



Ekonomická
fakulta
Faculty
of Economics

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Jihočeská univerzita v Českých Budějovicích

Ekonomická fakulta

Katedra aplikované matematiky a informatiky

Bakalářská práce

Predikce sázkových kurzů na základě logistické regrese

Vypracoval: Pavel Janda

Vedoucí práce: Mrkvička Tomáš, doc. RNDr. Ph.D.

České Budějovice 2018

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH

Fakulta ekonomická

Akademický rok: 2016/2017

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Pavel JANDA**
Osobní číslo: **E15765**
Studijní program: **B6208 Ekonomika a management**
Studijní obor: **Obchodní podnikání**
Název tématu: **Predikce sázkových kurzů na základě logistické regrese**
Zadávající katedra: **Katedra aplikované matematiky a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Cílem bakalářské práce je vytvoření modelu pravděpodobnosti výhry týmu ve sportu na základě logistické regrese, kde vysvětlující parametry budou pořadí týmu v tabulce, předchozí vzájemná utkání a domácí hřiště. Predikce bude srovnána s kurzy jednotlivých sázkových kanceláří.

Metodický postup:

1. Studium literatury - 1. semestr.
2. Sběr dat - 2. semestr.
3. Vyhodnocení informací - 2. semestr.
4. Zpracování a interpretace výsledků - 3. semestr.

Rozsah grafických prací: **dle potřeby**
Rozsah pracovní zprávy: **40 - 50 stran**
Forma zpracování bakalářské práce: **tištěná**


Seznam odborné literatury:

1. McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman and Halls.
2. Crawley, M. J. (2007). *The R Book*. Chichester: Wiley and son.
3. Petronius. (2015). *Kurzové sázení krok za krokem*. e-kniha, Petr Šístek.
4. Další časopisecká a knižní literatura dle zaměření práce.

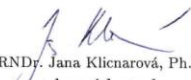
Vedoucí bakalářské práce: **doc. RNDr. Tomáš Mrkvička, Ph.D.**
Katedra aplikované matematiky a informatiky

Datum zadání bakalářské práce: **11. února 2017**

Termín odevzdání bakalářské práce: **13. dubna 2018**


doc. Ing. Ladislav Rolínek, Ph.D.
děkan

JIHOČESKÁ UNIVERZITA
V ČESKÝCH BUDĚJOVICÍCH
EKONOMICKÁ FAKULTA
STUDIJSKÁ 13 (26)
370 05 České Budějovice


RNDr. Jana Klícnarová, Ph.D.
vedoucí katedry

V Českých Budějovicích dne 14. března 2017

Prohlášení

Prohlašuji, že svoji bakalářskou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to – v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejich internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne

.....

Pavel Janda

Poděkování

Rád bych na tomto místě poděkoval vedoucímu mé bakalářské práce doc. RNDr. Tomášovi Mrkvičkovi, Ph.D. za obětavou spolupráci, za čas, který mi byl ochoten věnovat po celou dobu psaní mé bakalářské práce a také v neposlední řadě za řadu podnětných návrhů, které jsem mohl ve své práci využít.

Abstrakt

Jméno a příjmení: Pavel JANDA

Obor studia: Obchodní podnikání

Univerzita: Jihočeská univerzita v Českých Budějovicích

Fakulta: Ekonomická fakulta

Ročník studia: 3

Kurzové sázení v mnoha oblastech lidského života se stává velmi populárním. Tato bakalářská práce se zabývá predikcí sázkových kurzů založených na logistické regresi. Tato práce popisuje obecné informace o sázkových kurzech, historii sázek jak v České republice, tak i v zahraničí, formy sázek, kdo je bookmaker a jak jsou sázkové kurzy vytvářeny. K dispozici jsou 3 sady dat. Konkrétně se jedná o výhodu domácího hřiště, předzápasová tabulka a předchozí vzájemné zápasy. Data jsou zpracovány logistickou regresí za účelem predikce sázkových kurzů. Tři výše zmíněné podmínky jsou zahrnuty do programu STATISTICA, který počítá pravděpodobnost každé varianty na základě logistické regrese. Výsledku z programu STATISTICA jsou následně porovnány s vypsanými kurzy sázkové kanceláře.

Klíčová slova: Statistika, statistická regrese, sázkové kurzy

Abstract

Name and surname: Pavel JANDA

Field of study: Management of Commerce

University: University of South Bohemia in České Budějovice

Faculty: Faculty of Economics

Year of Study: 3

Betting odds in various areas of human life is becoming very popular. This Bachelor's Thesis deals with the prediction of betting odds based on logistic regression. This work describes general information about betting odds, the history of betting odds both here in the Czech Republic and abroad, the forms of betting odds, who is a bookmaker and how betting odds are created. There are 3 sets of data. Specifically, it is the advantage of the home ground, the pre-match table and previous matches. Data is processed by logistic regression to predict betting odds. The three above conditions are included in the Statistica program, which calculates the probability of each variant based on logistic regression. The results from the Statistica program are then compared to the actually listed betting odds.

Key words: Statistics, statistic regression, betting odds

Obsah

Obsah	1
1 Úvod	3
1.1 Výběr tématu	4
1.2 Cíle práce.....	4
Část první – Představení kurzového sázení.....	5
2 Co to vlastně kurzové sázení je?	5
3 Historie kurzového sázení.....	6
3.1 Historie kurzového sázení	6
3.1.1 V ČR.....	6
3.1.2 Ve světě.....	7
4 Co jsou to kurzové sázky?	8
4.1 Princip kurzového sázení.....	8
4.2 Základní typy sázkových příležitostí a druhy sázek v tuzemských sázkových provozovnách	8
4.3 Stanovení kurzu	10
4.4 Skutečnosti ovlivňující výši kurzu	11
4.5 Kdo je to vlastně bookmaker?.....	13
4.6 Formy kursového sázení v praxi	14
4.6.1 Internetové sázení	14
4.6.2 Výhody a nevýhody internetové sázení	14
4.6.3 Telefon	16
4.6.4 Mobilní telefon (Smartphone).....	16
4.7 Formát kurzu	16
4.8 Manipulační poplatek	16
Část druhá – Představení použité metody a výpočetního programu	18
5 Statistické metody a program STATISTICA	18
5.1 Generalizovaný (zobecněný) lineární model.....	18
5.2 Logit	19
5.2.1 Použití a vlastnosti Logitu.....	21
5.3 Logistická regrese	22
5.3.1 Příklady.....	23
5.3.2 Pojmy související s logistickou regresí	24
5.3.3 Vstupní data pro program STATISTICA.....	24
5.4 Program STATISTICA	26

5.4.1	Grafika	26
5.4.2	Uživatelské rozhraní.....	27
5.4.3	Moje verze programu.....	27
Část třetí – Praktické výsledky a výpočty.....		29
6	Praktická část	29
6.1	Vstupní data	29
6.1.1	Popis datového souboru	29
6.1.2	Příprava dat	29
6.1.3	Popis podmínek (proměnných)	30
6.2	Popis experimentů.....	30
6.2.1	Cíle experimentů.....	30
6.2.2	Zkoumaná soutěž.....	31
6.2.3	Sázková kancelář.....	32
6.2.4	Postup výpočtu	33
6.3	Experimenty	36
6.3.1	Kolika % odchylky v kurzech byli v mém výpočtu oproti sázkové kanceláři Tipsport	36
6.3.2	V kolika % se sázková kancelář Tipsport trefila svým nejnižším kurzem do výsledku utkání	37
6.3.3	Jaké kurzy vycházejí nejčastěji.....	38
6.3.4	V jaké části sezóny se vyplatí sázet dle předpokladů sázkové kanceláře.....	39
6.3.5	Vyplatilo by se sázet za 10 Kč vždy na nejnižší kurz spíše u specializované sázkové kanceláře nebo podle vytvořeného modelu skrze statistický software?	41
7	Závěr.....	42
8	Summary.....	43
9	Zdroje a citace.....	44
10	Seznam obrázků, tabulek a grafů.....	47

1 Úvod

Poslední dobou se neustále častěji využívá spojení dříve ne tak propojených dvou světů jako je sport a statistika. Statistika začíná pomáhat ve sportovním světě nejen hráčům, trenérům či různým činovníkům klubů ve všemožných sportech po celém světě, ale také získává stále důležitější místo v mysli sázejících na sportovní události a zápasy. V čím dál tím více digitalizovaném světě, zejména díky internetu a chytrým telefonům již není problém najít téměř jakýkoliv statistický ukazatel, který by mohl pomoci sázejícímu odhadnout výsledek zápasu a následně na něj vsadit s minimem rizika a poměrně vysokou šancí na vítězství. Toto všechno je navíc možné v podstatě odkudkoli na světě a také v jakoukoli denní dobu. Samozřejmě se jedná o sport a různé překvapení či nepředpokladatelné okolnosti k němu patří, ale asi každý by chtěl umět předpovědět budoucnost a umět si tak poměrně bez velké námahy s pomocí legálních sázek na sportovní zápasy trochu přivydělat. V této práci se tedy budu snažit za pomocí pár proměnných předpovědět pravděpodobnost výsledků jednotlivých zápasů 1. anglické fotbalové ligy, kterou můžete znát pod názvem Premier League. Kromě hlavního experimentu, kde budu porovnávat kurzy, které vypočítám díky logistické regresi se sázkovými kurzy společnosti Tipsport, jsem se pokusil odpovědět ještě na pár otázek s tímto tématem úzce souvisejících. Jedná se o procentuální odchylky mých kurzů od kurzů Tipsportu, v kolika procentech se sázková kancelář Tipsport trefila svým nejnižším kurzem do správného zápasového výsledku, jaké kurzy vycházejí nejčastěji, v jaké fázi soutěže se nejvíce vyplatí sázet na favority neboli na nejnižší vypsané kurzy a na závěr i to, jestli by se vyplatilo sázet 10 Kč na nejnižší kurzy od Tipsportu nebo případně na kurzy z mnou vytvořené logistické regrese.

1.1 Výběr tématu

Protože už od útlého věku jsem hrál fotbal aktivně a do současnosti sleduju víceméně veškeré sportovní dění ve světě, tak bylo vybrání tématu pro bakalářskou práci, které se alespoň částečně bude dotýkat sportovního světa nasnadě. Po přečtení několika článků na téma sport a statistika a zhlédnutí několika málo filmů na to samé téma, mě toto spojení zaujalo a rozhodl jsem se mu věnovat o něco hlouběji.

1.2 Cíle práce

Cílem mé práce bylo za pomoci několika málo proměnných zkusit vytvořit takový model logistické regrese, který by mohl být úspěšnější v odhadování fotbalových výsledků než systémy a schopnosti profesionálních bookmakerů a sázkových kanceláří. Pro porovnání jsem si vybral sázkovou společnost Tipsport, která se stala jedničkou na českém sázkařském poli.

Část první – Představení kurzového sázení

2 Co to vlastně kurzové sázení je?

Samotná činnost, která se nazývá kurzové sázení, je taková činnost při které se sázející snaží co možná nejpřesněji a nejpravděpodobněji odhadnout konečný stav či výsledek nějakého zápasu, turnaje či události. Tento svůj odhad poté musí podpořit svými vloženými penězi neboli sázkou. V případě správného odhadu se vklad (vložené peníze) mohou až několikanásobně zhodnotit na základě kurzu, který byl vypsán profesionálními bookmakery od sázkových kanceláří. Vložené peníze se násobí kurzem, ve kterém byla daná sázka uzavřena. Může se tedy stát, že vložené peníze se mohou i několikanásobně zhodnotit. Kurzové sázení jako takové existuje přibližně od 17. století. Od začátku bylo a stále v některých případech je považováno za podobu hazardu. Ale na rozdíl od různých hazardních her je možno u sportovního kurzového sázení ovlivňovat svoji úspěšnost vlastní znalostí případně zkušenostmi z minulosti. Kurzové sázení jako takové bude popsáno na dalších řádcích této práce. (Historie kurzového sázení, 2016)

3 Historie kurzového sázení

3.1 Historie kurzového sázení

3.1.1 V ČR

Sázkařské prostředí v české kotlině začínalo, tak jako to britské a světové na dostihovém závodišti ve Velké Chuchli v Praze. To vše samozřejmě v ilegalitě. Vše se začalo měnit kolem roku 1990. V tento rok založila skupina několika sázkařů sázkovou kancelář, kterou v současné době můžeme znát pod názvem Fortuna. Hned rok na to vznikla druhá sázková kancelář na českém trhu. Jednalo se o dnešní jedničku na našem trhu, tedy Tipsport. Vůbec celá devadesátá léta byla obdobím, kdy se kurzové sázky na našem území asi nejvíce rozrůstaly a s tím souvisí i to, že v tomto období byl zaznamenán velice prudký nárůst samotných sázkových kanceláří, ale také se to samozřejmě odrazilo i v počtu sázejících. Na českém trhu v současné době působí hlavně čtyři sázkové kanceláře. Jmenovitě se jedná o Tipsport (přibližně 1000 poboček), který jak jsem již zmínil je jasnou jedničkou na českém trhu. Druhé místo zaujímá Fortuna (více než 600 poboček) a také SynotTip a Chance, která ovšem již pár let patří pod Tipsport.

Obliba kurzového sázení v České republice neustále stoupá. Na základě údajů ministerstva financí se před 8 lety v roce 2010 prosázelo necelých 8 miliard Kč a na samotných výhrách bylo vyplaceno přes 5 miliard Kč. Prosázené peníze, ale byli mnohem větší, protože se některé druhy sázení po internetu zařadily do kategorie technických her.

Prvopočátek internetového sázení ve světě můžeme vypočítat již ke konci 90. let 20. století. Postupně se internetové sázení stává oblíbenějším než sázení na kamenné pobočce. V již zmíněných 90. letech 20. století si museli Češi, když chtěli sázet přes internet otevřít účet u některé ze zahraničních sázkových kanceláří. V Česku nastal průlom v internetovém sázení až v roce 2009. V tomto roce byla v novele loterijního zákona schválena možnost spuštění internetové aplikace na sázení u jednotlivých sázkových kanceláří i na území ČR. (Vincent, 2009; Tůma, 2011)

V poslední době zejména díky obrovské medializaci světového sportu a také díky neustále se zvyšující roli a významu internetu se neustále navyšuje počet sázejících na sportovní události. Tato skutečnost také tlačí na sázkové kanceláře, aby neustále rozšiřovali nabídku událostí, na které je možné si vsadit a také možnosti jak si na dané události vsadit. V dnešní době si může již v podstatě každý vsadit na různé sportovní politické či jiné události odkudkoli na světě v kteroukoli denní dobu. Zejména v posledních pár letech dokonce samotné sázkové kanceláře nabízejí na svých internetových stránkách přímé přenosy daných událostí. Sázející si tak může poměrně jednoduše v přímém přenosu zkontrolovat, zda jeho tip byl správný či nikoliv. (Historie kurzového sázení, 2016)

3.1.2 Ve světě

Již na rok 1635 připadá pravděpodobný začátek sázkařské historie. V Londýně proti sobě dva pánové postavili své koně do dostihového souboje a každý z nich vsadil na výsledek tohoto závodu stejné peníze. V současnosti by tato situace znamenala kurz ve výši 2,00 na každého z účastníků. V dalších letech právě zejména dostihové závody přitahovali sázející jedince. Postupem času lidé, ale začali čím dál tím více také sázet na různé organizované soutěže, ligy a poháry. Dlouhá léta bylo kurzové sázení, které je populární zejména v USA a na britských ostrovech, kde bylo úplné povolení kurzového sázení schváleno až v roce 1961 vnímáno jako hazard a pohybovalo se na hraně i za hranou zákona. Jednotliví pracovníci v tomto odvětví v tehdejší době měli pár svých klientů a dělali jim v podstatě osobní bookmakery. Až někdy kolem 30. let 20. století se rozjel byznys s hromadnými kurzovými sázkami. Starala se o něj tehdejší mafie z Ameriky. Jak již bylo zmíněno, tak nezbytností pro popularitu kurzového sázení byl bezesporu nástup internetu a internetových aplikací. V současnosti sídlí velké procento heren na britských ostrovech nebo různě po světě v daňových rájích jako je kupříkladu Malta a tak dále. (Historie kurzového sázení, 2016)

4 Co jsou to kurzové sázky?

4.1 Princip kurzového sázení

Principem činnosti kurzového sázení je snaha o co nejpřesnější odhad výsledku některé sportovní, společenské či třeba politické události samotným sázkařem. Za účelem finanční výhry musí být tento odhad podpořen sázkou – nejčastěji peněžním vkladem. Jednotlivé sázkové kanceláře a jejich bookmakeři zvažují pravděpodobnosti výskytu všech možných scénářů a výsledků u dané události. Na základě těchto posudků, názorů a zkušeností, pak vypisují na danou událost jednotlivé kurzy. V případě výhry, tedy správného odhadu sázkaře se vložená částka vynásobí s vypsáním kurzem od sázkové společnosti a tato výsledná částka se vyplatí sázejícímu. V případě prohry neboli špatného odhadu sázejícího propadá veškerý vsazený finanční obnos sázkové kanceláři. (Kurzové sázení, 2018)

4.2 Základní typy sázkových příležitostí a druhy sázek v tuzemských sázkových provozovnách

a) Základní druhy sázek

DUO – v tomto druhu sázek jsou možné jen dva typy výsledků (výhra nebo prohra). Tuto možnost splňuje například tenis či badminton.

KVINTA – tento druh sázek se vyskytuje u utkání, kde mohou nastat tři různé výsledky (výhra, remíza či prohra). Všechny sázkové kanceláře navíc podporují i sázky na tzv. neprohru. To znamená, že při sázce na neprohru domácích sázející vyhrává i v případě remízy. Tyto podmínky splňuje například fotbal, hokej nebo výjimečně i basketbal či házená.

VICTORIA – tento typ sázek se stává neustále populárnějším. Typ sázek VICTORIA umožňuje sázet na vítězství nebo na umístění jednotlivého sportovce či mužstva v jednorázových sportovních soutěžích nebo v disciplínách, které mohou trvat i celý rok. Jedná se například o tipování umístění nějakého týmu v ligové soutěži. (Podrobná pravidla sázek, 2016)

Na evropském kontinentu je nejčastější sázka na konečný stav či výsledek utkání. Tomuto typu sázky se jinak také říká *moneyline*. Oproti tomu na asijském kontinentu a v USA je nejčastější sázka tzv. *asijský handicap*. Asijský handicap neznačí tip na vítěze utkání, ale ukazuje o kolik gólů či bodů musí favorit utkání vyhrát. V opačném případě o kolik bodů či gólů může outsider prohrát. Samotný tip pak vypadá například následovně: Nymburk -10,5. Aby byla sázka úspěšná, tak je nutné, aby Nymburk vyhrál nad svým soupeřem alespoň o 11 bodů. Naopak v případě, že se vsadí na outsidera, tak to vypadá následovně: Dukla Praha + 1,5. V tomto případě musí Dukla Praha buď zvítězit, zápas zremizovat nebo podlehnout svému soupeři o maximálně jeden gól nebo bod. Dalším typem je over/under. Tato sázka značí kolik si sázkař myslí, že padne gólů či bodů. Když se vsadí například Over 5,5, tak k vítězství je třeba alespoň 6 gólů. Tento typ sázek je také zatím více populární mimo evropský kontinent. Zejména v poslední době se začíná čím dál tím více sázet na tzv. *futures*. Jedná se o dlouhodobé sázky kupříkladu na vítěze některé soutěže či poháru. Posledním z neznámějších typů je tzv. *props*. Jedná se o sázky typu, že některý hráč dá alespoň 2 góly, v zápase padne maximálně 6 žlutých karet nebo, že bude v zápase odpískaná penalta a tak dále. Velice oblíbenou sázkou tohoto typu je také, zda při losu mincí na začátku zápasu padne panna nebo orel. (České sázkové kanceláře a svět internetového sázení, 2009)

a) Nejčastější typy sázek

Sólo sázka – druh Sázky, která obsahuje jen jeden tip

AKU sázka – Sázka, která obsahuje 2 a více tipů. Na každém tiketu nelze mít 2 a více tipů na jeden zápas. Maximální počet tipů je 50. Aby byla sázka úspěšná, je nutné, aby vyšli beze zbytku všechny tipy na tiketu. Jak bylo již výše zmíněno, tak výhra se vypočítává, tak že se vynásobí finanční vklad se součtem kurzů na tiketu.

Například, když se vsadí 2 zápasy v kurzech 2 a 3 při vkladu 20 Kč, tak výhra bude činit 120 Kč ($2 * 3 * 20 = 120$ Kč).

Extratip – typ Sázky, kde je dovolena jedna chyba na tiketu. Ta chyba, ale může nastat jen v tzv. Extratipu. V případě, že tiket obsahuje minimálně 10 tipů, tak si může sázející tiket doplnit o Extratip. V případě, že celý tiket vyjde a Extratip jako jediný nevyjde, sázející má stále nárok na výhru do které se Extratip nepočítá. Hodnota Extratipu musí nabývat maximálně takové hodnoty, která se rovná nejnižšímu kurzu na zbytku tiketu. V případě, že Extratip vyjde, ale nevyjde jakýkoliv jiný tip, tak sázející prohrává. (Herní plán pro provozování internetových kurzových sázek, 2017)

4.3 Stanovení kurzu

Čím má dle mínění bookmakerů daná možnost nebo daný jev vyšší pravděpodobnost výskytu, tak tím nižší kurz je na tuto možnost vypsáný a s tím související možná výhra je také nižší. A naopak, když je nějaká možnost málo pravděpodobná, tak kurz vypsáný sázkovou kanceláří bude o to vyšší a potenciální výhra také. Kurzy musejí vždy nabývat hodnoty, které jsou vyšší než jedna. To znamená, že nejnižší kurz, na který lze vsadit má hodnotu 1,01. Nižší kurzy nejsou ani vypisovány a nelze na ně vsadit. Opačná hranice kurzů není omezena, proto zejména u společenských sázek lze vsadit i na kurz 1000 a vyšší. Bookmakeři jednotlivých sázkových kanceláří upravují vypsané kurzy až do začátku dané události. Činí tak zejména dle nejnovějších informací a také podle již vsazeného obnosu peněz na některou z možností. (Kurzové sázení, 2018)

Nejdříve je třeba si vysvětlit, co to vlastně ten vypsáný kurz je. Zjednodušeně řečeno se jedná o pravděpodobnost. Jedná se o vyjádření šance a pravděpodobnosti na uskutečnění některého z možných jevů. Například ve sportu se jedná o pravděpodobnosti na výhru jednotlivých hráčů či týmů. Samotné kurzy se ale nejčastěji neuvádějí v procentech, ale pomocí desetinného čísla. Tato hodnota je ještě upravena o

marži, kterou si ke kurzu přidá samotná sázková kancelář. (Sázkový kurz, 2018; Vincent, 2009)

4.4 Skutečnosti ovlivňující výši kurzu

Hodnota kurzu vždy odpovídá aktuálnímu dění či aktuálním informacím jako jsou třeba forma týmů nebo hráčů, minulé vzájemné zápasy, postavení v tabulce a tak dále. Kurz dále také ovlivňuje výhoda domácího hřiště, prostředí a fanoušků nebo například týmová souhra či připravenost hráčů po fyzické a psychické stránce. Navíc jak už bylo zmíněno, tak bookmakeři a sázkové kanceláře mohou výši kurzu ovlivňovat v podstatě až do začátku utkání na základě aktuálnějších informací.

Jako v každém jiném byznysu platí i u kurzových sázek, že kurz, který daná sázková kancelář nabízí, tak musí být konkurenceschopný vůči nabídkám jiných sázkových kanceláří. Rozdílnost kurzů od různých poskytovatelů je vždy úměrný výši tohoto kurzu. Čím nižší kurz a s tím spojená i větší pravděpodobnost naplnění, tím budou menší odchylky u různých sázkových kanceláří. Naopak u vysokých kurzů se můžeme setkat s opravdu znatelnými rozdíly.

Jedním z dalších faktorů, které ovlivňují výši vypsání kurzu je i výše vsazeného finančního obnosu na danou příležitost. Jednoznačným cílem sázkové kanceláře je, aby vklady na špatnou možnost u daného zápasu pokryli a spíše překonali výhry, které se musí vyplatit sázejícím, kteří se trefili správně. Rozdíl mezi těmito hodnotami je konečný zisk či případná ztráta.

Existují i situace, kdy jsou kurzy nastaveny od sázkové kanceláře tak, že při jakémkoliv výsledku dané události nemá kancelář z této události absolutně žádný zisk. Toto jednání má za cíl nabídnout výhodnější kurzy pro sázející než jiná sázková kancelář, což by mělo mít za následek získání nových zákazníků. Zmíněné jednání se obvykle používá při velkých sportovních událostech, kdy se rozhoduje nejvíce lidí, zda nezačít sázet. Jedním z hlavních činitelů při výběru právě bývá velmi často i výše nabízených kurzů od jednotlivých konkurenčních sázkových kanceláří. (Skutečnosti ovlivňující výši kurzu, 2018)

Ještě bych se vrátil k marži sázkových kanceláří. Samotná marže jinak také řečeno playback bývá definována následujícími okolnostmi: typem nebo druhem zápasu a také strategií dané kanceláře. Druh neboli typ zápasu v tomto případě vyjadřuje nakolik si je bookmaker jistý a nakolik je znalý v prostředí celé soutěže. U nejznámějších soutěží z celého světa, ať už se jedná o 1. fotbalovou ligu v Anglii, kterou budu studovat i já v této práci nebo například severoamerickou hokejovou NHL budou vypsány kurzy od bookmakerů o něco vyšší. Je to dané tím, že u největších, nejlepších a nejsledovanějších soutěží téměř není možnost, aby profesionálnímu bookmakerovi utekla nějaká novinka či zásadní informace, která by mohla ovlivnit výsledek zápasu. Naopak u druhé fotbalové argentinské ligy tyto informace ve většině případů bookmakeři k dispozici nemají, a proto se vypsány kurzy pohybují v nižších hladinách. Druhou okolností, která ovlivňuje výši vypsání kurzu je strategie dané sázkové kanceláře. Strategie určuje, jak se chce bookmaker a daná sázková kancelář prosadit na velmi drsném a konkurenčním trhu. Jsou tři možné cesty. První cestou by byly vysoké kurzy. Vysoké kurzy mají za následek menší zisky. Druhou možnou cestou je agresivní marketing. Pro agresivní marketing je nutné mít nějaký rozpočet. Na tento požadovaný rozpočet se ale nedosáhne pomocí vysokých kurzů, neboť jak jsem již zmínil, tak vysoké kurzy znamenají menší zisky a tudíž i menší rozpočet pro marketing. Většina sázkových kanceláří využívá zlatou střední cestu. (Vincent, 2009)

Na příkladu si můžeme ukázat, jak bookmaker vypisuje svoje kurzy. Vybereme si náhodný zápas z první české fotbalové ligy. Jako příklad lze použít pražské derby Sparta Praha – Dukla Praha. Poté co bookmaker vyhodnotí všechny informace, které o tomto konkrétním zápase má, může vytvořit svůj, tzv. autorský kurz. Bookmaker si může například myslet, že Sparta Praha zvítězí nad Duklou na 75 %. Remíze přisoudí 15 % šanci a tomu, že vyhraje Dukla Praha moc nevěří, a proto pravděpodobnost této varianty odhaduje na zbylých 10 %. Ideálně by se pak konečný kurz vypočítal následovně: Výhra Sparty Praha = $100 : 75 = 1,33$, remíza = $100 : 15 = 6,66$ a nakonec výhra hostující Dukly Praha = $100 : 10 = 10$. Jenže v ideálním světě se nenacházíme. Sázkové kanceláře, aby vůbec mohly existovat, tak si musí připočítat marži. Marže společně s tzv. manipulačními poplatky slouží k snížení rizika ztráty. V případě marže, která by dosahovala hodnoty 10 %, by se naše kurzy vypočítávaly následovně: Výhra Sparty Praha = $90 : 75 = 1,2$, remíza = $90 : 15 = 6$ a nakonec výhra hostující Dukly Praha = $90 : 10 = 9$. (Tůma, 2011)

Bookmakerova práce tedy spočívá v tom, aby všechny šance nebyly pravděpodobnosti, tedy celých 100 % rozdělil mezi všechny možné výsledky vypsaného utkání. (Vincent, 2009)

4.5 Kdo je to vlastně bookmaker?

Bookmakeři to jsou a musí být odborníci na oblast, sport, soutěž či tým, pro který dané kurzy vypisují. V sázkové kanceláři Fortuna pracovalo po třech bookmakerech pro první českou fotbalovou ligu a také pro nejznámější fotbalovou soutěž světa anglickou Premier League. Každý z bookmakerů chodí přibližně dvakrát do týdne na služby do sázkové kanceláře. Další dny již pracuje výlučně z domova. A co se dělají bookmakeři na službách v sázkových kancelářích? Hlavně autorizace sázek. Autorizace sázek představuje schvalování sázek za větší finanční obnos, případně se také mohou vyhodnocovat sázky, které se často opakovaly. Také existuje tzv. hlavní bookmaker. Hlavní bookmaker dozoruje na práci podřízených bookmakerů a zároveň jim musí tu práci naplánovat. (Šedivý, 2009)

Práce bookmakera možná vzdáleně může připomínat alchymii nebo značit postup pokus-omyl. V tomto oboru nejsou jasně daná pravidla nebo tabulky pro výpočet kurzů. Hlavně záleží na samotném bookmakerovi, jak samotný kurz nastaví. V případě, že by nastavil příliš vysoký kurz, je tu možnost velkého prodělku pro sázkovou kancelář. Naopak, kdyby kurz nastavil příliš nízko, tak si nikdo nevsadí, protože se to nevyplatí a konkurence bude mít určitě kurz výhodnější. Mezi hlavní parametry, které se musí vždy zohledňovat při tvorbě kurzů jsou následující: forma jednotlivých týmů a hráčů, dále se také posuzuje věhlas týmů a hráčů. Poté je také třeba vzít v úvahu postavení týmů v tabulce. Nakonec je tu prostor i pro osobní názor daného bookmakera. A nakonec se také kurzy porovnávají s konkurencí a případně se ještě drobně upravují. Kurzy, které z tohoto procesu vzejdou, se musí poslat bookmakerovi týdně. Bookmaker týdně kurzy uzavře a může ještě kurzy naposledy změnit. (Šedivý, 2009)

4.6 Formy kurzového sázení v praxi

4.6.1 Internetové sázení

Internetové sázení začíná stejně jako to na kamenné pobočce výběrem sázkové kanceláře. Po výběru sázkové kanceláře je nutno se registrovat v internetové aplikaci dané sázkové kanceláře a pochopitelně nabít konto penězi. V zámoří je u některých společností možnost sázet tzv. na dluh, ale v České republice se s tímto nesetkáte. Teď již přichází na řadu samotné sázení. V nabídce sázkové kanceláře se zvolí daná událost, na kterou chce sázející vsadit, a na kterou je vypsán kurz. Můžeme narazit na několik druhů kurzů, ale v našich končinách nejčastěji narazíme na tzv. evropský kurz neboli decimální kurz. V případě správného tipu sázející vyhrává objem peněz, který se rovná násobku vložených peněz a vypsaneho sázkového kurzu. Můžeme si to opět uvést na jednoduchém příkladu. V případě vsazených 100 Kč na vypsany kurz v hodnotě 3,5, je výhra sázejícího celých 350 Kč. Čistý zisk sázkaře po odečtení vkladu tedy činí 250 Kč. (České sázkové kanceláře a svět internetového sázení, 2009)

4.6.2 Výhody a nevýhody internetové sázení

Jako hlavní výhody kurzového sázení přes internet bych vyzdvihl těchto pár skutečností. Jedná se o výhodnější kurzy, možnost sázení v přímém přenosu a také lehká dosažitelnost a s tím spojený komfort.

- 1) Výhodnější kurzy – Neboť je na českém území, ale i po celém světě opravdu mnoho sázkových kanceláří, vytváří se konkurenční prostředí. Chce-li některá ze sázkových kanceláří přitáhnout k sobě nové zákazníky neboli sázkaře, je nutné jim nabídnout výhodnější kurzy než u konkurence nebo třeba na kamenné pobočce. Navíc se u online sázení vytrácí manipulační poplatek, který je vyžadován při sázení na kamenné pobočce.

- 2) Sázení během živého přenosu – V dnešní době již snad všechny sázkové kanceláře poskytují živé přenosy tzv. live streamy daných sportovních událostí. Během těchto přenosů si lze i vsadit. Na základě průběhu daného zápasu či dané akce se kurzy na každou příležitost dynamicky vyvíjí.

- 3) Dosažitelnost a komfort - Dle mého úsudku je největší předností internetového sázení dostupnost v podstatě kdekoli na světě a také jednoduchost spojená s tím, že v dnešní době již skoro každý vlastní chytrý telefon nebo notebook případně počítač, skrze které je velice snadné si vsadit. Jediné co je potřeba, tak je funkční připojení k internetu. Nemusí se řešit cesta na kamennou pobočku sázkové kanceláře a s tím spojené nepříznivé počasí a časová náročnost nebo třeba denní doba, kdy si chce člověk vsadit.

Mezi jiné prokazatelné přínosy internetového sázení zcela určitě patří i klid na sázení, přehlednost nabídky sázkové kanceláře, možnost získání vstupního bonusu, který může pomoci v začátcích při výběru sázkové kanceláře nebo souhrnný přehled všech sázek uskutečněných v minulosti. (Sázení po internetu, 2016)

U kurzového sázení přes internet, ale lze objevit i nějaké nevýhody. Jako hlavní nevýhoda na kterou bývá často upozorňováno je jakási odtažitost od fyzických peněz. Když máte peníze ve fyzické podobě v ruce, tak je těžší se jich vzdát, než když jsou jen v podobě čísel na nějakém účtu sázkové kanceláře. Toto může být poměrně velké nebezpečí a je třeba si dobře rozmyslet kolik peněz si vůbec na tento typ účtu vložit, aby nás jejich případné prohrání moc nebolelo. A druhým možným negativem sázení přes internet je to, že někteří berou sázení na kamenné pobočce jako událost společenského významu. Chtějí se na pobočce sejít s podobně smýšlejícími lidmi, rozvířit diskuzi a někdy i třeba nechat poradit. Tento nedostatek si sázkové kanceláře provozující sázení přes internet uvědomují, a proto do internetových aplikací nainstalovali všemožné diskuze, které mají alespoň částečně nahradit výše zmíněné diskuze na kamenných pobočkách.

4.6.3 Telefon

Kurzové sázení skrze telefon existovalo hlavně v době, která předcházela internetu. Tato služba již není moc rozšířená a nenachází se ve standardní nabídce všech tuzemských sázkových kanceláří. Podmínkou je registrace na kamenné pobočce sázkové kanceláře a také nabitě sázkařské konto penězi. Dříve to byl takový ekvivalent dnešního sázení po internetu. Lidé tuto možnost využívali při sázení z domova nebo při svých cestách do ciziny. V současnosti je již tato metoda překonaná internetem.

4.6.4 Mobilní telefon (Smartphone)

Poměrně novým způsobem sázení je využívání svého chytrého telefonu neboli smartphonu. Navzdory dřívějšího povolení internetového sázení byla tato metoda ještě nějaký čas zakázána.

4.7 Formát kurzu

Existuje hned několik různých formátů sázkových kurzů:

- *Kurz desetinný: 1,07*
- *Kurz zlomkový: 15/14*
- *Kurz americký: -200*
- *Kurz procentuální: 40 %*

(Kurzové sázení, 2018)

4.8 Manipulační poplatek

Příklad: V jakém případě se vyplatí vsadit u společnosti, která vyžaduje manipulační poplatek.

- Sázková kancelář X má kurz bez manipulačního poplatku v hodnotě 1,80.
- Sázková kancelář Y má kurz na stejnou možnost u stejného zápasu v hodnotě 2,00, zároveň vyžaduje manipulační poplatek 10 %.
- U sázkové kanceláře X bude při sázce 110 Kč možná výhra celých 198 Kč, zisk z této sázky tedy bude činit 88 Kč.
- U sázkové kanceláře Y se vsadí 100+10 Kč manipulační poplatek a možná výhra by byla 200 Kč, neboť manipulační poplatek se do vkladu nepočítá, zisk z této sázky by tedy činil 90 Kč.
- Porovnáním v tomto případě můžeme zjistit, že i přes manipulační poplatek v hodnotě 10 % je v této sázce lepší vsadit u kanceláře co vyžaduje manipulační poplatek. (Kurzové sázení, 2018)

Tento poplatek si kamenné pobočky sázkových kanceláří přičítají ke vkladu sázejícího.

Nejčastěji v hodnotě 10 % (tzn. v případě sázky za 100 Kč bude celkový vklad 110 Kč). Postupně se ale od tohoto poplatku upouští, jak při internetovém sázení, tak poslední dobou stále častěji i v kamenných pobočkách sázkových kanceláří. Kanceláře se tohoto poplatku vzdávají zejména kvůli konkurenčnímu boji o zákazníka.

Část druhá – Představení použité metody a výpočetního programu

5 Statistické metody a program STATISTICA

5.1 Generalizovaný (zobecněný) lineární model

V oblasti statistiky je zobecněný lineární model (GLM) flexibilní zobecnění běžné lineární regrese. GLM dovoluje na základě historie (x měření) předvídat systematickou složku za pomoci vybraných vysvětlujících veličin a současně respektuje náhodnost podkladového děje. Toto má souvislost s náhodnou distribucí závislé proměnné ve výzkumu (tzv. funkce distribuční) na systematickou část (nebo lineární prediktor) pomocí funkce nazvané "funkce link". GLM byly v minulosti formulovány Robertem Wedderburnem a také Johnem Nelderem coby způsob, jakým lze sjednotit více statistických modelů v rámci jednoho teoretického rámce, počítaje regrese Poissonovo, lineární a logistickou. Toto byl základ pro vyvinutí obecného algoritmu za účelem maximálního odhadu pravděpodobnosti ve veškerých výše zmíněných modelech. Také se to mohlo rozšiřovat do mnoha dalších modelů. Například metoda nejmenších čtverců za účelem odhadu maximální pravděpodobnosti vzorových parametrů. Tento odhad pro maximální pravděpodobnosti je stále velmi oblíbený a zůstává běžnou metodou mnoha balíčků pro různé statistické výpočty.

Předpokládejme, že hodnoty závisle proměnné je možné při modelování rozdělovat na náhodnou část (rezidua) a část systematickou. Rezidua, tedy část náhodná je v tom případě v obyčejném regresním modelu normálně rozdělená. GLM nám tedy umožňují modelovat výsledek rozdělený poissonovsky (případ počtů), binomicky (případě proměnných ano/ne) nebo jinak v případě rozdělení z tzv. třídy exponenciálních rozdělení. Pro druhou část, tedy **systematickou** je zaveden pojem lineární prediktor (zpravidla značen řeckým písmenem η – éta). Konstantní změna v prediktoru vede ke konstantní změně odezvy (tzv. model lineární odezvy). V rámci GLM lze předpokládat, že lineární prediktor symbolizuje transformovanou střední hodnotu (tzv. linkovací funkcí).

$$f(\mu_i) = \eta_i$$

Již není nutný ten předpoklad, který nám říká, že mezi prediktorem a střední hodnotou je rovnost. (Proč používáme zobecněné lineární modely, 2010; Generalized linear model, 2018)

Použití lineárního modelu má 4 základní podmínky:

- Náhodné chyby jsou nesystematické.
- Náhodné chyby jsou homogenní se stejným neznámým rozptylem.
- Náhodné chyby jsou nezávislé.
- Náhodné chyby mají normální rozdělení.

V případě, že nahradíme výše zmíněné 4 podmínky podmínkami obecnějšími, dostaneme se k zobecněnému lineárnímu modelu. (Proč používáme zobecněné lineární modely, 2010; Generalized linear model, 2018; Neubauer, 2018)

5.2 Logit

V situaci, kdy chceme zjistit současný vliv více různých faktorů na nějakou danou veličinu, ale tato veličina nabývá jen dvou hodnot (tzv. dichotomická či binární), nemůžeme užít klasický lineární model. V těchto situacích se musí použít logistická regrese. Ta má za úkol zjistit, na čem je závislá pravděpodobnost, že dotazovaný uvede nějakou danou odpověď (např. nesouhlas). Model mimo jiné také dovoluje odhadnout šance na určitou odpověď pro rozdílné osoby. Modelování pravděpodobnosti přímo z hlediska statistického není možné (jedná se pouze o hodnoty v rozmezí 0 až 1). Z hlediska matematického mají logity mnohem lepší vlastnosti oproti původní pravděpodobnosti. Jako příklad lze uvést, že mohou nabývat hodnot na celé reálné ose. Z tohoto důvodu, tak musí být dané pravděpodobnosti převedeny na tzv. logity:

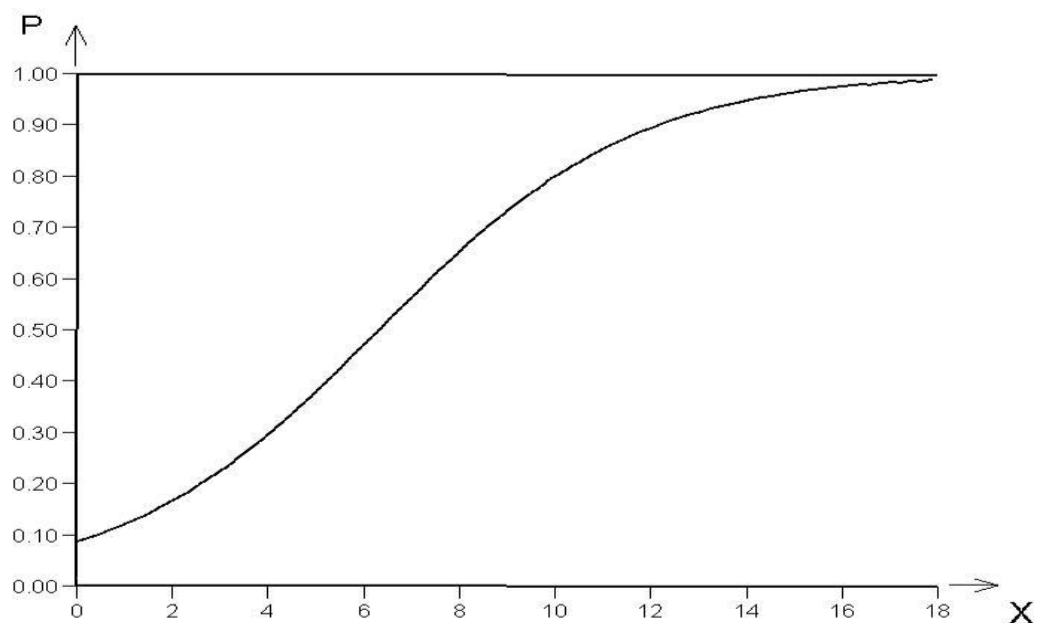
$$\log \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{x}$$

(Logistická regrese a logity, 2013)

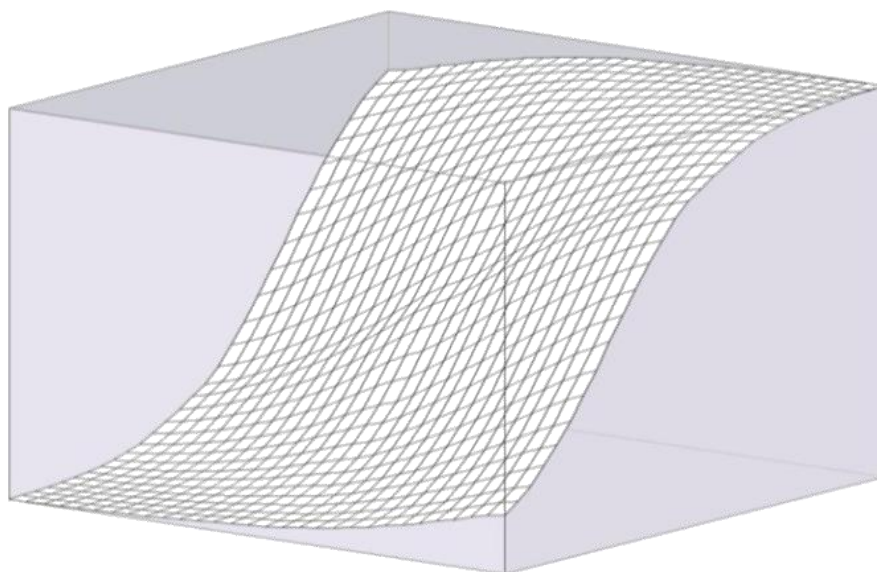
Výše uvedený výraz se jmenuje logit. Hodnoty α a β jsou regresními koeficienty. K odhadu těchto regresních koeficientů se používá metoda nejmenších čtverců. Metoda nejmenších čtverců získává odhady, které jsou maximálně věrohodné. Logistický regresní model můžeme vyjádřit sigmoidální křivkou:

$$\pi(x) = 1 / (1 + e^{-t})$$

Hodnoty této funkce se nachází mezi 0 a 1. Tvar křivky je rostoucí, s-ovitý, v nekonečnu se křivka přibližuje k 1 a v mínus nekonečnu k 0.



Obrázek 1 - *Logistický regresní model pro 1 proměnnou x*, dostupné z: <http://www.trilobyte.cz/downloadfree/qcemanual/logreg.pdf>



Obrázek 2 - Ukázka 3D logistického modelu s použitím 2 proměnných, dostupné z: <http://www.trilobyte.cz/downloadfree/qcemanual/logreg.pdf>

Tato křivka vyjadřuje odhad závislosti šance neboli pravděpodobnosti výskytu námi pozorovaného jevu v závislosti na hodnotě x . Nezávisle proměnná x může být taktéž vícerozměrná. Odpovídající model poté nabývá tvaru analogické lineární regrese (viz níže).

$$\pi(\mathbf{x}) = \frac{\exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)}{1 + \exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m)}$$

(Logistická regrese, 2018)

5.2.1 Použití a vlastnosti Logitu

- Logit je inverzní funkcí k funkci tzv. sigmoidální.
- Logit se používá v matematice a statistice.
- Logistická regrese - zvláštní případ funkce v zobecněném lineárním modelu. Logit má zásadní význam pro logistickou regresi.
- Funkce logit je ústřední bod pravděpodobnostního Raschova modelu měření (psychologické a vzdělávací hodnocení).

- Funkce logit se používá k přizpůsobení dat logistickému modelu.
- Funkce logit je negativní derivát funkce binární entropie.
- Regresní koeficienty si můžeme po transformaci přirozeným logaritmem vyložit jako poměry šancí.
- Logit se velice často používá pro vývoj tzv. skóringových modelů.
- Používá se k linearizaci logistických funkcí
(Logit, 2018)

5.3 Logistická regrese

Jedná se o statistickou metodu pro analýzu dat, což následně vede k odhadu pravděpodobnosti zkoumaného jevu. V tom případě jde o závislou proměnnou. Pravděpodobnost se odhaduje za pomoci nezávisle proměnných. Nezávisle proměnné jsou v tomto případě známé skutečnosti. Počet měření by měl přitom být pro každý poměr přibližně stejný. Zda vybraný jev, který zkoumáme nastal se utváří za pomoci náhodné veličiny. Náhodná veličina může nabývat pouze dvou hodnot. V případě, že námi zkoumaný jev nastal, tak veličina nabude hodnoty 1. Ale v opačném případě, když námi zkoumaný nenastal, tak hodnota veličiny bude 0.

Samotná metoda stojí na předpokladu, že za podmínek, které jsou dány vektorem \mathbf{x} , bude náhodná veličina $Y(\mathbf{x})$ nabývat hodnoty 1 s pravděpodobností, jejíž závislost na vektoru \mathbf{x} lze vystihnout díky tzv. logistické funkci. To můžeme zapsat následně:

$$P[Y(\mathbf{x}) = 1] = \frac{\exp(\beta' \mathbf{x})}{1 + \exp(\beta' \mathbf{x})}$$

Další vektor, který se označuje jako β je vektorem neznámých parametrů. Odhadem tohoto vektoru β se tím pádem odhaduje i celou dobu hledaná šance výskytu námi zkoumaného jevu.

Vektor s označením \mathbf{x} se zpravidla ukazuje v podobě $(1, z_1, z_2, \dots)$. Složka β_0 následně stanovuje dopad takzvaného absolutního členu. Z faktu, že pravděpodobnost výskytu námi zkoumaného jevu není závislá na námi zkoumaných nezávislých proměnných (tj. $\beta_1 = 0, \dots, \beta_k = 0$) vyplývá, že ho lze vyjádřit tímto tvarem:

$$P[Y(x) = 1] = P[Y(1, z') = 1] = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \text{ nezávisle na } z.$$

(Logistická regrese, 2014)

5.3.1 Příklady

1. *Sázková kancelář* - Tento příklad úzce souvisí s touto prací. Odhaduje se pravděpodobnost nějakého výsledku v daném sportovním klání. Musí se zde zadat veškeré známé skutečnosti, jež mohou mít vliv na výsledek duelu. Například v mé práci využívám tři údaje. Jedná se o postavení fotbalových týmů v tabulce před utkáním, poslední vzájemné zápasy a jako poslední proměnná je započítána výhoda domácího hřiště.
2. *Výskyt nemoci* - Nejdříve je třeba zadat, zda jedinec, kterého budeme zkoumat je žena či muž. Parametr může nabývat dvou hodnot. Jedná se o hodnoty 0 a 1. Hodnoty 0 nabude v případě, že zkoumaný jedinec bude žena a v případě, že zkoumaný jedinec je muž, tak parametr bude nabývat hodnoty 1. V případě, že výsledek vyjde nulový, tak to značí, že výskyt daného onemocnění absolutně nezávisí na pohlaví jedince. V případě, že odhad parametru beta bude kladný, znamená to, že výskyt této nemoci je o něco pravděpodobnější u mužů než u žen. Naopak, pokud odhad parametru beta bude záporný, tak výskyt této nemoci je pravděpodobnější u žen. Absolutní velikost nám nakonec ukazuje, kolikrát je pravděpodobnost výskytu nemoci u mužů či žen vyšší než u druhého pohlaví.
3. *Kreditní riziko* - Vypočítává se pravděpodobnost, že klient nebude schopen splatit bance úvěr. Jsou zde uvedené znaky či rysy klienta, který v tu dobu žádá banku o úvěr. Složky, které v daném modelu nabývají nulových hodnot, nemají na výslednou pravděpodobnost nesplacení žádný vliv. Složky s kladnými hodnotami mají plusový vliv na nesplacení úvěru. V sektoru bankovníctví případně v sektoru řízení rizik se tyto modely, které jsou založeny na logistické regresi velmi často nazývají pojmy jako skóring, skóringové funkce či skóringový model. (Logistická regrese, 2014)

5.3.2 Pojmy související s logistickou regresí

- Odds (neboli šance, případně riziko) – jedná se o podíl šance či pravděpodobnosti výskytu daného zkoumaného jevu ku šanci nevýskytu daného zkoumaného jevu. Např. šance 1:1, z 2 možných případů (např. panna a orel).
- Log odds – ukazuje přirozený logaritmus odds. Tato hodnota se občas nazývá skóre, zejména v sektoru bankovníctví a řízení rizik.
- Odds ratio (podíl šancí, případně rizik) – tato hodnota představuje podíl šancí nebo naopak rizika výskytu pro dvě odlišné hodnoty nezávisle proměnných.
- Log odds ratio – reprezentuje přirozený logaritmus výše zmíněného odds ratio.

(Logistická regrese, 2014)

5.3.3 Vstupní data pro program STATISTICA

Vstupní data do programu STATISTICA musí obsahovat jeden či více sloupců nezávisle proměnné (taková, jejíž hodnotu lze jakkoliv účelově ovlivňovat a měnit) a jeden sloupec závisle proměnné (u této proměnné očekáváme změny na základě změn nezávisle proměnné). V průběhu jakéhokoliv experimentu zkoumáme prvky daného výběrového souboru. Údaje, jež my sledujeme, se nazývají proměnné a hodnoty těchto proměnných se nazývají varianty. Nezávisle (neboli vysvětlující) proměnné, jinak řečeno prediktory mohou nabývat jakékoliv číselné hodnoty. Binární závisle proměnná (neboli vysvětlovaná) musí nabývat pouze hodnot 0 nebo 1. Tyto dvě hodnoty odpovídají výskytu či nevýskytu sledovaného jevu (rozbití, úraz, zisk klienta, smrt, apod.). Při tomto označování nezáleží na volbě, zda výskyt bude označován jako jednička a nevýskyt jako nula nebo naopak. Binární proměnná se používá jen v případě, když daná hodnota prediktoru má jen jeden jediný výsledek ve formátu ano-ne. To můžeme vidět na obrázku číslo 3. Další typ je poměrná, neboli frekvenční závisle proměnná. Tato proměnná se používá, když pro hodnotu prediktoru x provedeme z testů a z toho nám tedy vyjde z výsledků, z nichž r je pozitivních a $z - r$ negativních. Ve sloupci pro závisle proměnnou pak lze zapsat poměr r/z , případně $(z - r)/z$, to lze vidět

na obrázku číslo 4. Nezávisle proměnná může být vícerozměrná – tedy ve více sloupcích, např. na obrázku číslo 5. (Logistická regrese, 2018)

Nezávisle proměnná	Poměrná závisle proměnná
Koncentrace mg/l	Účinek (podíl z 10)
0	0
1	0
2	0.2
3	0.3
4	0.7
5	0.9
6	1

Obrázek 3 - *Data s binární závisle proměnnou*, dostupné z: <http://www.trilobyte.cz/downloadfree/qcmanual/logreg.pdf>

Nezávisle proměnná	Binární závisle proměnná
Zátěž dílu	Porucha dílu
0	0
0	0
0	0
1	0
1	1
2	0
3	0
4	1
4	1
4	0
.....

Obrázek 4 - *Data s poměrovou závisle proměnnou*, dostupné z: <http://www.trilobyte.cz/downloadfree/qcmanual/logreg.pdf>

Nezávisle proměnné			Binární závisle proměnná
Doba	Intenzita	Frekvence	Porucha
8	34	200	1
5	38	250	0
5	35	250	0
7	40	200	1
6	29	100	1
4	35	150	0
11	37	150	1
4	28	200	0
8	32	200	0
6	30	250	1

Obrázek 5 - *Data s více nezávisle proměnnými a binární závisle proměnnou*, dostupné z: <http://www.trilobyte.cz/downloadfree/qcmanual/logreg.pdf>

5.4 Program STATISTICA

Počítačový program STATISTICA je pokročilý softwarový balíček pro nejrůznější způsoby analýzy dat. STATISTICA existuje na trhu již 27 let. Díky mnoha tisícům zpětných vazeb a hodnocení od zákazníků bylo možné STATISTICU přizpůsobit pro mnoho různých odvětví. Statistica poskytuje mimo jiné již zmíněnou analýzu dat, správu dat, statistiky, dolování dat, strojové učení, textovou analýzu a vizualizaci dat. Aplikace je určena především pro podporu strategického a výkonného řízení ekonomiky a financí podniku, kvality a rizik, vztahů se zákazníky, vývoje apod. K dispozici jsou samozřejmě i přehledné grafické výstupy, grafy, reporty atd. STATISTICA byla původně vyvinuta firmou StatSoft, která byla získána společností Dell v březnu 2014. Následovalo ještě několik prodejů licencí a daného softwaru. Načež dne 15. května 2017 oznámila společnost TIBCO Software Inc., že koupila společnost Statistica a od té doby se o vývoj tohoto programu stará právě tato společnost. Pro konfiguraci programu STATISTICA jsou možné použít následující skriptovací jazyky: Visual Basic nebo R. Kompatibilní programovací jazyky jsou C#, C++, Java, VB.NET. (STATISTICA Features Overview, 2018)

Rozličné verze programu STATISTICA jsou lokalizovány do mnoha světových jazyků: čínština (tradiční i zjednodušená), čeština, angličtina, francouzština, němčina, italština, japonština, polština, ruština a španělština. Dokumentace k samotným programům je dostupná v arabštině, čínštině, češtině, angličtině, francouzštině, němčině, maďarštině, italštině, japonštině, korejštině, polštině, portugalsštině, ruštině, španělštině a dalších jazycích. Tolik lokalizací dělá z programu STATISTICA nástroj pro celosvětové použití. Různé balíčky analytických technik jsou k dispozici v šesti výrobních řadách. (Statistica, 2018)

5.4.1 Grafika

Statistický program STATISTICA v sobě ukrývá analytické a průzkumné grafy a navíc standardní dvou a tří dimenzionální grafy. Činnosti jako interaktivní štítkování,

značení a vyloučení dat umožňují vyšetření mimořádných hodnot a také analýzu průzkumných dat. (Statistica, 2018)

5.4.2 Uživatelské rozhraní

První co se v programu po jeho spuštění stane je načítání tabulky dat a také načítání samotných statistických funkcí do rozbalovacích nabídek. Následně je od uživatele vyžadováno vložení samotných proměnných nezbytných pro výpočet a zadání požadovaného typu analýzy. V programu STATISTICA se nemusí nic zadávat pomocí příkazového řádku, proto je tento program uživatelsky velmi přívětivý. Každá z analýz v tomto programu obsahuje tabulkový či grafický výstup, který je automaticky uložen na samostatném sešitu v programu. (Statistica, 2018)

5.4.3 Moje verze programu

Má verze programu STATISTICA se jmenovala STATISTICA Base Cz. Tato verze je přímo uzpůsobena pro základní práci s daty, pro testování hypotéz a také pro výpočet nezbytných popisných statistik v stále se měnících datových souborech. STATISTICA Base Cz nabízí velmi obsáhlou pro dnešní dobu již nezbytných statistických nástrojů pro nejčastější statistické analýzy. Tato verze programu obsahuje:

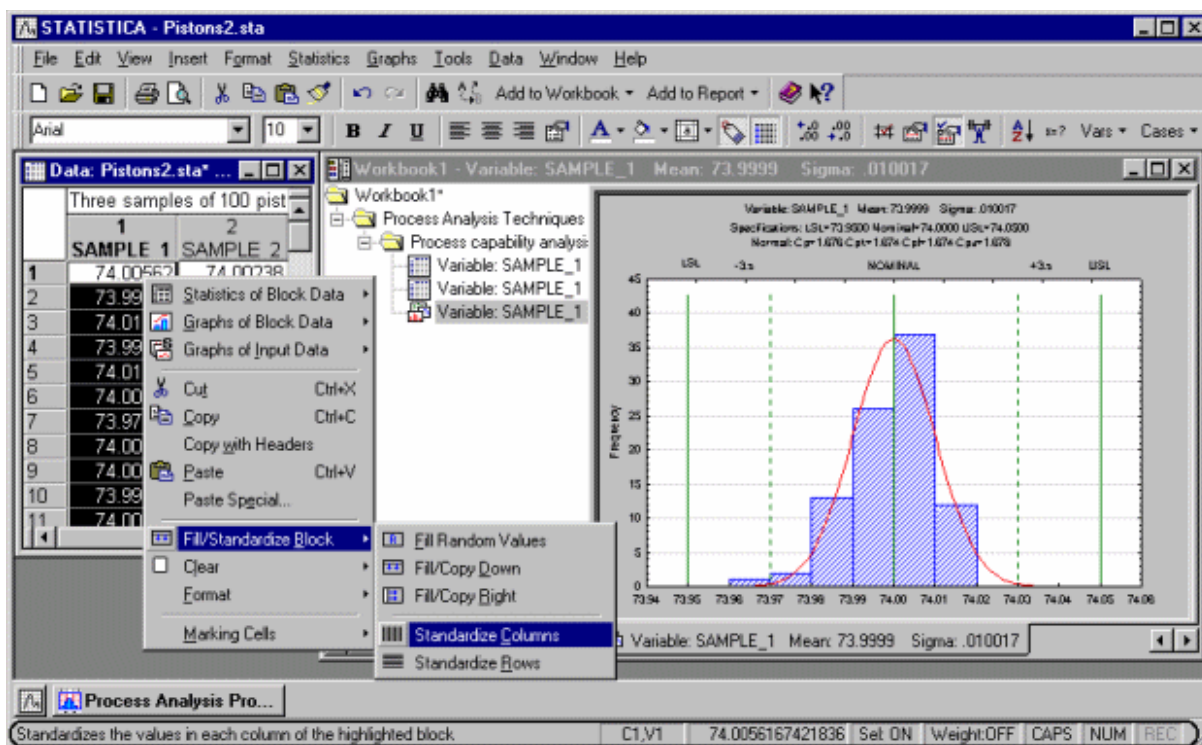
- Grafické nástroje,
- Základní tabulky a statistiky,
- Vícerozměrné lineární regrese,
- Analýzu rozptylu,
- Neparametrické statistiky,

Dále nabízené verze programu STATISTICA:

- STATISTICA Advanced Cz – oproti verzi Base přidává mnoho dalších analytických technik pro všechny typy dat.

- STATISTICA Reportovací tabulky – nástroj pro tvorbu kvalitních a propracovaných tabulkových výstupů.
- STATISTICA Interactive Visualizations & Dashboards – modul, který je přímo určený pro sledování dat v reálném čase.

(STATISTICA Base CZ, 2017; STATISTICA Base CZ – Detail, 2017)



Obrázek 6 - Program STATISTICA - uživatelské prostředí, dostupné z: <http://documentation.statsoft.com/STATISTICAHelp.aspx?path=Customizing/CustomizingSTATISTICA/Overview/ToolbarOverviews/StatusBar>

Část třetí – Praktické výsledky a výpočty

6 Praktická část

6.1 Vstupní data

6.1.1 Popis datového souboru

Celý datový soubor byl opravdu velmi obsáhlý. Byly třeba vyhledat výsledky všech dvaceti týmů ve všech 38 kolech daného ročníku mnou zkoumané soutěže. V další části se muselo jít ve vyhledávání i několik málo let do minulosti za účelem zjištění posledních vzájemných zápasů daných dvou týmů. Drtivou většinu těchto dat jsem získal z internetových stránek: www.livesport.cz, www.soccerassociation.com a www.worldfootball.net. Na posledních dvou jmenovaných stránkách bylo k dostání i pořadí jednotlivých týmů před vzájemným zápasem, což bylo enormně důležité, jelikož se jednalo o jednu z mých proměnných. V případě, že bych tyto informace o pozici jednotlivých týmů nenašel na uvedených stránkách, tak by bylo potřeba si tabulku kolo za kolem propočítávat.

6.1.2 Příprava dat

Veškeré zápasy daného ročníku Premier League jsem si opsal i s výsledky a vypsanými kurzy od sázkové kanceláře Tipsport a samozřejmě také s předzápasovým postavením obou týmů v tabulce Premier League. Dále jsem si našel předchozí vzájemné zápasy a přidal to mezi další proměnné na základě toho, jak se týmům vůči danému soupeři dařilo v ligových či pohárových zápasech.

6.1.3 Popis podmínek (proměnných)

Podmínek neboli proměnných bylo v této práci použito hned několik. Jako první byla použita výhoda domácího prostředí. Tato podmínka ukazovala, kdo měl v daném zápase výhodu svého vlastního hřiště, stadionu, svých fanoušků a také ušetřil síly, protože nemusel k danému zápasu cestovat. Druhou podmínkou byli předchozí vzájemné zápasy. Konkrétně se jednalo o tři předchozí zápasy. Každý tým má nějakého oblíbeného a také neoblíbeného soupeře, na které se jim výsledkově daří či naopak. Ve sportovním světě je tato možná psychologická výhoda velmi důležitá. A poslední podmínkou bylo postavení v tabulce před zahájením zápasu. Tady se jednalo o velice důležitou podmínku, protože pohled na tabulku je i pro bookmakery vypisující kurzy jedním z těch nejdůležitějších ovlivňujících faktorů.

6.2 Popis experimentů





















6.2.1 Cíle experimentů

V mojí práci se zabývám jedním stěžejním a několika dalšími a o něco menšími a méně důležitými experimenty. Hlavní experiment této bakalářské práce spočívá v tom, že se snažím porovnat mnou vypočítané sázkové kurzy ve statistickém programu STATISTICA s kurzy, které ve skutečnosti vypsala česká sázkařská společnost Tipsport. Tipsport je jedničkou na českém trhu a má poměrně dlouhou historii. Zejména z těchto dvou důvodů jsem upřednostnil právě tuto sázkovou společnost například před Fortunou či SynotTipem. Zkoumal jsem veškeré zápasy 1. anglické fotbalové ligy známé zejména jako Premier League. Z porovnání mých kurzů s kurzy od Tipsportu nám vyjdou hodnoty, které nám ukážou, jak moc se moje logistická regrese přiblížila odborníkům na slovo vzatých. Zároveň nám toto srovnání ukáže, jestli tři proměnné pro statistický software je dostatečný počet, nebo jestli by bylo nutné jich několik ještě přidat pro vyšší úspěšnost. Poté budou na následujících stranách ještě podrobněji rozepsány menší a méně významné experimenty. Nicméně i tyto menší experimenty nám mohou pomoci odhalit, kde se hlavně nacházejí rozdíly mezi sázkovou kanceláří a mojí logistickou regresí. Konkrétně se jedná o 4 další menší experimenty či výsledky

šetření. V prvním případě se jedná o to, jak často se sázková kancelář Tipsport trefila se svým nejnižším kurzem na vítěze daného utkání. Tyto hodnoty porovnávám i se svými kurzy a vzniká z toho další typ porovnání kurzů mých a profesionální sázkové kanceláře. V dalším malém výzkumu jde o to, které sázkové kurzy vycházejí nejčastěji. Zda je opravdu tak výhodné sázet na favority nebo se vyplatí sázet na o něco vyšší kurzy, kde se případná menší úspěšnost nahradí vyšším ziskem z jednotlivých sázek. Třetí menší experiment zkoumá, kdy bylo nejlepší sázet na nejnižší vypsany kurz od sázkové kanceláře. Jinak řečeno, v jaké fázi sezóny se sázková kancelář svými nejnižšími kurzy nejčastěji trefovala, a sázejícím se nejvíce vyplatilo kanceláři věřit. A ve čtvrté a poslední části těchto menších experimentů jsem vypočítal, jaký zisk potažmo ztrátu by člověk dosáhl v případě, že by na každý zápas daného ročníku dané soutěže vsadil na nejnižší kurz od sázkové kanceláře 10 Kč. Tuto hodnotu jsem nakonec porovnal s mými kurzy a v podstatě nám shrne úspěšnost dvou rozdílných metod výpočtu sázkových kurzů.

6.2.2 Zkoumaná soutěž

Mnou zkoumaná soutěž se jmenuje Premier League. Dříve zněl celý název soutěže Barclays Premier League. Jedná se o nejvyšší fotbalovou soutěž v Anglii. Premier League se účastní 20 týmů. Celkem se v soutěži odehraje 38 kol. Každý s každým doma a venku. To dělá celkem 380 zkoumaných zápasů. Pro moji bakalářskou práci jsem si vybral soutěžní ročník 2016/2017, který probíhal od srpna 2016 do května 2017. Soutěž ve zkoumaném ročníku vyhrála londýnská Chelsea před londýnským Tottenhamem a Manchester City. Níže přikládám konečnou podobu tabulky z mnou zkoumaného ročníku.

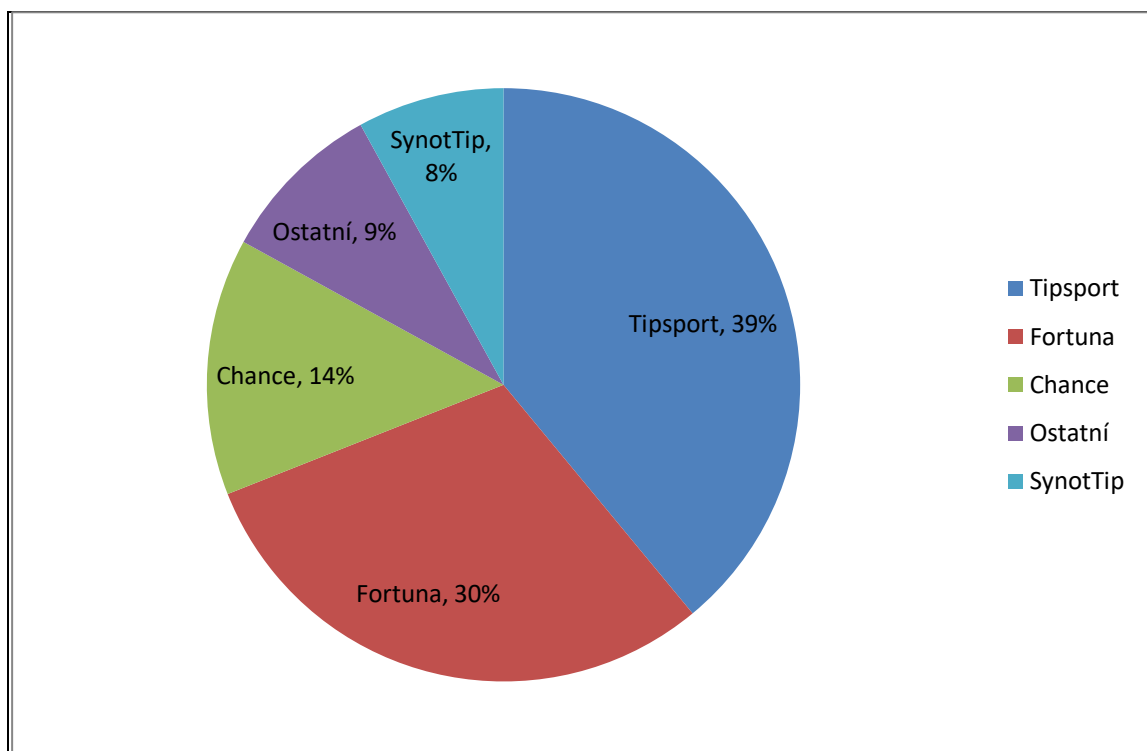
Position	Club	Played	Won	Drawn	Lost	GF	GA	GD	Points
1 ●	 Chelsea	38	30	3	5	85	33	+52	93
2 ●	 Tottenham Hotspur	38	26	8	4	86	26	+60	86
3 ●	 Manchester City	38	23	9	6	80	39	+41	78
4 ●	 Liverpool	38	22	10	6	78	42	+36	76
5 ●	 Arsenal	38	23	6	9	77	44	+33	75
6 ●	 Manchester United	38	18	15	5	54	29	+25	69
7 ●	 Everton	38	17	10	11	62	44	+18	61
8 ●	 Southampton	38	12	10	16	41	48	-7	46
9 ▲	 Bournemouth	38	12	10	16	55	67	-12	46
10 ▼	 West Bromwich Albion	38	12	9	17	43	51	-8	45
11 ▲	 West Ham United	38	12	9	17	47	64	-17	45
12 ▼	 Leicester City	38	12	8	18	48	63	-15	44
13 ▲	 Stoke City	38	11	11	16	41	56	-15	44
14 ▼	 Crystal Palace	38	12	5	21	50	63	-13	41
15 ▲	 Swansea City	38	12	5	21	45	70	-25	41
16 ▼	 Burnley	38	11	7	20	39	55	-16	40
17 ▼	 Watford	38	11	7	20	40	68	-28	40
18 ●	 Hull City	38	9	7	22	37	80	-43	34
19 ●	 Middlesbrough	38	5	13	20	27	53	-26	28
20 ●	 Sunderland	38	6	6	26	29	69	-40	24

Obrázek 7 – *Konečná tabulka Premier League pro sezónu 2016/2017*, dostupné z: <https://semper.soccersavings.com/201617-premier-league-season-review/>

6.2.3 Sázková kancelář

Výsledky mých výpočtů, tedy mnou vypočtené sázkové kurzy pomocí logistické regrese se porovnávali se sázkovou kanceláří Tipsport. Tato akciová společnost sídlí v Berouně a byla založena 5. dubna roku 1991. V současné době má sázkařská společnost Tipsport na území celé České republiky skoro 1000 kamenných poboček. Aktuálně Tipsport sponzoruje například českou a slovenskou nejvyšší hokejovou

soutěž. Tuto sázkovou kancelář jsem si vybral, protože se aktuálně jedná o jedničku na českém trhu, mají dlouholetou historii a sám jsem si u nich zkusil založit sázkařský účet. Mezi další známé sázkové kanceláře, které působí na českém trhu se zařazují: Chance (v současné době již patří pod sázkovou kancelář Tipsport), Fortuna (dvojka na českém trhu) a také SynotTip. Na níže uvedeném grafu můžete vidět procentuální rozdělení sázkových kancelářů na celém českém trhu.



Graf 1 - Tržní podíly sázkových kanceláří na českém trhu, vlastní tvorba, podklady dostupné z: https://www.penize.cz/spotrebitel/227043-kurzove-sazky-miliardovy-byznys#element_54_8643

6.2.4 Postup výpočtu

Nejdříve bylo nutné vytvořit model, kam budu dosazovat vypočtené hodnoty a samotné proměnné. Bylo třeba spočítat pravděpodobnost výhry domácích, poté hostů a zbytek do hodnoty 1 (neboli 100 %) činila pravděpodobnost remízy. Tedy vytvořit dva modely. Prvotní vzorec vypadal následovně: $a + bx_1 + cx_2$. Kde za X_1 byl dosazován rozdíl v pořadí a za X_2 vzájemné zápasy. Hodnoty a , b a také c byli čísla, která vyšli

v programu STATISTICA, kam byl dosazen i třetí parametr, který značil výhodu domácího hřiště. Výsledek z této rovnice se zasadil do dalšího vzorce:

$$\frac{\exp(\text{výsledek z předchozího vzorce})}{1 + \exp(\text{výsledek z předchozího vzorce})}$$

Jednotlivé pravděpodobnosti jsem si poté zápas po zápase převedl na kurzy, které se porovnávaly s kurzy od sázkové kanceláře Tipsport. Problém nastal pro předpověď prvních pěti kol soutěže mnou zkoumaného ročníku, protože tabulka mohla být lehce zkrácena malým množstvím zápasů. Tento problém jsem vyřešil tak, že pro prvních pět zápasů sezóny jsem použil tabulku po posledním zápase sezóny předcházející. Toto nahrazení skýtalo ještě jeden menší problém a to ten, že poslední tři týmy předcházející sezóny byli nahrazeni třemi postoupivšími z druhé nejvyšší soutěže. Proto jsem použil pro prvních pět zápasů (jen u třech týmů, kterých se to týkalo) klíč z následující tabulky, kde jsem prvnímu týmu z druhé ligy přidělil pořadí 18, druhému týmu ve druhé lize pořadí 19 a třetímu z druhé ligy jsem přiřadil pořadí 20:

Pozice v minulé sezóně v první lize	Pozice v minulé sezóně ve druhé lize
18	1
19	2
20	3

Tabulka 1 - Nahrazování pozic pro prvních pět kol – vlastní tvorba

K celému výpočtu jsem kromě statistického programu STATISTICA používal i program Microsoft Excel, kde jsem zejména využíval možnosti vzorců, které mi do značné míry usnadnilo práci.

vysledek - Parameter estimates (Spreadsheet1)
 Distribution : BINOMIAL, Link function: LOGIT
 Modeled probability that vysledek = 1

Effect	Level of Effect	Column	Estimate	Standard Error	Wald Stat	Lower CL 95.0%	Upper CL 95.0%	p
Intercept		1	-0,005689	0,110138	0,00267	-0,221555	0,210177	0,958806
poradirozdil		2	0,089668	0,014556	37,94782	0,061139	0,118198	0,000000
predvajzapasy		3	0,162737	0,064746	6,31754	0,035837	0,289636	0,011955
Scale			1,000000	0,000000		1,000000	1,000000	

Obrázek 8 – Výsledky z programu STATISTICA pro určení pravděpodobnosti výhry domácích, používal jsem hodnoty z prvního sloupce do výše uvedeného vzorce, vlastní tvorba za pomoci programu STATISTICA

6.3 Experimenty

6.3.1 Kolika % odchylky v kurzech byli v mém výpočtu oproti sázkové kanceláři Tipsport

Kurzy vypsané od společnosti Tipsport a ty z mnou nastavené logistické regrese na výhru domácích jsou skoro stejné a rozdíl v jejich výši činí ne příliš vysokých 22 %. O větší rozdíl se jednalo u kurzů na výhry hostů. Zdejší rozdíl činil přibližně 40 %. Na tak výrazném rozdílu u šance hostů mělo hlavní vliv to, že sázková kancelář přikládá domácímu prostředí vyšší hodnotu, než jsem tomu mohl přikládat já ve statistickém programu a proto také méně věřili hostům a více domácím. To vysvětluje výši rozdílů u domácích a hostujících kurzů. Na tyto výsledky, nebo přesněji řečeno, na výšku jednotlivých kurzů měl vliv také fakt, že se použilo jen několik málo proměnných. V této práci se konkrétně použily proměnné tři, místo třeba patnácti, které berou do úvahy povolání bookmakeři u sázkové kanceláře Tipsport. Pro snížení rozdílů a vypsaní přesnějších kurzů je potřeba vzít v úvahu daleko více proměnných. Některé z těchto proměnných, které by byli přínosné pro náš výpočet, ale jen stěží lze ohodnotit nějakou hodnotou a vložit je do nějakého výpočetního softwaru jako byl například v mém případě statistický program STATISTICA. Jedná se o proměnné, které jsou subjektivní, a každý by si je ohodnotil jinou hodnotou. Může se jednat například o jméno či slávu jednotlivých týmů. Každý by tento parametr nastavil odlišně. Výše uvedené rozdíly jsou průměrem absolutních hodnot odchylek. Souhrnně vzato byly mé kurzy na výhru domácích a hostů nižší než u specializované sázkové kanceláře Tipsport. Naopak kurzy na remízu byly o hodně vyšší ty, které pocházeli z mého modelu. Znamená to, že pro případy výhry domácích a hostů můj model logistické regrese výrazně nadhodnotil pravděpodobnosti úspěchu. Byly nadhodnocené více než u profesionální sázkové kanceláře, proto jsem měl kurzy nižší než Tipsport. S tím je spojeno to, že na remízu zbylo méně pravděpodobnosti a kurzy na tuto možnost byly o hodně vyšší než u Tipsportu.

	Výhra domácích - průměr	Remíza - průměr	Výhra hostů - průměr
Můj model	2,39	6,18	3,53
Tipsport	2,92	4,21	4,93
Rozdíl	22 %	46 %	40 %

Tabulka 2 – Porovnání rozdílů v průměrných kurzech na všechny varianty – vlastní tvorba

6.3.2 V kolika % se sázková kancelář Tipsport trefila svým nejnižším kurzem do výsledku utkání

Z celkového počtu 380 zkoumaných zápasů během celé sezóny Premier League v sezoně 2016/2017 se sázková kancelář Tipsport správně trefila svým nejnižším kurzem na vítězný tým (u žádného zápasu nebyl vypsán nejnižší kurz na remízu) v necelých 61 % zápasů. To znamená, že sázková kancelář svým nejnižše vypsaným kurzem trefila necelé 2/3 zápasů. Když zabrousíme do konkrétních čísel, tak můžeme zjistit, že Tipsport správně odhadl 231 zápasů z celkového počtu 380 zápasů. To je celkem slušná bilance a úspěšnost.

Můj model byl v porovnání se sázkovou kanceláří Tipsport přece jenom o něco méně úspěšný. V mém modelu založeném na logistické regresi bych úspěšně odhadl nejnižším kurzem z celkového počtu 380 zápasů celých 198 zápasů. Jedná se tedy o více než 52 % úspěšnost. Zároveň to je úspěšnost o 9 % nižší než u specializovaných bookmakerů profesionální sázkové kanceláře Tipsport. Správnost mých odhadů u jednotlivých zápasů kopírovala úspěšnost u Tipsportu i v průběhu celé soutěže. Odchyly až na minimum výjimek byli u jednotlivých kol v úspěšnosti odhadu jednoho zápasu. Rozdíl devíti procent v porovnání statistického programu, který bral do úvahy tři proměnné, s profesionálními bookmakery mi přijde jako velice uspokojivý. Model byl zcela určitě ovlivněn tím, že se v práci použilo jen pár proměnných.

	Úspěšně odhadnuté zápasy	Celkový počet zápasů	% úspěšnost odhadu
Tipsport	231	380	61 %
Já	198	380	52,1 %

Tabulka 3 – Porovnání úspěšnosti Tipsportu a logistické regrese – vlastní tvorba

6.3.3 Jaké kurzy vycházejí nejčastěji

V dalším menším experimentu jsem se zaměřil na to, jaké kurzy nebo přesněji řečeno, kurzy v jaké hodnotě nejčastěji vycházejí. Princip u kurzového sázení je takový, že čím menší kurz bookmaker a nakonec i celá sázková kancelář vypíše, tím větší pravděpodobnost úspěchu daného jevu v daném zápase očekává. Celkem jsem si vytvořil 9 kurzových rozmezí. Nejvícekrát přirozeně vyšly zápasy v rozmezích od 1 do 1,5 a také od 1,51 do kurzu 2. Ukazuje nám to, že nejvíce vyhrávají favorité nad outsidersy a níže postavenými celky v tabulce, což se dalo předpokládat. Na opačném konci, ale možná překvapivě nejsou kurzy s hodnotou nad 5, ale možná trochu překvapivě kurzy kolem 4,5. Z tabulky je jasné vidět, že nejvíce se vyplatí sázet na kurzy do hodnoty 2,5. Pro odvážnější je tu zajímavá možnost sázet na kurzy, které nabývají hodnot lehce nad 3. Vyšší kurzy již vycházejí velice sporadicky a s pravidelnými výhrami si příliš netykají.

Rozmezí kurzů	Počet výherních zápasů
1 – 1,5	76
1,51 – 2	86
2,01 – 2,5	58
2,51 – 3	26
3,01 – 3,5	61
3,51 – 4	16
4,01 – 4,5	16
4,51 – 5	10
5,01 a více	31

Tabulka 4 – Četnosti neúspěšnějších kurzů – vlastní tvorba

6.3.4 V jaké části sezóny se vyplatí sázet dle předpokladů sázkové kanceláře

Tato část experimentu navazuje na část výzkumu, ve které se snažím popsat, v kolika procentech se sázková kancelář Tipsport společně se svými bookmakery trefila při svých odhadech. Z procentuální úspěšnosti v jednotlivých kolech vyplývá, že bookmakeři Tipsportu nejlépe odhadli úplný závěr soutěže. Zajímavá úspěšnost se ukázala i zhruba v polovině sezóny mezi 16. a 20. kolem soutěže. Naopak nejmenší úspěšnost je celkově v první třetině soutěže, kde sázkové kanceláře do jisté míry tipují naslepo, protože na začátku soutěže ještě zdaleka nemají takový přehled o síle jednotlivých týmů po dlouhé letní pauze. Pokud tedy chcete brát sázkovou kancelář a její vypsané kurzy jako nějaké měřítko či pomocný ukazatel, kdo by měl uspět v daném zápase, tak bych doporučil počkat alespoň do poloviny soutěže. Nejvíce se můžeme nechat ovlivňovat samotnými kurzy v úplném finiši soutěže. Na tento experiment měl zcela jistě vliv také ten fakt, že pro prvních pět kol se použila tabulka z předchozího ročníku z důvodu malé vypovídající hodnoty tabulky po pár zápasech.

Rozmezí kol	Úspěšně odhadnuté zápasy	Procentuální úspěšnost
1 – 5	28	56 %
6 – 10	27	54 %
11 – 15	24	48 %
16 – 20	36	72 %
21 – 25	25	50 %
26 – 30	33	66 %
31 - 38	55	72,5 %

Tabulka 5 – Úspěšnost sázkové kanceláře Tipsport – vlastní tvorba

Mnou vypočtené kurzy nabývali nejvyšší úspěšnosti stejně jako u Tipsportu v rozmezí 16. – 20. kola a hlavně v rozmezí 31. – 38. kola. Rozdíl mezi mojí úspěšností a úspěšností sázkové kanceláře Tipsportu činil na jeden úsek průměrně 8,5 %, což poměrně přesně koresponduje s výše zmíněnými devíti procenty, o které se lišila úspěšnost sázkové kanceláře a mnou vytvořené logistické regrese.

Rozmezí kol	Úspěšně odhadnuté zápasy	Procentuální úspěšnost
1 – 5	25	50 %
6 – 10	24	48 %
11 – 15	20	40 %
16 – 20	32	64 %
21 – 25	23	46 %
26 – 30	24	48 %
31 - 38	50	62,5 %

Tabulka 6 – Úspěšnost logistické regrese – vlastní tvorba

6.3.5 Vyplatilo by se sázet za 10 Kč vždy na nejnižší kurz spíše u specializované sázkové kanceláře nebo podle vytvořeného modelu skrze statistický software?

V této části jsem zkusil imaginárně vsadit na každý nejmenší kurz u každého jednotlivého zápasu každého kola 10 korun českých. Do vsazené částky jsem nepočítal manipulační poplatky, neboť internetové sázení u společnosti Tipsport je od tohoto poplatku osvobozeno a ostatně i většina kamenných poboček většiny sázkových kanceláří od tohoto poplatku upouští. Celkově prosázená částka by byla tedy po 380 zápasech 3800 Kč. Kdybych se řídil kurzy od společnosti Tipsport, tak by příjem ze sázek na nejnižší kurzy, tedy na favority činil 4058,6 Kč z možných 7056,3 Kč. Zisk by tedy v tomto případě činil 258,6 Kč za celý soutěžní ročník anglické Premier League. Už jenom skutečnost, že by sázející skončil se ziskem, značí odbornost a kvalitu u bookmakerů Tipsportu.

V případě, že bych stejným způsobem věřil i mému modelu logistické regrese a řídil se jím při svých sázkách, tak bych skončil bohužel ve ztrátě. Celková možná výhra na základě mnou vypočtených kurzů by činila 5948 Kč. Tato celková možná výhra je o 1108,3 Kč nižší než u sázkové kanceláře Tipsport. Na základě výsledků bych vyhrál 3034,84 Kč, což je o 1023,76 Kč méně než u Tipsportu. To je dané tím, že mé veškeré kurzy byli někdy i o několik desítek procent nižší než u Tipsportu. Příjem 3034,84 Kč by znamenal ztrátu za jeden soutěžní ročník anglické Premier League 765,16 Kč.

7 Závěr

Závěrem lze říci, že k tomu aby se vypočítané kurzy pomocí logistické regrese opravdu přiblížily realitě, tak by bylo třeba zahrnout daleko více proměnných. Přesně tak, jak to dělají i profesionální bookmakeři například i u mnou zkoumané sázkové společnosti Tipsport, kteří mohou reagovat na desítky podnětů, které je velice obtížné zadat do nějakého programu, systému či vzorce ať je jakkoliv kvalitní. Výhoda bookmakerů je také v tom, že mohou upravovat kurzy v podstatě až do výkopu jednotlivých zápasů a mohou pružně reagovat na různé indicie, vypsání základní sestavy, zranění důležitých hráčů, možné šetření některých hráčů na důležitější zápasy a tak dále, které je těžko možné odhadnout případně odhalit dříve než několik minut či hodin před zápasem. Je třeba uznat, že mnou vytvořený model na predikci sázkových kurzů na základě logistické regrese nedosáhl požadované úspěšnosti, tedy takové aby byl úspěšnější než systémy používané v největší sázkové kanceláři na českém trhu nebo alespoň takové, aby ten, kdo by sázel na základě tohoto modelu, skončil v ziskových číslech. Mnou vytvořeným modelem vypočtené kurzy byly nižší než u sázkové kanceláře a také o přibližně 9 % méně přesné.

Ne všechno, ale bylo neúspěšné. Na to, že jsem použil jen pár vybraných proměnných, konkrétně se jednalo o výhodu domácího prostředí, dále se jednalo o čtyři předchozí vzájemné zápasy a také o postavení v předzápasové tabulce, tak nebyly rozdíly nikterak závažné. Postupným přidáváním proměnných by se rozdíly ve výsledcích postupně tenčily a i ztráta by se mohla postupem času překloupit do zisku. Proto si myslím, že nápad na predikci výsledků ve sportovních utkáních, v tomto případě v profesionální fotbalové soutěži na základě logistické regresi má potenciál a velmi pravděpodobně lze dojít k ještě zajímavějším výsledkům, než ke kterým jsem se dostal já osobně.

8 Summary

The bachelor thesis is aimed to create a probability model of the team's winning in a football match on the basis of the logistic regression where the explanatory variables include the ranking of individual teams in the table before the match, previous mutual matches and the advantage of the team's home playground. For calculating of the odds I used the STATISTICA statistical programme which computed the probability of every variant that may happen in the football match. The predicted odds were then compared with the actually published odds of the Tipsport professional betting company. The utilisation of statistics in the world of sport, e.g. the professional soccer, is increasingly frequent and, therefore, I decided to attempt to create a functional model for predicting the football match result using the statistical method. The theoretical part of this thesis introduces betting as a whole with all more or less important essential elements, such as the history of betting in the Czech Republic and abroad, the activity of bookmakers or how the odds are created by the professional betting agency. This part is followed by the introduction of the logistic regression. It is especially the logistic regression my model for estimating the football match result and the betting odd is based on. The last part of the thesis includes the final comparison of the model created by me and its results and the results of the betting agency. This section also contains other minor experiments which compare the results of the statistic model with the work of the real bookmaker in an interesting way.

Key words: Statistics, statistic regression, betting odds

9 Zdroje a citace

- [1] Historie kurzového sázení (2016). In: *Encyklopedie hazardu.cz* [online]. [cit. 2017-8-13]. Dostupné z: https://www.encyklopedie hazardu.cz/rubriky/kurzove-sazeni/historie-kurzoveho-sazeni_22.html
- [2] Kurzové sázení (2017). In: *Cs.wikipedia.org* [online]. [cit. 2017-8-13]. Dostupné z: https://cs.wikipedia.org/wiki/Kurzové_sázení
- [3] Herní plán pro provozování internetových kurzových sázek. (2017). In: *Chance.cz* [online]. [cit. 2017-8-14]. Dostupné z: <http://home.chance.cz/datafiles/hp-internet-chance.pdf>
- [4] Sázkový kurz (2016). In: *Hazardni-hry.eu*. [online]. [cit. 2017-8-14]. Dostupné z: <http://www.hazardni-hry.eu/sport/sazkovy-kurz.html>
- [5] Skutečnosti ovlivňující výši kurzu (2017). In: *Cs.wikipedia.org*. [online]. [cit. 2017-8-13]. Dostupné z: https://cs.wikipedia.org/wiki/Kurzové_sázení#Skutečnosti_ovlivňující_výši_kurzu
- [6] Podrobná pravidla sázek (2016). In: *Tipsport.cz*. [online]. [cit. 2017-8-14]. Dostupné z: https://home.tipsport.cz/doc_net/podrobna-pravidla-tipsport-net.pdf
- [7] Sázení po internetu (2016). In: *Cs.wikipedia.org*. [online]. [cit. 2017-8-15]. Dostupné z: https://cs.wikipedia.org/wiki/Sázení_po_internetu
- [8] Vincent, L. Škola sázení: Tipovat sportovní zápasy začali lidé už v 17. století (2009). In: *Sazeni.idnes.cz*. [online]. [cit. 2017-8-13]. Dostupné z: https://sazeni.idnes.cz/skola-sazeni-tipovat-sportovni-zapasy-lide-zacali-uz-v-17-stoleti-py0-/sazeni.aspx?c=A090703_090239_sazeni_pes
- [9] Šedivý, P. Práce bookmakerů (2009). In: *Sazeni.idnes.cz*. [online]. [cit. 2017-8-15]. Dostupné z: https://sazeni.idnes.cz/video-tvori-kurzy-a-sleduji-sport-presto-prace-bookmakeru-neni-med-1fp-/sazeni.aspx?c=A090707_120656_sazeni_pes

- [10] Vincent, L. Škola sázení: Kurz odráží šanci na výhru (2009). In: *Sazeni.idnes.cz*. [online]. [cit. 2017-8-14]. Dostupné z: https://sazeni.idnes.cz/skola-sazeni-kurz-odrazi-sanci-na-vyhru-fc8-/sazeni.aspx?c=A090723_094130_sazeni_pes
- [11] Tůma, O. Kurzové sázky: miliardový byznys (2011). In: *Peníze.cz*. [online]. [cit. 2017-8-13]. Dostupné z: https://www.penize.cz/spotrebitel/227043-kurzove-sazky-miliardovy-byznys#element_54_8643
- [12] České sázkové kanceláře a svět internetového sázení (2009). In: *Lupa.cz*. [online]. [cit. 2017-8-14]. Dostupné z: <https://www.lupa.cz/clanky/ceske-sazkove-kancelare-svet-internetoveho-sazeni/#zacatek>
- [13] Proč používáme zobecněné lineární modely (2010). In: *Portal.matematickabiologie.cz*. [online]. [cit. 2018-2-15]. Dostupné z: <http://portal.matematickabiologie.cz/index.php?pg=analyza-a-hodnoceni-biologickych-dat--regresni-modelovani--logisticky-regresni-model-a-jine-zobecnene-linearni-modely--proc-pouzivame-zobecnene-linearni-modely>
- [14] Generalized linear model (2017). In: *En.wikipedia.org*. [online]. [cit. 2018-2-15]. Dostupné z: https://en.wikipedia.org/wiki/Generalized_linear_model
- [15] Neubauer, J. Zobecněný lineární model (2017). In: *k101.unob.cz*. [online]. [cit. 2018-2-15]. Dostupné z: http://k101.unob.cz/~neubauer/pdf/ekon_zobecneny_linearni_model.pdf
- [16] Logistická regrese a logity (2013). In: *scio.cz*. [online]. [cit. 2018-2-15]. Dostupné z: https://www.scio.cz/o-vzdelavani/teorie-a-metodika-testu/statisticke-pojmy/#logisticka_regrese_logity
- [17] Logistická regrese (2017). In: *trilobyte.cz*. [online]. [cit. 2018-2-17]. Dostupné z: <http://www.trilobyte.cz/downloadfree/qcemanual/logreg.pdf>
- [18] Logit (2017). In: *En.wikipedia.org*. [online]. [cit. 2018-2-17]. Dostupné z: <https://en.wikipedia.org/wiki/Logit>
- [19] Logistická regrese (2014). In: *En.wikipedia.org*. [online]. [cit. 2018-4-20]. Dostupné z: https://cs.wikipedia.org/wiki/Logistická_regrese

- [20] STATISTICA Features Overview (2018). In: *Statsoft.com*. [online]. [cit. 2018-4-20]. Dostupné z: <http://www.statsoft.com/Products/STATISTICA-Features>
- [21] Statistica (2018). In: *En.wikipedia.org*. [online]. [cit. 2018-4-20]. Dostupné z: <https://en.wikipedia.org/wiki/Statistica>
- [22] STATISTICA Base CZ (2017). In: *Statsoft.cz*. [online]. [cit. 2018-4-21]. Dostupné z: <http://www.statsoft.cz/produkty/2-statistica-professional/12-statistica-base-cz/>
- [23] STATISTICA Base CZ - Detail (2017). In: *Statsoft.cz*. [online]. [cit. 2018-4-21]. Dostupné z: <http://www.statsoft.cz/produkty/2-statistica-professional/12-statistica-base-cz/detail/>
- [24] McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman and Halls.
- [25] Crawley, M. J. (2007). *The R Book*. Chichester: Wiley and son.
- [26] Petronius. (2015). *Kurzové sázení krok za krokem*. E-kniha, Petr Šístek.

10 Seznam obrázků, tabulek a grafů

Seznam obrázků

Obrázek 1: <i>Logistický regresní model pro 1 proměnnou x</i>	20
Obrázek 2: <i>Ukázka 3D logistického modelu s použitím 2 proměnných</i>	21
Obrázek 3: <i>Data s binární závisle proměnnou</i>	25
Obrázek 4: <i>Data s poměrovou závisle proměnnou</i>	25
Obrázek 5: <i>Data s více nezávisle proměnnými a binární závisle proměnnou</i>	25
Obrázek 6: <i>Program STATISTICA - uživatelské prostředí</i>	28
Obrázek 7: <i>Konečná tabulka Premier League pro sezónu 2016/2017</i>	32
Obrázek 8: <i>Oblíbenost spořicíh nástrojů mezi obyvateli</i>	35

Seznam tabulek

Tabulka 1: <i>Nahrazování pozic pro prvních pět kol</i>	34
Tabulka 2: <i>Porovnání rozdílů v průměrných kurzech na všechny varianty</i>	36
Tabulka 3: <i>Porovnání úspěšnosti Tipsportu a logistické regrese</i>	37
Tabulka 4: <i>Četnosti nejúspěšnějšíh kurzů</i>	38
Tabulka 5: <i>Úspěšnost sázkové kanceláře Tipsport</i>	39
Tabulka 6: <i>Úspěšnost logistické regrese</i>	40

Seznam grafů

Graf 1: <i>Tržní podíly sázkových kanceláří na českém trhu</i>	33
--	----