

ZÁPADOČESKÁ UNIVERZITA V PLZNI

FAKULTA EKONOMICKÁ

Diplomová práce

Statistická analýza podnikových dat

Statistical analysis of business data

Michal Elhota

Plzeň 2023

Čestné prohlášení

Prohlašuji, že jsem diplomovou práci na téma

„Statistická analýza podnikových dat“

vypracoval samostatně pod odborným dohledem vedoucího diplomové práce za použití pramenů uvedených v příložené bibliografii.

Plzeň dne 24. 4. 2023

v. r. Michal Elhota

Zásady pro vypracování práce

1. Formulujte výzkumné úkoly (otázky, hypotézy).
2. Zpracujte metodiku pro řešení formulovaných úkolů.
3. Získejte potřebná data.
4. Analyzujte data, vyhodnoťte výzkumné úkoly.
5. Formulujte závěry.

Poděkování

Touto cestou bych velice rád poděkoval vedoucímu mé diplomové práce Ing. Mgr. Milanu Svobodovi, Ph.D. za cenné rady, vstřícné jednání a mnoho poznatků, nezbytných pro vypracování této práce, a to nejen v rámci vedení této práce, ale i v rámci celého navazujícího studia.

Obsah

Úvod	6
1 Charakteristika problému daného podniku.....	8
2 Metodologie výzkumu.....	11
2.1 Dělení výzkumu	11
2.2 Kvalitativní výzkum.....	14
2.3 Kvantitativní výzkum.....	15
2.3.1 Vývoj statistiky a jejího významu	15
2.3.2 Charakteristické rysy	17
2.3.3 Výhody a nevýhody kvantitativního výzkumu.....	17
2.4 Metody sběru externích dat.....	18
2.4.1 Pozorování	18
2.4.2 Experiment.....	19
2.4.3 Měření.....	19
2.4.4 Dotazování	19
2.5 Kvalita měření.....	21
2.6 Fáze statistického výzkumu	22
2.6.1 Vymezení problému a volba přístupu.....	22
2.6.2 Plán výzkumu	23
2.6.3 Sběr dat	23
2.6.4 Zpracování dat, interpretace, zpráva o výzkumu	24
3 Metody statistického zpracování dat.....	25
3.1 Statistické usuzování (hypotézy).....	25
3.1.1 Základní pojmy z testování statistických hypotéz	25
3.1.2 Typologie statistických testů	30

3.2	Analýza závislostí	32
3.2.1	Regresní analýza	33
3.2.2	Korelační analýza	36
3.3	Analýza časových řad	38
4	Vybraný postup řešení problému	40
4.1	Sestavení hypotéz	41
4.2	Zpracování a analýza dat	41
4.3	Tvorba dotazníkového šetření	42
5	Analýza primárních interních dat	44
5.1	Analýza časových os a predikce	44
5.2	Hypotéza o nakupování voucherů	51
6	Analýza primárních externích dat	55
6.1	Hypotéza o tendenci šetřit	55
6.2	Hypotéza o nové úrovni	58
6.3	Hypotéza o marketingové kampani	59
6.4	Hypotéza o úrovni kvalit služeb	60
7	Výsledná interpretace	62
	Závěr	64
	Seznam použitých zdrojů	66
	Seznam tabulek	68
	Seznam obrázků	69
	Seznam příloh	70
	Přílohy	
	Abstrakt	
	Abstract	

Úvod

V dnešní době, kdy se firmy musí rychle přizpůsobovat neustále se měnícímu trhu a zvyšovat svou konkurenceschopnost, je důležité mít přesný a aktuální přehled o stavu podnikových procesů. K tomu se využívá statistiky neboli nauky, jak získat informace z numerických dat. Pomáhá přípravě a provedení výzkumu a při vyhodnocování získaných výsledků. Statistická analýza podnikových dat se stává stále důležitějším nástrojem pro řízení podnikových aktivit a rozhodování na základě datových zdrojů.

Diplomová práce se zaměřuje na využití statistických metod a technik pro analýzu a interpretaci podnikových dat, s cílem odhalit klíčové faktory ovlivňující výkon firmy a navrhnout strategie pro zlepšení podnikových procesů. Práce se také zabývá problematikou sběru a zpracování dat a popisuje konkrétní příklady použití statistických metod v praxi.

Práce je rozdělena do sedmi hlavních kapitol. První část je spíše věnovaná teorii řešené problematiky, poté na ni navazuje část praktická.

První kapitola představuje podnik, s jejíž spoluprací byla tato diplomová práce vytvořena. Čtenářům je vysvětlena nejen činnost podniku a podnikatelského prostředí, ale také problematika, se kterou se společnost snaží vypořádat. Taktéž je vysvětlena speciální situace, co se konkurence týče, a kvůli které zde nelze aplikovat klasické metody uvažování.

V druhé kapitole je vysvětlen pojem výzkum a jeho metodologie, především rozdílnosti mezi výzkumem kvalitativním a kvantitativním. Taktéž se zaměřuje na metodu sběru externích dat, dosažení jejich kvality a nastíní fáze statistického výzkumu, jež bude prováděn v praktické části diplomové práce. Třetí kapitola pojednává o třech hlavních možných metodách průzkumu statistické analýzy, jmenovitě statistické usuzování prostřednictvím hypotéz, analýza závislosti a analýza časových řad. Každá z těchto metod je podrobně rozepsána v jednotlivých podkapitolách, aby čtenář snadno rozpoznal, pro jakou skutečnost v praxi je vhodná daná metoda. Obě kapitoly se zaměřují na přehled možných metod a postupů řešení týkajících se problému společnosti popsané v první kapitole.

Zbylé čtyři kapitoly se věnují praktické části práce. Čtvrtá kapitola detailně popisuje vybrané metody k řešení stanoveného problému a důvody k jejich využití. Dojde ke stanovení výzkumných otázek a hypotéz.

Výzkum, který bude v rámci čtvrté kapitoly představen, je rozdělen na dvě části, kde ta první se zabývá analýzou primárních interních dat, čemuž je věnována pátá kapitola, a druhá část výzkumu analyzuje sesbíraná externí data v rámci kapitoly číslo šest.

Závěrečná kapitola poté dává zjištěné jevy do shrnutí a managementu společnosti předává své doporučení pro následné strategické kroky na základě těchto zjištění.

1 Charakteristika problému daného podniku

Tato kapitola blíže představí společnost ESCG games s.r.o., na kterou je zaměřena praktická část této diplomové práce, a také problematiku, kterou se autor snaží vyřešit.

Společnost byla založena Michalem Repelem v roce 2021 a její aktuální hlavní činností je provoz únikové hry Chernobyl v centru města Plzeň. Navazuje tak na činnost předchozí společnosti provozující tuto únikovou hru od roku 2017. Jedná se o typ zábavy, ve které účastníci neboli hráči vstupují do místnosti nebo budovy a jsou tam zamčeni. Mají poté určitý časový limit (obvykle 60 minut), během kterého musí najít cestu ven či splnit předem zadané úlohy, řešit různé hádanky a úkoly. Cílem hry je splnit požadovaný cíl (tj. dostat se ven z místnosti či splnit zadané úlohy) v rámci časového limitu. Únikové hry jsou oblíbené jako zábava pro skupiny přátel, narozeninové oslavy, tzv. rozlučky se svobodou nebo jako teambuildingová aktivita pro firemní akce.

ESCG games s.r.o. provozuje únikové hry s tematikou výbuchu Černobylské jaderné elektrárny v roce 1986 a nabízí autentickou atmosféru protiatomového krytu a silný příběh s promyšlenými hádankami. V současnosti nabízí dvě hry: Dozimetrická věž (obtížnost: normální), určená pro menší skupiny lidí, tudíž je vhodná pro začátečníky či nepříliš zkušené hráče, a poté Záchrannou misi, která se chlubí reputací jedné z nejtěžších, ne-li úplně nejtěžší únikové hry v západních Čechách. Jednodušší verze hry stojí 999 Kč, složitější vychází na 1 700 Kč. Na místě je taktéž možné zakoupit občerstvení (ESCG games s.r.o., 2023).

Úniková hra Chernobyl je v provozu již od roku 2017 a dle slov majitele si až do pandemické situace způsobené onemocněním COVID-19 udržovala rovnoměrnou a pravidelnou návštěvnost. Po zrušení protipandemických opatření, které zasáhly mnohé podniky poskytující zábavní služby, návštěvnost prudce vzrostla a po tomto vrcholu se opět ustálila na hodnotách před pandemií. Od září 2022 však začal přísun nových návštěvníků rapidním způsobem klesat a přes mnohé snahy a kroky, např. reklamní kampaň, slevové akce apod., se společnosti nedaří tuto hodnotu opětovně navýšit. Vedení společnosti přesně neví, jakým způsobem má problematiku řešit, neboť nezná její základní příčinu.

Důvodem neznalosti příčiny je fakt, že možností je vícero. Následující důvody byly jmenovány přímo majitelem společnosti ESCG games s.r.o.:

1. ekonomická krize z důvodu v té době stále rostoucí inflace a zdražování základních potravin, energií či pohonných hmot a s tím spojený pocit nejistoty a potřeby šetřit,
2. „vyčerpání“ obou her, znamená, že už byly navštíveny příliš vysokým množstvím hráčů, kteří již nemají důvod se vracet a hru opakovat, jelikož postup hrou se nemění,
3. nesprávně zvolená marketingová strategie (nedostatečný rozpočet, špatné nastavení klíčových slov, využívání nevhodných platforem apod.),
4. snížení kvality poskytovaných služeb, nespokojenost zákazníků, snižující se preference.

Uvedené důvody jsou poměrně specifické a nepočítá se s možností alternativního scénáře, jelikož prostředí společností provozující únikové hry je poměrně vysoce specifické a atypické. Nabízejí jednorázový produkt/službu, jenž pro zákazníka představuje unikátní zážitek, ale neměnný čili kdyby se zákazník rozhodl svůj zážitek zopakovat, dojde ke ztrátě unikátnosti a výrazného poklesu užitku. Do jisté míry se tento jev dá srovnat s filmovým a kinematografickým průmyslem, neboť valná většina lidí v kině zhlédne jeden konkrétní film za stejných podmínek (stejná volba jazyka, způsob projekce – 2D, 3D; apod.) pouze jednou. Podnikové prostředí je tedy omezeno především na získávání nových zákazníků, lákání minulých zákazníků na další produkty (v tomto případě únikové hry) nabízené společností (za předpokladu existence více než jedné hry), pronájem prostor, údržbu rekvizit, přístrojů apod. a zajištění dostatečné pracovní síly.

Z výše uvedených důvodů není mezi možné příčiny poklesu návštěvnosti uvedená konkurence, neboť dle veřejně dostupných informací v západních Čechách (a především v Plzni) nevznikly žádné nové únikové hry a marketingové oddělení nezaznamenalo žádné neobvyklé kroky (slevové akce, reklamní kampaně apod.) u největších konkurentů. Konkurenční prostředí únikových her je navíc taktéž poněkud neobvyklé, neboť jak již bylo výše zmíněno, návštěvníci ve valné většině případů nemají důvod únikové hry opakovat, neboť by se jednalo o opakování již prožitého zážitku, tudíž zde příliš nedochází k věrnosti zákazníka ke konkrétní značce, ba naopak příznivci této volnočasové aktivity navštěvují vysoké množství různých her od vysokého množství provozovatelů a tvůrců.

Management tak stojí před zásadním rozhodnutím, zdali investovat do masivnější a profesionální marketingové kampaně, do tvorby nových únikových her, které by mohly znovu přilákat jak nové, tak předchozí hráče, anebo nepodstupovat riziko, uchovat si rezervy a počkat, až se ekonomická a společenská situace stabilizuje. K řešení tohoto problému pomůže autorem navržený statistický výzkum a analýza dat.

2 Metodologie výzkumu

K vyřešení problémů společnosti ESCG games s.r.o. bude využito statistického výzkumu, pomocí kterého budou formulovány hypotézy, sesbírána potřebná data, jež budou následně analyzována. Výzkum sám o sobě znamená systematické zkoumání přírodních či sociálních jevů za účelem získání poznatků, jež vysvětlují a popisují svět kolem nás. Jedná se o systematický proces shromažďování údajů/dat, který problematizuje dosavadní znalosti, zahrnuje kritickou analýzu a výstupem musí být zvyšování znalostí. Také by měl být řízen dosavadními teoriemi a z nich odvozenými hypotézami (Hendl, 2006; Peck et al., 2015).

2.1 Dělení výzkumu

Výzkum lze rozdělit dle použité metodologie na kvalitativní a kvantitativní. Tyto přístupy zahrnují různé výzkumné metody, avšak každý z nich poskytuje odlišný pohled na celkový obraz zkoumané problematiky. Finn et al. (2000, s. 37) tvrdí: *„První se pokouší vysvětlit lidské chování skrz příčiny a následky, zatímco ten druhý má za cíl porozumět a interpretovat lidské aktivity skrze realitu každého individuálně.“*

Pojmy kvalitativní a kvantitativní výzkum úzce souvisí s pojmem znak, jenž představuje společnou vlastnost jednotek (v některých literaturách označované také jako statistické proměnné, anglicky „variable“) statistického souboru. Statistické jednotky jsou základními jednotkami, na které se vztahuje statistické pozorování. Příkladem mohou být osoby (např. zaměstnanci podniku při zkoumání mezd), organizace (např. podniky, provozy, obchodní jednotky apod. při zkoumání výše produkce), věci, události, zvířata apod. Vlastnosti statistických jednotek vyjadřují právě statistické znaky (v některých literaturách také jako vlastnosti, resp. hodnoty, anglicky „attributes“). Kupříkladu statistickou jednotku pracovníka průmyslového podniku lze charakterizovat např. znaky: mzda, stáří, platové zařazení, nejvyšší dosažené vzdělání, počet let praxe a jiné (Hindls a kol., 2006, str. 13-14;).

Tab. 1: Příklady statistických jednotek a znaků

Jednotka	Znak
Pohlaví	muž, žena
Věk	18, 26, 38, 41, 56, 78, ...
Vzdělání	různá kategorizace stupňů vzdělání – např. ZŠ, SŠ, VŠ
Volební preference	výčet politických stran, nebo rozdělení na pravici a levici
Postoj (k názoru)	zcela souhlasím – spíše souhlasím – neutrální postoj – spíše nesouhlasím – zcela nesouhlasím

Zdroj: Vondroušová (2019, s. 22)

Statistické znaky se dělí podle několika kritérií, zejména podle toho, jak lze vyjádřit jejich obměny. Různí odborníci využívají mnoha odlišných způsobů dělení typů statistických znaků, nejpodstatnější ale zůstává rozdělení na kvalitativní a kvantitativní, kdy kvalitativní znaky jsou vyjádřeny slovně (např. muž/žena, kuřák/nekuřák, typ zaměstnání apod.), zatímco kvantitativní jsou vyjádřené číselně. Kvantitativní znaky lze taktéž rozdělit na nespojité (diskrétní) a spojité, kdy nespojité statistické znaky nabývají pouze některých číselných hodnot, nejčastěji celých přirozených či nezáporných čísel (počet studentů ve třídě, počet zmetků ve výrobě, zatímco spojité znaky mohou nabývat libovolných hodnot v rámci určitého intervalu (spotřeba pohonných hmot, doba mezi jednotlivými výrobními intervaly) (Vondroušová, 2017; Zháněl a kol., 2014).

Kvalitativní výzkum v sociálních vědách se zaměřuje na to, jak jednotlivci a skupiny vnímají, chápou a interpretují svět. Podle jiných kritérií se výzkum považuje za kvalitativní, pokud nevyužívá statistických metod a technik. V tomto pojetí je kvalitativní výzkum v opozici k výzkumu kvantitativnímu. Kvantitativní výzkum je metoda vědeckého výzkumu, která se zaměřuje na popis jevů pomocí proměnných (znaků). Tyto proměnné jsou definovány tak, aby měřily určité vlastnosti a výsledky takového měření jsou následně zpracovány a interpretovány, například s pomocí statistických metod. Smíšený výzkum vychází z propojení přístupů kvalitativního a kvantitativního výzkumu (Zháněl, 2014).

Podrobnější charakteristiky kvalitativního, kvantitativního a smíšeného výzkumu jsou uvedeny v tabulce 2:

Tab. 2: Charakteristiky kvalitativního, kvantitativního a smíšeného výzkumu

	Kvalitativní výzkum	Smíšený výzkum	Kvantitativní výzkum
Vědecká metody	Induktivní, abduktivní, generování teorie zakotvené v datech	Deduktivní a induktivní	Deduktivní, testování hypotéz a teorie
Pohled na chování člověka	Chování je dynamické, situační, sociální, kontextuální	Chování je někdy predikované	Chování je určeno faktory a má predikovatelný charakter
Cíle výzkumu	Popis, explorace, objevování	Mnoho cílů	Popis, explanace a predikce
Zaměření	Široká a hloubková perspektiva zacílená na hloubku a šířku případu	Více perspektiv	Úzká perspektiva, testování hypotéz
Povaha pozorování	Zkoumání chování v přirozených každodenních podmínkách	Zkoumání chování ve více kontextech	Pokus zkoumat chování za kontrolovaných podmínek
Povaha reality	Subjektivní, osobní a sociálně konstruovaná (realismus)	Realismus a pragmatismus	Objektivní, různí pozorovatelé se shodnou na tom, co pozorují
Forma dat	Sběr kvalitativních dat výzkumníkem	Mnoho forem	Strukturovaná; validovaná data, objektivní instrumenty měření
Povaha dat	Slova, obrazy, kategorie	Směs proměnných, slov a obrazů	Proměnné
Analýza dat	Hledání vzorců, témat a historických vlastností	Kvantitativní/kvalitativní	Identifikace statistických vztahů
Výsledky	Pohledy účastníků, návrh teorie, zdůrazněná mnohost perspektiv	Směs různých výsledků	Zobecněné výsledky
Výzkumná zpráva	Narativní zpráva s kontextuálními popisy a přímými citacemi výpovědí.	Eklektická, pragmatická	Statistický zpráva, korelace, predikční rovnice

Zdroj: vlastní zpracování dle Hendl & Blahuš (2012)

2.2 Kvalitativní výzkum

Kvalitativní výzkum zahrnuje proces zkoumání jevů a problémů v autentickém prostředí s cílem získat komplexní obraz těchto jevů na základě hlubokých dat a specifického vztahu mezi badatelem a účastníkem výzkumu. Výzkumník provádějící kvalitativní výzkum se snaží rozkrýt a reprezentovat to, jak lidé chápou, prožívají a vytvářejí sociální realitu pomocí celé řady postupů a metod. Snaží se získat integrovaný pohled na předmět studie, na jeho kontextovou logiku a na explicitní a implicitní pravidla, která fungují v dané oblasti. Výzkumníci používají relativně málo standardizovaných metod získávání dat a využívají různé typy dat, jako jsou přepisy terénních poznámek z pozorování a rozhovorů, fotografie, audio a videozáznamy, deníky, osobní komentáře, poznámky, úřední dokumenty, úryvky z knih a všechno, co nám přibližuje všední život zkoumaných lidí. Účelem je izolovat určitá témata, projevy a datové konfigurace, ale obvykle je ponechávají co nejdéle v kontextu ostatních dat (Zháněl a kol., 2014).

Úkolem kvalitativního výzkumu je objasnit, jak se lidé v daném prostředí a situaci snaží porozumět tomu, co se děje, proč jednají určitým způsobem a jak organizují své každodenní aktivity a interakce. Výzkumník získaná data induktivně analyzuje a interpretuje. Nevytváří předem daný konečný tvar skládanek, ale spíše konstruuje obraz, který získává kontury během sběru a poznávání jednotlivých částí. Výzkumník se snaží podrobně popsat to, co pozoruje a zaznamenává, a nevynechává nic, co by mohlo pomoci vyjasnit situaci. Kvalitativní výzkum je v současnosti považován za samostatný druh výzkumu s vlastním zaměřením, metodami získávání a zpracování dat a metodami interpretace výsledků. Hlavní pole působnosti je zejména v společenských vědách (filozofie, sociologie, psychologie, pedagogika) a může se uplatnit i v kinantropologii. Zásadní výhodou kvalitativního výzkumu je získání hloubkového popisu případů v přirozeném prostředí. Cílem je získat popis zvláštností případů, generovat hypotézy a rozvíjet teorie o fenoménech světa. V poslední době se při řešení výzkumných projektů často využívá kombinace kvalitativního a kvantitativního přístupu (Hendl, 2006; Zháněl a kol., 2014).

Pro snazší pochopení a využití v praxi lze nalézt v tabulce č. 3 komparaci předností a nevýhod kvalitativního výzkumu:

Tab. 3: Výhody a nevýhody kvalitativního výzkumu

Výhody	Nevýhody
Podrobný popis jedince, fenoménu	Získané znalosti nemusí být zobecnitelné na populaci a na jiné prostředí
Zkoumání v přirozeném prostředí	
Možnost studování procesu	Obtížné kvantitativní predikce
Možnost navrhování teorií	Obtížné testování hypotéz a teorií
Reakce na lokální situace a podmínky	Sběr a analýza dat časově náročná
Hledání lokálních příčinných souvislostí	Výsledky snadno ovlivnitelné výzkumníkem
Pomoc při počáteční exploraci fenoménu	

Zdroj: vlastní zpracování dle Zháněl (2014)

2.3 Kvantitativní výzkum

Kvantitativní výzkum, který spadá do objektivního paradigmatu, se zaměřuje na sběr dat, která jsou následně analyzována pomocí různých statistických technik. Jeho cílem je získat relativně malé množství informací od velkého počtu respondentů. Využívá přesného procesu hypotetické formulace, detašovaného pozorování, sběru dat, analýzy dat a následné akceptace nebo odmítnutí hypotézy. Experimentální výzkum, který je jedním z typů kvantitativního výzkumu, spočívá v manipulaci s vybraným vzorkem a následném měření dopadů (tato metoda je často využívána v přírodních vědách a medicíně, kde se jedná o tzv. klinický výzkum). Druhým typem kvantitativního výzkumu je šetření (observace), při kterém výzkumník aktivně neovlivňuje realitu, pouze ji pozoruje (např. statistické šetření, v medicíně tzv. klinické studie) (Weaver a Lawton, 2002).

2.3.1 Vývoj statistiky a jejího významu

Kvantitativní výzkum se silně opírá o statistiku a statistické zpracování dat. Dá se definovat jako nauka, jak získat informace z numerických dat. Statistika má své kořeny v dávných civilizacích, kdy se prováděly soupisy obyvatelstva, hlavně pro účely daní. Avšak první skutečné statistické analýzy událostí, které se odehrávaly ve společnosti, začaly vznikat až v 17. století. Slovo "statistika" jako název pro samostatný vědní obor se objevilo v 18. století. Původ tohoto slova se dá vystopovat k latinskému "status", což

znamená stav, ze kterého bylo odvozeno italské slovo "stato", neboli stát. V Itálii se z tohoto výrazu odvozoval další pojem "státistico", ze kterého později vznikl termín "statistika" (Cyhelský a kol., 2001; Hindls a kol., 2006, s. 12).

V dalším vývoji této disciplíny v 19. a 20. století se již využívaly základy teorie pravděpodobnosti. Statistika se stala samostatným vědním oborem již od začátku 20. století a rychle se rozvíjela. Nicméně od 70. let minulého století se její vývoj ještě více urychlil, což bylo způsobeno rozmachem výpočetní techniky a vznikem statistických softwarových nástrojů, bez nichž by dnes moderní statistické analýzy nebyly možné. Náročné potřeby podnikání v oblasti extrakce informací z dat, které byly ihned využitelné ke zvýšení příjmů, vyžadovaly nové analytické techniky, které umožnily analýzu nelineárních vztahů v obrovských souborech dat s neznámým rozložením (Hindls a kol., 2006, s. 12; Nisbet et al., 2009, s. 11).

V současnosti již neexistuje vědní obor, jež by nepoužíval statistické metody pro vyhodnocení velkých objemů dat. Mezi obory, které tyto metody běžně využívají, patří medicína, fyzika, biologie a další přírodní i technické obory. Významnou roli hraje statistika v oblasti sociálně-ekonomické analýzy. Její použití v ekonomické oblasti je tak rozsáhlé, že existuje mnoho specializovaných oblastí statistiky, které byly vytvořeny na základě potřeb ekonomické teorie a praxe (Hindsl, 2006).

Slovo statistika má různé významy. Může označovat vyplněné statistické formuláře nebo číselné údaje, které se objevují v médiích. Také může znamenat organizaci sběru a zpracování statistických údajů nebo různé metody, které se v této oblasti používají. Statistiku lze tedy chápat jako číselné údaje, praktickou činnost nebo teoretickou disciplínu, která se zabývá metodami popisu a odhalování zákonitostí hromadných jevů, tzn. jevů vyskytují se v masovém měřítku u velkého počtu jedinců. Nicméně, číselné údaje o jednotlivých skutečnostech nebo o několika málo jednotlivostech nejsou dostatečné pro vytvoření statistiky. K získání informací a formulování závěrů o zkoumaných jevech je potřeba sbírat pozorování hromadného jevu (Hindls a kol., 2006).

Statistika hraje významnou roli v současném světě a ekonomice, zejména při správném a cílevědomém chodu ekonomiky s cílem maximalizace její efektivnosti. Kvalitní informační systém je zásadní pro úspěšné řízení ekonomiky, a zde sehraává statistika mimořádně významnou roli. Statistika poskytuje číselné informace o celém národním

hospodářství i jeho subsystémech a je klíčovým nástrojem při realizaci změn v ekonomice. Schopnost porozumět statistice je nezbytná nejen pro statistiky, ale i pro ekonomy, kteří se jí v procesu tvorby manažerských rozhodnutí, analýz trhu, řízení jakosti a dalších oblastech neustále opírají. Vývoj statistiky a statistických metod probíhá neustále i v současnosti, kdy se díky hromadění stále více dat objevují stále chytřejší způsoby, jak co nejvíce simulovat činnost nejkompexnějšího učícího stroje ve vesmíru – lidského mozku (Hindls a kol., 2006; Nisbet et al., 2009, s. 13).

2.3.2 Charakteristické rysy

Podle Pavlici a kol. (2000, s. 27-28) jsou charakteristickými rysy kvantitativního výzkumu v oblasti společenských věd následující:

- Nezávislost – výzkumník není ovlivněn zkoumanými jevy.
- Hodnotová svoboda a autonomie vědy – výběr a volba témat a metod výzkumu by měly být určovány objektivními kritérii, jako jsou výsledky předchozích studií nebo identifikace problémů.
- Kauzalita – cílem výzkumu je identifikovat kauzální vztahy a zákonitosti, které vysvětlují pravidelnost lidského chování.
- Hypoteticko-deduktivní přístup – věda se vyvíjí formulací a testováním hypotéz o obecných pravidlech a zákonitostech.
- Operacionalizace – vědecké pojmy jsou převedeny na konkrétní projevy a fakta, aby bylo umožněno kvantitativní měření skutečností, které charakterizují danou oblast.
- Redukcionismus – pro lepší porozumění problémům je výhodné je redukovat na jednodušší prvky (nicméně ne všichni pozitivisté s tímto předpokladem souhlasí).
- Generalizace – možnost zobecnění zjištěných zákonitostí lidského a sociálního chování musí být zajištěna dostatečně velkými vzorky zkoumaných osob.
- Průřezová analýza – obecná pravidla a zákonitosti lze nejlépe identifikovat porovnáním variací napříč různými vzorky.

2.3.3 Výhody a nevýhody kvantitativního výzkumu

Stejně jako u kvalitativního výzkumu lze porovnat výhody a nevýhody výzkumu kvantitativního, a to v tabulce č. 4:

Tab. 4: Výhody a nevýhody kvalitativního výzkumu

Výhody	Nevýhody
Testování a validizace teorií	Kategorie a teorie nemusí odpovídat lokálním zvláštnostem
Možnost zevšeobecnit populaci	
Možnost eliminace rušivých vlivů	Soustředění jen na určitou teorii a testování
Rychlý a přímočarý sběr dat	
Přesná numerická data	Nerozvíjí teorii
Rychlá analýza dat (využití softwaru)	Riziko abstraktní a všeobecné znalosti
Výsledky relativně nezávislé na výzkumníkovi	Výsledky a znalosti těžko aplikovatelné v lokálních podmínkách
Užitečnost při zkoumání velkých skupin	Omezenost způsobem získávání dat

Zdroj: vlastní zpracování dle Zháněl (2014)

2.4 Metody sběru externích dat

Primární data lze získat pozorováním (měřením), experimentem či dotazováním. Každá z těchto možností představuje určité výhody i nevýhody a výběr závisí především na podstatě zkoumaného problému.

2.4.1 Pozorování

Pozorování je nezbytnou součástí většiny kvalitativních výzkumných prací, kde je jednání a chování lidí pozorováno přirozeným způsobem. Na rozdíl od rozhovorů, kde jsou zahrnuty názory respondentů, je pozorování zaměřeno na skutečné projevy chování a jednání. Výzkumné pozorování se může vyskytovat v různých formách, jako například skryté nebo otevřené, zúčastněné nebo nezúčastněné, strukturované nebo nestrukturované, v umělé nebo přirozené situaci a může být zaměřeno na pozorování sebe sama nebo někoho jiného. Výhodou jsou objektivní a ověřitelné údaje. Z kvantitativního hlediska lze pozorovat počet vyrobených kusů za časovou hodinu apod. (Zháněl a kol., 2014).

2.4.2 Experiment

Výzkumný experiment je speciální typ pozorování, kde výzkumník záměrně mění podmínky a sleduje, jak se to projeví u jedné nebo více skupin jedinců nebo jiných subjektů. Tyto změny mohou být dosaženy pomocí intervencí nebo ošetření. Přirozené experimenty se naopak zaměřují na nezáměrné změny podmínek. V případě porovnávání rozdílností mezi skupinami se jedná o komparativní experiment. Zkoumá se, zda rozdíly mezi skupinami jsou náhodné nebo způsobené nezávislými proměnnými. Je důležité si uvědomit, že výzkumník sám zařazuje jednotky do různých skupin a volí hodnoty nezávisle proměnných. V případě výběrového šetření nebo jiné observační studie jsou hodnoceni jedinci ti, kteří si sami volí, do jakých skupin se zařadí a výzkumník nemá na tuto volbu vliv. Pokud je například zkoumán vliv kouření na zdraví, chtěl by výzkumník přiřadit některé nekuřáky do skupiny kuřáků, ale z etických důvodů to nemůže udělat. V takových případech se musíme spokojit s observačními studii (Hendl, 2006, s. 59-61).

2.4.3 Měření

Měření je metodou, kterou kvantitativně srovnáváme vlastnosti podobných objektů. Jedná se o specifický statistický proces, který zahrnuje tři základní postupy:

- Počítání: Jedná se o určování jednoznačných vztahů mezi zkoumanými prvky a procesy pomocí pořadového určení.
- Škálování: Tento postup zahrnuje uspořádání zkoumaných prvků, stavů a vztahů do skupin a tříd na základě určeného kritéria či znaku. Třídění vyjadřuje stupeň rozvoje daného znaku, nikoliv však jeho velikost.
- Vlastní měření: Tato fáze zahrnuje určování parametrů sledovaných prvků a procesů na základě standardních jednotek měření pomocí měřících prostředků a přístrojů (Zháněl a kol., 2014.)

2.4.4 Dotazování

Kvantitativní metoda výzkumu, která se používá především v oblasti společenských věd, jako je psychologie, sociologie, demografie a marketing. Provedení dotazování zahrnuje použití nástrojů, jako jsou dotazníky a záznamové archy, a vhodně zvolené komunikace výzkumníka s informačním nositelem, aby řešitel výzkumného projektu získal žádoucí primární údaje. Metoda dotazování se dělí podle způsobu kontaktování respondenta na:

- osobní dotazování,
- písemné dotazování,
- internetové (online) dotazování,
- telefonické dotazování (Walker, 2013).

Dotazník je hlavním nástrojem pro sběr kvantitativních primárních dat. Umožňuje sbírat kvantitativní data standardizovaným způsobem, aby byla data interně konzistentní a koherentní pro analýzu. Dotazníky by vždy měly mít určitý účel, který souvisí s cíli výzkumu, a je třeba již zpočátku jasně stanovit, jak budou zjištění použita. Používá se, pokud jsou zdroje omezené, protože náklady na návrh a administraci dotazníku mohou být poměrně nízké a čas je důležitým zdrojem, který dotazník spotřebuje do svého maximálního rozsahu. Taktéž je nutné myslet na ochranu soukromí účastníků, neboť účastníci budou reagovat upřímně pouze tehdy, pokud je skryta jejich identita a zachována důvěrnost. Dále se používají pro potvrzení s dalšími zjištěními, jelikož dotazníky mohou být užitečnými nástroji pro ověření, když jsou potvrzeny jinými studii, které mají zdroje pro provádění jiných strategií sběru dat (Roopa & Rani, 2012).

Dotazníky se dělí na strukturované (s uzavřenými otázkami) a nestrukturované (s otevřenými otázkami). Základními fázemi dotazování jsou úvod, střední část a závěr. Úvod slouží k vysvětlení cíle a způsobu práce, střední část se skládá z postupného zadávání otázek a závěr umožňuje respondentům vyjádřit svůj názor a případně se dále kontaktovat. Při tvorbě dotazníku se často využívá škálování, které umožňuje měřit míru intenzity sledovaného jevu. Validita škál je důležitá, aby se zajišťovalo, že měří, co tvrdí, že měří. Pro kontrolu validity se používají různé nástroje, jako je zdánlivá validita, prediktivní validita a měřítko validity. Spolehlivost (reliabilita) škál a celého dotazníku je také důležitá, aby se zajišťovala stálost a přesnost opakovaně zjišťovaných výsledků (Walker, 2013).

Příkladem škálování je Likertova škála sloužící pro zjištění a síly postojů v dotaznících a také hojně využívaná v psychologických testech. Respondentům je položen určitý výrok, např: „Jsem spokojen s dostupností zubní péče ve svém městě“ a je jim nabídnuto určité množství odpovědí (běžně 5 či 7) na škále „nesouhlasím“ po „souhlasím“. V případě 5 možných odpovědí mohou být škály nastavené jako: 1 – silně nesouhlasím; 2 – nesouhlasím; 3 – nevím; 4 – souhlasím; 5 – silně souhlasím (Roopa & Rani, 2012).

2.5 Kvalita měření

O úspěšnosti výzkumu rozhoduje etapa pořizování údajů. Pokud již v tomto kroku dojde k chybám, nelze jakýmkoliv způsobem zachránit výzkum skrze zpracování dat. Ze špatných dat nelze získat platné závěry. Z tohoto důvodu je nutné dodržovat tři základní pravidla kvality měření, kterými jsou objektivita, spolehlivost (reliabilita) a validita (Odom & Morrow, 2009).

Měřicí **objektivita** se týká míry, do jaké jsou výsledky nezávislé na zkoumajícím nebo jedinci, který je měřen, vzhledem k subjektivnímu zaujetí nebo náhodné chybě v měření. Při měření fyzikálních veličin v laboratoři se otázka objektivitativy zpravidla nevyskytuje, nicméně v sociologii a psychologii musí být objektivita pečlivě zvažována při hodnocení měření (Hendl, 2006, s. 47).

Spolehlivost (reliabilita) měření se týká konzistence výsledků měření jednoho subjektu nebo objektu při opakovaném měření za stejných podmínek. U testů s mnoha položkami se zkoumá konzistence odpovědí různých podmnožin položek mezi sebou. Například, pokud jsou dvě sady skóre měřeny na kontinuální stupnici, tak pro zjištění spolehlivosti a konzistence měření bude použit koeficient korelace Pearsona, nazývaný též koeficient spolehlivosti mezi třídami. Nespolehlivost (nízká reliabilita) měření může mít různé příčiny. Jedním zdrojem nespolehlivosti je subjektivní chyba, kterou způsobuje individuální variabilita měřeného subjektu, jako je únava, klesající zájem atd. Pozorovací chyba je dalším zdrojem chyb a závisí na provedení měření hodnotitelem. Mezi další příčiny nespolehlivosti patří například chyby přístroje (např. selhání hardwaru). Existuje mnoho postupů pro určení spolehlivosti měření, například opakované měření (test-retest reliabilita) pro posouzení konzistence opakovaných měření oddělených určitým časovým intervalem, měření paralelních testů pro posouzení konzistence mezi dvěma ekvivalentními měřeními stejného konstrukt (například verze A i B téhož testu), nebo půlení testu (split-half reliabilita) pro posouzení konzistence různých částí jednoho testu. V posledním případě se jedná o metodu posuzování interní konzistence, která nevyžaduje opakované použití měřicí procedury u jedince (Hendl, 2006, s. 48; Odom & Morrow, 2009).

Validita nám zjednodušeně odpovídá na otázku, zda je zkoumáno to, co opravdu výzkumník chce, neboť se týká pravdivosti měření. Existují tři běžné metody určování validity. Jedna metoda je založena na logickém hodnocení odborníků na obsah (obsahová

validita), druhá na empirických důkazech ze statistické analýzy (kriteriální validita) a třetí na složité řadě testů hypotéz a statistických analýz (konstruktivní validita). Druhá metoda je zaměřena na tento diskusní příspěvek a podobně jako spolehlivost se zakládá na korelačním koeficientu Pearsonova typu. Kriteriální validita vyžaduje kritérium, "pravdivé" měření výkonu na nějaké proměnné, a skóre z nástroje, pro který je validace požadována (Odom & Morrow, 2009).

2.6 Fáze statistického výzkumu

Hindls a kol. (2006) tvrdí, že statistický proces v praxi zahrnuje několik etap. Obvykle začíná statistickým šetřením, tedy sběrem dat, následovaným jejich statistickým zpracováním, a nakonec jejich analýzou. Ačkoliv by se mohlo zdát, že nejdůležitější je poslední etapa, ve skutečnosti jsou všechny tři úkoly rovnocenně důležité. Data, která nebyla zpracována a analyzována, mají malou hodnotu. Navíc, použití sofistikovaných statistických metod nepřinese smysluplné informace, pokud jsou data špatně shromážděna.

Etapy a prvky výzkumu se dají rozdělit na následující kroky:

- vymezení problému a volba přístupu – volba oblasti výzkumu, formulace hypotéz,
- plán výzkumu – určení předmětu a způsobu analýzy,
- provedení – sběr dat,
- zpracování dat a analýza – explorace dat,
- interpretace výsledků,
- zpráva o výzkumu – publikování výsledků výzkumu (Hendl, 2006; Peck et al., 2015).

2.6.1 Vymezení problému a volba přístupu

V rámci vymezení problému a volby přístupu je nutné vycházet z určitého **problému**, který je pro zkoumanou oblast závažný, ať už se jedná o zkušenost v předmětné oblasti či o nějakou nevyjasněnou, spornou, neřešenou situaci. Taktéž je důležité provést podrobnou **rešerši** a vyhodnotit relevantní literaturu. Při rozhodování o **metodologii** je nutné zvážit kvantitativní a kvalitativní přístup, které jsou obvykle používány v sociálním výzkumu. Tuto volbu je nutné přesně specifikovat výzkumnou otázkou a hypotézami. Formulace **výzkumné otázky** je klíčová pro definování tématu, na které se výzkum chce zaměřit a získat odpověď. Například: „Jaký je vztah mezi procesy myšlení učitele během

plánování a dobou, kterou žák tráví při provádění přímých učebních aktivit?“ **Hypotéza** je tvrzení o podstatě určité situace ve zkoumané oblasti, které se ověřuje na základě empirických dat. Hypotézy mohou být navrženy na základě předběžného výzkumu a pilotní studie. Například:

- Počet naplánovaných pohybových úkolů k procvičení je pozitivně korelován s aktivní učební dobou.
- Počet rozhodnutí o řízení výuky v plánu výuky má pozitivní vztah k aktivní učební době.
- Celkový počet rozhodnutí v plánu výuky nemá vztah k aktivní učební době (Hendl, 2006; Peck et al., 2015).

2.6.2 Plán výzkumu

V etapě plánování se provádí konkrétní specifikace jednotlivých volby a aspektů výzkumu. Navrhnu se způsoby výběru zkoumaných objektů a typ experimentu nebo observační studie. Dále se upřesní volby metod pro získání dat a určí jednotlivé kroky, jak tyto metody využít. Vyjasní se administrativní předpoklady výzkumného projektu. Také je nutné se zabývat volbou metod pro analýzu dat. V této fázi je třeba identifikovat a naplánovat vše, co se bude ve studii provádět. V mnoha případech je plán popsán v projektu výzkumu, který student nebo výzkumník obhajuje před odbornou komisí školy nebo grantovou agenturou (Hendl, 2006, s. 21).

2.6.3 Sběr dat

Ve fázi sběru dat se provádí statistické zkoumání, kdy jsou získávány hodnoty sledovaných statistických charakteristik v podobě numerických nebo slovních dat. Existují dva způsoby, jak získat tato data. Pokud jsou již někde k dispozici, nazývají se **sekundární data**. Pokud jsou shromažďována přímo výzkumníkem, jedná se o **primární data**. Bez ohledu na zdroj dat platí, že pokud jsou data nízké kvality, bude i kvalita informací získaných z jejich statistické analýzy nízká (Cyhelský a kol, 2001; Hendl, 2006).

Primární data jsou obvykle shromažďována pro určitý účel a dělí se na interní a externí. V rámci podnikové ekonomiky jsou interní data generována v průběhu běžných podnikových aktivit a jsou proto relativně snadno dostupná z firemních databází a relativně nenákladná na shromažďování. Externí data pocházejí z externích zdrojů

mimo podnik, jako jsou jiné podniky, státní instituce, univerzity a výzkumné ústavy. Mohou být placená nebo zdarma, v závislosti na zdroji. Při využívání sekundárních dat je důležité pečlivě zvážit, jak byla tato data získána a pro jaký účel byla použita, neboť kontext je nejdůležitějším aspektem dat, nicméně je často opomíjen (Wegner, 2013).

2.6.4 Zpracování dat, interpretace, zpráva o výzkumu

Během procesu sběru dat se provádí průběžná analýza (např. pro kontrolu dat), nicméně hlavní zpracování dat se uskutečňuje až po dokončení experimentu, dotazování nebo pozorování. Nejprve se provede kontrola dat a vyhledávání odlehlých hodnot. Poté se vypočítají popisné charakteristiky a aplikují se statistické testy. Dále se upravuje model dat a navrhují se vhodné metody grafické a numerické analýzy. Výsledky se interpretují s ohledem na formulované hypotézy, vztahují se k výsledkům předchozích výzkumů a posuzují se v kontextu teorií. Cílem je najít odpovědi na otázky, které byly formulovány v prvním kroku a často se objevují nové otázky, které vedou k dalším cyklům výzkumu (Hendl, 2006; Peck et al., 2015).

Výsledky, závěry a doporučení získané v předchozích krocích se publikují v nějaké formě, například jako příspěvek na konferenci, článek v odborném časopise, diplomová nebo disertační práce nebo kniha. Sestavení publikace je poslední intelektuální aktivitou v celém procesu (Hendl, 2006, s. 22).

3 Metody statistického zpracování dat

Jelikož je praktická část diplomové práce zaměřena především na statistickou analýzu a zpracování získaných a sesbíraných podnikových dat, je úkolem následující kapitoly přiblížit tři hlavní metody zpracovávání dat. Nejprve budou vysvětleny principy statistických hypotéz, dojde k seznámení s veškerými základními pojmy potřebné k uchopení této tematiky a k uvedení jejich typologie. Dále se kapitola zaměří na tematiku regresní a korelační analýzy. V závěru je představena analýza speciálního případu regresního modelu, a to časových řad.

3.1 Statistické usuzování (hypotézy)

Pomocí statistiky se dají řešit rozhodovací situace na základě náhodného výběru. Na začátku je tvrzení o určité vlastnosti pravděpodobnostního rozdělení náhodné veličiny, které sledujeme. Úkolem výzkumníka je posoudit, zda je toto tvrzení správné s přijatelnou spolehlivostí a jistou přesností. Příkladem takovéhoto tvrzení může být, zdali nová strategie, investice do marketingu, nově navázaná spoluprací apod. vedly ke změně sledovaných parametrů výrobku, obratu nebo tržeb. Taktéž lze tímto způsobem otestovat, zda se preferovaná značka automobilu liší podle nejvyššího dosaženého vzdělání kupujícího. Proces ověřování správnosti těchto tvrzení se nazývá testování hypotéz.

3.1.1 Základní pojmy z testování statistických hypotéz

V rámci této podkapitoly jsou vysvětleny následující pojmy: statistická hypotézy, chyby I. a II. druhu, testové kritérium, kritický obor, p-hodnota a další s nimi související.

- **Statistická hypotéza**

Definována jako předpoklad o parametrech nebo tvaru rozdělení zkoumaného znaku. Předpoklad, že průměr základního souboru μ se rovná konkrétní hodnotě μ_0 , vysloví hypotézu o průměru základního souboru. Vyčerpávající šetření celého základního souboru je obvykle neekonomické nebo dokonce technicky neproveditelné, proto se obvykle používá výběrový soubor. Pomocí výsledků získaných náhodným výběrem se poté rozhoduje o správnosti nebo nesprávnosti hypotézy (Hindls a kol, 2006, s. 133).

Předpoklad vyslovený o určité charakteristice či tvaru rozdělení se nazývá **nulová hypotéza**, značená H_0 . Poté je definována další hypotéza, nazvaná **alternativní hypotéza**, která určitým způsobem popírá to, co je uvedeno v hypotéze nulové. Alternativní hypotézu označujeme jako H_A nebo H_1 .

V případě předpokladu ohledně průměru základního souboru se nulová hypotéza zapíše jako:

$$H_0: \mu = \mu_0$$

Proti ní může být stanovena hypotéza alternativní v následujících podobách:

- a) $H_A: \mu \neq \mu_0$
- b) $H_A: \mu < \mu_0$
- c) $H_A: \mu > \mu_0$

V případě a) je alternativní hypotéza zformulována tak, že popírá platnost hypotézy H_0 a tvrdí, že hodnota parametru μ se liší od hodnoty udávané H_0 . Tento typ hypotézy se nazývá dvoustranná nebo **oboustranná hypotéza** a testování takové hypotézy se označuje jako oboustranný test (anglicky two-tailed test). V případech b) a c) alternativní hypotéza popírá platnost nulové hypotézy H_0 a zároveň specifikuje jednostranný obor hodnot parametru v základním souboru. V případě b) tvrdí, že hodnota parametru je menší než hodnota daná hypotézou H_0 , v případě c) je tomu naopak. Tyto typy hypotéz se nazývají **jednostranné hypotézy** a testování se označuje jako jednostranný test (anglicky one-tailed test), případ b) jako levostranný test (anglicky lower tail test) a případ c) jako pravostranný test (anglicky upper tail test) (Anderson et al., 2017, s. 393-399; Hindls a kol, 2006, s. 134). Jako mnemotechnická pomůcka může posloužit dokreslení vodorovné čáry do středu symbolu „je menší než“ či „je větší než“, čímž vznikne šipka určující směr pro levostranný test (\leftarrow) a pravostranný test (\rightarrow).

Správná formulace nulové a alternativní hypotézy nemusí být vždy zřejmá. Je třeba pečlivě strukturovat hypotézy tak, aby závěr testování hypotéz poskytoval požadované informace pro výzkumníka nebo rozhodovatele. Kontext situace je velmi důležitý při stanovování, jak by měly být hypotézy formulovány. Všechny aplikace testování hypotéz zahrnují sběr vzorku a použití výsledků vzorku k poskytnutí důkazů pro vyvození závěru. Užitečné otázky při formulaci nulové a alternativní hypotézy jsou: Jaký je účel sběru vzorku? Jaké závěry se snažíme vyvodit? Mnoho aplikací testování hypotéz se zabývá pokusem získat důkazy podporující výzkumnou hypotézu. V těchto situacích je

nejlepší začít s alternativní hypotézou a udělat ji závěrem, který výzkumník doufá podpořit (Anderson et al., 2017, s. 386-387; Hindls a kol, 2006, s. 135-137).

Závěrem statistické hypotézy je totiž vždy rozhodnutí, zdali:

- byla nulová hypotéza zamítnuta, což potvrdí hypotézu alternativní,
- byla nulová hypotéza nezamítnuta, čímž nebyla alternativní hypotéza potvrzena

Při testování hypotéz může dojít k mnoha chybám, v literaturách jsou často vyzdvihovány tři konkrétní.

- **Chybné závěry hypotéz**

První chybou, prováděnou především studenty či nezkušenými výzkumníky, je při nezamítnutí nulové hypotézy prohlásit, že tím byla buď nulová hypotéza potvrzena či byla alternativní hypotéza vyvrácena. Je nutné dbát zřetel na to, že nezamítnutí H_0 pouze znamená nepotvrzení H_A a nedělat žádné další závěry. Další chyba je označována jako **chyba I. druhu** (anglicky type I error), kdy dojde k zamítnutí nulové hypotézy H_0 , ačkoliv ve skutečnosti platí. Pravděpodobnost této chyby se označuje α a nazývá se **hladinou významnosti** (anglicky significance level). Hodnotu α si výzkumník stanovuje dle vlastního uvážení, nejčastější užívanými je 0,1 (10 %); 0,05 (5 %) a 0,01 (1 %). Nezamítnutí nulové hypotézy H_0 i přes to, že je chybná, se označuje jako **chyba II. druhu** (anglicky type II error), jejíž pravděpodobnost se značí písmenem β . Pravděpodobnost $1 - \beta$ se nazývá **síla testu** (anglicky power of the test) a vyjadřuje pravděpodobnost zamítnutí nulové hypotézy H_0 . (Anderson et al, 2017, s. 390-391; Rice, 2006, s. 331).

Vztah mezi chybami obou druhů je vyjádřen v tabulce č. 4:

Tab. 5: Chyby I. a II. druhu a jejich pravděpodobnosti

Skutečnost Úsudek o H_0	H_0 je pravdivá	H_0 je nepravdivá
Nezamítá se	správné rozhodnutí = $1 - \alpha$	chyba II. druhu = β
Zamítá se	chyba I. druhu = α	správné rozhodnutí = $1 - \beta$

Zdroj: vlastní zpracování dle Hindls a kol. (2006, s. 134)

- **Testové kritérium**

Dle Hendla a kol. (2006) je po stanovení hypotéz a hladiny významnosti potřeba zpracovat nasbíraná data a spočítat z nich hodnotu zvanou testové kritérium či testová statistika (anglicky test statistic), v českých literaturách obvykle označováno písmenem T . Existuje mnoho typů testovacích statistik, jejichž výpočet závisí na povaze dat a testované hypotéze. Například pro testování průměru, relativních četností a v mnoha dalších případech se používá jako testovací statistika standardizovaná vzdálenost odhadu od nulové hypotézy H_0 . Obecný tvar této testovací statistiky je definován následovně:

$$T = \frac{\text{bodový odhad} - \text{hypotetická hodnota}}{\text{směrodatná odchylka odhadu}}$$

- **Kritický obor**

Se spočítaným testovým kritériem lze formulovat závěr testování. Jedním ze dvou způsobů je určení tzv. kritické hodnoty pro dané testovací kritérium, které může dosahovat libovolné hodnoty z dané množiny hodnot S (odvozené dle typu rozdělení testového kritéria, např. normovaného normálního). Tento výběrový prostor lze rozdělit na dva podprostory:

- podprostor V obsahující hodnoty nezamítající H_0 , tzv. obor přijetí,
- podprostor W obsahující hodnoty zamítající H_0 , tzv. kritický obor.

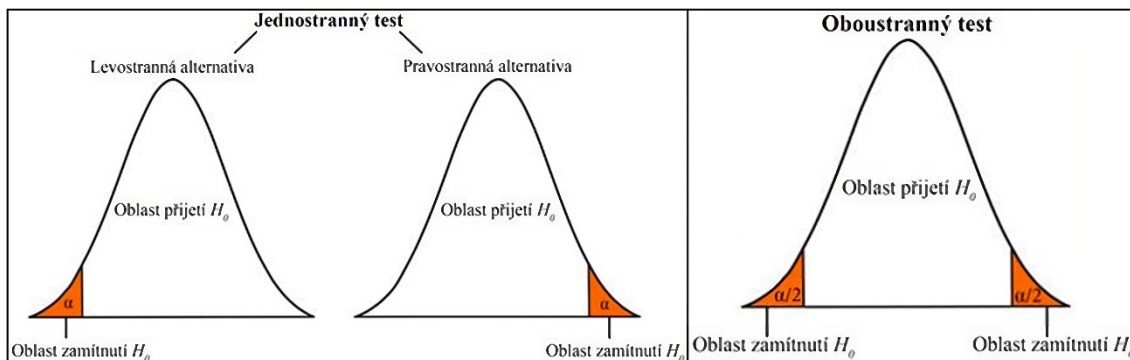
Velikost kritického oboru zajišťuje, že chyby I. druhu bude dopuštěno pouze v 100α procentech případů, tj. že pravděpodobnost hodnoty testové statistiky ležící v kritickém oboru je rovna předem stanovené hladině významnosti α . Hranici oddělující kritický obor od oboru přijetí se nazývá kritická hodnota, tzn. v případě pravostranného testu slouží kritická hodnota jako hraniční hodnota pro určení toho, zda je hodnota testovací statistiky dostatečně velká na to, aby byla zamítnuta nulová hypotéza (Anderson et al, 2017; Hindls a kol, 2006, s. 135-136).

Kritické hodnoty se spočítají jakožto kvantily rozdělení testového kritéria při platnosti H_0 , konkrétně:

- levostranný test: $W = (T_{min}; T_{\alpha})$,
- pravostranný test: $W = (T_{1-\alpha}; T_{max})$,
- oboustranný test: $W = (T_{min}; T_{\alpha/2}) \cup (T_{1-\alpha/2}; T_{max})$ (Hindls a kol, 2006).

Pro lepší vizualizaci kritického oboru lze nahlédnout na obrázek č. 9.

Obr. 1: Vizualizace kritického oboru



Zdroj: Drozdová & Homola (2017)

- **p-hodnota**

Druhým způsobem formulace závěru testování je hodnota významnosti p čili pravděpodobnost, že výzkumníkem vypočítaná testovací statistika dosáhne právě takové nebo ještě extrémnější hodnoty, pokud platí nulová hypotéza. Malá p -hodnota znamená, že byla získána hodnota testovací statistiky, která je neobvyklá, pokud nulová hypotéza platí. Konkrétněji nízká p -hodnota svědčí proti nulové hypotéze, naopak vysoká p -hodnota připouští (nicméně nepotvrzuje!) její platnost. Interpretace statistického testu pomocí p -hodnoty zahrnuje porovnání p -hodnoty s hladinou významnosti:

- $p\text{-hodnota} \leq \alpha \dots H_0$ zamítnuta; potvrzena H_A ,
- $p\text{-hodnota} > \alpha \dots H_0$ nezamítnuta.

Protože p -hodnota závisí na hladině významnosti, je důležité hodnotu α stanovit předem. Čím nižší hladina významnosti, tím bude těžší zamítnout nulovou hypotézu, ale výsledky testu budou více spolehlivé. Stejně tak vyšší hladina významnosti znamená snazší zamítnutí nulové hypotézy za cenu nižší spolehlivosti. Vyjma statistického software se dá p -hodnota vypočítat za pomoci distribuční funkce $F(U)$ pro danou testovou statistiku v rámci množiny rozdělení testové statistiky, konkrétně:

- levostranný test: $F(U)$,
- pravostranný test: $1 - F(U)$,
- oboustranný test: $2\min(F(U), 1 - F(U))$.

Pro některé testy může být výpočet pro levostranný a pravostranný test obrácený, např. u kvantilového testu (Hendl, 2006, s. 177-178; Rice, 2007, s. 335).

Shrnutí: obecně lze kroky při testování hypotézy definovat jako:

1. formulace výzkumné otázky ve formě nulové a alternativní hypotézy,
2. zvolení přijatelné hladiny významnosti α ,
3. výpočet testové statistiky,
4. porovnání testové statistiky s kritickým oborem či vypočtenou p-hodnotou,
5. formulace závěru, doporučení.

3.1.2 Typologie statistických testů

Detailní a rozsáhlý popis postupu zpracování všech možných statistických testů je nad rámec této práce a bude jej využito až při konkrétních testech v praktické části. Podkapitola uvádí pouze stručný přehled mnoha užívaných testů vč. účelů a předpokladů.

Dle počtu zkoumaných souborů se testy dělí na testy pro jeden výběr, dva výběry a více výběrů. Při testování parametrů populace (střední hodnota, rozptyl, relativní četnost) se testy taktéž dělí na:

- parametrické – jsou vázané na splnění určitých předpokladů o parametrech a určitém typu rozdělení znaku v základním souboru (např. předpoklad normality),
- neparametrické – nejsou závislé na typu rozdělení základního souboru, používají se především k analýze dat nevyhovující požadavku na rozdělení v parametrických testech, jejich nevýhodou je především nižší síla testu, tedy schopnost správně zamítnout H_0 (Drápela, 2012, s. 103).

Testy v rámci jednoho souboru rozhodují o vztahu mezi neznámým parametrem populace a konkrétní číselnou hodnotou:

- z -test o střední hodnotě – předpoklad normality dat, znám rozptyl populace σ^2 ,
- t -test o střední hodnotě – předpoklad normality dat,
- test o rozptylu – předpoklad normality dat,
- test o parametru π – zkoumání relativní četnosti určitého jevu k určitému číslu, předpoklad výběru dat s alternativním rozdělením a jevů s binomickým rozdělením,
- kvantilový test – neparametrický,
- Wilcoxonův test pro jeden výběr – neparametrický, test mediánu či střední hodnoty (Anderson et al., 2017; Hendl, 2006; Hindls a kol., 2006).

Pro dva výběry se testy zaměřují na porovnávání parametrů dvou navzájem nezávislých populací (s výjimkou párových testů). Tyto testy mohou taktéž být parametrické či neparametrické.

- F-test – test shody rozptylu dvou populací, předpoklad normality dat,
- z-test pro dva výběry – předpoklad normality dat, znány rozptyly obou populací,
- t-test pro dva výběry – předpoklad normality dat, předpoklad shody rozptylů obou populací,
- Aspinové-Welchův test – univerzální test pro střední hodnotu dvou výběrů, předpoklad normality dat,
- Mannův-Whitneyův test – neparametrický, test shody mediánů dvou výběrů,
- párový t-test – závislá pozorování, předpoklad normality dat,
- párový Wilcoxonův test (znaménkový test) – závislá pozorování (Anderson et al., 2017; Hendl, 2006; Hindls a kol., 2006).

Další skupinou testů je pro porovnání více výběrů. Pro tyto účely se využívá analýzy rozptylu, zkráceně ANOVA (**A**nalysis **O**f **V**ariance). Nulová hypotéza se stanovuje jako rovnost daného parametru (např. střední hodnoty) napříč všemi pozorovanými soubory, alternativní hypotéza jako „non H_0 “ neboli jakýkoliv jiný stav, než rovnost parametru všech souborů (stačí, aby pouze mezi dvěma soubory byla jedna nerovnost).

- Jednofaktorová ANOVA (one-way ANOVA) – porovnání více než dvou středních hodnot, předpoklady vzájemné nezávislosti uvnitř skupin i mezi nimi, normality dat v každé skupině a shodného rozptylu v každé skupině,
- Kruskal-Wallisův test – neparametrická ANOVA, porovnání více než dvou mediánů,
- Jonckheere-Terpstra test – neparametrická ANOVA, testování předpokladu rostoucího účinku, alternativní hypotéza ve tvaru $\tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \dots \leq \tilde{\mu}_m$,
- Friedmanův test – závislé výběry (Hendl, 2006; Rice, 2007).

V rámci testů ANOVA se taktéž při zamítnutí nulové hypotézy využívají post-hoc testy pro zkoumání rozdílů mezi skupinami. Jinými slovy pomáhají identifikovat, mezi kterými soubory nastala nerovnost.

- Scheffeho test – kontrola chyb I. druhu,
- Fisherův LSD test – porovnávání všech dvojic, nejliberálnější z post-hoc testů (Anderson et al., 2017; Hendl, 2006).

Poslední skupinou testů je analýza kategoriálních dat. Proměnné jsou často nominálního, resp. kvalitativního typu. Získaná data jsou zachycena pomocí jedno-, dvou- nebo vícerozměrných tabulek četností (absolutních či relativních). Některé proměnné mají závislý charakter (cílové proměnné), jiné se považují za nezávislé.

- χ^2 -test dobré shody¹ - přezkoumání, zdali tvar pravděpodobnostního rozdělení kategoriální proměnné má specifikovanou podobu, využíváno např. pro test normality,
- χ^2 test nezávislosti – posuzování závislosti v kontingenčních tabulkách,
- McNemarův test – analýza dichotomických proměnných při dvou měřeních,
- Cochranův Q -test – analýza dichotomických proměnných ve více závislých výběrech,
- Kappa koeficient shody – analýza míry shody ve čtvercových tabulkách (Anderson et al., 2017; Hendl, 2006).

3.2 Analýza závislostí

Statistická analýza se obvykle nezabývá pouze jednou izolovanou proměnnou, ale spíše porovnáváním několika rozdělení, sledováním změn proměnné v čase nebo zkoumáním vztahů mezi proměnnými. Tato kapitola se věnuje základním metodám pro studium závislosti mezi statistickými znaky. Tyto metody jsou užitečné, pokud nás zajímají například:

- predikce budoucích zisků z prodeje produktu v závislosti na jeho ceně,
- závislost redukce váhy na počtu týdnů, kdy se jedinec podrobí dietnímu režimu s příjmem určitého energetické hodnoty,
- jak výška dítěte v šesti letech predikuje jeho výšku v šestnácti letech,
- jak spotřeba alkoholu ovlivňuje snížení tělesné teploty a podobně (Hendl, 2006, s. 237).

Tyto vztahy obvykle nemají čistě funkčně deterministický charakter, a proto je nutné použít pro jejich analýzu statistické metody. Příslušná oblast statistiky se nazývá regresní a korelační analýza, přičemž regresní analýza se zabývá jednostrannými závislostmi (kdy nezávislá proměnná ovlivňuje závislou) a korelační analýza je zaměřena na intenzitu vzájemného vztahu proměnných (Hindls a kol., 2006).

¹ Výslovnost χ^2 zní „chí kvadrát“

3.2.1 Regresní analýza

Cílem regresní analýzy je formulovat vztah mezi nezávislou proměnnou x a závislou proměnnou y pomocí matematického modelu – rovnicí **regresní funkce**, jež obsahuje několik neznámých parametrů β . Pokud tato funkce obsahuje pouze lineární parametry, mluví se o **lineárním regresním modelu**, v opačném případě se jedná o modely nelineární. Zcela lineární model se zapisuje pomocí rovnice:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \varepsilon$$

kde ε znamená náhodnou chybu (Hendl, 2006; Hindls, 2006).

V následující tabulce jsou zachyceny údaje ohledně počtu studentů na 10 různých univerzitách a kvartálních tržeb 10 přílehlých poboček pizzerie:

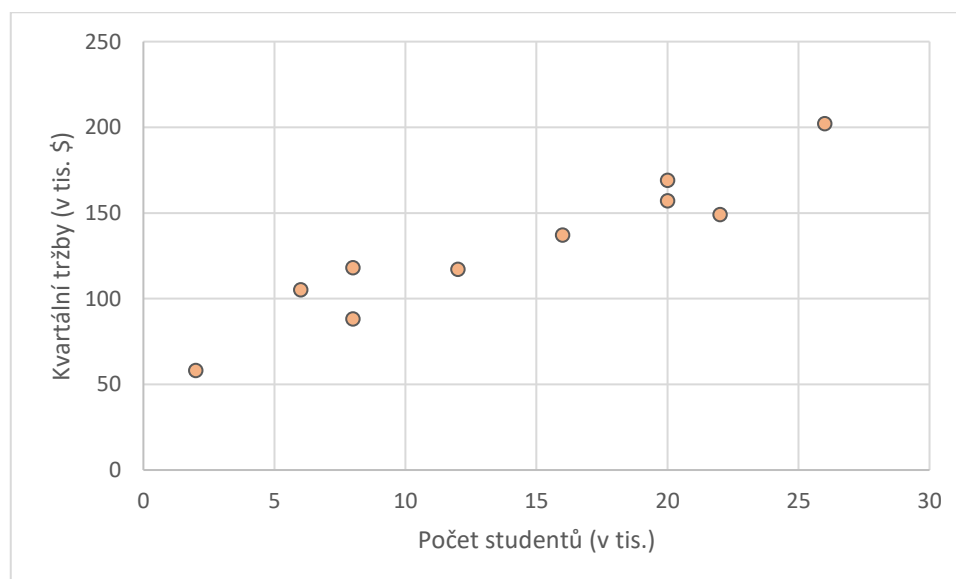
Tab. 6: Počet studentů a kvartální tržby poboček pizzerie

počet studentů (v tis.)	2	6	8	8	12	16	20	20	22	26
kvartální tržby (v tis. \$)	58	105	88	118	117	137	157	169	149	202

Zdroj: vlastní zpracování dle Anderson et al. (2017, s. 603)

Tyto údaje lze poté převést do jednoduchého XY bodového grafu, kde počet studentů je nezávislou proměnnou x a tržby jsou proměnnou závislou y :

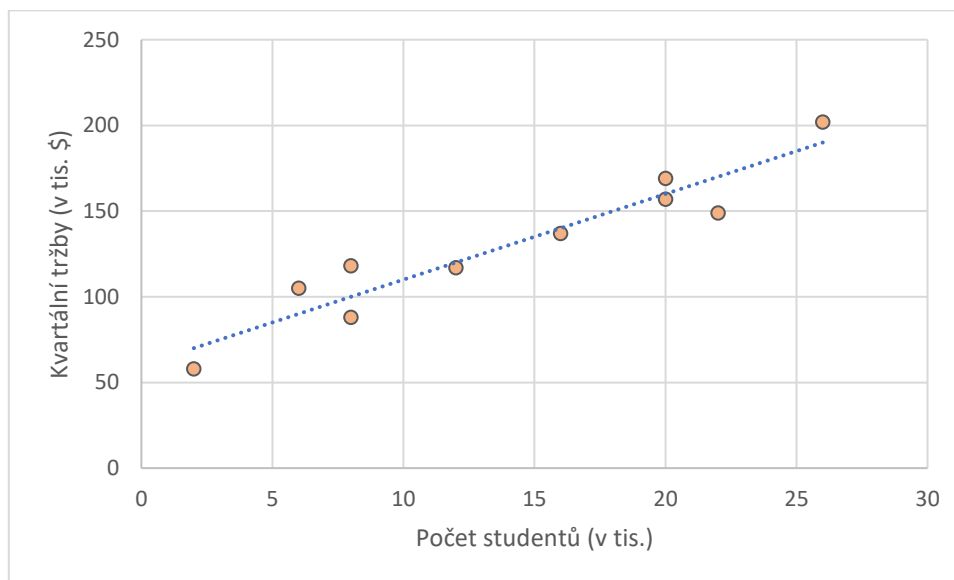
Obr. 2: Vizualizace počtu studentů a kvartálních tržeb poboček pizzerie



Zdroj: vlastní zpracování v MS Excel (2023)

Pro zjištění vhodné lineární regresní funkce se využívá metody nejmenších čtverců, jež proloží body přímkou tak, aby byla všem bodům co nejbližší, čehož dosáhne takovým tvarem, kdy bude součet čtverců residuí (odchylek) co nejnižší (Rice, 2007).

Obr. 3: Proložení bodů přímkou pomocí metody nejmenších čtverců



Zdroj: vlastní zpracování v MS Excel (2023)

Jedná o lineární závislost s pouze jednou nezávislou proměnnou, model tedy vypadá jako:

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

Nicméně skutečné parametry β_0 a β_1 nejsou známy, nahradí se tedy jejich odhady:

$$y = b_0 + b_1 x$$

Ačkoliv se parametry b_0 a b_1 dají spočítat ručně skrze parciální derivace, pro účely této kapitoly se využije softwarového řešení pro získání následující funkce:

$$y = 60 + 5x$$

Pokud by se tedy pizzerie rozhodla otevřít novou pobočku poblíž univerzity s 23 tis. studenty, lze pomocí této regresní funkce predikovat, že kvartální tržby dosáhnou hodnoty 175 tis. \$² (Anderson et al., 2017; Rice, 2007).

Kvalitu regresní funkce udává **koeficient determinace r^2** (v některých literaturách také nazvaný index determinace), neboť jeho hodnota vyjadřuje procentuální část rozptylu empirických hodnot vysvětlené daným regresním modelem (Hendl, 2006).

² $60 + 5 \times 23 = 175$

Definice a vzorec pro koeficient determinace je následující:

$$\text{koeficient determinace } r^2 = \frac{\text{variabilita vysvětlená modelem}}{\text{celková variabilita}} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

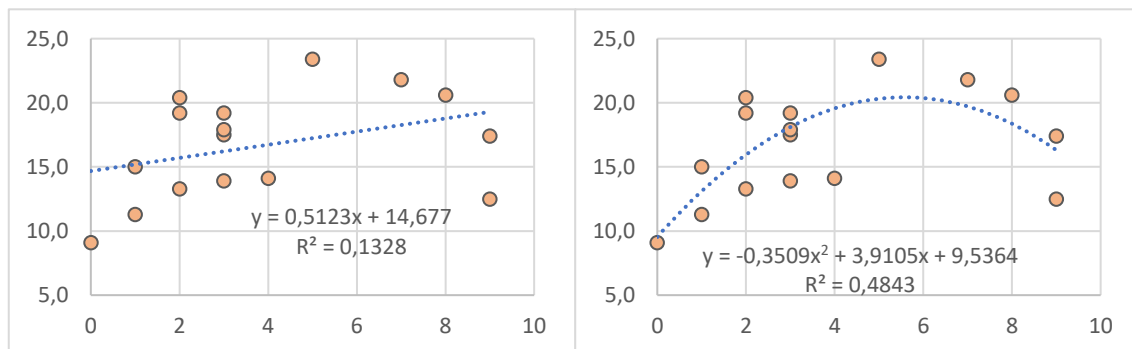
Hodnota r^2 se vždy nachází v intervalu $\langle 0;1 \rangle$ a čím více se přibližuje k 1, tím je daná regresní funkce kvalitnější. Mnoho literatur udává vlastní hranice mezi úrovněmi kvality, pro účely této práce budou využity hranice:

- 0 – 0,5 ... nízká kvalita regresní funkce,
- 0,5 – 0,8 ... střední kvalita regresní funkce,
- 0,8 – 1 ... vysoká kvalita regresní funkce.

Ve výše zmíněném příkladu s počtem studentů a pizzerií je koeficient determinace spočítán jako 0,9027; lze tedy tuto lineární funkci považovat za vysoce kvalitní a vhodně zvolenou (Anderson et al., 2017).

Ne vždy je lineární regresní funkce tou nejvhodnější pro zjištění vztahů mezi závislou a nezávislou proměnnou, jak ilustruje následující obrázek:

Obr. 4: Porovnání lineární a parabolické regresní funkce



Zdroj: vlastní zpracování v MS Excel (2023)

Na další straně jsou uvedeny další možné regresní funkce s jednou nezávislou proměnnou:

Tab. 7: Příklady regresních funkcí s jednou nezávislou proměnnou:

Název funkce	Empirická regresní funkce
Parabolická	$Y = b_0 + b_1x + b_2x^2$
Exponenciální	$Y = b_0e^{b_1x}$
Logaritmická	$Y = b_0 + b_1 \ln x$
Hyperbolická	$Y = b_0 + b_1 \frac{1}{x}$
Polynomická n-tého stupně	$Y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n$

Zdroj: vlastní zpracování dle Hindls a kol. (2006); Rice (2007)

V případě existence více než jedné nezávislé proměnné se hovoří o **vícenásobné regresi** (případně mnohonásobné). Je důležité poznamenat, že volba vhodného typu regresní funkce pro vícenásobnou závislost je obtížná, protože není možné graficky zobrazit celý průběh závislosti ani logicky posoudit, který typ regresní funkce je nejvhodnější. Proto se při hledání vhodného typu vícenásobné regresní funkce spoléhá převážně na matematicko-statistická kritéria, která pomáhá definovat především statistický software, jenž následně vybere nejvhodnější typ regresní funkce ze širokého spektra možností. Velmi často se při hledání vhodného typu vícenásobné regresní funkce postupuje tak, že se analyzuje zvláště závislost mezi závislou proměnnou y a jednotlivými vysvětlujícími proměnnými x_1, x_2, \dots, x_p a výsledná regresní funkce je pak konstruována jako součet jednoduchých regresních funkcí (Hindls a kol, 2006, s. 213).

Příkladem lineární vícenásobné regrese může být vztah mezi dvěma druhy příjmů studenta vysoké školy (příspěvek od rodiny – x_1 ; výplata z brigádní činnosti – x_2) a jeho výdaji (y) závislé na výši zmíněných příjmů. Regresní funkce by v takovémto případě vypadala následovně:

$$y = b_0 + b_1x_1 + b_2x_2$$

3.2.2 Korelační analýza

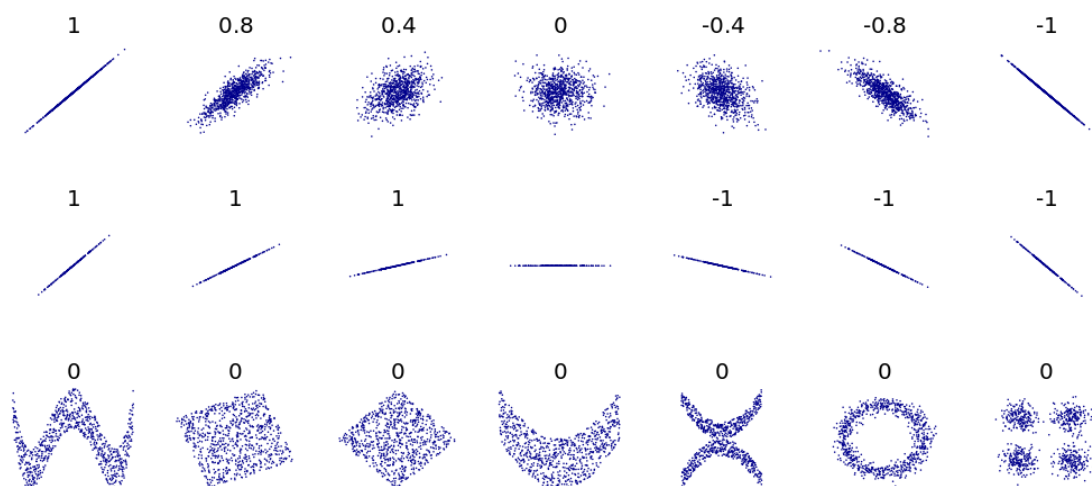
V obecném smyslu se slovo "korelace" používá k vyjádření míry spojitosti mezi dvěma či více proměnnými. Dvě proměnné jsou považovány za korelované, pokud určité hodnoty jedné proměnné mají tendenci se vyskytovat společně s určitými hodnotami druhé proměnné. Míra této tendence se může lišit od absence korelace (všechny hodnoty proměnné y se vyskytují stejně pravděpodobně s každou hodnotou proměnné x)

až po absolutní korelaci (s danou hodnotou proměnné x se vyskytuje právě jedna hodnota proměnné y). Existuje mnoho korelačních koeficientů, které se liší v závislosti na typech proměnných, pro které se používají. Statistické odhady korelačních koeficientů se opírají o teorii pravděpodobnosti pro společné rozdělení dvou nebo více náhodných proměnných. Při zkoumání korelačních vztahů je kvalitativní rozbor příslušného materiálu rozhodující. Nemá smysl měřit závislost tam, kde na základě logických úvah nemůže existovat. Měření korelace může být také zbytečné, pokud je způsobena:

- a) formálními vztahy mezi proměnnými,
- b) nehomogenitou studovaného základního materiálu,
- c) působením společné příčiny (Anderson et al., 2017; Hendl, 2006).

Rozlišuje se kladná korelace, kdy se zvyšováním hodnot znaků jedné proměnné se zvyšují i hodnoty druhé proměnné, a poté záporná korelace, kdy se zvyšováním hodnot jedné proměnné se hodnoty druhé proměnné zmenšují. I přes některé nedostatky patří mezi nejdůležitější míry síly vztahu dvou náhodných veličin **Pearsonův korelační koeficient** r neboli míra intenzity lineární závislosti. Oproti koeficientu determinace u regresní analýzy mohou korelační koeficienty dosahovat hodnot $\langle -1; 1 \rangle$, kdy opětovně bližší hodnota koeficientu k jedné z hranic znamená silnější vazbu, v případě -1 klesající, v případě 1 rostoucí (Anderson et al., 2017; Hendl, 2006).

Obr. 5: Hodnoty Pearsonova korelačního koeficientu



Zdroj: Ed's Scribbles (2012)

Pro zachycení jiných, než lineárních vztahů se využívá **Spearmanův korelační koeficient**, taktéž nazýván korelace pořadí (Hendl, 2006).

3.3 Analýza časových řad

Časovou řadou se označuje sled věcně a prostorově srovnatelných pozorování, která jsou uspořádána chronologicky, tzn. od minulosti přes přítomnost k budoucnosti. Jedná se o speciální regresní model s vysvětlující proměnnou v podobě času, ve kterém dochází k autokorelaci a u proměnných se může vyskytnout sezónní charakter. Metody analýzy a prognózy časových řad slouží k popisu chování těchto sledů a jejich předpovídání. Chronologicky uspořádaná data jsou běžná v mnoha oblastech života, např. ve fyzice, biologii, seismologii a meteorologii. V ekonomii nabývá stále většího významu práce s časovými řadami, např. v makroekonomii nebo při sledování vývoje kurzů měn, cen akcií nebo průmyslové produkce. V posledních letech se rozvíjí metody analýzy a prognózy ekonomických časových řad, které jsou dnes poměrně rozmanité a nabízejí různé nástroje a techniky. S rozvojem výpočetní techniky a statistického softwaru jsou tyto metody stále přesnější a rychlejší a umožňují sériové ověřování v praxi (Hindls a kol., 2007, s. 246).

Základní druhy časových řad se dělí podle:

- rozhodného časového hlediska
 - intervalové (rozdílné dny v měsících),
 - okamžikové (určitý časový okamžik),
- periodicity
 - dlouhodobé (roční),
 - krátkodobé (čtvrtletní, měsíční, týdenní apod.),
- druhu sledovaných ukazatelů
 - primární (prvotní),
 - sekundární (odvozené),
- způsobu vyjádření údajů
 - naturální (v naturálních jednotkách),
 - peněžní (v peněžních jednotkách) (Hindls a kol., 2006; Kropáč, 2007).

Tradiční princip modelování časových řad se nazývá jednorozměrný model:

$$y_t = f(t, \varepsilon_t),$$

kde y_t je hodnota modelovaného ukazatele v čase ($t = 1, 2, \dots, n$) a ε_t je hodnota náhodné složky (poruchy) v čase t .

K tomuto modelu se dá přistupovat třemi způsoby:

1. klasickým (formálním modelem),
2. Boxovo-Jenkinsovou metodologií,
3. spektrální analýzou.

Tato práce se bude zabývat pouze prvním způsobem, také někdy nazývaným jako dekompozice časové řady. Jde pouze o popis různých způsobů, jak se časová řada pohybuje, a ne o vysvětlení skutečných příčin dynamiky časové řady. Tento model je založen na rozložení řady na čtyři různé složky (formy) pohybu v čase, a to:

- trendovou složku T_t ,
- sezónní složku S_t ,
- cyklickou složku C_t ,
- náhodnou složku ε_t ,

přičemž vlastní tvar zápisu může být dvojího typu:

- aditivní: $y_t = T_t + S_t + C_t + \varepsilon_t \quad t = 1, 2, \dots, n$,
- multiplikativní: $y_t = T_t S_t C_t \varepsilon_t, \quad t = 1, 2, \dots, n$.

Tyto formy představují systematickou část průběhu časové řady a naším cílem je najít nástroje, které co nejlépe vysvětlí tento systematický pohyb sledovaného procesu. Avšak ne všechny čtyři formy jsou vždy přítomny, a to záleží na konkrétním charakteru sledovaného ukazatele (někdy může chybět sezónní složka a podobně) (Hindls, 2006; Kropáč, 2007).

V čase se **trendem** rozumí hlavní směr dlouhodobého vývoje hodnoty analyzovaného ukazatele. Trend může být rostoucí, klesající nebo konstantní. **Sezónní složka** se vyskytuje u časových řad s periodou kratší než jeden rok a je to pravidelně se opakující odchylka od trendu způsobená například změnou ročních období, měsíčním nebo pracovním cyklem nebo různými společenskými zvyklostmi. **Cyklickou složkou** se rozumí kolísání okolo trendu způsobené dlouhodobým cyklickým vývojem s délkou vlny delší než jeden rok. **Náhodnou složku** nelze popsat funkcí času a je to složka, která zbývá po vyloučení trendu, sezónní a cyklické složky. Její chování se často musí prověřovat prostřednictvím testů (Hindls, 2006; Kropáč, 2007).

4 Vybraný postup řešení problému

Společnost ESCG games s.r.o. se dle majitele dlouhodobě potýká s poklesem návštěvnosti únikových her. Majitel si přeje ověřit, zdali důvodem onoho poklesu není jeden ze čtyř důvodů jmenovaných v kapitole č. 1, tzn. ekonomická krize a s ní související tendence lidí šetřit, vyčerpání kapacit obou únikových her, nesprávně zvolený marketing či snížení kvality poskytovaných služeb.

Za tímto účelem autor navrhne a provede ve spolupráci s managementem společnosti ESCG games s.r.o. analýzu interních dat a smíšený výzkum napříč zákazníky únikových her Chernobyl z minulosti skrze dotazníkové šetření, ve kterém budou sesbírána jak kvantitativní, tak kvalitativní data.

Cílem výzkumu je nalezení důvodů klesající návštěvnosti únikové hry Chernobyl.

K dosažení tohoto cíle je třeba provést několik kroků, například:

1. Získat data o návštěvách únikové hry z minulých období.
2. Definovat výzkumné otázky a na jejich základě sestavit statistické hypotézy k otestování.
3. Sesbírat potřebná data od minulých návštěvníků únikové hry Chernobyl.
4. Analyzovat data a hledat možné příčiny poklesu návštěvnosti.

Management na základě těchto kroků poté navrhne opatření ke zlepšení návštěvnosti únikové hry a/nebo ke zvýšení atraktivity hry pro hráče a po následném zhodnocení možných rizik a přínosů investic rozhodne o dalším postupu.

Výzkum se tedy bude zabývat spokojeností zákazníků. Podle Kotlera a Kellera (2013, s. 164), je spokojenost pocitem potěšení nebo zklamání jedince v porovnání s jeho očekáváním, tedy pokud je očekávání překonáno, je velmi spokojen nebo i potěšen. Taktéž uvádějí, že se spokojeností zákazníků jsou spojeny i značné výhody, např. že velmi spokojený zákazník obvykle zůstává věrný značně delší dobu a kupuje další produkty a je méně vnímavý k nabídce konkurenčních společností. Jak již bylo vysvětleno, tak tato výhoda u únikových her příliš nefunguje, nicméně věrný zákazník může přinést jiné výhody v podobě pozitivních recenzí či doporučení. Tuto tezi potvrzuje Světlík (2018, str. 47): „*Spokojenost zvyšuje jednak pravděpodobnost, že se nákup bude v budoucnosti opakovat. Zároveň bude o výrobku a firmě pozitivně informovat své okolí.*“

4.1 Sestavení hypotéz

Podle Egera a Egerové (2014, s. 38) jsou výzkumné otázky základním kamenem celého návrhu výzkumu. Ze všeho nejdříve je vhodné položení široce vymezených, abstraktních, obecných otázek, od kterých se další, složitější otázky, začnou odvíjet.

Nová publikace Egera a Egerové (2022) uvádí, že hlavní výzkumné otázky umožňují vyjádřit jaký problém bude výzkumný projekt řešit a na jaké otázky bude díky němu možné odpovědět.

Jelikož se společnost snaží zjistit možné příčiny úpadku návštěvnosti, je v primární části výzkumu položena **obecná výzkumná otázka**:

- Jaká je spokojenost zákazníků se zážitkovou hrou?

Z této obecné otázky následně vychází **otázky specifické**:

- Jak byste popsali Váš zážitek s únikovou hrou Chernobyl?
- Jak současná ekonomická krize ovlivnila Vaše výdaje?
- Pokud byste se rozhodli účastnit únikové hry, jakým způsobem byste zjišťovali informace a vybírali?

Po vzájemné dohodě s managementem společnosti ESCG games s.r.o., jehož představitelé si prostřednictvím výzkumu chtějí ověřit své předpoklady a tvrzení, bylo sestaveno následujících 5 hypotéz:

- Příznivci únikových her v současnosti utrací méně za zábavní průmysl.
- Minulí návštěvníci vítají možnost otevření nové úrovně.
- Zákazníci se o hře dozvěděli převážně pomocí reklamní kampaně.
- Zákazníci nakupují vstupenky rovnoměrně napříč využívanými portály.
- Hráči z roku 2022 jsou s kvalitami služeb stejně spokojeni jako hráči z předchozích let.

4.2 Zpracování a analýza dat

Samotný výzkum bude proveden ve dvou fázích: v první řadě dojde k analýze návštěvnosti z minulých období v porovnání s rokem 2022 za účelem ověření skutečného poklesu frekvence rezervací. Toho bude dosaženo za pomoci analýzy časové osy a také následné predikce pro rok 2023, taktéž dojde k ověření statistické hypotézy ohledně nákupu voucherů. Tato data jsou primárního interního charakteru, protože budou

poskytnuta přímo společností, a to formou přístupu do vnitřní databáze společnosti, jež ukládá údaje o současných i minulých rezervacích a jejich účastnících. Náhled databáze s ošetřenými osobními údaji účastníků je k dispozici v příloze č. 1.

Pokud bude potvrzen klesající trend, přejde výzkum do druhé fáze, kdy dojde ke kontaktování předchozích návštěvníků únikové hry Chernobyl a sběru dat potřebných k ověření hypotéz stanovených s managementem společnosti. Jedná se tedy o primární externí data. Metodou sběru dat pro potvrzení či vyvrácení hypotéz bylo zvoleno dotazníkové šetření v elektronické podobě pro zajištění snadné přehlednosti a upravitelnosti, snadného přístupu a rychlejšího zpracování výsledků. Kontakt s minulými účastníky bude možný pomocí podnikové databáze, ve které jsou uvedeny potřebné kontaktní údaje veškerých návštěvníků únikové hry Chernobyl.

Celý výzkum potrvá od 17. 3. 2023 do 23. 4. 2023, tzn. 38 dní.

Tab. 8: Časový plán výzkumu pro společnost ESCG games s.r.o.

Proces	Datum
Vytvoření dotazníku	17. 3. 2023 – 19. 3. 2023
Dotazování	20. 3. 2023 – 2. 4. 2023
Zpracování výsledků a analýza dat	3. 4. 2023 – 16. 4. 2023
Vyhodnocení výzkumu	17. 4. 2023 – 23. 4. 2023

Zdroj: Vlastní zpracování (2023)

Detailní postupy při zpracování jednotlivých kroků budou sepsané při provádění jednotlivých šetření a testů.

4.3 Tvorba dotazníkového šetření

Pro druhou fázi šetření bylo autorem připraveno dotazníkové šetření návrh byl představen managementu společnosti i s následným vysvětlením, jak bude na základě získaných odpovědí probíhat statistická analýza. Po několika připomínkách ze strany managementu a rozdílnosti v názorech o tom, co mají data přinést za informace, byla vytvořena finální podoba:

1. V jakém roce jste poprvé navštívili únikovou hru Chernobyl?
 - 2018
 - 2019
 - 2020
 - 2021
 - 2022
2. Jakou hru jste vyzkoušeli?
 - Dozimetrická věž
 - Záchranná mise
 - Obě
3. Uvítali byste možnost otevření třetí úrovně?
 - Ano
 - Ne
4. Jakým způsobem jste se dozvěděli o únikové hře Chernobyl?
 - Ústní sdělení (od známých)
 - Reklama na sociálních sítích
 - Kulturní weby (GoOut, Slevomat apod.)
 - Jiné:
5. Jak moc jste byli spokojeni s úrovní poskytnutých služeb (výběr prostředí, srozumitelnost hádanek, přístup obsluhy apod.)?

Silně nespokojeni 1 2 3 4 5 Silně spokojeni
6. Jak moc ovlivnila současná ekonomická krize Váš postoj k zábavnímu průmyslu?

Nijak 1 2 3 4 5 Výrazně
7. Za jakou maximální cenu byste byli ochotni zaplatit za únikovou hru pro 6 lidí s vícero místnostmi před současnou ekonomickou situací?

(otevřená otázka, možnost zadat jakékoliv číslo)
8. Za jakou maximální cenu byste byli **nyní** ochotni zaplatit za únikovou hru pro 6 lidí s vícero místnostmi?

(otevřená otázka, možnost zadat jakékoliv číslo)

5 Analýza primárních interních dat

Dle majitele společnost ESCG games s.r.o. se hra potýká s klesající návštěvností. Toto tvrzení bude prozkoumáno za pomoci analýzy časové osy. a také analýzou rozptylu.

K analýze návštěvnosti únikové hry Chernobyl v minulých obdobích bude využito dat z vnitřní podnikové databáze, která shrnuje veškeré údaje o návštěvnících a konaných rezervacích. Jedná se tedy o průzkum primárních interních dat. Důležité upozornění: v roce 2021 byl provoz únikových her na několik měsíců pozastaven z důvodu vládních restrikcí proti COVID-19. Z tohoto důvodu bude rok 2021 z analýz vyloučen, neboť by došlo ke zkreslování výsledků, a sledovanými roky tak budou pouze roky 2018, 2019, 2020 a 2022.

5.1 Analýza časových os a predikce

Jelikož společnost ESCG games s.r.o. nabízí dvě únikové hry, budou návštěvnosti těchto her analyzovány samostatně. Časové řady budou stavěny podle měsíčních intervalů, jedná se tedy o intervalový krátkodobý druh.

Hra: Dozimetrická věž

Nejdříve je nutné z databáze zjistit počet odehraných her v jednotlivých měsících. Jelikož databáze pouze ukládá seznam rezervací a údajů hráče, jenž rezervaci vytvořil, a ve společnosti není žádný automatizovaný analytický systém, je nutné tento počet zjistit vyfiltrováním a seřazením dat pomocí softwaru Microsoft Excel, dále rozdělením dat na jednotlivé měsíce a následně využitím funkce POČET2. Výsledkem je následující tabulka:

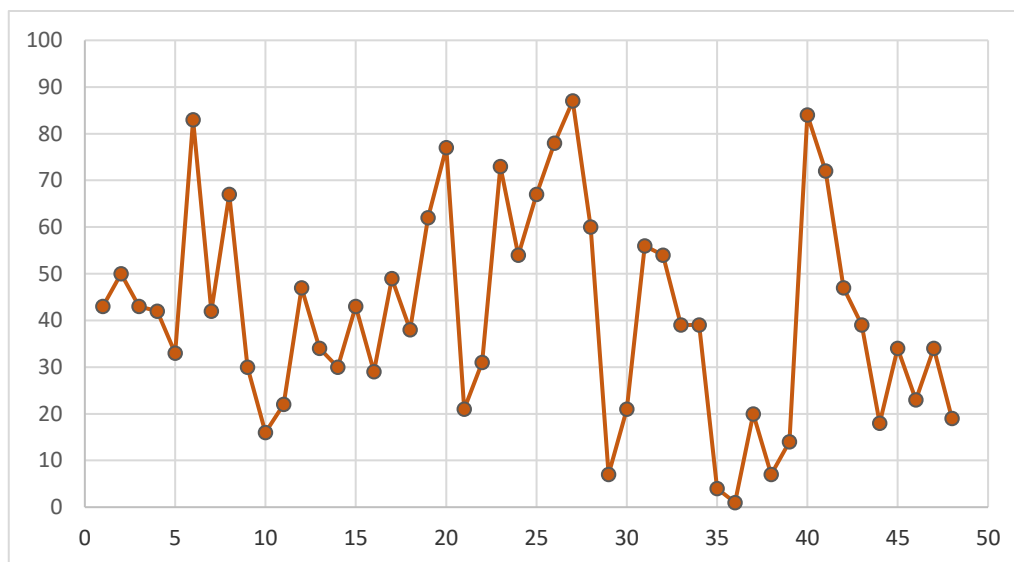
Tab. 9: Počet měsíčně odehraných her Dozimetrická věž 2018–2022 (bez 2021)

Dozimetrická věž	2018	2019	2020	2022
Leden	43	34	67	20
Únor	50	30	78	7
Březen	43	43	87	14
Duben	42	29	60	84
Květen	33	49	7	72
Červen	83	38	21	47
Červenec	42	62	56	39
Srpen	67	77	54	18
Září	30	21	39	34
Říjen	16	31	39	23
Listopad	22	73	4	34
Prosinec	47	54	1	19
Suma	518	541	513	411

Zdroj: vlastní zpracování dle ESCG games s.r.o. (2023)

Na základě těchto dat se následně sestrojí graf:

Obr. 6: Počet měsíčně odehraných her Dozimetrická věž 2018–2022 (bez 2021)



Zdroj: vlastní zpracování v MS Excel (2023)

Z tohoto grafu je potřebné vybrat vhodný model pro časovou osu. Zde se jedná o nelehkou úlohu, neboť tvar této křivky je natolik specifický, že žádná regresní funkce

nedokáže tuto křivku vhodně reprezentovat či predikovat. Nejsilnější index determinace má polynomičká funkce n-tého stupně, v tomto případě šestého, nicméně autor tuto funkci zavrhl z důvodu tendence polynomičké funkce predikovat extrémní hodnoty, ať už rostoucí či klesající. Byla proto zvolen lineární (přímkový) trend. Z grafu lze vyčíst sezónnost směřující nahoru v měsíci dubnu a během letních měsíců (červen, červenec, srpen) a naopak klesající sezónnost v posledních 3 měsících v roce. Variabilita této sezónní složky se časem zvětšuje, proto zvolen multiplikativní typ. Cyklická složka není uvažována. Funkce trendu je tedy stanovena jako:

$$y_t = T_t \cdot S_t \cdot \varepsilon_t$$

Pomocí MS Excel a funkce LINREGRESE zjistíme trendovou funkci:

$$y = 48,992 - 0,3134x,$$

pomocí které se následně spočtou hodnoty T_t na trendové přímce, a to dosazením konkrétního počtu odehraných her za měsíc za proměnnou x_t . V případě ledna 2018 se tedy jedná o: $48,992 - 0,3134 \times 43 = 48,6786$. Přehled o všech hodnotách se nachází v tabulce č. 8.

Dále se určí velikosti složek $S_t \cdot \varepsilon_t$, které se vypočítají poměrem T_t / x_t . Tato složka se poté zprůměruje na jednotlivé měsíce, aby došlo k eliminaci náhodné složky. Jednotlivé měsíční výkyvy se musí vykompenzovat, aby se jejich součet rovnal 12. Tímto se získají hodnoty S_t .

Následně se spočítají očekávané hodnoty $T_t \cdot S_t$, tyto hodnoty spočítáme také v rámci predikce na následující rok 2023. Posledním krokem je určení náhodné složky, a to podílem $x_t / (T_t \cdot S_t)$.

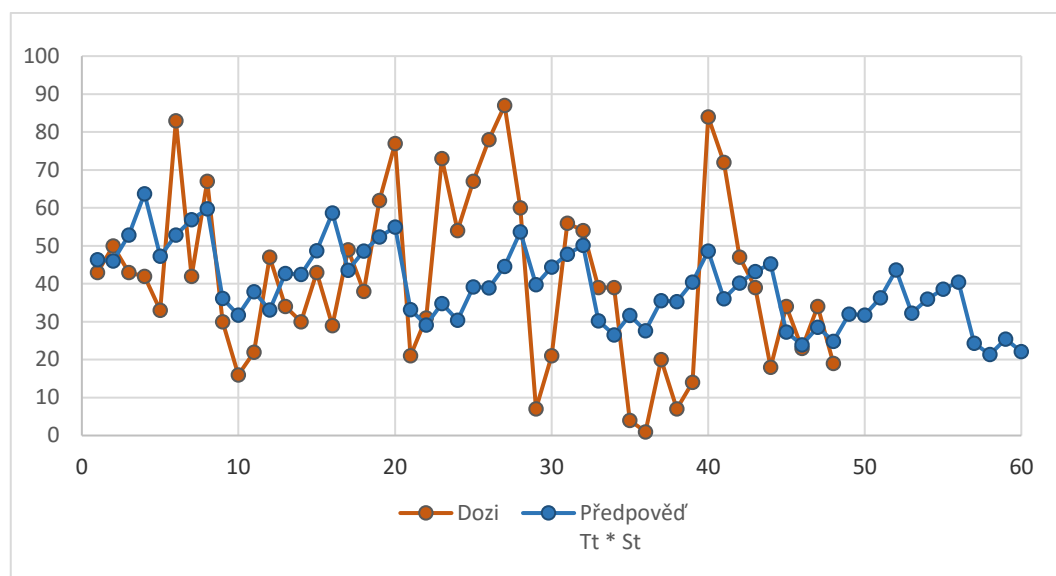
Tab. 10: Výpočet hodnot pro časovou osu Dozimetrické věže

Měsíc (x)	Pořadí (t)	Počet her (x_t)	Trend (T_t)	$S_t \cdot \varepsilon_t$	Průměry měsíců	Úprava (S_t)	Predikce ($T_t \cdot S_t$)	ε_t
01-18	1	43	48,6786	0,8833	0,9508	0,9516	46,3228	0,9283
02-18	2	50	48,3651	1,0338	0,9512	0,9521	46,0470	1,0858
03-18	3	43	48,0517	0,8949	1,0983	1,0992	52,8205	0,8141
04-18	4	42	47,7382	0,8798	1,3339	1,3350	63,7323	0,6590
05-18	5	33	47,4248	0,6958	0,9964	0,9973	47,2967	0,6977
06-18	6	83	47,1113	1,7618	1,1202	1,1212	52,8189	1,5714
07-18	7	42	46,7979	0,8975	1,2155	1,2166	56,9341	0,7377
08-18	8	67	46,4844	1,4413	1,2852	1,2864	59,7963	1,1205
09-18	9	30	46,1710	0,6498	0,7822	0,7828	36,1444	0,8300
10-18	10	16	45,8575	0,3489	0,6920	0,6926	31,7603	0,5038
11-18	11	22	45,5441	0,4830	0,8320	0,8327	37,9236	0,5801
12-18	12	47	45,2306	1,0391	0,7319	0,7325	33,1323	1,4186
01-19	13	34	44,9172	0,7569		0,9516	42,7435	0,7954
02-19	14	30	44,6037	0,6726		0,9521	42,4659	0,7064
03-19	15	43	44,2903	0,9709		1,0992	48,6858	0,8832
04-19	16	29	43,9768	0,6594		1,3350	58,7107	0,4939
05-19	17	49	43,6634	1,1222		0,9973	43,5455	1,1253
06-19	18	38	43,3499	0,8766		1,1212	48,6018	0,7819
07-19	19	62	43,0365	1,4406		1,2166	52,3580	1,1842
08-19	20	77	42,7230	1,8023		1,2864	54,9577	1,4011
09-19	21	21	42,4096	0,4952		0,7828	33,1998	0,6325
10-19	22	31	42,0961	0,7364		0,6926	29,1552	1,0633
11-19	23	73	41,7827	1,7471		0,8327	34,7916	2,0982
12-19	24	54	41,4692	1,3022		0,7325	30,3770	1,7777
01-20	25	67	41,1558	1,6280		0,9516	39,1641	1,7108
02-20	26	78	40,8423	1,9098		0,9521	38,8848	2,0059
03-20	27	87	40,5289	2,1466		1,0992	44,5511	1,9528
04-20	28	60	40,2154	1,4920		1,3350	53,6890	1,1175
05-20	29	7	39,9020	0,1754		0,9973	39,7942	0,1759
06-20	30	21	39,5885	0,5305		1,1212	44,3847	0,4731
07-20	31	56	39,2751	1,4258		1,2166	47,7819	1,1720
08-20	32	54	38,9616	1,3860		1,2864	50,1192	1,0774
09-20	33	39	38,6482	1,0091		0,7828	30,2553	1,2890
10-20	34	39	38,3347	1,0174		0,6926	26,5501	1,4689
11-20	35	4	38,0213	0,1052		0,8327	31,6595	0,1263
12-20	36	1	37,7078	0,0265		0,7325	27,6217	0,9283
01-22	37	20	37,3944	0,5348		0,9516	35,5847	0,5620
02-22	38	7	37,0809	0,1888		0,9521	35,3037	0,1983
03-22	39	14	36,7675	0,3808		1,0992	40,4164	0,3464
04-22	40	84	36,4540	2,3043		1,3350	48,6674	1,7260
05-22	41	72	36,1406	1,9922		0,9973	36,0430	1,9976
06-22	42	47	35,8271	1,3119		1,1212	40,1676	1,1701
07-22	43	39	35,5137	1,0982		1,2166	43,2058	0,9027
08-22	44	18	35,2002	0,5114		1,2864	45,2806	0,3975
09-22	45	34	34,8868	0,9746		0,7828	27,3107	1,2449
10-22	46	23	34,5733	0,6653		0,6926	23,9450	0,9605
11-22	47	34	34,2599	0,9924		0,8327	28,5275	1,1918
12-22	48	19	33,9464	0,5597		0,7325	24,8664	0,7641
01-23	49		33,6330			0,9516	32,0054	
02-23	50		33,3195			0,9521	31,7226	
03-23	51		33,0061			1,0992	36,2817	
04-23	52		32,6926			1,3350	43,6458	
05-23	53		32,3792			0,9973	32,2918	
06-23	54		32,0657			1,1212	35,9505	
07-23	55		31,7523			1,2166	38,6297	
08-23	56		31,4388			1,2864	40,4420	
09-23	57		31,1254			0,7828	24,3661	
10-23	58		30,8119			0,6926	21,3399	
11-23	59		30,4985			0,8327	25,3954	
12-23	60		30,1850			0,7325	22,1111	

Zdroj: vlastní zpracování (2023)

Predikované hodnoty vložíme do grafu spolu s křivkou návštěvnosti:

Obr. 7: Návštěvnost Dozimetrické věže 2018–2022 (bez 2021) s predikcí pro 2023



Zdroj: vlastní zpracování v MS Excel (2023)

Z grafu lze vyčíst klesající trend, nicméně není příliš strmý.

Hra: Záchranná mise

Postup je principálně stejný, jako u předchozí hry. Prvním krokem je sesbírat potřebná data z databáze a zformátovat je do čitelné a snadno zpracovatelné podoby. Zformátovaná data lze zhlédnout v tabulce č. 11.

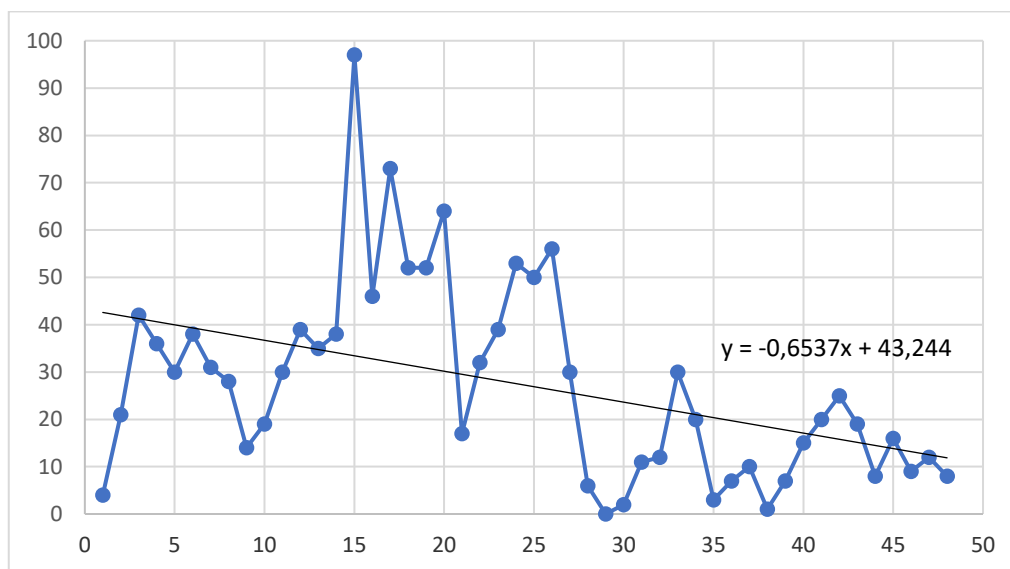
Tab. 11: Počet měsíčně odehraných her Záchranná mise 2018–2022 (bez 2021)

Záchranná mise	2018	2019	2020	2022
Leden	4	35	50	10
Únor	21	38	56	1
Březen	42	97	30	7
Duben	36	46	6	15
Květen	30	73	0	20
Červen	38	52	2	25
Červenec	31	52	11	19
Srpen	28	64	12	8
Září	14	17	30	16
Říjen	19	32	20	9
Listopad	30	39	3	12
Prosinec	39	53	7	8
Suma	332	598	227	150

Zdroj: vlastní zpracování dle ESCG games s.r.o. (2023)

Na obrázku č. 8 je vykreslena časová osa pro návštěvnost hry Záchranná mise. S pomocí nástrojů v MS Excel jsou vyobrazeny i parametry funkce trendu. V tabulce 12 jsou zachyceny vypočtené údaje pro jednotlivé složky a predikci.

Obr. 8: Počet měsíčně odehraných her Záchranná mise 2018–2022 (bez 2021)



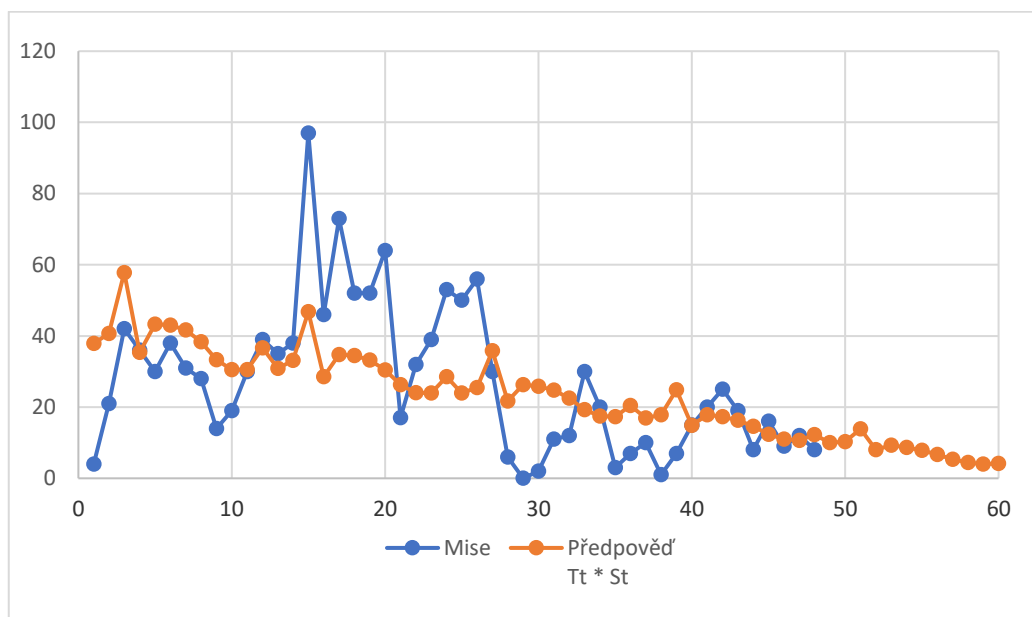
Zdroj: vlastní zpracování v MS Excel (2023)

Tab. 12: Výpočet hodnot pro časovou osu Záchrané mise

Měsíc (x)	Pořadí (t)	Počet her (x_t)	Trend (T_t)	$S_t \cdot \varepsilon_t$	Průměry měsíců	Úprava (S_t)	Predikce ($T_t \cdot S_t$)	ε_t
01-18	1	4	42,5901	0,0939	0,8711	0,8896	37,8894	0,1056
02-18	2	21	41,9365	0,5008	0,9508	0,9710	40,7196	0,5157
03-18	3	42	41,2828	1,0174	1,3712	1,4003	57,8076	0,7265
04-18	4	36	40,6292	0,8861	0,8518	0,8698	35,3412	1,0186
05-18	5	30	39,9755	0,7505	1,0597	1,0822	43,2602	0,6935
06-18	6	38	39,3218	0,9664	1,0716	1,0943	43,0306	0,8831
07-18	7	31	38,6682	0,8017	1,0556	1,0781	41,6871	0,7436
08-18	8	28	38,0145	0,7366	0,9869	1,0079	38,3142	0,7308
09-18	9	14	37,3609	0,3747	0,8730	0,8915	33,3074	0,4203
10-18	10	19	36,7072	0,5176	0,8152	0,8325	30,5600	0,6217
11-18	11	30	36,0536	0,8321	0,8301	0,8477	30,5622	0,9816
12-18	12	39	35,3999	1,1017	1,0136	1,0351	36,6420	1,0644
01-19	13	35	34,7462	1,0073		0,8896	30,9113	1,1323
02-19	14	38	34,0926	1,1146		0,9710	33,1033	1,1479
03-19	15	97	33,4389	2,9008		1,4003	46,8240	2,0716
04-19	16	46	32,7853	1,4031		0,8698	28,5182	1,6130
05-19	17	73	32,1316	2,2719		1,0822	34,7717	2,0994
06-19	18	52	31,4779	1,6520		1,0943	34,4469	1,5096
07-19	19	52	30,8243	1,6870		1,0781	33,2308	1,5648
08-19	20	64	30,1706	2,1213		1,0079	30,4085	2,1047
09-19	21	17	29,5170	0,5759		0,8915	26,3145	0,6460
10-19	22	32	28,8633	1,1087		0,8325	24,0297	1,3317
11-19	23	39	28,2097	1,3825		0,8477	23,9130	1,6309
12-19	24	53	27,5560	1,9234		1,0351	28,5229	1,8582
01-20	25	50	26,9023	1,8586		0,8896	23,9331	2,0892
02-20	26	56	26,2487	2,1334		0,9710	25,4870	2,1972
03-20	27	30	25,5950	1,1721		1,4003	35,8403	0,8370
04-20	28	6	24,9414	0,2406		0,8698	21,6952	0,2766
05-20	29	0	24,2877	0,0000		1,0822	26,2833	0,0000
06-20	30	2	23,6340	0,0846		1,0943	25,8632	0,0773
07-20	31	11	22,9804	0,4787		1,0781	24,7745	0,4440
08-20	32	12	22,3267	0,5375		1,0079	22,5028	0,5333
09-20	33	30	21,6731	1,3842		0,8915	19,3216	1,5527
10-20	34	20	21,0194	0,9515		0,8325	17,4994	1,1429
11-20	35	3	20,3658	0,1473		0,8477	17,2638	0,1738
12-20	36	7	19,7121	0,3551		1,0351	20,4038	0,3431
01-22	37	10	19,0584	0,5247		0,8896	16,9549	0,5898
02-22	38	1	18,4048	0,0543		0,9710	17,8707	0,0560
03-22	39	7	17,7511	0,3943		1,4003	24,8566	0,2816
04-22	40	15	17,0975	0,8773		0,8698	14,8722	1,0086
05-22	41	20	16,4438	1,2163		1,0822	17,7949	1,1239
06-22	42	25	15,7901	1,5833		1,0943	17,2794	1,4468
07-22	43	19	15,1365	1,2552		1,0781	16,3182	1,1643
08-22	44	8	14,4828	0,5524		1,0079	14,5970	0,5481
09-22	45	16	13,8292	1,1570		0,8915	12,3288	1,2978
10-22	46	9	13,1755	0,6831		0,8325	10,9691	0,8205
11-22	47	12	12,5219	0,9583		0,8477	10,6147	1,1305
12-22	48	8	11,8682	0,6741		1,0351	12,2846	0,6512
01-23	49		11,2145			0,8896	9,9768	
02-23	50		10,5609			0,9710	10,2544	
03-23	51		9,9072			1,4003	13,8729	
04-23	52		9,2536			0,8698	8,0492	
05-23	53		8,5999			1,0822	9,3065	
06-23	54		7,9462			1,0943	8,6957	
07-23	55		7,2926			1,0781	7,8619	
08-23	56		6,6389			1,0079	6,6913	
09-23	57		5,9853			0,8915	5,3359	
10-23	58		5,3316			0,8325	4,4388	
11-23	59		4,6780			0,8477	3,9655	
12-23	60		4,0243			1,0351	4,1655	

Zdroj: vlastní zpracování (2023)

Obr. 9: Návštěvnost Záchrané mise 2018–2022 (bez 2021) s predikcí pro 2023



Zdroj: vlastní zpracování v MS Excel (2023)

I zde lze vidět klesající trend, tentokrát ale výrazně strmější. Predikce se přibližuje k hodnotě 0, což by znamenalo ukončení návštěvnosti této únikové hry.

5.2 Hypotéza o nakupování voucherů

Formulovaná hypotéza: „Zákazníci nakupují vstupenky rovnoměrně napříč využívanými portály“.

Aby si mohl kdokoliv vytvořit rezervaci na webových stránkách www.chernobyl86.cz/#rezervace, musí mít nejprve zakoupený voucher, ať už z portálu GoOut či Slevomat. Třetí možností, jež se využívá v případě speciálních a předem domluvených akcí, jako jsou teambuildingy či školní zájezdy, poskytne společnost své vlastní, systémem vygenerované kódy.

Databáze společnosti sice uvádí u každé rezervace kód voucheru, ale neuvádí, z jakého portálu byl tento voucher pořízen. Naštěstí, každý autor kódu (GoOut, Slevomat, Chernobyl) mají svůj vlastní formát, a to:

- GoOut: XXXXXXXXXXXXXXX – 13 znaků,
- Slevomat: XXXX-XXXX-XXX-XXX – 17 znaků,
- Chernobyl: XXXX:XXXX – 9 znaků.

Díky těmto originálním formátům pro každý portál lze snadno identifikovat původ každého kódu, a to za pomoci funkce DÉLKA v MS Excel a následné funkci KDYŽ.

Management si přeje ověřit, zdali jsou nákupy voucherů z obou externích portálů, tzn. GoOut a Slevomat rovnoměrné. Neopomíná na skutečnost, že společnost sama výjimečně generuje vlastní vouchery. Vedení společnosti se vyjádřilo, že by bylo ideální, kdyby maximální hodnota vlastně vygenerovaných voucherů bylo 4–5 % a zbytek byl zakoupen na ostatních dvou webových portálech.

Jelikož se jedná o analýzu kategoriálních (kvalitativních) dat a cílem hypotézy je ověřit, zdali skutečné hodnoty odpovídají hodnotám očekávaným, využije se χ^2 test dobré shody.

Postup:

Tento test pracuje se dvěma typy četností: pozorované (observed – O_i) a očekávané (expected – E_i). Pozorované četnosti jsou ty, které se získají sběrem dat, v tomto případě přes interní databázi a analýzou kódů voucherů. Očekávané hodnoty se vypočítají pomocí vzorce:

$$E_i = n\pi_i,$$

kde π_i je očekávaná relativní četnost.

Z databáze byly zjištěny následující četnosti:

Tab. 13: Četnosti pořízení voucherů

Autor voucheru	Počet
GoOut	1 790
Slevomat	1 884
Vlastní	135
Celkový součet	3 809

Zdroj: vlastní zpracování dle ESCG games s.r.o. (2023)

Nulová a alternativní hypotéza byla stanovena jako:

H_0 : Očekávané a pozorované četnosti se shodují

H_A : Očekávané a pozorované četnosti se neshodují

Testová statistika se následně vypočítá jako:

$$\chi^2 = \sum \frac{(\text{pozorované četnosti} - \text{očekávané četnosti})^2}{\text{očekávané četnosti}},$$

neboli:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i},$$

Pro výpočet testové statistiky je tedy potřebné dosadit hodnoty π_i , vypočítat hodnoty E_i a provést sumu dle výše uvedeného vzorce:

Tab. 14: Postup výpočtu testové statistiky

Autor voucheru	O_i	π_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
GoOut	1 790	0,48	1 828,32	0,8032
Slevomat	1 884	0,48	1 828,32	1,6957
Vlastní	135	0,04	152,36	1,9780
Celkový součet	3 809	1,00	3 809	4,4769

Zdroj: vlastní zpracování (2023)

Test dobré shody má jednu podmínku, a to, aby každá hodnota $E_i > 5$, což je v tomto případě splněno.

Hodnota testové statistiky: $\chi^2 = 4,4769$

p-hodnota se vypočítá pomocí pravostranného testu s χ^2 rozdělením, kdy počet stupňů volnosti = $k - 1$, kde k představuje počet pozorovaných prvků. V tomto případě $k = 3$, jsou tedy 2 stupně volnosti. Hladina významnosti je stanovena na 5 %.

p-hodnota = 0,1066

H_0 nezamítáme: Nebylo prokázáno, že by se očekávané a pozorované četnosti lišily.

Managementu bude tedy předána zpráva, že jejich tvrzení nebylo vyvráceno, nicméně p-hodnota není příliš vysoká. Sice by i při hladině významnosti na 10 % nedošlo k zamítnutí nulové hypotézy, nicméně by se jednalo už o minimální rozdíl.

6 Analýza primárních externích dat

Předchozí kapitola potvrdila tvrzení majitele společnosti ESCG games s.r.o., že návštěvnost únikové hry Chernobyl je klesající, minimálně u hry Záchranná mise. Nicméně nijak nepomohla určit důvod tohoto klesajícího trendu. Identifikace těchto důvodů je možná pouze za pomoci prozkoumání primárních externích dat neboli názorů účastníků her v minulosti. V kapitole 4 byly vyzdvíženy čtyři hlavní možné důvody, které si majitel společnosti přeje prozkoumat, a to ekonomická krize spojená s tendencí šetřit, vyčerpání potenciálu hry, nedostatečná marketingová strategie a snížená kvalita poskytovaných služeb. Pro každý z těchto možných důvodů je stanovena hypotéza:

- 1) Příznivci únikových her v současnosti utrací méně za zábavní průmysl.
- 2) Minulí návštěvníci vítají možnost otevření nové úrovně.
- 3) Zákazníci se o hře dozvěděli převážně pomocí kampaně na sociálních sítích.
- 4) Hráči z roku 2022 jsou s kvalitami služeb stejně spokojeni jako hráči z předchozích let.

Metodou sběru dat pro potvrzení či vyvrácení hypotéz bylo zvoleno dotazníkové šetření v elektronické podobě prostřednictvím Google Forms pro zajištění snadné přehlednosti a upravitelnosti, snadného přístupu a rychlejšího zpracování výsledků. Kontakt s minulými účastníky byl možný pomocí podnikové databáze, ve které jsou uvedeny potřebné kontaktní údaje veškerých návštěvníků únikové hry Chernobyl.

V databázi bylo uloženo celkem 2 848 unikátních emailových adres. Během dvou týdnů dotazník nasbíral 364 odpovědí.

6.1 Hypotéza o tendenci šetřit

Zdali se lidé usmysleli více šetřit se autor rozhodl ověřit dvěma způsoby – pomocí porovnání míry vlivu ekonomické krize na postoj dotazovaných k využívání služeb zábavního průmyslu se stanovenou konstantou a párovým testem.

Na otázku, zdali respondenty ovlivnila současná ekonomická situace postoj k zábavnímu průmyslu, bylo možné odpovědět na škále od 1 do 5, kde nejnižší hodnota reprezentuje odpověď „nijak“ a nejvyšší „výrazně“.

Jelikož hodnoty mohly dosahovat pouze hodnot 1–5, je s jistotou vyloučena normalita dat, což potvrdil Lillieforsův test v SW Statistica, nicméně pro větší výběry ($n > 30$) není

normalita požadována, neboť dle limitní věty má výběrový průměr pro velké výběry normální rozdělení, i když výběr pochází z libovolného rozdělení. Aby mohlo být ověřeno tvrzení, že se lidé rozhodli využívat méně zábavních služeb, bude soubor porovnán s hodnotou 3, která na škále zastupuje možnost „nevím“. Po konzultaci s managementem byl přesto za sledovaný parametr zvolen nikoliv průměr, ale medián. Zvoleným testem tedy je **jednovýběrový Wilcoxonův test**.

K výpočtu bude využito SW Statistica. Do jednoho sloupce proměnné se vloží celý soubor odpovědí na otázku a do druhého sloupce se podél vloží hodnota 3. Poté se zvolí Statistika – Neparametrické statistiky – Porovnání dvou závislých vzorků – Wilcoxonův párový test.

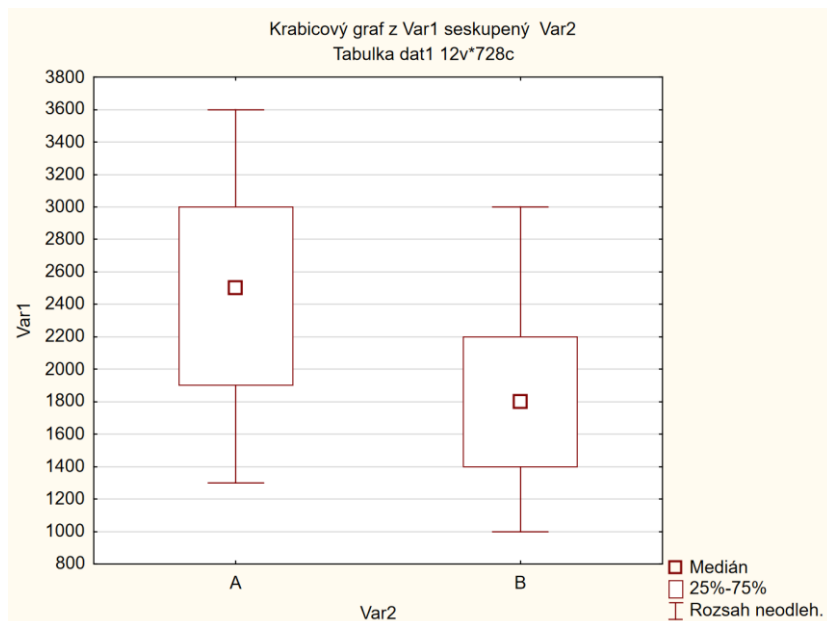
Tab. 15: Průběh a výsledek jednovýběrového Wilcoxonova testu

Test	Jednovýběrový Wilcoxonův test
Hypotézy	$H_0: x_{0,5} = 3$ $H_A: x_{0,5} > 3$
Typ testu	Pravostranný
Testová statistika	1,0923
p-hodnota	0,1374
p-hod $< \alpha$?	NE
Závěr	H_0 nezamítnuta: Nebylo prokázáno, že by současná ekonomická krize ovlivnila postoj respondentů k využívání služeb zábavního průmyslu.

Zdroj: vlastní zpracování (2023)

Dále byli respondenti dotázáni, jakou nejvyšší hodnotu by byli zaplatit za únikovou hru dle určitých požadavků před ekonomickou krizí a kolik by zaplatili nyní. Byly tak získány dva sloupce závislých dat. Tato data byla prozkoumána krabicovým grafem kvůli možné přítomnosti podezřelých či odlehlých dat nutných k ošetření. Naštěstí se zde žádné takové hodnoty nenacházejí, viz obrázek č. 23.

Obr. 10: Krabicový graf maximální ceny



Zdroj: vlastní zpracování v SW Statistica (2023)

Krabicové grafy totiž slouží k vizuálnímu znázornění extrémních hodnot a kvartilů statistického souboru, dovoluje posoudit a porovnat jak centrální tendence dat, tak jejich rozptýlenost. Pokud je potřeba porovnat více statistických znaků jednoho souboru, nebo jeden statistický znak v různých souborech, je možné využít souběžnou analýzu pomocí více krabicových grafů. Toto řešení umožňuje snadné a přehledné porovnání analyzovaných vlastností. Ačkoliv se jedná o velmi oblíbený prostředek pro zobrazení dat, je implementován pouze v některých solidních statistických programových systémech.

Oba soubory nesplnily předpoklady normality (ověřeno Lillieforsovo testem), ovšem opětovně platí pravidlo, že pro výběr $n > 30$ není normalita dat vyžadována. Zvoleným testem je **párový t-test**. Výpočet bude proveden za pomoci nástroje Analýza dat v MS Excel, kde se zvolí „Dvouvýběrový párový t-test pro střední hodnotu“, vyberou oblasti pro 1. a 2. soubor a zpracují výsledky, zaznamenané v tabulce č. 16.

Tab. 16: Průběh a výsledek párového t-testu

Test	Párový t-test
Hypotézy	$H_0: \mu_1 = \mu_2$ $H_A: \mu_1 > \mu_2$
Typ testu	Pravostranný
Testová statistika	16,5632
p-hodnota	0,0000
p-hod < α ?	ANO
Závěr	H_0 zamítnuta: Respondenti jsou ochotni za specifickou únikovou hru zaplatit méně než před ekonomickou krizí.

Zdroj: vlastní zpracování (2023)

6.2 Hypotéza o nové úrovni

Návštěvníci únikové hry Chernobyl byli dotázáni, zdali by uvítali možnost otevření nové, konkrétně třetí únikové hry. Na otázku bylo možné odpovědět pouze „Ano“ nebo „ne“.

Z celkových 364 respondentů odpovědělo na otázku kladně 287, neboli 78,846 %. Vedení společnosti si přeje, aby zájem o novou únikovou hru dosáhl nejméně hodnoty 75 %.

Pro ověření hypotézy bude využit **test o parametru π alternativního rozdělení**.

Tab. 17: Průběh a výsledek testu o parametru π alternativního rozdělení

Test	Test o parametru π alternativního rozdělení
Hypotézy	$H_0: \pi = 0,75$ $H_A: \pi > 0,75$
Typ testu	Pravostranný
Předpoklad	$n > \frac{9}{p(1-p)}$
Splněn předpoklad?	ANO
Vzorec	$\frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}} \cdot \sqrt{n}$
Testová statistika	1,6946
p-hodnota	0,0451
p-hod $< \alpha$?	ANO
Závěr	H_0 zamítnuta: Respondenti mají zájem o novou únikovou hru.

Zdroj: vlastní zpracování (2023)

6.3 Hypotéza o marketingové kampani

Reklamní kampaň vedená přes sociální sítě by se projevila jako neefektivní, pokud by to nebyl jeden z hlavních kanálů sloužící k šíření povědomí o společnosti a jejích službách. Po konzultaci byla požadovaná minimální hodnota poměru lidí, kteří se o únikové hře Chernobyl dozvěděli právě skrze sociální sítě, nastavena na 50 %. Zbylé hodnoty mají být opětovně rovnoměrné. V dotazníku měli respondenti na výběr ze tří možností plus mohli zanechat vlastní odpověď v rámci odpovědi „jiné“, čehož však nikdo nevyužil. Z celkového počtu 364 respondentů označilo reklamu na sociálních sítích jako způsob prvního povědomí o značce 106.

Tab. 18: Průběh a výsledek testu dobré shody

Test	Test dobré shody
Hypotézy	H_0 : Očekávané a pozorované četnosti se shodují H_A : Očekávané a pozorované četnosti se neshodují
Typ testu	Pravostranný
Předpoklad	$E_i > 5$
Splněn předpoklad?	ANO
Vzorec	$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
Testová statistika	107,97
p-hodnota	0
p-hod $< \alpha$?	ANO
Závěr	H_0 zamítnuta: Očekávané a pozorované četnosti se neshodují.

Zdroj: vlastní zpracování (2023)

6.4 Hypotéza o úrovni kvalit služeb

Poslední hypotéza na prozkoumání se zabývá spokojeností respondentů s kvalitou služeb, především těch, co se hry zúčastnili v roce 2022. Pro účely této hypotézy museli respondenti nejen odpovědět na otázku, jak moc byli spokojeni s úrovní poskytnutých služeb na škále 1–5, ale také rok, kdy únikovou hru navštívili. Následuje rozdělení na dva samostatně nezávislé soubory podle roku, kdy odpovědi od účastníků z let 2018-2021 tvoří jeden soubor a odpovědi od účastníků v roce 2022 soubor druhý. Vzhledem k tomu, že data jsou opět na škále od 1–5, bude pozorovaným parametrem medián, bude tedy využit **Mannův-Whitneyův** test. Účelem je zjistit, zdali jsou návštěvníci z roku 2022 únikové hry stejně spokojeni s nabídkou služeb jako zákazníci z let předchozích.

Tab. 19: Průběh a výsledek Mann-Whitneyova testu

Test	Mannův-Whitneyův test
Hypotézy	$H_0: x_{0,5} = y_{0,5}$ $H_A: x_{0,5} \neq y_{0,5}$
Typ testu	Oboustranný
Testová statistika	11,4225
p-hodnota	0,6071
p-hod $< \alpha$?	NE
Závěr	H_0 nezamítnuta: Nebylo potvrzeno, že by byli návštěvníci únikové hry z roku 2022 odlišně spokojeni oproti návštěvníkům z předchozích let.

Zdroj: vlastní zpracování (2023)

7 Výsledná interpretace

Na základě analýzy dat a zjištěných závěrů z kapitol 5 a 6 autor předává společnosti ESCG games s.r.o. následující poznatky:

- Dle analýzy časové osy a predikce na rok 2023 bude návštěvnost nadále klesat, pokud nedojde k adekvátním nápravám, a to především u hry s názvem Záchranná mise, kde se návštěvnost blíží k nule.
- Zákazníci si zakupují vouchery z portálu GoOut a Slevomat téměř rovnoměrně, rozdílnost mezi počtem vytvořených rezervací pomocí kódů pořízených z těchto dvou platforem není statisticky významná.
- Dle průzkumu respondenti mají pocit, že se jejich nákupní preference a postoj k placení za služby v rámci zábavního pocitu statisticky významně nezměnily (přesněji řečeno – nebyla prokázána statisticky významná odlišnost), nicméně;
- V současnosti by ti a samí zákazníci, kteří již únikovou hru Chernobyl navštívili, byli ochotni zaplatit za specifikovanou hru méně než v předchozích letech.
- Dle průzkumu mají předchozí zákazníci zájem o novou úroveň/hru.
- Očekávané četnosti o úspěšnosti šíření jména společnosti přes sociální sítě (přesněji četnosti o tom, jak se respondenti o únikové hře Chernobyl dozvěděli) se neshodují, významný rozdíl mezi očekávanou a skutečnou četností lze vidět právě u testované kampaně přes sociální sítě.
- Výzkum nepotvrdil, že by se kvalita poskytovaných služeb v průběhu let zhoršila.

V rámci těchto zjištění, která byla předána managementu společnosti, doporučuje autor zvážit následující kroky:

- 1) Naplánování nové únikové hry.
- 2) Počáteční nastavení ceny nové únikové hry na nižší cenovou hladinu a její následné navýšení v souvislosti s vývojem ekonomické situace v ČR, především s vývojem inflace.
- 3) Zaslání informace o nové únikové hře všem předchozím návštěvníkům únikové hry Chernobyl.
- 4) Zaslání slevového kupónu na novou únikovou hru všem respondentům dotazníkového šetření.

- 5) Navrhnutí nové marketingové strategie, zvážit využití nejen sociálních sítí, ale také možnost billboardů či spolupráce s osobnostmi a weby zabývající se tematikou Černobylské jaderné havárie, urbexu (prozkoumávání měst, továren či jiných objektů, které vznikly lidskou činností, nespádají pod muzea a nejedná se o snadno přístupné objekty, ale naopak o objekty, které lidé běžně nemohou vidět) apod.

Majitel společnosti ESCG games s.r.o. se vyjádřil, že nad podobnými kroky již uvažoval a že výsledky statistické analýzy podnikových dat ho o jeho nápadu přesvědčila.

Závěr

Cílem diplomové práce bylo navrhnout statistický výzkum pro společnost ESCG games s.r.o. zabývající se provozem únikových her s tematikou výbuchu Černobylské jaderné elektrárny a následně zanalyzovat získaná data za účelem pokusu o odhalení příčin poklesu návštěvnosti zmíněných únikových her, ověření tvrzení managementu o daném snížení rezervací a dodat vedení společnosti doporučení pro následující strategické kroky.

Z důvodu atypického prostředí firem podnikajících v oblasti únikových her byly managementem společnosti identifikovány 4 hlavní podezření vedoucí ke snížení návštěvnosti:

- 1) ekonomická krize a snaha lidí šetřit,
- 2) vyčerpání kapacit únikových her,
- 3) nesprávně nastavená marketingová strategie,
- 4) snížená kvalita nabízených služeb.

Z tohoto důvodu byl navržen smíšený výzkum s výzkumnou otázkou „Jaká je spokojenost zákazníků se zážitkovou hrou?“, na základě které bylo stanoveno pět hypotéz, jež měl statistický výzkum vyjma analýzy časové osy ověřit, a to:

- 1) Příznivci únikových her v současnosti utrací méně za zábavní průmysl.
- 2) Minulí návštěvníci vítají možnost otevření nové úrovně.
- 3) Zákazníci se o hře dozvěděli převážně pomocí reklamní kampaně.
- 4) Zákazníci nakupují vstupenky rovnoměrně napříč využívanými portály.
- 5) Hráči z roku 2022 jsou s kvalitami služeb stejně spokojeni jako hráči z předchozích let.

Analýza interních dat potvrdila klesající trend návštěvnosti obou únikových her, daleko strmější byl identifikován u druhé, dražší z nabízených. Taktéž nevyvrátila hypotézu č. 4 neboli že by si zákazníci nekupovali voucheru rovnoměrně ze dvou aktuálně možných portálů.

Na základě provedeného dotazníkové šetření, kdy z 2 848 unikátních emailových adres bylo obdrženo 364 odpovědí bylo dáno managementu společnosti doporučení, že by se firma měla zaměřit na vybudování nové únikové hry, nastavit na ni nižší cenu a tu postupně zvedat s vývojem ekonomické situace, pozvat na novou hru všechny předešlé zákazníky únikové hry Chernobyl, respondentům nabídnout slevový poukaz

a vymyslet novou marketingovou strategii, ideálně zvážit spolupráci s osobnostmi a jinými společnostmi, které se také zabývají tematikou Černobylu či sportu s názvem urbex.

Výsledná zjištění a doporučení byla předána majiteli společnosti ESCG games s.r.o., který vyjádřil nadšení, neboť ho tyto výsledky potvrdily o krocích, které se mají v následujících měsících uskutečnit.

Seznam použitých zdrojů

- Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D. & Cochran, J. J. (2017). *Statistics for Business & Economics* (13th ed.). Cengage Learning
- Cyhelský, L., Kahounová, J. & Hindls, R. (2001). *Elementární statistická analýza*. Management Press
- Drápela, K. (2012). *Statistické metody I*. Mendelova zemědělská a lesnická univerzita
- Drozdová, J., & Homola, V. (2017). *Statistika pro Geovědní a montánní turismus*. Technická univerzita Ostrava
- Ed's Scribbles (2012). *Correlation*. Dostupné 2. 4. 2023 z <https://scribbles.dot-ed.com/post/34776727747/correlation>
- Eger, L., & Egerová, D. (2014). *Základy metodologie výzkumu*. Západočeská univerzita v Plzni.
- Eger, L., & Egerová, D. (2022). *Metodologie výzkumu*. Západočeská univerzita v Plzni.
- ESCG games s.r.o. (2023). *Historie rezervací*. Interní dokument podniku ESCG games s.r.o. se sídlem v Praze.
- Finn, M., White, M. E. & Walton, M. (2000) *Tourism and Leisure Research Methods: Data Collection, Analysis, and Interpretation*. Pearson.
- Hendl, J. (2006). *Přehled statistických metod zpracování dat: analýza a metaanalýza dat*. Portál
- Hendl, J., & Blahuš, P. (2012). *Metodologie výzkumné práce. Jak na to?* Dostupné 20. 4. 2023 z <http://www.ftvs.cuni.cz/hendl/metodologie/index1.htm>
- Hindls, R., Hronová, S, Seger, J. & Fischer, J. (2006). *Statistika pro ekonomy*. Professional Publishing
- Kotler, P. & Keller, K. (2013). *Marketing Management*. Grada.
- Kropáč, J. (2007). *Statistika B: jednorozměrné a dvourozměrné datové soubory, regresní analýza, časové řady*. Vysoké učení technické v Brně
- Nisbet, R., Elder, J. & Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*. Elsevier Inc.
- Odom, L. R. & Morrow, J. R. Jr. (2009). What's this r? A Correlational Approach to Explaining Validity, Reliability and Objectivity Coefficients. *Measurement in Physical Education and Exercise Science*, 10(2), 137-145. https://doi.org/10.1207/s15327841mpee1002_5
- Pavlica, K. a kolektiv (2000). *Sociální výzkum, podnik a management*. Ekopress
- Peck, R., Olsen, C., & Devore, J. (2015). *Introduction to statistics and data analysis*. Cengage Learning
- Rice, J. A. (2007). *Mathematical Statistics and Data Analysis* (3rd ed.). Thomson Brooks/Cole
- Roopa, S., & Rani, M. S. (2012). Questionnaire Designing for a Survey. *J Ind Orthod Soc*, 46(4), 273-277. <https://doi.org/10.5005/jp-journals-10021-1104>

- Světlík, J. (2018). *Marketing: Cesta k trhu*. Vysoká škola podnikání a práva, a. s.
- Vondroušová, K. (2019). *Statistická analýza dat pro kvantitativní výzkum, 1. díl*. Ostravská univerzita
- Walker, I. (2013). *Výzkumné metody a statistika*. Grada
- Weaver, D., & Lawton, L. (2002). *Tourism Management*. Wiley
- Wegner, T. (2013). *Applied Business Statistics: Methods and Excel-based Applications*. Juta and Company
- Zháněl, J., Hellenbrandt, V. & Sebera, M. (2014). *Metodologie výzkumné práce*. Masarykova univerzita

Seznam tabulek

Tab. 1: Příklady statistických jednotek a znaků	12
Tab. 2: Charakteristiky kvalitativního, kvantitativního a smíšeného výzkumu	13
Tab. 3: Výhody a nevýhody kvalitativního výzkumu.....	15
Tab. 4: Výhody a nevýhody kvalitativního výzkumu.....	18
Tab. 5: Chyby I. a II. druhu a jejich pravděpodobnosti	27
Tab. 6: Počet studentů a kvartální tržby poboček pizzerie	33
Tab. 7: Příklady regresních funkcí s jednou nezávislou proměnnou:.....	36
Tab. 8: Časový plán výzkumu pro společnost ESCG games s.r.o.....	42
Tab. 9: Počet měsíčně odehraných her Dozimetrická věž 2018–2022 (bez 2021).....	45
Tab. 10: Výpočet hodnot pro časovou osu Dozimetrické věže	47
Tab. 11: Počet měsíčně odehraných her Záchranná mise 2018–2022 (bez 2021).....	49
Tab. 12: Výpočet hodnot pro časovou osu Záchranné mise	50
Tab. 13: Četnosti porřízení voucherů.....	52
Tab. 14: Postup výpočtu testové statistiky.....	53
Tab. 15: Průběh a výsledek jednovýběrového Wilcoxonova testu.....	56
Tab. 16: Průběh a výsledek párového t-testu.....	58
Tab. 17: Průběh a výsledek testu o parametru π alternativního rozdělení.....	59
Tab. 18: Průběh a výsledek testu dobré shody.....	60
Tab. 19: Průběh a výsledek Mann-Whitneyova testu	61

Seznam obrázků

Obr. 1: Vizualizace kritického oboru.....	29
Obr. 2: Vizualizace počtu studentů a kvartálních tržeb poboček pizzerie.....	33
Obr. 3: Proložení bodů přímkou pomocí metody nejmenších čtverců	34
Obr. 4: Porovnání lineární a parabolické regresní funkce	35
Obr. 5: Hodnoty Pearsonova korelačního koeficientu.....	37
Obr. 6: Počet měsíčně odehraných her Dozimetrická věž 2018–2022 (bez 2021).....	45
Obr. 7: Návštěvnost Dozimetrické věže 2018–2022 (bez 2021) s predikcí pro 2023	48
Obr. 8: Počet měsíčně odehraných her Záchranná mise 2018–2022 (bez 2021).....	49
Obr. 9: Návštěvnost Záchranné mise 2018–2022 (bez 2021) s predikcí pro 2023.....	51
Obr. 10: Krabicový graf maximální ceny	57

Seznam příloh

Příloha A: Ukázka dat v databázi Chernobyl se zakrytými osobními údaji

Příloha A: Ukázka dat v databázi Chernobyl se zakrytými osobními údaji

Historie rezervací

Datum	Voucher	Hra	Jméno a příjmení	Email	Telefon	Poznámka
středa 29.9.2021						
21:00	1555-2068-00A-987	Záchranná mise				Slevomat 2afk3892
19:00	3678-4952-30A-104	Záchranná mise				
17:00	8766506649991	Dozimetrická věž				4 osoby
17:00	8942067168335	Záchranná mise				6 osob
neděle 26.9.2021						
19:00	7225-4719-10A-541	Dozimetrická věž				
15:00	7548-0489-20A-879	Záchranná mise				
sobota 25.9.2021						
19:00	7717-9227-70A-674	Dozimetrická věž				Slevomat
19:00	8939386160713	Záchranná mise				
17:00	8202-8446-20A-837	Dozimetrická věž				
17:00	6610-5650-30A-168	Záchranná mise				
15:00	3922-2506-60A-511	Záchranná mise				

Abstrakt

Elhota, M. (2023). *Statistická analýza podnikových dat* [Diplomová práce, Západočeská univerzita v Plzni].

Klíčová slova: statistický výzkum, kvantitativní výzkum, statistická analýza, podniková data, primární data, dotazníkové šetření, predikce

Diplomová práce na téma „Statistická analýza podnikových dat“ se zabývá společností provozující únikovou hru a jejím problémem se snižující se návštěvností. V úvodní části práce je společnost i s problematikou detailně popsána. Další část se věnuje metodologii kvalitativního i kvantitativního výzkumu, sběru dat, zajištění kvality měření a sběru dat, jednotlivým fázím statistického šetření a metodice statistické analýzy, především prostřednictvím testování hypotéz. V praktické části dochází ke zvolení konkrétních metod, stanovení výzkumných otázek a následnému sestavení a provedení statistického šetření. Interní podniková data a sesbíraná externí data jsou poté analyzována a testována dle stanovených hypotéz za účelem odhalení důvodů k poklesu návštěvnosti. Pozorované skutečnosti jsou shrnuty v závěru práce v podobě doporučení následujících strategických kroků.

Abstract

Elhota, M. (2023). *Statistical analysis of business data* [Master's Thesis, University of West Bohemia].

Key words: Statistical research, quantitative research, statistical analysis, business data, primary data, questionnaire survey, prediction.

The Master's Thesis on "Statistical Analysis of Company Data" deals with an escape room company and its problem of decreasing attendance. The introductory part of the thesis provides a detailed description of the company and its issues. The following section focuses on the methodology of qualitative and quantitative research, data collection, quality assurance, the phases of statistical investigation, and the methodology of statistical analysis, primarily through hypothesis testing. In the practical part, specific methods are selected, research questions are established, and a statistical investigation is designed and conducted. Internal company data and external data are analyzed and tested based on established hypotheses to uncover the reasons for the attendance decline. The observed facts are summarized in the conclusion of the thesis in the form of recommendations for strategic actions to be taken.