

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 Counter-Strike: Global Offensive (CSGO) a predikovat výhry vybraného teamu. Použitý datový soubor je z internetového portálu kaggle.com¹ 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	2
2	Představení esportu	3
2.1	Historie esportu	3
2.2	Zasazení do dnešní doby	3
2.3	Představení titulu Counter-Strike: Global Offensive	4
2.4	Propojení práce a titulu Counter-Strike: Global Offensive	6
2.4.1	Vysvětlení hry	6
3	Teoretická část	8
3.1	Vizualizace dat	8
3.1.1	Bodový graf	8
3.1.2	Histogram	9
3.1.3	Boxplot	12
3.2	Logistický regresní model	13
3.2.1	Jednoduchý logistický regresní model	13
3.2.2	Vícerozměrný logistický regresní model	13
3.2.3	Optimalizace parametrů	13
3.2.4	Vyhodnocení modelu	14
4	Praktická část	15
4.1	Metodika	15
4.2	Představení data setu	15
4.3	Příprava data setu	15
4.4	Omezení dat	15
4.5	Sestavení modelu	15
4.5.1	Interpretace	15
4.6	Vyhodnocení modelu	15
4.6.1	Statistika 1	15
5	Závěr	16
5.1	Závěrečné vyhodnocení modelu	16
5.2	Interpretace modelu do reálného světa	16
5.3	Použití modelu v reálném světě	16
5.4	Místo pro budoucí vylepšení	16
	Seznam použité literatury	17
	Seznam elektronických zdrojů	18
	Seznam obrázků	19

Seznam tabulek	20
Seznam použitých zkratek	21
I Přílohy	22

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 - střílečky, Multiplayer online battle arena (MOBA)¹, karetní hry, First-person shooter (FPS) či Battle Royale (BR).

Práce se zaměřuje primárně na esportový titul Counter-Strike: Global Offensive (CSGO). Je to jeden z nejdéle hraných esportových titulů, boří mnohé divácké rekordy² 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 módů³ a hráč může strávit mnoho hodin pouze objevováním komunitních serverů, hraním klasických zápasů či trénováním na offline mapách.

Finální cíl práce je vytvořit logistický model, který předpovídá výsledek zápasů. 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 - kdo vyhraje zápas, kdo vyhraje mapu, jaký hráč bude mít nejlepší statistiky, či zda si hráč koupí určitou zbraň.

Tento finální model je vyhodnocen různými klasifikacemi pro vyhodnocení kvality. Práce také popisuje grafické metody vizualizace dat a teorii k tvorbě a vyhodnocení logistických modelů. Také zde najdeme popis datového souboru a postup, jakým byli vybráni nejvýznamnější prediktory - ať už statisticky, či čistě ze znalosti esportu.

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

²<https://www.invglobal.com/articles/15619/csgo-major-breaks-viewership-records-overtakes-the-international>

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

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 první 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 LAN¹ party či organizovat BYOC² 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 třetím rokem, 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 považováni za experty v oblasti esportu. Tento průzkum byl následně podpořen růstem že tržní hodnoty esportu a mezi lety 2019 a 2020 vzrostla o téměř 50 % (Gough 2021).

¹Hráči hrají v jedné místnosti na lokální počítačové síti.

²z 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 investovat více peněz (Professeur 2021) (liquipedia 2021). S větším počtem diváku roste i marketingový potenciál, investiční příležitost a kariérní růst.

V dnešní době má esport mnoho titulů, proto jsou představeny jen ty nejvýznamnější. Největší esport rivalita je mezi herním titul League of Legends a Dota 2. Oba tituly jsou žánru MOBA, díky čemuž mají podobnou, avšak velmi rozdílnou fanouškovskou základnu. Historie mezi tituly je velmi složitá a mimo rozsah této práce.

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 trochu rozdílné oproti žánru FPS. Hlavním titulem této kategorie je hra Fortnite, která žánru dominuje. Stal se z ní jak esport 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 dlouhou historii. Ne vždy se to ovšem jmenovalo stejně. Úplně první iterace hry se jmenovala čistě Counter-Strike a byl to pouze mód⁵ do hry Half-Life. Half-Life byl vyvinutí společností Valve, tehdy primárně společností zaměřenou na vyvíjení 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.

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

⁵upravení 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 jeho jednoduchosti. Hra se dá velmi dobře popsat pořekadlem „Lehké hrát, těžké vypilovat“. Hra má mechaniky⁶, 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 můžou, hru velmi lehce a rychle začít hrát - tento formát se totiž za posledních 20 let 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 kotelná 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 major⁷ 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 - CSGO

CSGO se snažilo sjednotit oba tábory - 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).

⁶herní prvky či unikátní vlastnosti

⁷turnaj pořádaný přímo Valve, který má největší prestiž

2.4 Propojení práce a titulu Counter-Strike: Global Offensive

2.4.1 Vysvětlení hry

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 vybuchnout. Naopak cíl Teroristů (T) je právě bombu položit a „bomboviště“ nechat vybuchnout. 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 - 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 finance, 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 pomocníka - 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 - 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 - odstřelovací puškou AWP) 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 vyhraje dvě, vyhrál zápas. Turnaje se pak odehrávají v tradičních formátech, jako je pavouk. Ten se charakterizuje tím, že vypadá jak pavučina, jde zleva 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.

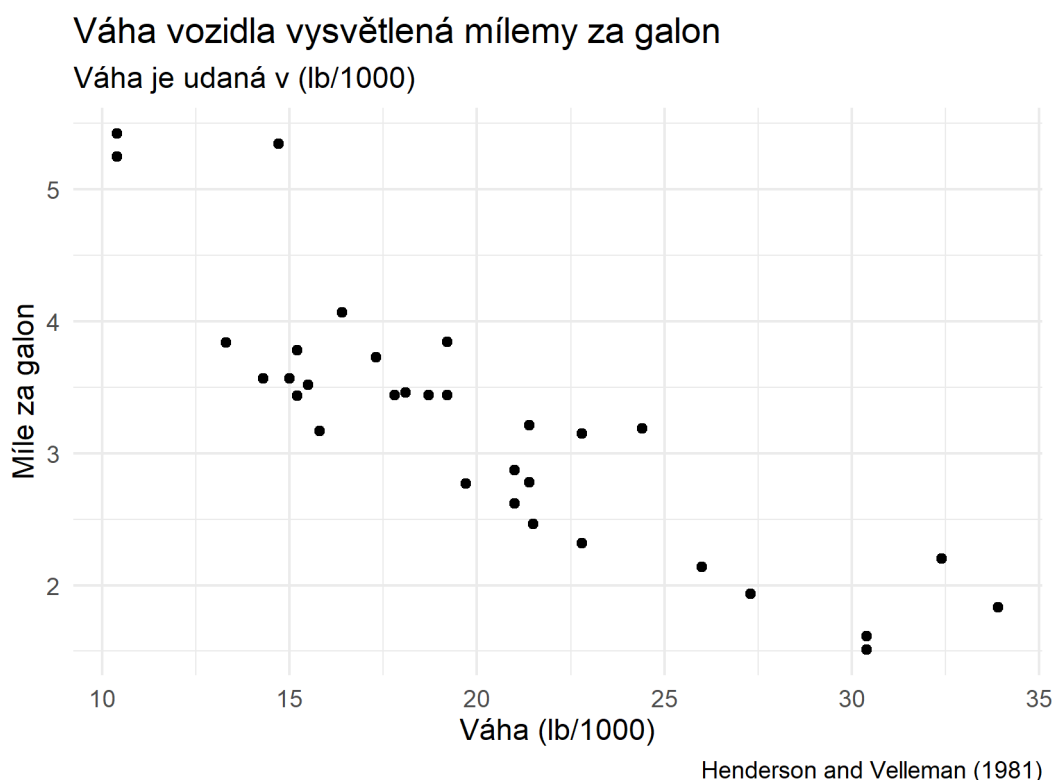
3. Teoretická část

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. U regresního modelu je následně dán důraz na teorii, jelikož model je následně prakticky použit v celé třetí části. Důležité je si uvědomit, že nástrojů pro vizualizaci dat je mnoho a grafů či souhrnných statistik je nespočet. V této práci se ovšem věnujeme vizualizacím, které jsou pro regresní model vhodné. Testovací citace: (Hebák 2015), (Kleinbaum 2010)

3.1 Vizualizace dat

3.1.1 Bodový graf

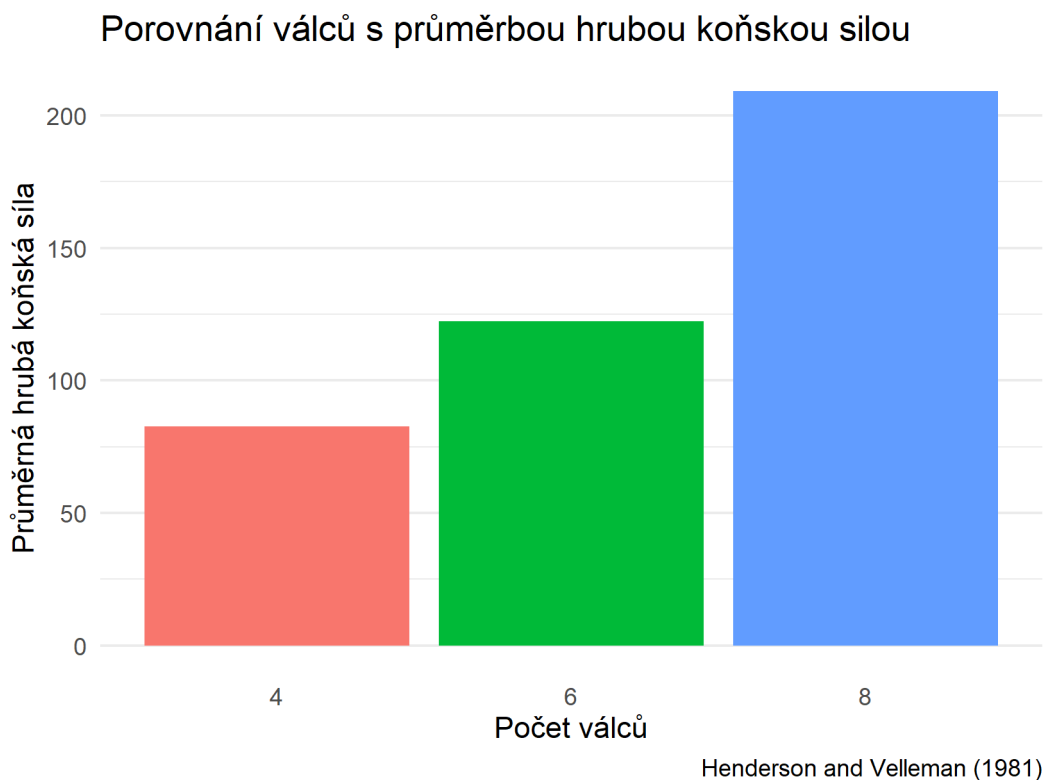
Bodový graf slouží pro zobrazení vztahu dvou numerických proměnných. Z pravidla se vysvětlovaná proměnná dává na osu Y, zatímco proměnná vysvětlující se nachází na ose X. Graf 3.1 zobrazuje negativní vztah mezi váhou vozidla a mílemi ujetými za galon.



Obrázek 3.1: Bodový graf váhy a míly za galon z data setu mtcars

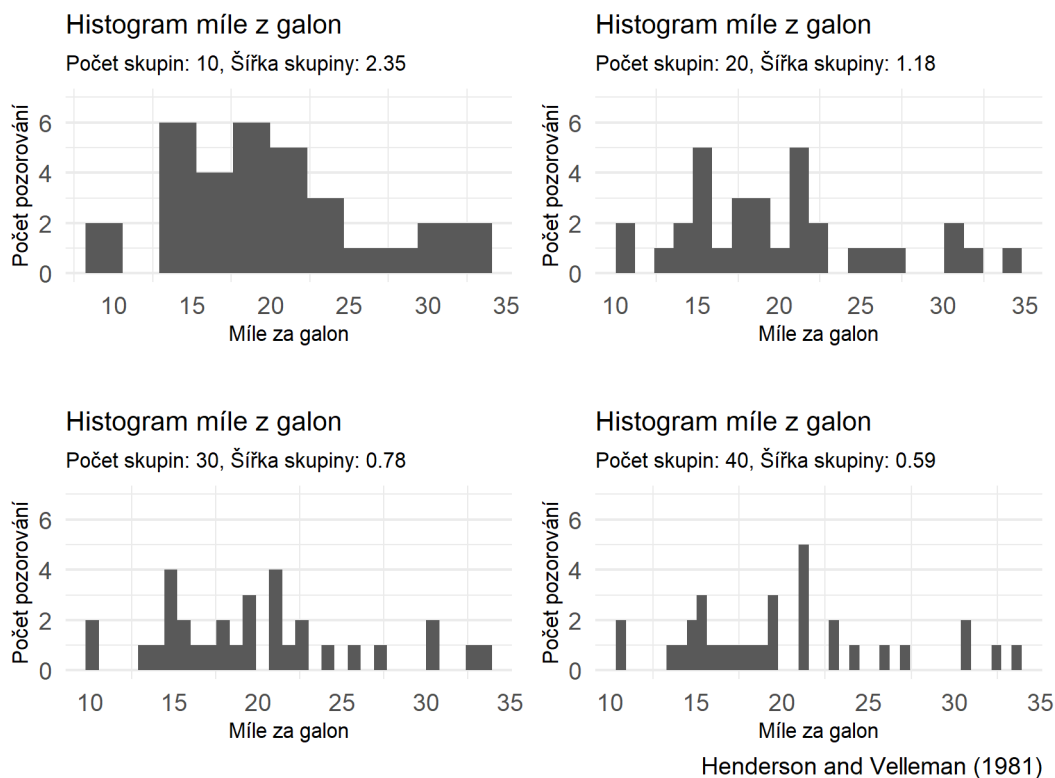
3.1.2 Histogram

Histogram je speciální typ sloupcového grafu. Je specifický tím, že zobrazuje počet jednotlivých pozorování ve skupině. Skupiny je nutné vytvořit v případě, že je zobrazovaná proměnná spojitá. Zároveň histogram oproti klasickému sloupcovému grafu, který zobrazuje kategorie, nemá mezi sloupci žádné mezery. Na grafu 3.2 je příklad sloupcového grafu. Na ose X se nachází kategorická vysvětlovací proměnná. Vysvětlovaná proměnná Y je průměrná hrubá koňská síla.



Obrázek 3.2: Sloupcový graf z data setu mtcars

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 moc 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 grafu 3.3.



Obrázek 3.3: Porovnání histogramů s různým počtem skupin

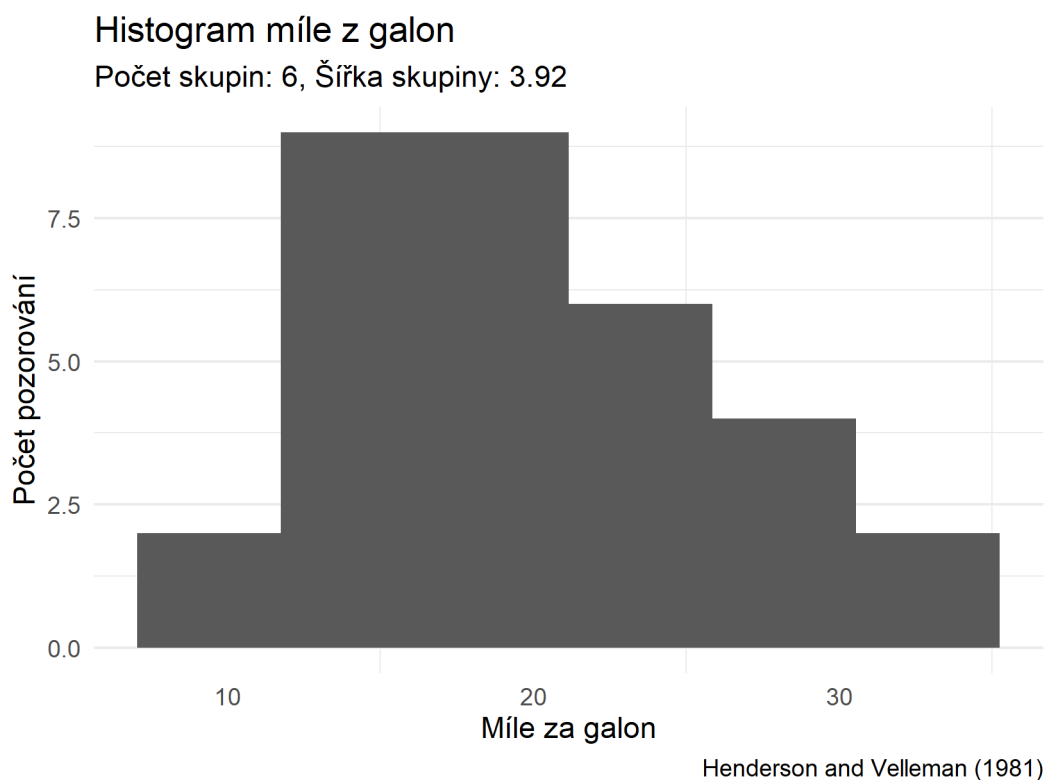
Pro vhodný počet skupin existuje mnoho způsobů. Nejznámější je takzvané Sturgesovo pravidlo (3.1), které se spočítá následujícím vztahem:

$$k \doteq 1 + 3,22 * \log_{10}(n) \quad (3.1)$$

kde k je výsledný zaokrouhlený počet skupin 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:

$$w = \left\lceil \frac{\max(x) - \min(x)}{k} \right\rceil \quad (3.2)$$

kde x je zobrazovaná proměnná, k je počet skupin a w je výsledná šířka intervalu, která je zaokrouhlena nahoru. To je z toho důvodu, aby se každé pozorování promítlo právě do jedné kategorie. Pokud na stejný dataset, jako na grafu 3.4, použije sturgesovo (rovnice 3.1) pravidlo, graf vypadá následovně:



Obrázek 3.4: Histogram s počtem skupin dle Sturgesova pravidla

Histogram lze samozřejmě rozepsat. Lze vytvořit tzn. tabulku četností (3.1), která mimo jiné obsahuje spodní hranici intervalu, horní hranici intervalu a počet pozorování. Důležité je, aby přechody mezi intervaly byli jasné.

Spodní hranice intervalu	Horní hranice intervalu	Počet pozorování
7.05	11.75	2
11.76	16.45	9
16.46	21.15	9
21.16	25.85	6
25.86	30.55	4
30.56	35.25	2

Tabulka 3.1: Tabulka četností

3.1.3 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 percentily. 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 specifitější 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 tabulka 3.2

$Q_0(Q_0)$	$Q_{0.25}(Q_1)$	$Q_{0.50}$	$Q_{0.75}(Q_3)$	$Q_{1.00}$
1.513	2.58125	3.325	3.61	5.424

Tabulka 3.2: Five-number summary tabulka hmotnosti vozidla (lb/1000)

Q_0 a $Q_{1.00}$ označují minimum a maximum číselné řady. Kvartily Q_1 a Q_3 jsou čísla, která rozděluji časovou řadu 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.

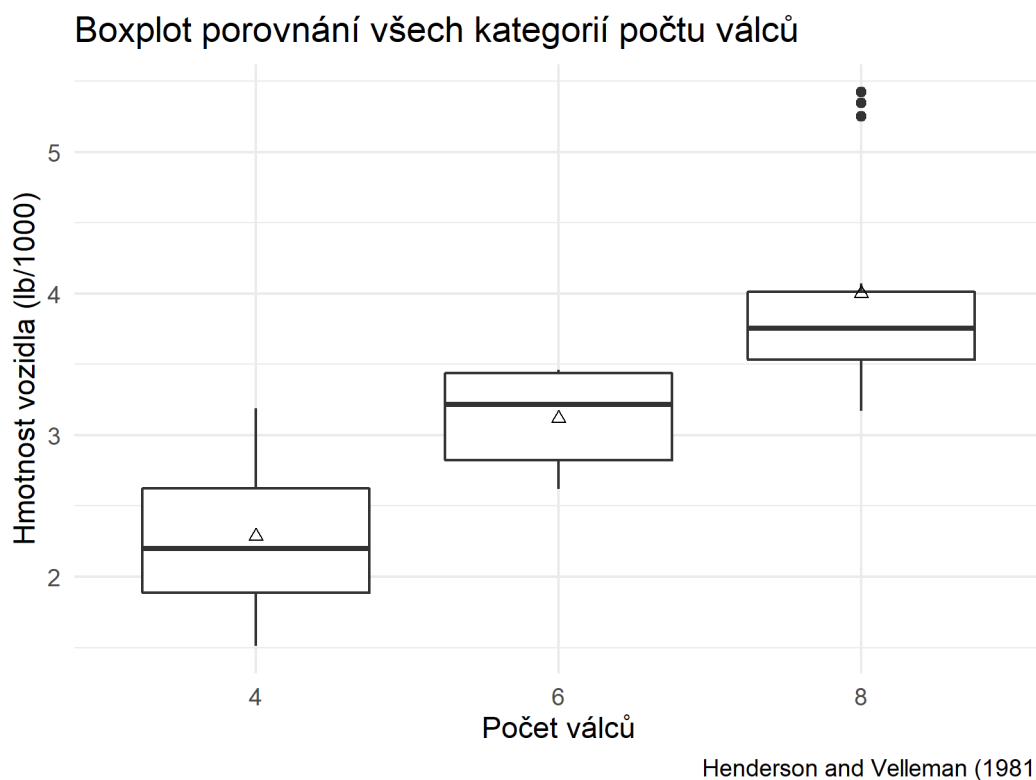
Boxplot

Boxplot je grafické zobrazení a rozšíření Five-number summary tabulky. Krom grafické zobrazení pěti percentilů obvykle ukazuje statistiky jako průměr a extrémní hodnoty. V boxplotu se také nachází obdélník, který ukazuje interkvartilní rozsah (IQR), tedy prostředních 50% dat. V obdélníku se také nachází černá čára, která značí medián. Z prostředního obdélníku vedou oběma směry čáry, které značí teoretické minimum a maximum (ne hodnoty Q_0 a Q_1). Obě dve hranice se počítají následovně

$$min = Q_1 - 1.5 * IQR \quad (3.3)$$

$$max = Q_3 + 1.5 * IQR \quad (3.4)$$

Hodnoty, které spadají do intervalu $< Q_1 - 1.5 * IQR; Q_1 >$ a $< Q_3, Q_3 + 1.5 * IQR >$ se nazývají jako odlehlé. Hodnoty které leží mimo tento vztah, tedy hodnoty menší než $Q_1 - 1.5 * IQR$ či větší než $Q_3 + 1.5 * IQR$, mají speciální název a označují se jako hodnoty extrémní. V boxplotu jsou z pravidla znázorněny nějakým speciálním znakem, např. kolečkem. Díky grafickému zobrazení lze lehce porovnávat distribuci jedné vysvětlované proměnné vůči několika kategoriím.



Obrázek 3.5: Boxplot váhy auta pro různé počty válců

Trojúhelník v grafu 3.5 značí průměr pro danou kategorii. Černé tečky v kategorii osmi válců značí extrémní hodnoty, t. j. hodnoty větší než $< Q_3, Q_3 + 1.5 * IQR >$ jsou extrémní.

3.2 Logistický regresní model

... Dopsat

3.2.1 Jednoduchý logistický regresní model

... Dopsat

3.2.2 Vícerozměrný logistický regresní model

... Dopsat

3.2.3 Optimalizace parametrů

... Dopsat

Metoda nejmenších čtverců

... Dopsat

Metoda největší pravděpodobnosti

... Dopsat

3.2.4 Vyhodnocení modelu

... Dopsat

4. Praktická část

4.1 Metodika

... Proč jsem si vybral právě vybrané metody a statistiky

4.2 Představení data setu

... Popis dat a struktury

4.3 Příprava data setu

... Popis PROČ a jak byl data set transformován, žádný kód, pouze slovně

4.4 Omezení dat

... Restrikce data, výběr jednoho konkrétního teamu

4.5 Sestavení modelu

... Slovní popis modelu a jeho proměnných, rovnice

4.5.1 Interpretace

... Matematická interpretace modelu a výstupu z Rka (koeficienty, odchylky, významnost...)

4.6 Vyhodnocení modelu

4.6.1 Statistika 1

... Různé statistiky, které ještě musím vybrat

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_dNU8PM71eUY1VKzNUZkp5_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

3.1	Bodový graf váhy a míly za galon z data setu mtcars	8
3.2	Sloupcový graf z data setu mtcars	9
3.3	Porovnání histogramů s různým počtem skupin	10
3.4	Histogram s počtem skupin dle Sturgesova pravidla	11
3.5	Boxplot váhy auta pro různé počty válců	13

Seznam tabulek

3.1	Tabulka četnostní	11
3.2	Five-number summary tabulka hmotnosti vozidla (lb/1000)	12

Seznam použitých zkratek

CSGO Counter-Strike: Global Offensive

BR Battle Royale

MOBA Multiplayer online battle arena

FPS First-person shooter

TGNS Twin Galaxies National Scoreboard

Část I

Přílohy