowment effect). Kahneman (1991)[9] jej definoval jako jev, při kterém lidé přeceňují hodnotu věcí, které vlastní, oproti stejným věcem, které nevlastní. Jedná se tedy o asymetrii ceny mezi subjekty, která závisí na statusu vlastnictví k dané věci a zároveň vysvětluje jednu z příčin averze ke ztrátě.

Nakonec si uveďme vybrané příklady ze světa finančního trhu.

- **Averze ke ztrátě a krátkozrakost investorů**: Studie provedená Benartzi a Thalerem (1995)[10] zkoumala, jak averze ke ztrátě a krátkozrakost investorů ovlivňuje jejich investiční rozhodnutí, a snažila se vysvětlit paradoxní výskyt vysokých výnosů z akcií (angl. equity premium). Na základě analýzy reálných finančních dat zjistili, že investorům se zdá, že ztráta na investicích je pro ně bolestivější než zisk stejné výše, což vede k opatrnému rozhodování a výběru investic s nižšími výnosy.To může následně způsobit vysoký výnos z akcií ve srovnání s méně rizikovými aktivy.

- **Averze ke ztrátě v hodnocení portfolia**: Favilukis a spol (2017)[11] tvrdí, že investoři mají tendenci více pociťovat ztrátu než se těšit z výhry, a zároveň vysvětluje iracionální chování investorů rychle prodávat ze ziskového portfolia za účelem zvýšení výnosnosti a naopak držet pozice ve ztrátovém portfoliu v naději na zlepšení situace.

- **Averze ke ztrátě v oblasti pojištění**: Sydnor a spol (2010)[12] z výzkumu zjistili, že lidé se nadměrně pojišťují v případě relativně malých rizik, což může mít za následek neefektivní alokace zdrojů a zvýšené náklady pro spotřebitele a dále narušuje trh.

Kapitola 3

Seznámení s podkladovou studií

Tato kapitola se věnuje samotnému podkladovému článku od Novemsky (2005)[1]: jakým způsobem byla data získána, co bylo měřeno a za jakých podmínek.

3.1 Formulace zadání

Všechny experimenty probíhaly ve skupinách v kontrolovaném prostředí, některé ve třídách a zbylé v laboratořích, přičemž byli testováni studenti a zaplacení respondenti. Bylo zavedeno celkem 5 disjunktních možností/rolí, avšak ne všechny byly použity ve všech studiích. Proběhlo celkem 8 experimentů, kde subjektům byla přiřazena nejméně jedna role. Ve studii byly použity levné předměty typu hrníček, pero či čokoláda. Článek se zabývá dvěma tématy: majetnickým efektem a porovnání jiných averzí vzhledem k averzi ke ztrátě. Tato práce se věnuje jenom prvnímu tématu, a proto si uvedeme pouze relevantní role k naší výzkumné otázce.[1]

- Prodejce (angl. Seller) - WTA (Willingness To Accept)
- Volitel (angl. Chooser) - CE (Choice Equivalent)
- Kupec (angl. Buyer) - WTP (Willingness To Pay)

3.2 Průběh experimentu

Všechny experimenty byly zahájeny náhodným přidělením formulářů s barevným kódováním. Zboží bylo přiděleno účastníkům, kteří obdrželi formulář *prodejce*. Následně vyplnili dotazník, který se ptal na jejich preference ohledně zboží a souvisejících cen. Poté byli informováni, že výsledek určí pouze jejich preference vůči tázaným cenám. Tento formát byl použit, aby se zabránilo tzv. strategickému odpovídání[2] (Becker a spol, 1964)[13].
Volitelům byla nabídnuta volba mezi zbožím a penězi. *Prodejci* obdrželi zboží a bylo jim řečeno, že jej vlastní.

[1] Pojmy WTA, CE, WTP jsou převzaté z tradičních ekonomických modelů.

[2] Strategické odpovídání se týká chování jedinců, kteří ve svých rozhodnutích zohledňují očekávání o chování ostatních subjektů. Typickým příkladem je vězňovo dilema z teorie her.

Cena	Souhlas	Nesouhlas
25.0 $	**Ano**	Ne
24.0 $	**Ano**	Ne
23.0 $	**Ano**	Ne
...		
...		
2.5 $	**Ano**	Ne
2.0 $	**Ano**	Ne
1.5 $	**Ano**	Ne
1.0 $	Ano	**Ne**
0.5 $	Ano	**Ne**

Tabulka 3.1: Ukázka vyplněného formuláře pro roli *volitele*. K formuláři byly přiloženy tyto instrukce: *Prosím označte svojí volbu pro každou uvedenou hodnotu. „Ano“ představuje: „Preferuji tuto částku nad [hrníčkem]“.*

Dále jim byla položena otázka, zda by uvažovali o prodeji za uvedené ceny. *Kupcům* byla nabídnuta možnost nákupu zboží za vlastní peníze. Pokud odmítli obchodovat, opustili experiment. Všem subjektům byl předložen formulář, jehož podobu jsme demonstrovali v tabulce 3.1. Pro každou roli se lišil pouze položenou otázkou. Po vyplnění byly formuláře vybrány a proběhly všechny obchody.

Měřené veličiny

Nechť P značí množinu všech přípustných cen z formuláře (viz tabulka 3.1), $p \in P$ cenu udanou jedním účastníkem experimentu, $N \subseteq P$ množinu cen, kde bylo zakroužkováno „Ne“ a $Y \subseteq P$ množinu cen, kde bylo zakroužkováno „Ano“. Měřené veličiny definujeme následovně.

- Veličinu CE_j (př. ekvivalent volby) j-tého *volitele* můžeme definovat jako průměr nejnižší přijaté nabídky a nejvyšší odmítnuté. Tento průměr je spočten z hodnot přímo nad čárkovanou čarou a přímo pod čárkovanou čarou (viz tabulka 3.1). V kontextu kapitoly 1 se jedná o prvek náhodného výběru $\mathbf{Y}$. Symbolicky:

$$Y_j = CE_j \equiv \frac{1}{2}(\max_{p \in N} p + \min_{p \in A} p) \text{ pro } j = 1, \ldots, m.$$

- Veličina WTP_i (př. ochota prodat) i-tého *prodejce* se liší od CE_j pouze v tom, zda bylo zboží účastníku na počátku experimentu darováno či nikoliv. V kapitole 1 je veličina ekvivalentní prvku náhodného výběru $\mathbf{X}$. Symbolicky:

$$X_i = WTP_i \equiv \frac{1}{2}(\max_{p \in N} p + \min_{p \in A} p) \text{ pro } i = 1, \ldots, n.$$

- Veličina WTA_k (př. ochota zaplatit) k-tého *kupce* definujeme jako průměr nejvyšší přijaté poptávky a nejnižší odmítnuté. Opět bychom si mohli představit horizontální čáru, která tvoří „hranici“ mezi množinami A a N. Veličinu WTA_k jsme zavedli v kapitole 1 jako prvek náhodné výběru $\mathbf{Z}$. Symbolicky:

$$Z_k = WTA_k \equiv \frac{1}{2}(\max_{p \in A} p + \min_{p \in N} p) \text{ pro } k = 1, \ldots, l.$$

Kapitola 4

Diskuze nad výsledky

4.1 Výsledky testování

V kapitole 1 jsme si představili metody, které nyní použijeme na poskytnutá data. Pro dvouvýběrové testování skupin *volitel* a *prodejce* byly vhodné pouze 4 z 8 poskytnutých studií, jelikož u zbylých se buďto testoval jiný jev nebo měl uchazeč přidělené všechny role, tedy jednotlivé výběry nebyly nezávislé.

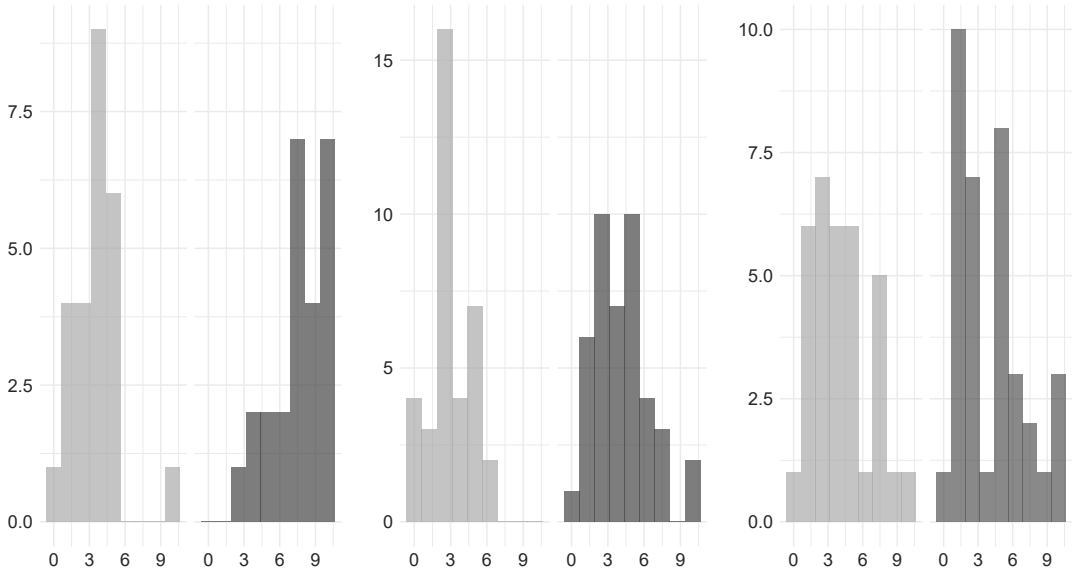
Značení studií bylo převzato z původního článku. Studie jsou označeny písmenem „S“ následované pořadovým číslem od 1 do 8. Pro studii sedm byla navíc přidána písmena „A“ až „E“, aby se odlišily různé varianty této studie.

Grafická analýza

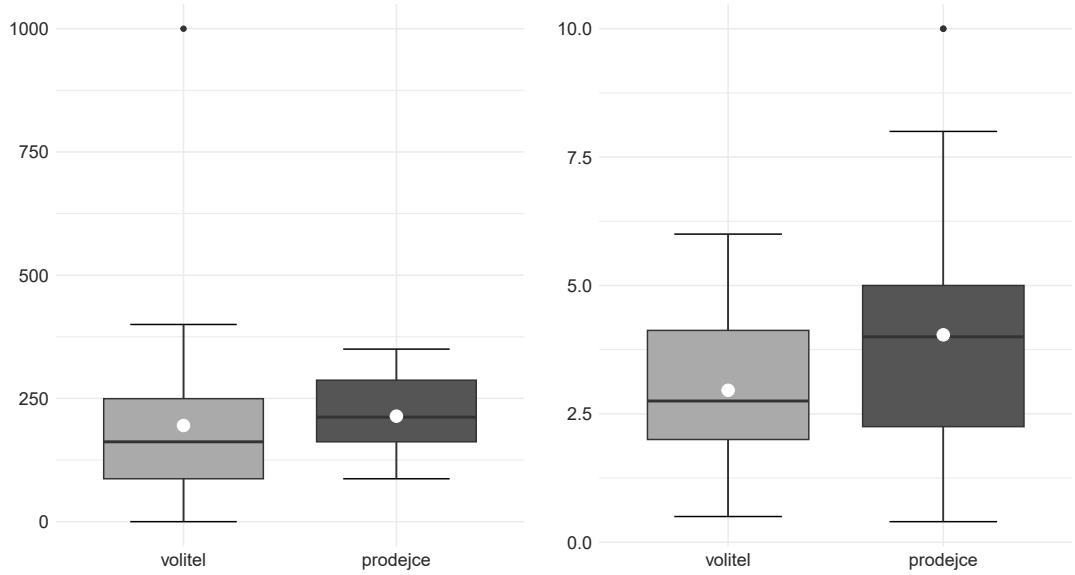
Pro názornou ukázku rozdělení veličin se podívejme na graf histogramů 4.1. Jak už bylo několikrát zmíněno, v datech se nachází nemalé množství extrémních hodnot, což lze zpozorovat například na histogramu první studie **S1**. Zdá se, že významná část *prodejců* své zboží zdráhala prodat a souhlasila pouze s nejvyšší částkou. Tím mohlo dojít ke zkreslení a účinky majetnického efektu mohly být mnohem vyšší. Na druhou stranu to pro naší analýzu neznamená fatální překážku. Přesuneme-li se na grafické znázornění pomocí boxplotů 4.2, tak zpozorujeme viditelný kontrast mezi zamítnutými a nezamítnutými hypotézami dvou odlišných studií. U studie **S7A** vidíme patrný rozdíl mezi měřenými charakteristikami, kdežto u studie **S4** nikoliv. Naši vizuální analýzu podpořily výsledky testu, které jsou k nahlédnutí v tabulce 4.1, kde na hladině α nulovou hypotézu H_0 pro **S7A** zamítáme a pro **S4** zamítnout nemůžeme.

Studie 1,4 a 8

Nyní si shrneme výsledky ze **S1**, **S4** a **S8**. První studie **S1** je prvotní experiment provedený Thalerem (1990)[14], kde byl zpozorován majetnický efekt. Osmá studie **S8** je replikací **S1**, provedená autorem článku Novemským[1]. V tabulce 4.1 nahlédneme, že p-hodnoty k testům podílu a rozdílu průměrů **S1** a **S8** vyšly ve prospěch alternativních hypotéz H_1. Řečeno slovy, máme evidenci, která podporuje existenci majetnického efektu u daných subjektů. U studie **S4** nelze dle p-hodnot zamítnout ani jednu nulovou hypotézu.



Obrázek 4.1: Histogramy náhodných výběrů **CE** (vpravo), **WTP** (vlevo) z jednotlivých studií dle rolí. Zleva vykreslujeme **S1**, **S7A** a **S7C**. Na x-ové ose se nachází cena zboží v dolarech a na y-ové počet subjektů, které zvolili danou cenu.



Obrázek 4.2: Boxploty náhodných výběrů **CE** (vpravo), **WTP** (vlevo) z jednotlivých studií. Zleva: **S4**, **S7A**. Na x-ové ose máme role. Na y-ové ose cenu zboží. Bílé tečky značí výběrový průměr.

Tento výsledek může být zapříčiněný v mírně odlišném přístupu při sběru dat[1] či vcelku nepřehlednými daty na vstupu.

Studie 7

Poslední sada je specifická tím, že subjekty nestanovovaly cenu jen jednoho objektu, nýbrž rovnou pěti (avšak stále ve stejné roli). Jednalo se o hrníček „A“, pero „B“, pero „C“, 50$ kupón na večeři „D“ a 50$ kupón na obuv „E“. Pouze v situaci „B“ se jednalo o skutečný obchod, zbytek byl pouze hypotetický. V případě hrníčku a kupónu na obuv máme důvod uvažovat existenci majetnického efektu. Je dobré zmínit, že data obsahovala signifikantní množství chybějících hodnot, které byly při testování vynechány.

Výsledné rozhodnutí

Na grafu 4.3 jsme zobrazili jednotlivé konfidenční intervaly pro relativní veličinu: podíl průměrů představený v sekci 1.2. Tím jsem docílili zachycení velikosti účinku v každém experimentu. Největší efekt jsme zaznamenali v první studii. Z vizualizace konfidenčních intervalů 4.3 lze také zpozorovat, v jakých případech jsme zamítli nulovou hypotézu.

V tabulce 4.1 vidíme, že zatímco některé studie s přehledem prokazují majetnický efekt, jiné ho neprokazují vůbec (případně ho prokazují pouze těsně). To může souviset s různými rozsahy výběrů. Jelikož však ve většině případů příliš odlišně nejsou, tak to spíše poukazuje na skutečnost, že prokazatelnost majetnického efektu je značně závislá na detailech provedení studie.

Těmito detaily myslíme subjektivní faktory, které se nepodaří izolovat či je izolovat vůbec nelze. Typickým příkladem je povaha jedince, jeho nálada během experimentu nebo jeho subjektivní vnímání obchodovaného předmětu či procesu samotného. Ukazuje to také, že tyto studie jsou natolik rozdílné, že se jeví jako poměrně problematické je jakkoliv sumarizovat do jedné hodnoty. K hlubšímu pochopení majetnického efektu by mohl přispět další výzkum.

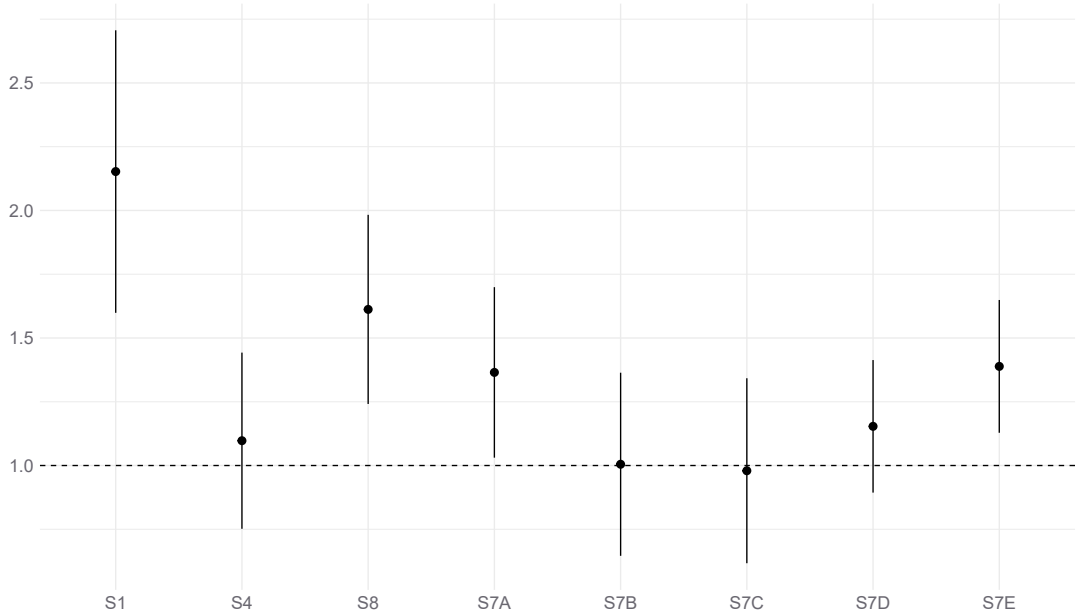
4.2 Rozdíly oproti původnímu článku

V této části prodiskutujeme metody, které použil autor původního článku a porovnáme je s těmi v této práci. Autor zformuloval čtyři domněnky, z čehož první dvě se zaměřují na testování majetnického efektu a zbylé dvě na hledání rizika mimo averzi ke ztrátě. Poněvadž se tato práce zabývala pouze první otázkou, tak vezmeme v úvahu pouze první dvě domněnky. K analýze byl zvolen medián, resp. podíl mediánů. Po nahlédnutí do struktury dat lze tvrdit, že medián je vhodnější než průměr, jelikož pocházejí z rozdělení s těžšími chvosty. Na druhou stranu, testování podílu mediánů není triviální problém. Na odhad hustoty neexistuje univerzální postup a většinou se obchází bootstrapem. Všimněme si, že u druhé domněnky se snažíme potvrdit rovnost. Tato problematika byla zmíněna na začátku kapitoly 1 a není náplní této práce.

[1] Novemsky (2005)[1] tyto odlišnosti komentuje na konci svého článku.

S|P	n/m	Logpodíl	Welch	Mann-Whitney
S1	25/25	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**
S4	31/36	0.563	0.553	0.109
S8	68/67	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**
S7A	43/36	**0.012**	**0.040**	**0.040**
S7B	43/35	0.978	0.978	0.572
S7C	36/31	0.912	1.000	0.761
S7D	35/27	0.213	0.204	0.278
S7E	35/28	$<$ **0.001**	$<$ **0.001**	$<$ **0.001**

Tabulka 4.1: Tabulka p-hodnot pro všechny zvolené parametry a testy. Ve sloupcích máme rozsahy a jednotlivé testované parametry. Rozsah výběru $\mathbf{X}$ (resp. $\mathbf{Y}$) jsme označili n (resp. m). *Logpodílem* myslíme zlogaritmizovaný podíl odpovídající parametru v sekci 1.2, *Welch* značí Welchovu aproximaci pro rozdíl průměrů zmíněné v sekci 1.4.1. *Mann-Whitney* náleží neparametrickému testu zmíněném v sekci 1.4.2. V řádcích máme jednotlivé studie.



Obrázek 4.3: Konfidenční intervaly sestrojené pro parametr $\theta_2 = \mu_X/\mu_Y$ (viz sekce 1.3). Na x-ové ose máme jednotlivé studie, na y-ové možné hodnoty podílu μ_X/μ_Y.

Hypotézy	Článek	Správný zápis
Efekt	$H_1 : WTA/CE > 1$	$H_0 : m_{WTA}/m_{CE} = 1$ vs $H_1 : m_{WTA}/m_{CE} > 1$
NLIB I	$H_2 : CE/WTP = 1$	$H_0 : m_{CE}/m_{WTP} \neq 1$ vs $H_1 : m_{CE}/m_{WTP} = 1$

Tabulka 4.2: Matematicky správný zápis domněnek z článku. Značení veličin je vysvětlené v kapitole 3. Symbolem m značíme medián. **NLIB I** označil autor domněnku, že *kupec* nevykazuje majetnický efekt.

Proces testování

Na rozdíl od běžného testování hypotéz, které bylo představeno v předchozích kapitolách, se autor rozhodl deterministicky spočítat mediány náhodných výběrů **WTA**, **WTP** a **CE** a jejich poměry, které jsou poskytnuty k nahlédnutí v tabulce 4.2. Dále spočítal vážený průměr těchto poměrů, kde váhy volil dle počtu účastníků v daném výzkumu pro danou podmínku, resp, veličinu.

Pomocí bootstrapu byl získán přibližný konfidenční interval poměru pro každý experiment, jakožto i pro ten agregovaný. Nakonec byly tyto agregované poměry porovnány dle příslušných domněnek, tedy zda byly větší nebo rovné jednotce.

Zmíněný postup byl použit jak na data párová, tak dvouvýběrová. V článku jsou dále diskutované jednotlivé výsledky i agregované poměry, které vyšly 1.85 a 1.07. Hodnota $1.85 > 1$ mluví ve prospěch existence majetnického efektu u **WTA** a $1.07 \approx 1$ ve prospěch „dostatečné podobnosti“ náhodných výběrů **CE** a **WTP** (tedy, že *prodejci* nevykazují majetnický efekt).

Studie	m_{WTA}	m_{CE}	m_{WTP}	N	m_{WTA}/m_{CE}	m_{CE}/m_{WTP}
S1	7.5	3	2.75	24.3	2.5	1.09
S2	3.75	1.75	1.25	53.6	2.14	1.4
S3	2.25			68.5		
S4	2.12	1.62		34.5	1.31	
S5	4.25	2.75		51.7	1.55	
S6	4.75	3		22.7	1.58	
S7	3.25	1.75	3.25	37.4	1.86	0.54
S8	10	4.75	4.25	33.6	2.11	1.12
Σ				1125	**1.85**	**1.07**

Tabulka 4.3: Tabulka hodnot napočítaných autorem článku. Značení m_{WTA}, m_{CE}, m_{WTP} odpovídá mediánům pro jednotlivé náhodné výběry představené v kapitole 3. N je průměrný počet účastníku na jednu skupiny pro danou studii a zbylé dva sloupce odpovídají poměrům mediánů.

Závěr

Věcnou otázkou této bakalářské práce bylo vyšetření významného jevu v behaviorální ekonomii: averze ke ztrátě a s ní spojený majetnický efekt. Namísto tvorby vlastního experimentu jsme se obrátili již na existují článek, *Boundary of Loss Aversion*, odkud jsme čerpali data a hlavní myšlenku analýzy. Identifikovali jsme, že se převážně jedná o dvouvýběrové problémy, a naše snahy směřovaly tímto směrem.

V první kapitole jsme uvedli některé dvouvýběrové testy vhodné pro naše data. Začali jsme testováním rozdílu průměrů s Welchovou aproximací. Následně byla odvozena testová statistika pro podíl průměrů, a to v běžné i v logaritmické variantě. Kapitola byla zakončená Wilcoxonovým testem, konkrétně Mann-Whitneyho formulací. Vše bylo provázeno praktickou ukázkou na části dat. Druhá kapitola čtenáře seznámila s pojmem averze ke ztrátě a majetnický efekt a jejich výskytem ve finančním sektoru. Hlavní náplní třetí kapitoly bylo obeznámení s podkladovým článkem Novemsky (2005)[1], kde jsme vysvětlili, jakým způsobem experiment probíhal, jakou formou se sbírala data a jak byly původně formulované měřené veličiny. Tyto veličiny byly taktéž sjednoceny, aby odpovídaly námi zavedeným značením v první kapitole. Poslední kapitola se věnovala vlastní analýze dat používajíc metody představené dříve. Nejprve byla provedena popisná grafická analýza, poté následovala aplikace vybraných testů na všechny vhodné studie. U čtyř studií vyšly testy ve prospěch existence majetnického efektu, u zbylých čtyř však efekt prokázán nebyl. Tento nejasný výsledek mohl být zapříčiněn mnoha faktory jako odlišné preference účastníků experimentu či chybný design experimentu. Na konec jsme naší metodiku porovnali s tou použitou autorem.

Práci bychom mohli obohatit hned několika způsoby. Jelikož se v poskytnutém datové souboru vyskytovaly studie (například **S3**), kde měl dotyčný více rolí, mohli bychom metody rozšířit o párové testy, a ty poté využít k vyhodnocení vynechaných studií. Dále by bylo zajímavé prozkoumat testování podílu či rozdílu mediánů, které by mohlo být vhodnější než testování průměrů, protože mediány nejsou tak citlivé na odlehlá pozorování. To by však vyžadovalo mimo jiné vhodné ošetření problematiky shod, které se v nemalé míře v datech vyskytovaly.

Seznam použité literatury

1. NOVEMSKY, Nathan; KAHNEMAN, Daniel. The boundaries of loss aversion. *Journal of Marketing research.* 2005, roč. 42, č. 2, s. 119–128.
2. WELCH, Bernard L. The significance of the difference between two means when the population variances are unequal. *Biometrika.* 1938, roč. 29, č. 3/4, s. 350–362.
3. ANDĚL, Jiří. *Statistické metody, 4. upravené vyd.* Praha: MATFYZPRESS, 2007.
4. KULLA, Filip. *Dvojvýberový Wilcoxonov test v prípade existencie zhôd.* Praha, 2020. Bakalářská práce. Univerzita Karlova, Matematicko-fyzikální fakulta, Katedra pravděpodobnosti a matematické statistiky.
5. VAART, Aad W Van der. *Asymptotic statistics.* Sv. 3. Cambridge university press, 2000.
6. WELLEK, Stefan. *Testing statistical hypotheses of equivalence and noninferiority.* CRC press, 2010.
7. TVERSKY, Amos; KAHNEMAN, Daniel. Judgment under uncertainty: Heuristics and biases. *Science.* 1974, roč. 185, č. 4157, s. 1124–1131.
8. CLARK, Andrew E; FRIJTERS, Paul; SHIELDS, Michael A. Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles. *Journal of Economic literature.* 2008, roč. 46, č. 1, s. 95–144.
9. KAHNEMAN, Daniel; KNETSCH, Jack L; THALER, Richard H. Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives.* 1991, roč. 5, č. 1, s. 193–206.
10. BENARTZI, Shlomo; THALER, Richard H. Myopic loss aversion and the equity premium puzzle. *The quarterly journal of Economics.* 1995, roč. 110, č. 1, s. 73–92.
11. FAVILUKIS, Jack; LUDVIGSON, Sydney C; VAN NIEUWERBURGH, Stijn. The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium. *Journal of Political Economy.* 2017, roč. 125, č. 1, s. 140–223.
12. SYDNOR, Justin. (Over) insuring modest risks. *American Economic Journal: Applied Economics.* 2010, roč. 2, č. 4, s. 177–199.
13. BECKER, Gordon M; DEGROOT, Morris H; MARSCHAK, Jacob. Measuring utility by a single-response sequential method. *Behavioral science.* 1964, roč. 9, č. 3, s. 226–232.

14. KAHNEMAN, Daniel; KNETSCH, Jack L; THALER, Richard H. Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*. 1990, roč. 98, č. 6, s. 1325–1348.