Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky

Modely logistické regrese v oblasti esportových dat

BAKALÁŘSKÁ PRÁCE

Studijní program: Aplikovaná informatika
Studijní obor: Aplikovaná informatika

Autor: Michal Lauer
Vedoucí práce: Ing. Zdeněk Šulc, Ph.D.

Praha, Duben 2022

Prohlášení

Prohlašuji, že jsem bakalářskou práci *Modely logistické regrese v oblasti esportových dat* vypracoval samostatně za použití v práci uvedených pramenů a literatury.

V Praze dne DD. Dubna 2021 .....

Podpis studenta

Poděkování

Rád bych poděkoval panu doktoru Zdenku Šulcovi, který mou bakalářskou práci podpořil a vedl, i přes odlišný studijní obor. Dále děkuji autorům citovaných knih za poskytnutou příležitost se ve logistických modelech zlepšit. Bez nich by se práce psala velmi složitě.

Abstrakt

Cílem bakalářské práce je kvantitativně zanalyzovat esportové zápasy ze hry Coutner-Strike: Global Offensive (CSGO) a predikovat výhry vybraného teamu. Použitý datový soubor je z internetového portálu kaggle.com[1] a obsahuje data od roku 2017 až do roku 2020. Práce je rozdělena do tří částí. V první části je představen esport jako takový, je zde shrnutý jeho vývoj a jsou zde definovány důležité pojmy a termíny. Druhá část obsahuje popis metod, které jsou pro analýzu a predikci použity. Pro analýzu data setu jsou zobrazeny grafy jako boxplot či čárový graf. Predikce jsou založené na vícerozměrném logistickém regresním modelu. V závěrečné praktické části jsou metody použité pro analýzu data setu, predikci výhry daného teamu a model je vyhodnocen jak kvantitativně, tak i v kontextu reálného využití.

Klíčová slova

Model, logistická regrese, predikce, esport

[1] https://www.kaggle.com/mateusdmachado/csgo-professional-matches

Abstract

– Bude přeložen po odsouhlasení abstraktu v češtině

Keywords

Model, logistic regression, prediction, esport

Obsah

1. Úvod

Esport je označení pro elektronický sport. Obsahuje všechny důležité oblasti jako klasický sport (např. turnaje, trénování, investice, stadiony, či sázení) s tím rozdílem, že se hraje na nějakém zařízení (počítač, konzole, mobil). Je to jedno z nejrychleji rostoucích odvětví v dnešní době. V roce 2021 se tržní hodnota esportu pohybovala kolem jedné miliardy dolarů - skoro 50% nárůst oproti roku 2020. Lze předpovídat, že v roce 2024 esport překročí hodnotu 1,5 miliardy dolarů (Gough 2021). Dalo by se spekulovat, že za takový velký nárůst je zodpovědná aktuální pandemie. Většina populace je nucena zůstat doma. Toto otevřelo dveře se s esportem přirozeně seznámit a nějakým způsobem se ho účastnit (online divák, soutěžící, organizátor, fanoušek...). Hrají se různé kategorie her jako např. střílečky, Multiplayer online battle arena (MOBA)[1], karetní hry, First-person shooter (FPS) či Battle Royale (BR).

Práce se zaměřuje primární na esportový titul Coutner-Strike: Global Offensive (CSGO). Je to jeden z nejdéle hraných esportových titulů, boří mnohé divácké rekordy[2] a je aktuálně nejhranějším FPS esport titulem. CSGO vyniká nejen detailní herní mechanikou, ale i bohatou a zajímavou historií. Hra je unikátní i tím, že obsahuje mnoho různých módu[3] a hráč může strávit mnoho hodin pouze objevováním komunitních serverů, hraním klasických zápasů či trénováním na offline mapách.

Finální cíl práce je vytvořit logistický regresní model, který předpovídá výsledek zápasů. Pro tvorbu kvalitního modelu bude kritické zvolit vhodné prediktory. Použitý data set[4] obsahuje čtyři soubory, které podávají informace jak už o zápase (např. datum, výsledek zápasu, výsledek jednotlivých map, typ zápasu), hráčích (např. statistiky za zápas, statistiky za mapy, statistiky za team), tak o vývoji celého zápasu (především ekonomika týmu). V práci bude tedy vytvořeno více specializovaných modelů pro každý vybraný týmů a následně je pro každý tým vybrán nejlepší model. Výsledné modely jsou v závěru mezi sebou porovnány.

Logistický model je preferován kvůli své lehké interpretaci a dobré aplikaci v reálném životě. Výsledky, statistiky a pravděpodobnosti mohou být použity např. v sázkových kancelářích, kdy se výsledky modelu dají využít na nejrůznější sázky a lze předpovídat, kdo vyhraje zápas, kdo vyhraje mapu, jaký hráč bude mít nejlepší statistiky, či zda si hráč koupí určitou zbraň.

Práce je tedy rozdělená do tří částí. V první části je kladen důraz na esport, jeho vývoj, a na esportový titul CSGO. Jsou zde také představená pravidla, podle kterých se hra hraje. V druhé části jsou popsány popisné a statistické metody. Jsou zde definované grafické nástroje pro popis datasetu, logistický regresní model, a evaluační nástroje pro model. Třetí část se zaměřuje na praktickou tvorbu modelů, jejich interpretaci, a vzájemné porovnání.

[1] tzn. MOBA, kde hráči hrají v jedné online aréně proti sobě

[2] https://www.invenglobal.com/articles/15619/csgo-major-breaks-viewership-records-overtakes-the-international

[3] rozšíření, jak hru hrát. Každý mód má svá vlastní pravidla, mapy, či herní fanoušky

[4] https://www.kaggle.com/mateusdmachado/csgo-professional-matches

2. Představení esportu

2.1 Historie esportu

I přes fakt, že esport není obecně známý pojem mezi širokou veřejností, má přes 70 let bohaté historie. Za jeho počátky by se daly považovat arkádové automaty, kde hráči z počátku soutěžili sami proti sobě. Největší rozvoj arkádových automatů se děl kolem 70 let minulého století. Nejen za tímto účelem byla 9. 2. 1982 založena Twin Galaxies National Scoreboard (TGNS). TGNS měla na starosti nejen udržování výsledkové tabulky (scoreboard), ale i tvorbu prvotních pravidel pro férovou hru. Za tímto účelem byla vydána kniha Twin Galaxies' Official Video Game & Pinball Book of World Records.

Na přelomu osmdesátých let minulého století se začal esport vyvíjet již více profesionálním směrem. V roce 1972 pořádala Stanfordská Universita historicky první esportový turnaj v arkádové hře Spacewar!. Výherce si mohl odnést předplatné magazínu Rolling Stones. Dále v roce 1983 byl založen první esportový profesionální tým, který se nacházel ve Spojených státech. Všechno toto se stalo díky podnikateli Walteru Day, který je zakladatel společnosti TGNS a založil již zmíněný prvních esportový tým. Ač se Walter považuje za jednoho z hlavních pionýrů esportu, v roce 2010 TGNS opustil kvůli své vášni pro hudbu.

Další důležitou kapitolou ve vývoji esportu je příchod internetu a výkonných počítačů. Hráči se dostali k rychlejším sestavám, stolní počítače se stali cenově dostupnějšími a díky tomu se zpřístupnili k více lidem. Klesala cena hardwaru, vývoj nové technologie a her se zrychloval. Díky rozvoji počítačových sítí se mohli hrát LAN[1] party či organizovat BYOC[2] turnaje. Dále už esport potřeboval jen čas na organický růst a dnes má tržní hodnotu přes jednu miliardu amerických dolarů (Gough 2021), (Larch 2019).

2.2 Zasazení do dnešní doby

V dnešní době je esport téměř miliardová záležitost. Díky pandemii, která trvá již od r. 2019, si esport ještě přilepšil. Dle průzkumu[3] z října roku 2020 si 73 % dotázaných myslelo, že se úroveň zájmu a obchodní činnost esportu v Q4 2020 a Q1 2021 zvětší. Respondenti, kteří se průzkumu zúčastnili, jsou považování za experty v oblasti esportu. Tento průzkum byl následné podpořen růstem že tržní hodnoty esportu a mezi lety 2019 a 2020 vzrostla o téměř 50 % (Gough 2021).

[1] Hráči hrají v jedné místnosti na lokální počítačové síti.
[2] z ang. Bring Your Own Computer, kde si hráči si na akci donesou vlastní počítač
[3] https://www.statista.com/statistics/1247902/covid-impact-esports-investments

K takto prudkému růstu tržní hodnoty esportu z velké části přispěla právě pandemie. Mladá generace byla nucena zůstat doma, což dovolilo i esportem nedotčeným jedincům do tohoto světa proniknout. Větší zájem o esport přinesl i větší tržby herním studiím, která začala do esportových turnajů více investovat (Professeur 2021), (liquipedia 2021). S větším počtem diváku roste i marketingový potenciál, investiční příležitost a kariérní růst.

Druhý dominantní žánr je FPS. V této kategorii jsou nejvýznamnější hry CSGO a Valorant. V tomto žánru proti sobě hrají dva týmy, většinou složené z pěti hráčů. Každý hráč pak má v týmu různou roli, jako např. velitel či odstřelovač. Jeden tým má obvykle za úkol něco zničit (položit bombu, unést rukojmí) a druhý tým jim v tom musí zabránit (ochránit oblast proti bombě, záchrana rukojmí).

Poslední žánr který zmíním je Battle Royale (BR). V těchto hrách hraje buď každý hráč sám za sebe, ve dvojici, nebo ve skupině po čtyřech. Zde hráči padají na začátku kola na velkou mapu. Jejich úkolem je získat vybavení, aby mohl porazit ostatní hráče a kolo sami, nebo s týmem vyhrát. Nacházejí se zde různé role, avšak trošku rozdílné oproti žánru FPS. Hlavním titulem této kategorie je hra Fortnite, která žánru dominuje. Stal se z ní jak esportový titul, tak perfektní marketingové místo pro teenagery. Hráči si zde mohou koupit oblečky různých filmových či komiksových postav. Pokud vychází nový film, ve hře se může objevit „event“ (událost), který daný film propaguje. Toto lze vidět například na propagaci Avengers: Endgame[4].

2.3 Představení titulu Counter-Strike: Global Offensive

CSGO, jak ho známe dnes, má bohatou a dlouho historii. Ne vždy se to ovšem jmenovalo stejně. Úplně první iterace hry se jmenovala čistě Counter-Strike a byl to pouze mód[5] do hry Half-Life. Half-Life byl vyvinutí společností Valve, tehdy primárně společností zaměřenou na vývoj her. Mód byl vytvořen studenty vysoké školy, panem Minh Le a Jess Cliffe. Toto rozšíření začali programovat v roce 1999. Jelikož mód byl neoficiálním rozšířením, Valve o něj neprojevovalo veliký zájem. Až po pěti betaverzích hry Counter-Strike si společnost Valve všimla rozšíření, její komunity, ale především jejich autorů. Minh a Jess se v roce 2000 stali oficiálními zaměstnanci Valve a duševní vlastnictví módu prodali. Autoři, nově jako zaměstnanci Valve, roku 2000 vydávají první oficiální verzi hry Counter-Strike. I přes toto „oficiální“ datum vydání je většina komunity přesvědčena, že výročí má CSGO v den svého úplně první vydání, a to 18. června 1999.

[4]Trailer pro propagaci události: https://www.youtube.com/watch?v=TanGK9o_d24

[5]upravení či rozšíření hry

Hra je z žánru FPS a hraje se primárně online proti skutečným hráčům. Counter-Strike se v herní komunitě rychle rozrostl díky své jednoduchosti. Hra se dá velmi dobře popsat pořekadlem „Lehké hrát, těžké vypilovat“. Hra má mechaniky[6], které jsou lehké na pochopení, ale velmi těžké na vypilování k dokonalosti. Spolu s touto vlastností je hra vlastně velmi jednoduchá a hráč hraje buď za policisty, nebo za teroristy. Hráči tak mohli, a stále mohou, hru velmi lehce a rychle začít hrát, jelikož se tento formát od roku 2000 nijak extrémně nezměnil.

Hra tedy rostla zejména díky své komunitě. Hráči hru různě upravovali, přidávali další módy, typy her, zbraně, mapy či audiovizuální obsah. Tento trend se přenášel přes mnoho různých verzí hry. První velký „průlom“ udělala verze 1.6, tedy Counter-Strike 1.6. Ta vynikala jak esportem, tak komunitním obsahem. Jen v České a Slovenské republice bylo několik herních serverů, na kterých se mohlo sejít sta tisíce hráčů. Např. na česko-slovenském herním portálu kotelna hrálo celkem přes 1,5 milionu unikátních hráčů (csko 2021). Hra byla populární nejen mezi obyčejnými hráči, ale i profesionály.

Counter-Strike 1.6 je pionýrem esportu pro FPS žánr. Za podpory Valve se hráli první major[7] turnaje, kde hráči mohli ukázat svůj um za tehdy relativně velkou sumu peněz. Hra se časem vyvíjela, hráči nalézali nové strategie či triky a Valve vydalo novou verzi — Counter-Strike: Source. Tato nová verze získala nepříliš pozitivní ohlas, jelikož velmi rozdělila herní komunitu. Představila nové mechaniky, staré mechaniky změnila a hráčům, zejména v esportu, se nechtělo učit něco úplně nového. Valve se rozhodlo sjednotit herní komunitu, a proto vydalo novou verzi hry s názvem CSGO

CSGO se snažilo sjednotit oba tábory z her Counter-Strike 1.6 a Counter-Strike: Source. Hra vyšla 21. srpna 2012 a z počátku nebyla tolik úspěšná, ale díky přidání různých skinů (Valve 2013) na zbraně hra přilákala úplně nové publikum. Díky novému a velkému publiku se začali hrát menší esportové turnaje právě ve hře CSGO, ke kterým se později přidali i profesionálové z předchozích dvou verzí. Díky tomuto organickému růstu má Counter-Strike velmi silnou komunitu, která se o hru i nadále stará. I přes netradiční interakci mezi Valve a herní komunitou hra stále roste. CSGO se díky své dlouhé historii, bohaté komunitě a různým možnostem, jak hru hrát, dostala na špičku esportu. I přes několik titulů, které se s hrou snaží soutěžit, je hra stále největším a nejsledovanějším esport titulem v rámci FPS žánru (Henningson 2020).

[6] herní prvky či unikátní vlastnosti

[7] turnaj pořádaný přímo Valve, který má největší prestiž

2.4 Propojení práce a titulu Counter-Strike: Global Offensive

Práce se zaměřuje na identifikování významných prediktorů a následně vytvoření regresního modelu. Před jakoukoliv prací s daty je ale nutné pochopit, jak se hra vlastně hraje a jaká jsou její pravidla. Ve hře CSGO hraje pět hráčů proti pěti (dále jen 5v5). Hra se většinou hraje online, avšak velké esportové turnaje se hrají offline, tedy v nějaké např. aréně. Hra má v základu 30 kol a po prvních patnácti se mění strany. Jedna strana jsou policisté (Counter-Terrorists či CT), kteří mají za úkol chránit „bomboviště“ - část mapy, která má vybouchnout. Naopak cíl Teroristů (T) je právě bombu položit a „bomboviště“ nechat vybouchnout. Vyhrává tým, který první vyhraje 16 kol. Pokud ovšem po první 30 kolech je stav nerozhodný, tedy 15:15, hraje se prodloužení. Tento formát není standardizovaný pro všechny turnaje, proto zmíním pouze pravidla, která se týkají turnajů od společnosti Valve (již zmíněné a nejvíc prestižní Majory). Zde se hraje prodloužení ve formát Bo6, tedy kdo první získá 4 body, vyhraje zápas. Takto může jít zápas teoreticky do nekonečna. Nejdelší semi-profesionální zápas, který se ovšem neodehrál na Majoru, se stal mezi týmem exceL a XENEX(hltv.org 2015). Zápas pokračoval do úctyhodných 88 kol.

V každém kole má tým určitý počet peněz. Každá hráč začíná polovinu (ted v první a šestnácté kolo) s $800. Finance každého hráče pak záleží na mnoha faktorech, jako kolik vyhrál jeho tým kol v řadě, kolik nakoupil zbraní, kolik zabil nepřátel, kolik peněz dostane hráč za zabití či jak kolo skončí. V profesionálním týmu je velmi obtížné pracovat s financemi, jelikož všichni musí být v tomto ohledu jednotní. V tuto chvíli přichází na řadu tzn. In-Game Leader (velitel týmu). Tuto roli má většinou jeden hráč v každém týmu. Je to ta nejdůležitější role ze všech Má na starosti např. finance týmu, rozhoduje kdy se koupí a kdy půjde tzn. eco (hráči nekoupí nic, aby ušetřili peníze), jaké se budou hrát mapy či jaká se půjde v daném kole strategie.V dnešní době k tomu In-Game Leader má i trenéra. Ten hru nehraje, ale pozoruje hráče a dává jim různé typy a triky.

Role trenéra není nijak silně definovaná a každý esportový tým má trošku jiného trenéra. V jednom případě může být trenér čistě jako podpora a pomáhá hráčům když se nedaří a řeší interní problémy. V jiném týmu může ovšem mít velký zásah do hry, pomáhat In-Game Leaderovi se strategiemi, obelstění soupeře či sledováním předchozích zápasů pro kontinuální zlepšování týmu. Další role v týmu jsou například Entry Fragger (má za úkol získat první zabití pro tým), support (podporuje svůj tým za pomoci různých granátů nebo se často pro svůj tým obětuje), AWP hráč (hráč je specifický tím, že hraje primárně s jednou zbraní) a Lurker (chodí po mapě sám a snaží se nepřítele odchytnout ze stran, které by nečekali)

Zápasy se pak hrají ve formátech „Best of“. Best of 3 například znamená, že se hrají tři mapy. Kdo první vyhraje dvě mapy, vyhrál celý zápas. Turnaje se pak odehrávají v tradičních formátech, jako je pavouk. Ten se charakterizuje tím, že vypadá jak pavučina, jde z leva doprava a každý tým může prohrát pouze jednou. Následně tu máme Upper/Lower bracket formát, který je v podstatě pavoučí formát, akorát jsou zde dvě „sítě“ a každý tým může prohrát maximálně jednou, jelikož druhá prohra znamená vyřazení z turnaje. Specifičtější formát pro CSGO je například swiss, který se počítá přes různé body a statistiky výsledných zápasů.

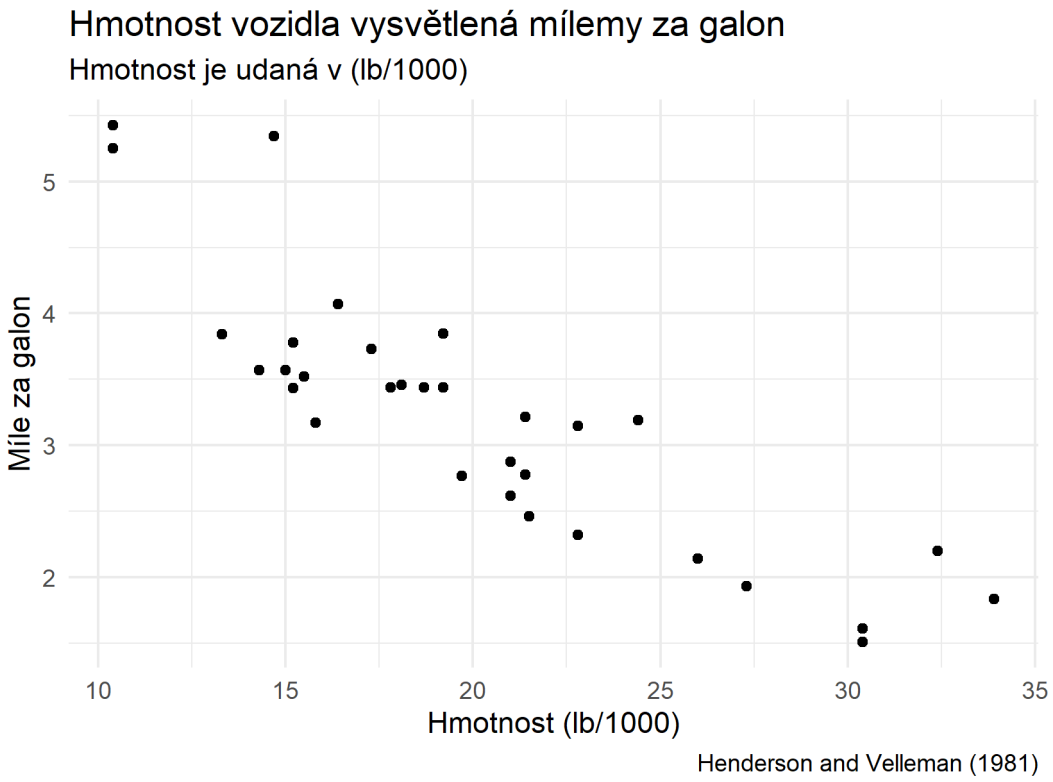
3. Teoretická část

V následující části jsou popsány jak teoretické metody pro vizualizaci dat, tak i tvar, forma a vyhodnocení logistického regresního modelu. Ke každé části, která se věnuje popisu dat pomocí nějakého grafu, je přidána praktická ukázka s popisem a praktickým vysvětlením. vhodné. Testovací citace: (Hebák 2015), (Kleinbaum 2010)

3.1 Vizualizace dat

3.1.1 Bodový graf

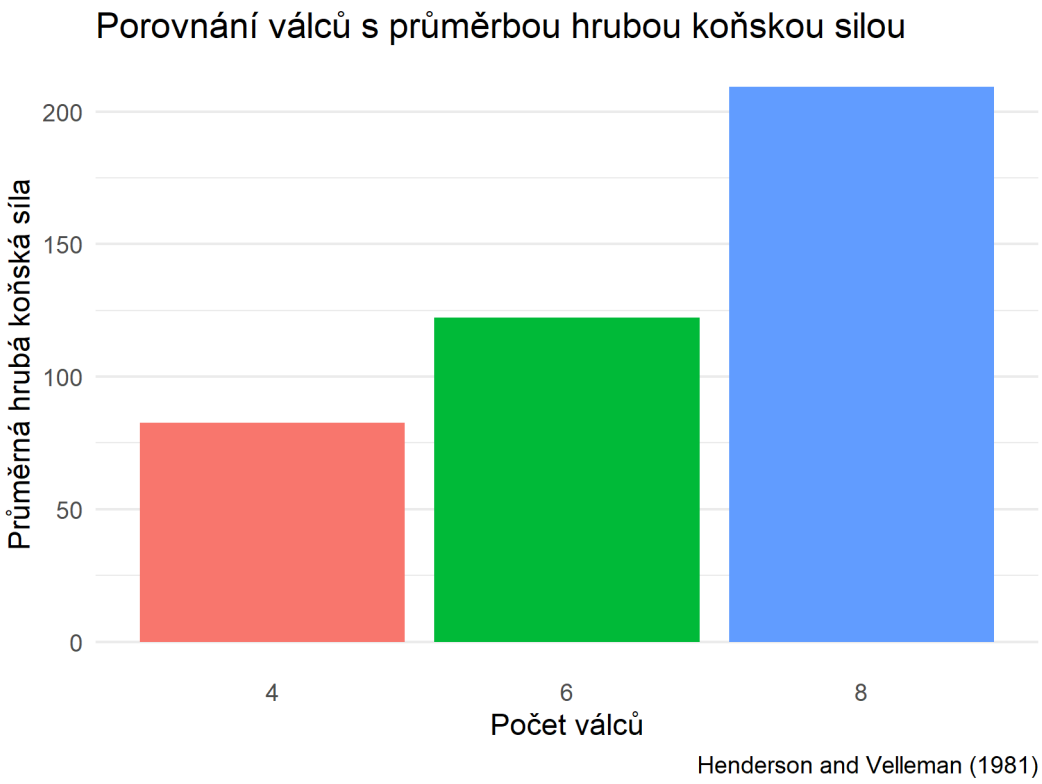
Bodový graf slouží pro zobrazení vztahu dvou kvantitativních proměnných. Z pravidla se vysvětlovaná proměnná dává na osu Y, zatímco proměnná vysvětlující se nachází na ose X. Vysvětlovaná (nezávislá) proměnná je ta proměnná, která má být předvídaná. Vysvětlující proměnná se naopak snaží vysvětlovanou proměnnou předpovědět či popsat. Propojením vysvětlované a vysvětlující proměnné na bodovém grafu lze vidět např. sílu korelace nebo vztah mezi proměnnými (např. lineární, kvadratický, logaritmický). Obrázek 3.1 zobrazuje negativní korelaci mezi hmotností vozidla a mílemi ujetými za galon.



Obrázek 3.1: Bodový graf hmotnosti a míly za galon z datasetu mtcars

3.1.2 Sloupcový graf

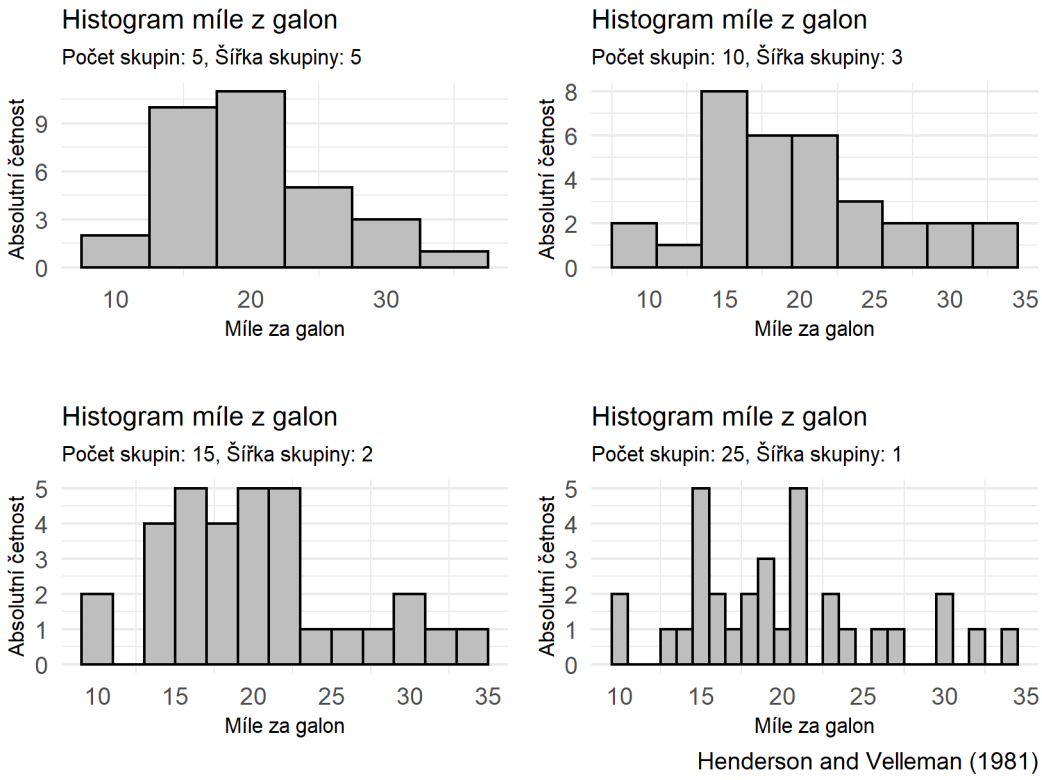
Sloupcový graf slouží k porovnání četnosti kategorií. Na jednu osu (z pravidla osu X) se položí možné kategorie. Na druhou osu se pak položí sledovaná statistika. Sledovat můžeme nejen četnost, ale i průměr či kteroukoli jinou statistika, kterou bude možné na ose zobrazit. Pokud je typ statistické proměnné nominální, tedy data pouze popisná, lze pomocí sloupcového grafu pouze porovnávat sledovanou statistiku. Lze tedy vnímat a porovnávat jednotlivé kategorie, ale nemůžeme z toho usoudit nějaký vztah. Mají-li ovšem kategorie nějaké přirozené řazení, je možné sledovat i určitý vztah mezi nimi a usoudit, že s změnou kategorie roste i sledovaná statistika. Příklad sloupcového grafu je zobrazen na obrázku, který porovnává průměrnou hrubou koňskou sílu s počtem válců. Je na něm také vidět vztah, kdy s vyšším počtem válců stoupá průměrná koňská síla.



Obrázek 3.2: Sloupcový graf z datasetu mtcars

3.1.3 Histogram

Histogram je speciální typ sloupcového grafu. Hlavní rozdíl je v tom, že popisuje rozdělení spojité proměnné a mezi sloupci není žádná mezera. Pro histogram je třeba data sloučit do skupin *(bins)* o určité šířce. Správný výběr počtu skupin je kritický, jelikož může velmi silně ovlivnit interpretaci dat. Pokud se vybere příliš malý počet skupin, data se seskupí a může se ztratit důležitý vztah. Pokud se ovšem vybere moc velký počet skupin, v datech bude obtížné najít nějaký obecný vztah či trend. Tento efekt je znázorněn na obrázku 3.3.



Obrázek 3.3: Porovnání histogramů s různým počtem skupin

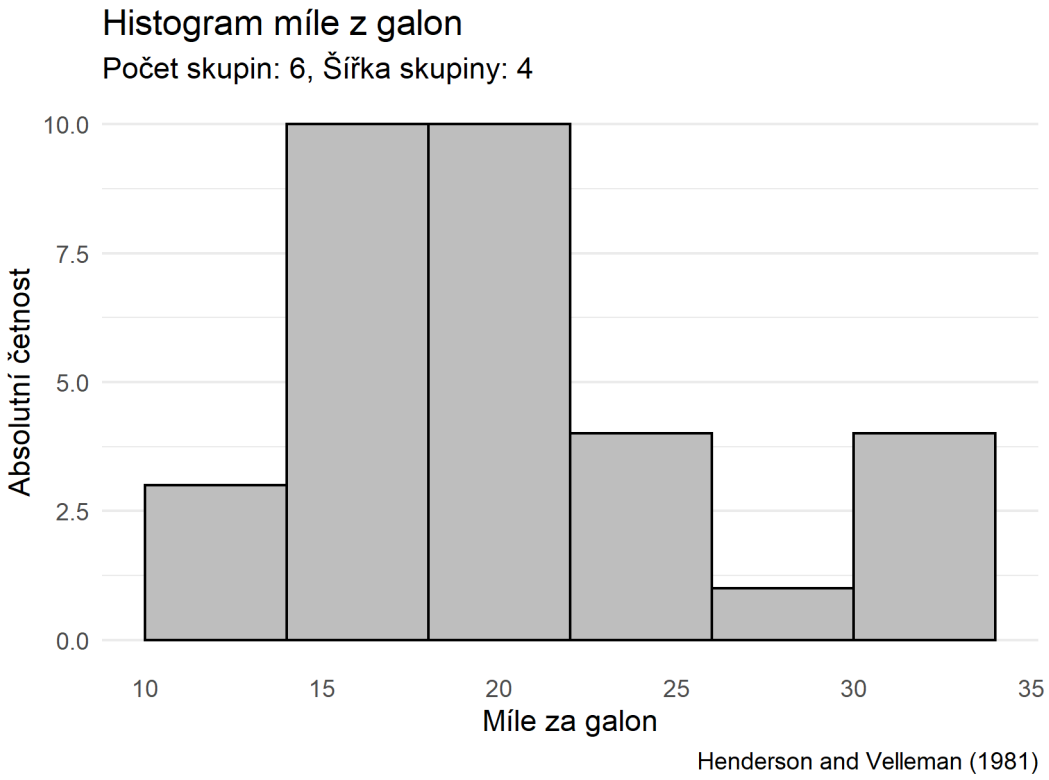
Pro vhodný počet skupin existuje mnoho způsobů. Nejznámější je takzvané Sturgesovo pravidlo, které se spočítá následujícím vztahem:

$$k \doteq 1 + 3{,}3 * log_{10}(n) \tag{3.1}$$

kde k je výsledný zaokrouhlený počet skupin nahoru, a n je počet pozorování. Druhý parametr, který je pro tvorbu histogramu potřeba, je šířka skupiny. Ta by měla být ideálně stejná pro všechny skupiny. Pokud tomu tak není, histogram může být zavádějící a čtenář mu nemusí plně rozumět. Pro vypočtení počtu skupin má šířka skupiny následující tvar:

$$w = \frac{max(x) - min(x)}{k} \tag{3.2}$$

kde x je zobrazovaná proměnná, k je počet skupin a w je výsledná šířka intervalu. Uplatněním rovnic 3.1 a 3.2 na dataset z obrázku 3.3 lze zobrazit representativnější sloupcový graf. Nutné je však podotknout, že není pravidlem se danými výpočty řídit a výsledný sloupcový graf je nutné přizpůsobit jednotlivým datům.



Obrázek 3.4: Histogram s počtem skupin dle Sturgesova pravidla

3.1.4 Boxplot

Five-number summary

Five-number summary je číselná tabulka, která pomocí pěti různých čísel shrnuje seřazenou číselnou řadu. Základní statistický nástroj pro vytvoření takové tabulky jsou kvantily. Hodnota P-tého percentilu označuje číslo, které rozděluje seřazenou číselnou řadu na dva intervaly. První interval obsahuje $P * 100\%$ číselné řady a druhý analogicky $(1 - P) * 100\%$. Různé hodnoty percentilů mohou mít specifičtější pojmenování a značí se Q_P. Percentil $P = 0{,}5$ se označuje jako medián a rozděluje seřazenou číselnou řadu na polovinu. Percentily, kde $P = 0{,}25$ nebo $P = 0{,}75$, se označují jako kvartily a značí se Q_1 a Q_3. Oba tyto typy kvartilů jsou použité při tvorbě Five-number summary tabulky. Jako příklad je uvedena následující tabulka

$Q_0(Q_0)$	$Q_{0,25}(Q_1)$	$Q_{0,50}$	$Q_{0,75}(Q_3)$	$Q_{1,00}$
1,513	2,58125	3,325	3,61	5,424

Tabulka 3.1: Five-number summary tabulka hmotnosti vozidla (lb/1000)

kde Q_0 a $Q_{1,00}$ označují minimum a maximum číselné řady. Kvartily Q_1, Q_2 (medián) a Q_3 jsou čísla, která rozděluji časovou řadu na na čtvrtiny. V prvním případě, tedy $Q_1 = Q_{0,25}$, je 25% čísel menší než 1,513 a 75% dat větší. Pro kvantil $Q_3 = Q_{0,75}$ je 75% čísel menších než 3,61 a 25% větších. $Q_{0,50}$ označuje medián.

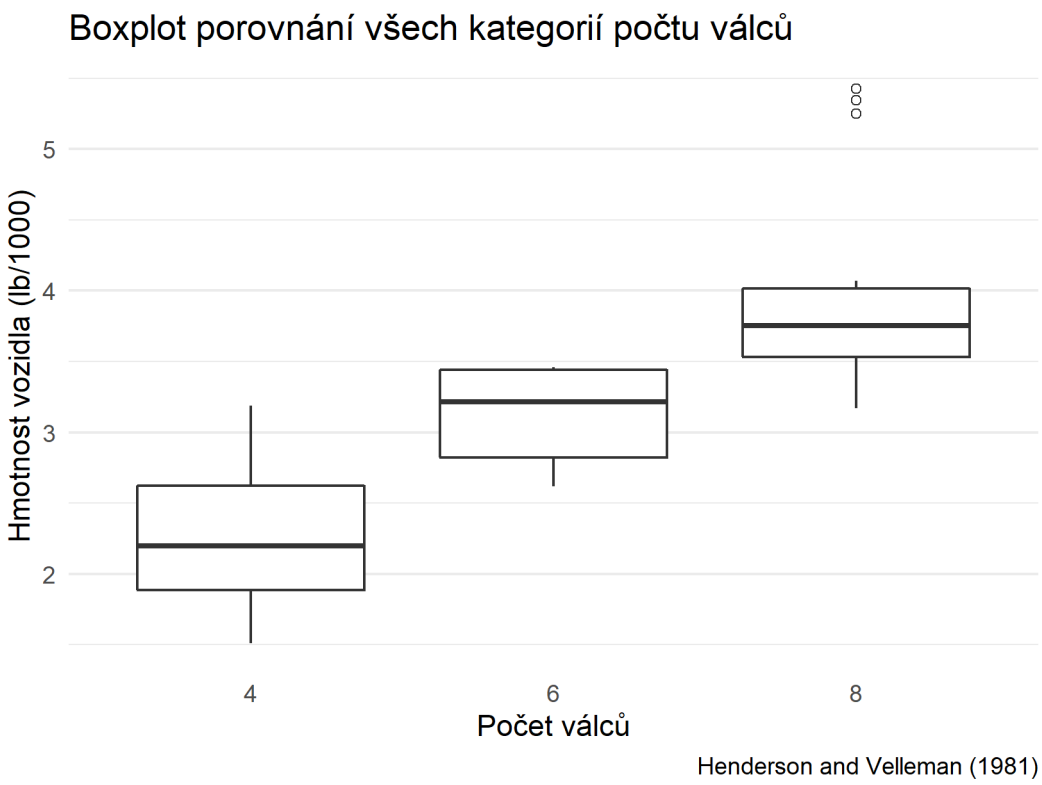
Boxplot

Boxplot je grafické zobrazení a rozšíření Five-number summary tabulky. Kromě grafického zobrazení pěti kvantilů ukazuje odlehlé a extrémní hodnoty. V boxplotu se také nachází obdélník, který ukazuje mezikvartilové rozpětí (IQR), tedy prostředních 50 % dat. V obdélníku se také nachází černá čára, která značí medián. Z prostředního obdélníku vedou oběma směry čáry, jejichž konce značí hranici pro odlehlá pozorování. Pozorování, která jsou buď větší než horní hranice, nebo menší než spodní hranice, označujeme jako odlehlé nebo extrémní.

$$\textit{Spodní hranice} = Q_1 - 1{,}5 * IQR \tag{3.3}$$

$$\textit{Horní hranice} = Q_3 + 1{,}5 * IQR \tag{3.4}$$

Hodnoty, které spadají do intervalu $\langle Q_1 - 1{,}5IQR; Q_1 - 3IQR\rangle$ a $\langle Q_3 + 1{,}5IQR, Q_3 + 3IQR\rangle$ se nazývají jako odlehlé. Hodnoty které leží mimo tento vztah, tedy hodnoty menší než $Q_1 - 3IQR$ nebo větší než $Q_3 + 3IQR$ se nazývají jako hodnoty extrémní a v boxplotu jsou z pravidla vyznačeny nějakým speciálním znakem, např. kolečkem. Díky grafickému zobrazení lze lehce porovnávat rozdělení jedné vysvětlované kvantitativní proměnné tříděné přes několik kategorií.

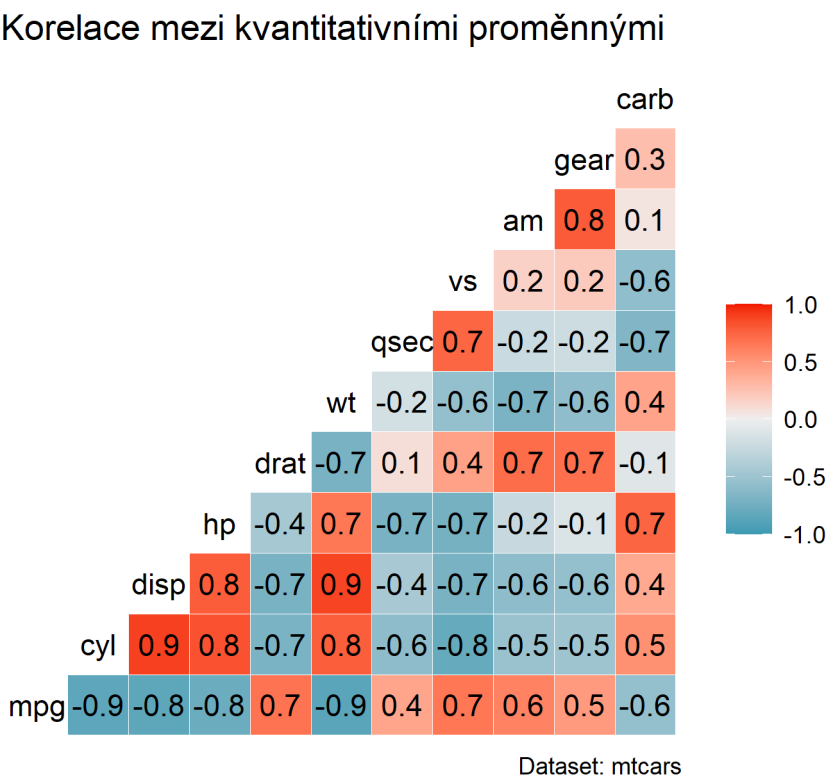


Obrázek 3.5: Boxplot hmotnosti auta pro různé počty válců

Průhledná kolečka v obrázku 3.5 v kategorii osmi válců značí odlehlé hodnoty, t.j. hodnoty v intervalu $\langle Q_3 + 1{,}5IQR, Q_3 + 3IQR \rangle$.

3.1.5 Korelační matice

Korelační matice je způsob, jak zobrazit korelaci mezi více jak dvěma proměnnými. Matice může být zobrazená jako tabulka nebo graf. Korelační matice je velmi užitečná v regresní analýze kvůli předpokladu nezávislosti. Pokud jsou dvě hodnoty vysoce korelované, musí se na to při tvorbě regresního modelu brát ohled.



Obrázek 3.6: Korelace

Graf korelační matice může mít mnoho podob. V příkladu obrázku 3.6 je zobrazená korelační matice jako teplotní mapa. Z obrázku je možné pozorovat vysokou pozitivní korelaci mezi proměnnými cyl, disp a hp. naopak skoro žádná korelace není mezi proměnnou qsec a proměnnou drat.

3.2 Logistická regrese

Logistická regrese je způsob, jak popsat vztah mezi jedním či několika prediktory a jednou binární vysvětlovanou proměnnou. K tomu slouží spojovací funkce, která transformuje lineární kombinaci prediktorů na index z. V případě logistické regrese se tato funkce nazývá logistická a je definovaná jako

$$f(z) = \frac{1}{1 + e^{-z}}. \tag{3.5}$$

Obor hodnot funkce je interval $\langle 0, 1 \rangle$. Proměnná z je lineární kombinace prediktorů $X_1, X_2, ..., X_k$, jejich koeficientů $\beta_1, \beta_2, ..., \beta_k$ a parametru α.

$$\begin{aligned} z &= \alpha + \beta_1 X_1 + ... + \beta_2 X_2 + \beta_k X_k \\ &= \alpha + \sum_{i=1}^{k} \beta_i X_i \end{aligned} \tag{3.6}$$

Mějme tedy binární vysvětlovanou proměnnou Y, u které hodnota 1 značí výskyt jevu. Pravděpodobnost, že jev nastane vzhledem k definovaným prediktorům lze zapsat jako

$$P(Y = 1 \mid X_1, X_2, ..., X_k) = \frac{1}{1 + e^{-\left(\alpha + \sum_{i=1}^{k} \beta_i X_i\right)}} \tag{3.7}$$

kde α a β_i jsou parametry odhadnuté z datového souboru.

3.2.1 Interpretace parametrů

Parametry α a β_i značí logaritmus šance. α je logaritmus šance v případě, že všechny prediktory jsou teoreticky rovné 0. Parametr β_i značí logaritmus šance pro prediktor X_i V případě, že všechny prediktory jsou konstantní a prediktor X_i se změní o jednotku, přirozený logaritmus šance se změní o B_i. Toto lze pozorovat například u binárních prediktorů, kdy typicky přítomnost daného prediktoru, značená jedničkou, změní výslednou šanci právě o odhadnutý parameter β. Pro přechod z přirozeného logaritmu šance na šanci lze využít vztahu

$$šance \ = e^{\beta_i}. \tag{3.8}$$

Šance je podíl dvou pravděpodobností. Pokud bychom měli šanci jevu A oproti jevu B $2:1$, značí to, že výskyt jevu A je dvakrát tak pravděpodobný jako výskyt jevu B a jev A se vyskytuje ve $\frac{2}{3}$ případů. Šance e^{β_i} tedy značí vztah mezi prediktorem X_i a vysvětlovanou proměnnou Y. Pokud je šance kladná, značí to, že s vyšší hodnotou prediktoru X_i se zvyšuje šance že $P(Y = 1)$. Pokud je naopak nižší, pravděpodobnost se zmenšuje. Pokud je potřeba interpretovat pravděpodobnost jako šanci, použije se logitová funkce

$$\textit{šance jevu } A = \frac{p}{1-p} \tag{3.9}$$

kde p je pravděpodobnost výskytu jevu A.

3.2.2 Maximální pravděpodobnost

Parametry logistického modelu v rovnici 3.7 jsou pouze teoretické a je třeba je odhadnout. Již vypočtené odhady se proto neznačí pouze β, ale $\hat{\beta}$. Pro odhad parametrů se při logistické regresi používá metoda maximální věrohodnosti. Pro výpočet maximální věrohodnosti se počítá pravděpodobnostní funkce $L(\theta)$ kde θ jsou parametry logistického modelu $\alpha, \beta_1, ..., \beta_k$. Pro logistickou regresi má věrohodnostní funkce tvar

$$L(\theta) = \Pi_{l=1}^{m_1} P(X_i) \Pi_{l=m_1+1}^{n} 1 - P(X_i), \tag{3.10}$$

kde n je počet pozorování a m_1 je počet příznivých $(Y = 1)$ jevů. Funkce předpokládá, že datový soubor je seřazen tak, že prvních m_1 výskytů jsou jevy příznivé. $P(X_i)$ poté značí logistickou funkci 3.5. Pro vypočtení optimálního parametru β_i je nutné vypočítat maximum funkce $L(\theta)$ vzhledem k parametru β_i. Parametr β_i lze tedy získat derivací funkce $L(\theta)$ vzhledem k parametru β_i

$$\frac{\partial L(\theta)}{\partial \beta_i} = 0 \tag{3.11}$$

3.2.3 Matice záměn

Matice záměn je nástroj pro vyhodnocení predikcí modelu. Matice je o velikosti nxn. Pro potřeby logistické regrese se matice skládá ze dvou řádku a dvou sloupců. V řádkách se nachází původní hodnoty, tedy hodnoty, které chceme předpovídat. Ve sloupcích se pak nachází předpovědi.

		Pozitivní predikce	Negativní predikce
		1	0
Původní pozitivní	1	True Positive	False Negative
Původní negativní	0	False Positive	True Negative

Tabulka 3.2: Matice záměn

Pro sestavení matice je potřeba množina dat, u kterých známe predikovanou proměnnou. Na datech pak provedeme predikci, díky čemuž získáme predikované hodnoty. Porovnáním původním a predikovaných hodnot vznikne matice 3.2. Každá ze čtyř vnitřních buněk má vlastní označení a interpretaci

- **True Positive** - počet správných predikcí, které byli rovné jedné
- **False Positive** - počet predikcí rovných jedné, kde byla původní hodnota rovná nule
- **True Negative** - počet správných predikcí, které byli rovné nule
- **False Negative** - počet predikcí rovných nule, kde byla původní hodnota rovná jedné

Z matice lze následně vypočítat mnoho statistik. Pro vyhodnocení regresního modelu lze použít např. přesnost, která se vypočítá jako počet všech správných predikcí nad počtem všech provedených predikcí n.

$$\mathit{Přesnost} = \frac{\mathrm{TP} + \mathrm{TN}}{n} \tag{3.12}$$

Může se stát, že je logistický model použít na predikce, kde je kritické předpovídat pozitivní výsledek (např. predikce nemoci). V tomto případě lze použít statistika zvaná citlivost. Ta se rovná poměru správných pozitivních predikcí a úhrnu všech pozitivních predikcí

$$\mathit{Citlivost} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3.13}$$

3.2.4 Testování hypotéz

... Dopsat

3.2.5 Waldův test

Waldův test ověřuje, zda je parametr populaci β_i významný či nikoliv. Definice testu hypotézy je tedy

H_0 : Koeficient β_i je rovný nule
H_A : Koeficient β_i je různý od nuly.

Pro vyhodnocení hypotézy se používá kritická hodnota Z, který se vypočítá jako poměr testovaného parametru $S_{\hat{\beta}_i}$ a směrodatné chyby koeficientu β_i

$$Z = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}} \tag{3.14}$$

Kritická hodnota Z má normální rozdělení Z $N(0, 1)$ a její mocnina, Z^2, má chi-kvadrát rozdělení s jedním stupněm volnosti.

3.2.6 Test poměru věrohodností

Test poměru věrohodnosti slouží jako alternativa k Waldovu testu. K vypočtení Z indexu slouží přirozený logaritmus poměru parciálních věrohodnostních funkcí. V praxi to znamená, že věrohodnostní funkce ve jmenovateli obsahuje méně prediktorů, než věrohodnostní funkce v čitateli. Pokud se tedy věrohodnostní funkce liší o prediktor X_i, jsou hypotézy definovány následovně

H_0 : Koeficient β_i je rovný nule
H_A : Koeficient β_i je různý od nuly.

Do výpočtu kritické hodnoty pak v poměru vstupují hodnoty věrohodnostní funkce 3.10, kde $\hat{L}_1$ značí hodnotu věrohodnostní funkce pro model se všemi prediktory a $\hat{L}_2$ hodnotu věrohodnostní funkce pro model bez prediktoru X_i

$$Z = -2ln(\frac{\hat{L}_1}{\hat{L}_2}) \tag{3.15}$$

při velké hodnotě n má Z zhruba chi-kvadrát rozdělení s jedním stupněm volnosti.

4. Praktická část

V následující části je popis dat, transformace dat, a tvorba logistického regresního modelu. Prve jsou představené datové soubory, se kterými se pracuje. Následně jsou požité grafy, které jsou představené v sekci 3.1. Následně jsou vytvořené logistické regresní modely, jejich výstup je interpretován a různé modelu jsou mezi sebou porovnány. V závěru se nachází každé podsekce se nachází zamyšlení, jak by daný model mohl být vylepšen a jaké je jeho použití v reálném životě.

4.1 Datové soubory

Dataset[1] obsahuje čtyři souboru, které popisují zápasy ve hře CSGO. K potřebám této bakalářské práce budou použity pouze soubory *players.csv* a *results.csv*. Soubor *economy.csv* obsahuje informace o vývoji ekonomiky v daném zápase. Tato informace by byla velmi užitečná v případě, že se snažíme předpovídat výsledek zápasu, který aktuálně probíhá. Tato informace není dostupná před začátkem zápasu a tudíž je tento dataset pro tuto bakalářskou práci nepoužitelný. Druhý soubor, *picks.csv*, obsahuje postup výběr map mezi dvěma týmy v daném zápase. Tato práce je omezená na zápasy, které jsou typu Bo1 nebo Bo3. Bo1 znamená, že první tým co vyhraje mapu vyhrál i zápas. Bo3 značí, že se hrají celkem tři mapy Kdo první vyhraje dvě mapy, vyhrál celý zápas. Tento postup by se dal využít v případě simulace zápasu mezi dvěma týmy. Dalo by se předpovědět, vzhledem k historickým postupům, jaká mapa má jakou procentní šanci být vybrána či zabanována[2]. Jelikož je práce zaměřená pouze na předpověď zápasu s již známou mapou, není příležitost dataset využít.

4.1.1 soubor players.csv

Soubory *players.csv* obsahuje statistiky jednotlivých hráčů v daném zápase. Původní dataset obsahuje 101 sloupců a 379 680 záznamů. Názvy všech sloupců je možné vidět v přiložené tabulce A.1. Pro potřeby logistické regrese je nutné datový soubor transformovat do stavu, kdy se jeden řádek rovná statistikám jednoho hráče na jedné mapě. Transformovaný dataset má 10 sloupců a 643 620 řádků. Příklad záznamu v transformovaném datasetu je v přiložené tabulce A.2.

[1] https://www.kaggle.com/datasets/mateusdmachado/csgo-professional-matches

[2] tým ji v zápasu zakáže a nemůže si ji vybrat druhý tým

Transformovaný dataset má 10 sloupců, které unikátně identifikují statistiky každého hráče na určité mapě v jednom zápase. Interpretace je následující:

- **match_id** - identifikátor zápasu
- **player_id** - identifikátor hráče
- **team** - jméno týmu
- **map** - název hrané mapy
- **kills** - počet zabití hráče v zápase na dané mapě
- **assists** - počet asistencí hráče v zápase na dané mapě
- **deaths** - počet smrtí hráče v zápase na dané mapě
- **hs** - procento zabití, které lze označit jako headshot[3]
- **fkdiff** - rozdíl, kolikrát hráč zabil jako první nepřítele versus kolikrát byl jako první zabit
- **rating** - shrnutí mnoha statistik za zápas, díky kterým lze hráč ohodnotit[4]

4.1.2 soubor results.csv

Druhý datový soubor, který je pro analýzu použit, obsahuje výsledky daných zápasů. Dataset se původně skládá z 45 773 řádků a 19 sloupců. Dataset obsahuje na rozdíl od datového souboru *players.csv* chybné záznamy, které značí, že tým hrál sám proti sobě. Také jsou zde uvedené názvy týmu jako „?“, které nelze interpretovat. Tyto záznamy jsou proto odstraněny. Po transformacích vznikne tabulka o 8 sloupcích a 91 436 řádcích. Každý záznam identifikuje výsledek jednoho týmu v jednom zápasu na jedné mapě. Příklad je zobrazen v přiložené tabulce A.3. Jednotlivé sloupce lze interpretovat následovně:

- **date** - datum, kdy se hrál zápas
- **match_id** - identifikátor zápasu
- **team** - jméno týmu
- **map** - název hrané mapy
- **map_winner** - binární značení, zda tým vyhrál (1) či prohrál (0)
- **starting_ct** - binární značení, zda tým začal zápas na straně Counter-Teroristů (1) či Teroristů (0)
- **team_rank** - rank týmu v okamžik, kdy se zápas hrál[5]
- **run_mean_3_months** - klouzavý průměr týmu za poslední tři měsíce

[3] hráč zabil nepřítele střelou do hlavy
[4] https://www.hltv.org/news/20695/introducing-rating-20
[5] https://www.hltv.org/news/16061/introducing-csgo-team-ranking

4.1.3 Omezení datasetu

Dataset obsahuje záznamy o zápasech a statistikách od konce roku 2015 do začátku roku 2020. Jelikož máme mnoho záznamu, není problém se takovýchto záznamu zbavit a odstranit je. Dále jsou ze souboru smazány polo vyplněné statistiky, kde máme např. statistiku kills, ale ne statistiku deaths. Toto je z důvodu vývoje stránky[6], která data sbírá, a vývoje hry samotné. Ne vždy je možné data získat ať už kvůli jiným verzím či retrospektivní kompatibilitě.

Jak již bylo zmíněno, z datového souboru *results.csv* jsou odstraněné záznamy, kde tým hrál sám proti sobě, nebo je název týmu „?“. Z datového souboru *players.csv* jsou pak odstraněné záznamy o zápasech, kde nemáme údaje o všech deseti hráčích. Občas se může stát, že tým hrál pouze ve čtyřech, např. z důvodu nedochvilnosti. Při finálním spojení obou souborů se pak může stát, že jsou specifické záznamy pouze v jednom datovém souboru. Může se tedy stát, že existují záznamy o statistikách pro zápas, pro který nemáme finální výsledek. Opačně také může nastat situace, kdy existuje záznam o výsledku zápasu, ale nejsou záznamy o statistikách, nebo jsou statistiky nekompletní. V obou těchto případech se záznamů zbavíme, jelikož i po kombinaci existuje velké množství záznamů.

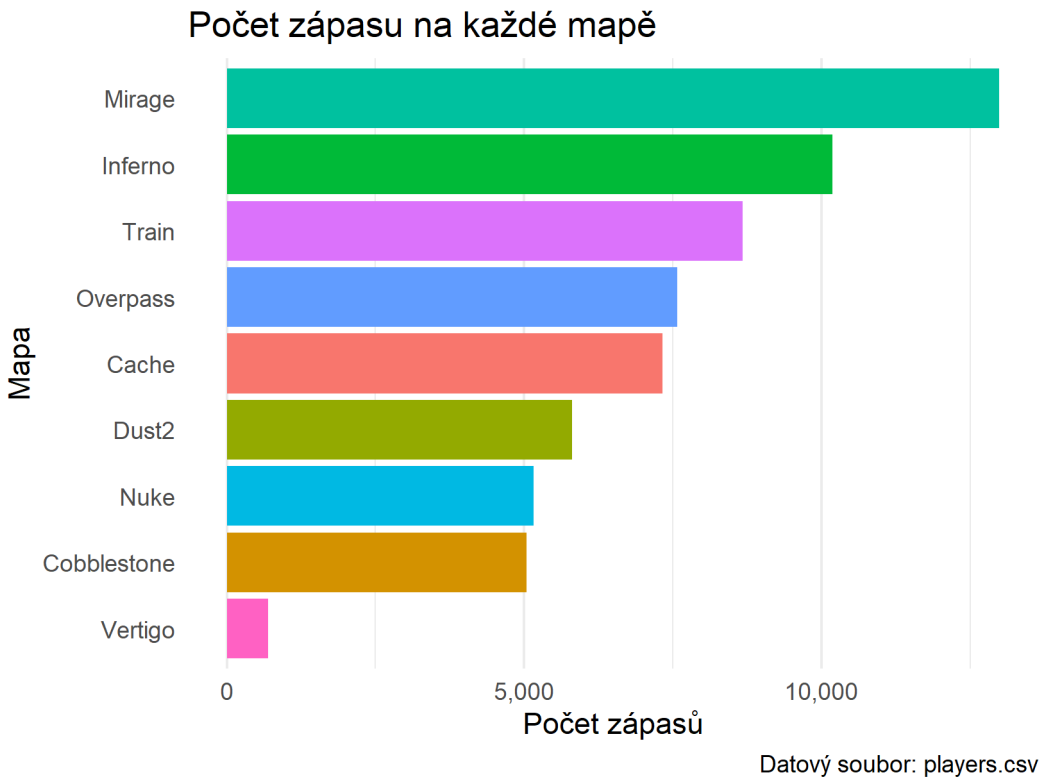
[6] https://www.hltv.org/

4.2 Explorační analýza dat

Explorační analýza dat je část, kdy se datasety vizualizují, hledají se různé vztahy, závislosti a zajímavé interpretace. K exploraci datasetu jsou použité nástroje představené v sekci 3.1.

4.2.1 Počet zápasů přes kategorie map

První zajímavý údaj může být, kolikrát se daná mapa hrála. Z výsledného grafu lze zjistit, zda je nějaká mapa výrazně preferovanější než jiná, nebo se mapy hrají v přibližně stejném poměru.

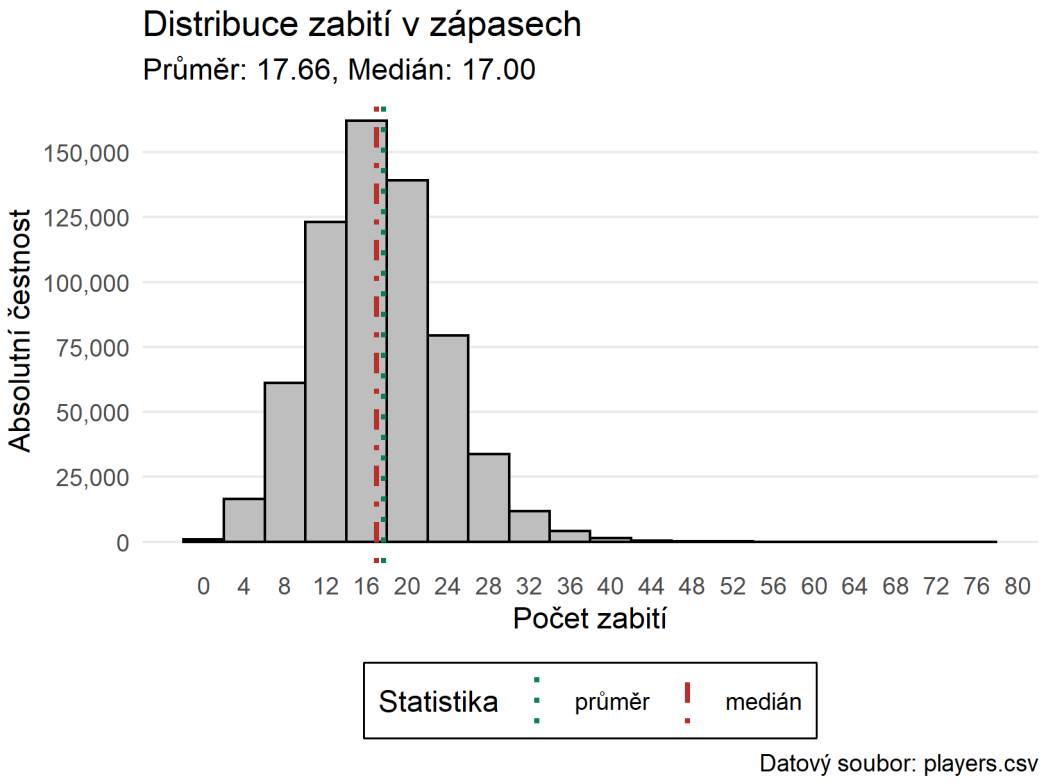


Obrázek 4.1: Počet zápasů na každé mapě

Z obrázku 4.1 lze odvodit, že mapy se ve stejném poměru nehrají. Nejhranější mapa je mapa s názvem Mirage. Ta je v map poolu[7] nejdéle bez žádné velké aktualizace vzhledu mapy, struktury mapy či rozložení. Naopak nejméně hraná mapa je Vertigo. Ta byla do map poolu přidán relativně nedávno[8] a mezi stálice se teprv dostává. Za svou existenci prošla mnoha aktualizacemi, které se implementují ze zpětné vazby hráčů. Jelikož je mapa nová, hráči stále objevují nové způsoby a strategie, jak mapu hrát. Druhý důvod proč je mapa méně preferovaná je částečně spojený s tím faktem, že je mapa nová. Pro nezkušené týmy je na mapě lehké porazit týmy, které jsou lepší. To je z toho důvodu, že na klasických mapách (jako např. mapa Mirage či Inferno), mají zkušené týmy velikou výhodu - hráli mapu již mnohokrát. Na nové mapě tato výhoda mizí a díky tomu mají nezkušené týmy určitou výhodu překvapení.

4.2.2 Histogram zabití

Statistika zabití na úrovni hráčů značí, kolik zabili v daném kole nepřátel. Obecně se dá říct, že čím víc hráč zabije nepřátel, tím více pomohl svému týmu. Samozřejmě že existuje mnoho faktorů, které malý počet zabitých nepřátel vysvětli. Může to být například role. Pokud je hráč velitel týmu nebo podpora týmu a má malý počet zabitých nepřátel, neznamená to nutně, že zhoršují šanci týmu vyhrát.



Obrázek 4.2: Počet zápasů na každé mapě

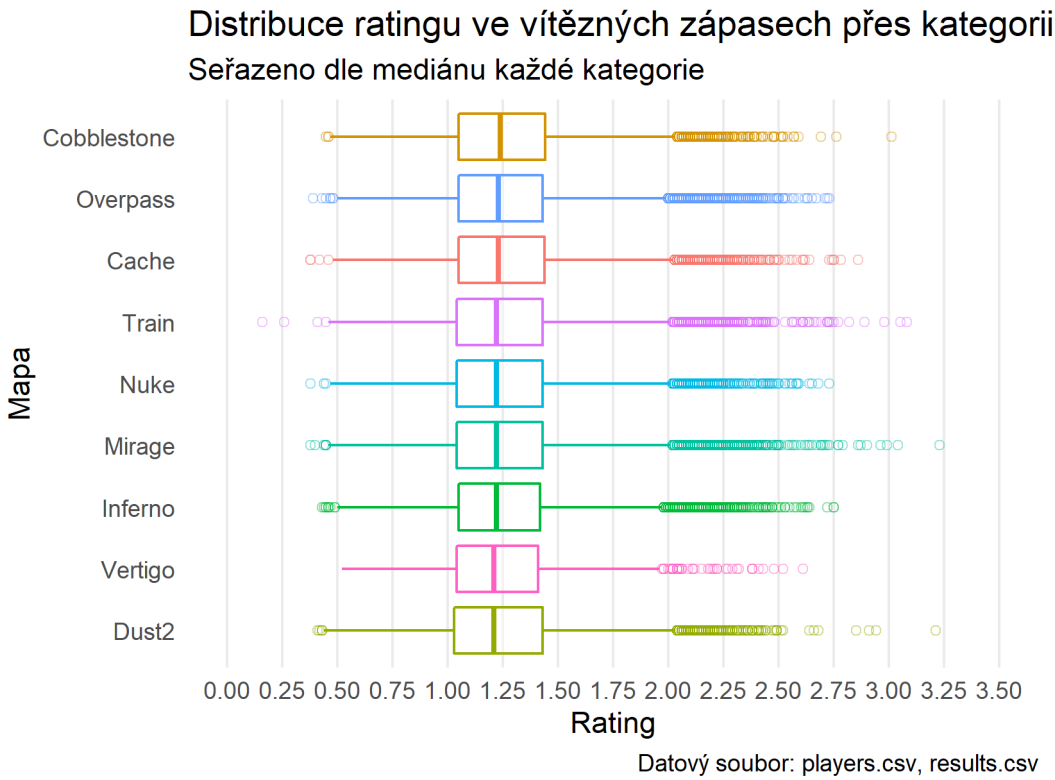
[7] množina map, které se hrají

[8] https://www.hltv.org/news/26418/vertigo-added-to-active-duty-map-pool-in-new-update

Obrázek histogramu počtu zabití naznačuje, že počet zabití v zápase má normální rozdělení s malým sklonem doprava. Průměrný počet zabití za zápas a mediánový počet zabití za zápas se tedy pohybuje kolem hodnoty 17 zabití za zápas.

4.2.3 Boxplot ratingu přes kategorie map

Pro statistiku rating se lze podívat na její distribuci přes různé kategorie mapy. Pokud je medián ratingu na nějaké mapě výrazně vyšší, lze usoudit, že je mapa strukturovaná pro individuální hráče. Takový typ mapy bude odměňovat spíše individuální výkon než souhru týmu a strategický plán. Obrázek níže zobrazuje boxploty hráčů přes mapy **pouze v případě, že zápas vyhráli.**

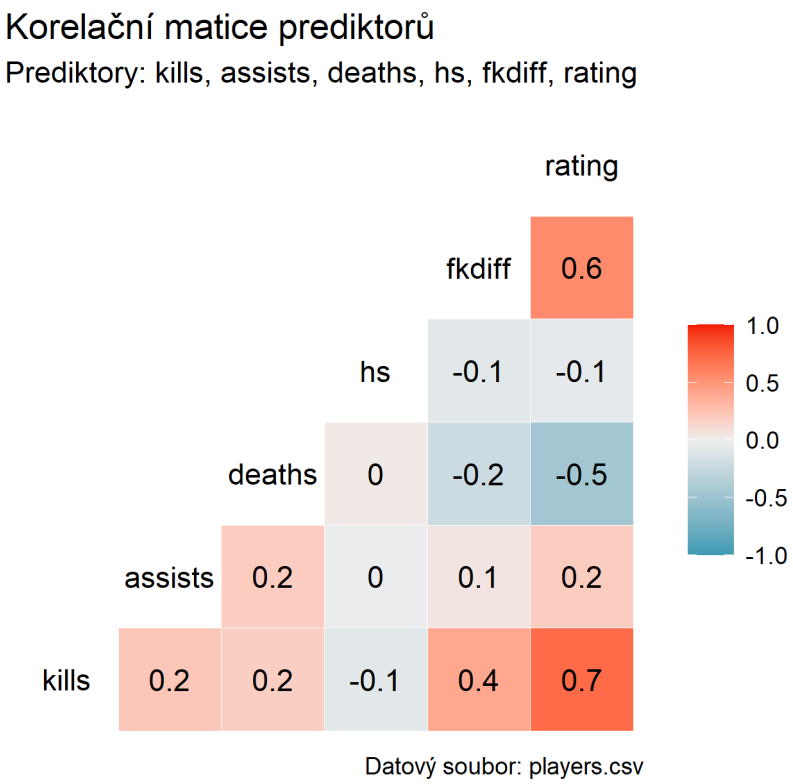


Obrázek 4.3: Počet zápasů na každé mapě

Z obrázku boxplotů 4.3 lze usoudit, že se mezikvartilové rozpětí a medián ratingu je na všech mapách podobný. Obrázek tedy naznačuje, že na žádné mapě není výrazná výhoda pro individuální rating hráče. Lze udělat závěr, že na výhru v zápase mají jiné statistiky větší vliv.

4.2.4 Korelace mezi statistikami

Pro logistickou regresi je důležité, aby prediktory nebyli korelované. Toto lze zjistit z korelační matice. Ta vyobrazí korelaci mezi individuálními prediktory.



Obrázek 4.4: Korelace

Korelační matice 4.4 ukazuje, že korelace mezi prediktorem rating a prediktory kills, fkdiff a deaths existuje relativně vysoká korelace. Jelikož je rating hodnocení hráče v daném zápase, dává smysl, že je pozitivně korelované s počtem zabití, statistikou fkdiff a záporně korelované s počtem úmrtí. Z tohoto důvodu je statistika z modelů odebrána.

4.3 Model pro hráče na určité mapě

První model, který je vytvořen, předpovídá pravděpodobnost výhry zápasu pouze na základě statistik jednoho hráče. Agregace dat je tedy na úrovni jednotlivých hráčů. Jsou vytvořeny dva modely se stejnými prediktory. První model je vytvořen pro mapu Dust 2 a druhý pro mapu Inferno. Prediktory, které jsou pro modely použité, jsou: kills, assists, deaths, hs, fkdiff, team_rank a run_mean_3_months.

4.3.1 model pro mapu Dust 2

První model je vytvořený pro mapu Dust 2. Mapa je zaměřená spíše na individuální výkon hráče než na strategie či týmové taktiky.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.2821	0.0788	28.98	0.0000
kills	0.1791	0.0028	64.57	0.0000
assists	0.2602	0.0056	46.29	0.0000
deaths	-0.3611	0.0042	-85.21	0.0000
hs	0.2154	0.0826	2.61	0.0091
fkdiff	-0.0100	0.0062	-1.63	0.1040
team_rank	-0.0024	0.0006	-3.96	0.0001
run_mean_3_months	-0.0008	0.0006	-1.38	0.1683

Tabulka 4.1: Výpis z funkce glm() v jazyku R

Pro model jsou signifikantní všechny prediktory s výjimkou fkdiff a run_mean_3_months. Zajímavé je, že na pravděpodobnost výhry nejvíce zvyšuje statistika assists má větší pozitivní vliv statistika assists než kills. Nejvíce pravděpodobnost výhry snižuje prediktor deaths. Nevýznamné prediktory fkdiff a run_mean_3_months naznačují, že pravděpodobnost výhry hráče není daná jeho týmem (run_mean_3_months) ani jeho výkonem v prvních vteřinách mapy (fkdiff)

4.3.2 model pro mapu Inferno

Druhý model se zaměřuje na mapu Inferno. Ta je oproti mapě Dust 2 více taktická. Je zde důležité, jak tým začne kolo a jak spolu tým hraje strategicky.

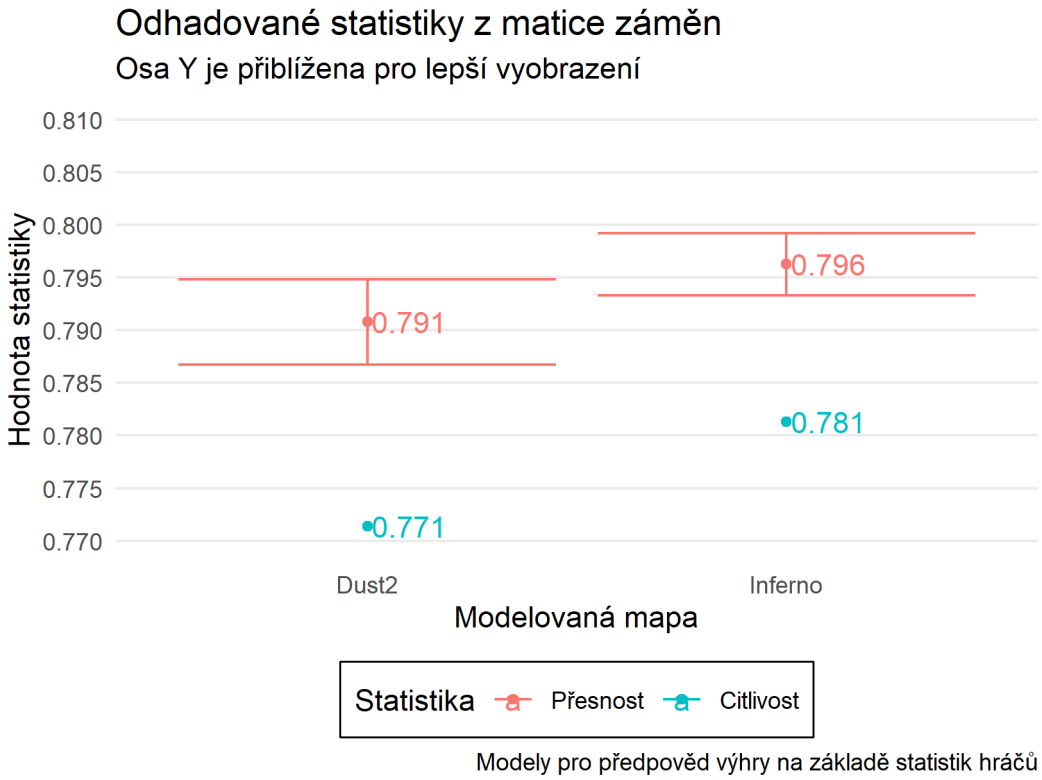
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.4010	0.0587	40.89	0.0000
kills	0.1840	0.0021	87.30	0.0000
assists	0.2807	0.0042	66.69	0.0000
deaths	-0.3728	0.0032	-117.01	0.0000
hs	-0.1507	0.0631	-2.39	0.0168
fkdiff	0.0194	0.0046	4.18	0.0000
team_rank	-0.0018	0.0004	-4.43	0.0000
run_mean_3_months	-0.0009	0.0004	-2.11	0.0352

Tabulka 4.2: Výpis z funkce glm() v jazyku R

Pro mapu jsou významné ($\alpha = 0.05$) všechny prediktory. Oproti modelu pro mapu Dust 2 jsou významné i statistiky fkdiff a run_mean_3_months. To značí, že je kladen větší důraz na týmovou souhru a sehranost hráčů. Čím větší mají klouzavý průměr za poslední tři měsíce, tím se očekává větší sehranost a zkušenosti. Důležitost týmové sehranosti též značí významnost prediktoru fkdiff. Čím je prediktor větší, tím více tým hraje spolu a získává první zabití nepřítele bez opětování smrti.

4.3.3 Vyhodnocení matice záměn

Díky vytvořeným maticím záměn je možné pro modely vytvořit statistiky Přesnost a Citlivost. Ty se dají mezi modely vzájemně porovnat.



Obrázek 4.5: Korelace

První porovnávaná statistika je přesnost. Ta je pro model na mapě Inferno větší o zhruba 0.05 jednotek. Zároveň je pro přesnost zobrazen intervalový odhad, který je pro mapu Inferno výrazně užší. Druhá sledovaná statistika je citlivost. Ta je pro mapu Inferno větší o zhruba jeden procentní bod. Z obrázku 4.5 lze usoudit, že model pro mapu Inferno je pro předpověď výsledku spolehlivější.

4.3.4 Vyhodnocení Waldova testu

Druhý test pro posouzení kvality modelu je použít takzvaný Waldův test. Ten je vypočten pro všechny prediktory a v případě rozdílu jsou dané statistiky zobrazeny na obrázku.

Obrázek 4.6: Korelace

Na obrázku 4.6 je vidět, že prediktory fkdiff a run_mean_3_months jsou významné pouze pro model mapy Inferno. Pro model mapu Dust 2 jsou oba ba dva prediktory nesignifikantní, Jak bylo zmíněno v sekci 4.3.2, toto může být z důvodu rozdílných hracích stylů na mapách.

4.4 Stejná struktura pro ostatní modely...

...

5. Závěr

... Uzavření bakalářské práce

5.1 Závěrečné vyhodnocení modelu

... Výsledné vyhodnocení modelu pomocí všech statistik

5.2 Interpretace modelu do reálného světa

... Přenesení modelu do reálného světa

5.3 Použití modelu v reálném světě

... Použití modelu v reálném světě

5.4 Místo pro budoucí vylepšení

...

Seznam použité literatury

Hebák, Petr (2015). *Statistické myšlení a nástroje analýzy dat*. 2. vyd. Informatorium. ISBN: 978-80-7333-118-4.

Kleinbaum, David (2010). *Logistic regression : a self-learning text*. 3. vyd. Springer. ISBN: 978-1-4939-3697-7.

Seznam elektronických zdrojů

csko (2021). „základní statistiky“. In: *csko.cz*. URL: https://stats.csko.cz/statsx/hlstats.php.

Gough, Christina (srp. 2021). „esports market revenue worldwide from 2019 to 2024“. In: *statista*. URL: https://www.statista.com/statistics/490522/global-esports-market-revenue/.

Henningson, Joakim (2020). „the history of counter-strike“. In: *redbull*. URL: https://www.redbull.com/se-en/history-of-counterstrike.

hltv.org (2015). „xenex vs. excel at esl uk premiership season 1“. In: *hltv.org*. URL: https://www.hltv.org/matches/2295340/xenex-vs-excel-esl-uk-premiership-season-1?__cf_chl_jschl_tk__=82S_Uc_dNU8PM7leUY1VKzNUZkp5_ArTb69qteh2wBI-1641459294-0-gaNycGzNChE.

Larch, Florian (led. 2019). „the history of the origin of esports“. In: *ispo*. URL: https://www.ispo.com/en/markets/history-origin-esports.

liquipedia (2021). „pgl major stockholm 2021“. In: *liquipedia*. URL: https://liquipedia.net/counterstrike/PGL/2021/Stockholm.

Professeur (2021). „esea increase prize pool and number of seasons for 2021; simplify path to pro league“. In: *hltv*. URL: https://www.hltv.org/news/30926/esea-increase-prize-pool-and-number-of-seasons-for-2021-simplify-path-to-pro-league.

Valve (2013). „counter-strike: the arms deal update“. In: URL: http://counter-strike.net/armsdeal.

Seznam obrázků

Seznam tabulek

Seznam použitých zkratek

CSGO Coutner-Strike: Global Offensive
BR Battle Royale
MOBA Multiplayer online battle arena
FPS First-person shooter
TGNS Twin Galaxies National Scoreboard

Část I

Přílohy

A. Datové soubory

A.1 Původní datový soubor players.csv

Tabulka A.1: Sloupce v původní struktuře datového souboru players.csv

date	m1_hs	m3_rating	m2_deaths_ct
player_name	m1_flash_assists	kills_ct	m2_kddiff_ct
team	m1_kast	deaths_ct	m2_adr_ct
opponent	m1_kddiff	kddiff_ct	m2_kast_ct
country	m1_adr	adr_ct	m2_rating_ct
player_id	m1_fkdiff	kast_ct	m2_kills_t
match_id	m1_rating	rating_ct	m2_deaths_t
event_id	m2_kills	kills_t	m2_kddiff_t
event_name	m2_assists	deaths_t	m2_adr_t
best_of	m2_deaths	kddiff_t	m2_kast_t
map_1	m2_hs	adr_t	m2_rating_t
map_2	m2_flash_assists	kast_t	m3_kills_ct
map_3	m2_kast	rating_t	m3_deaths_ct
kills	m2_kddiff	m1_kills_ct	m3_kddiff_ct
assists	m2_adr	m1_deaths_ct	m3_adr_ct
deaths	m2_fkdiff	m1_kddiff_ct	m3_kast_ct
hs	m2_rating	m1_adr_ct	m3_rating_ct
flash_assists	m3_kills	m1_kast_ct	m3_kills_t
kast	m3_assists	m1_rating_ct	m3_deaths_t
kddiff	m3_deaths	m1_kills_t	m3_kddiff_t
adr	m3_hs	m1_deaths_t	m3_adr_t
fkdiff	m3_flash_assists	m1_kddiff_t	m3_kast_t
rating	m3_kast	m1_adr_t	m3_rating_t
m1_kills	m3_kddiff	m1_kast_t	date
m1_assists	m3_adr	m1_rating_t	player_name
m1_deaths	m3_fkdiff	m2_kills_ct	team

A.2 Transformovaný datový soubor players.csv

Tabulka A.2: Záznam z transformovaného datového souboru players.csv

match_id	player_id	team	map	kills	assists	deaths	hs	fkdiff	rating
2339385	8738	Liquid	Overpass	15	3	12	0.6	3	1.32

A.3 Transformovaný datový soubor results.csv

Tabulka A.3: Příklad záznamu z transformovaného datového souboru results.csv

date	match_id	team	map	map_winner	starting_ct	team_rank	run_mean_3_months
2019-11-07	2337454	100 Thieves	Nuke	0	1	8	8