

UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Analýza návštěvnosti na českých hokejových
stadionech



Katedra matematické analýzy a aplikací matematiky

Vedoucí bakalářské práce: **Mgr. Ondřej Vencálek Ph.D.**

Vypracoval: **Martin Vavruša**

Studijní program: B1103 Aplikovaná matematika

Studijní obor Aplikovaná statistika

Forma studia: prezenční

Rok odevzdání: 2017

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Martin Vavruša

Název práce: Analýza návštěvnosti na českých hokejových stadionech

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Ondřej Vencálek Ph.D.

Rok obhajoby práce: 2017

Abstrakt: Mezi nejpopulárnější sporty v České republice patří bezesporu hokej, jehož nejvyšší soutěž se odehrává na čtrnácti různých zimních stadionech. Ve kterém městě jsou v posledních letech nejlépe zaplněné tribuny a co všechno ovlivní návštěvnost v příštím zápase, to je tématem této bakalářské práce.

Klíčová slova: Hokej, extraliga, návštěvnost

Počet stran: 41

Počet příloh: 0

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Martin Vavruša

Title: Analysis of visit rate at ice hockey stadiums in the Czech Republic

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: Mgr. Ondřej Vencálek Ph.D.

The year of presentation: 2017

Abstract: One of the most popular sports in the Czech Republic is ice hockey. The top league is called Extraliga and it is played at 14 stadiums around the country. This thesis is analysing which stadium's ice hockey matches have seen the highest attendance in the recent years and what different factors can influence the number of spectators in future fixtures.

Key words: Ice hockey, Extraliga, attendance

Number of pages: 41

Number of appendices: 0

Language: Czech

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně pod vedením pana Mgr. Ondřeje Vencálka Ph.D. a všechny použité zdroje jsem uvedl v seznamu literatury.

V Olomouci dne

.....

podpis

Obsah

Úvod	9
1 Česká hokejová extraliga	10
1.1 Základní část	10
1.1.1 Bodování	10
1.2 Play-off	11
1.3 Play-out	11
1.4 Analyzovaná data	12
2 Analýza	15
2.1 Vývoj návštěvnosti v čase	15
2.2 Faktory ovlivňující návštěvnost	20
2.2.1 Úspěšnost týmu	20
2.2.2 Město a velikost stadionu	23
2.2.3 Den v týdnu	25
2.2.4 Televizní utkání	26
2.2.5 Play-off, play-out	28
2.2.6 Atraktivnost soupeře, derby	32
Závěr	38
Literatura	40

Seznam obrázků

2.1	Histogram návštěvnosti – vlevo absolutní počet diváků, vpravo procentuální zaplnění stadionu	15
2.2	Průměrné zaplnění stadionů pro jednotlivá kola spolu s klouzavým průměrem	16
2.3	Krabicové grafy prokazující symetrii dat pro test vlivu příchodu Olomouce a Mladé Boleslavi na návštěvnost	18
2.4	Průběh jednotlivých sezon – zelenou čarou jsou vykresleny klouzavé průměry se zkracující se délkou okna na začátku a na konci, červenou odhady počátečních a koncových bodů pomocí lineárního trendu	19
2.5	Procentuální návštěvnost pro jednotlivé týmy	23
2.6	Zaplnění stadionu podle dne v týdnu	25
2.7	Graf návštěvnosti podle toho, zda utkání přenášela Česká televize – vlevo všechny zápasy sezony 2015/2016, vpravo zápasy vybraných dvojic týmů.	27
2.8	Krabicový graf zaplnění stadionu pro jednotlivé části soutěže	28
2.9	Krabicový graf zaplnění stadionu pro jednotlivé části soutěže u vybraných týmů	29
2.10	Zaplnění stadionu v závislosti na části play-off a odhadnutý lineární trend	30
2.11	Krabicový graf návštěvnosti – nahoře rozdělená podle stavu série, dole podle náskoku či ztráty domácího týmu	31
2.12	Krabicový graf návštěvnosti – vlevo skutečná derby, vpravo ostatní zápasy zahrnuté podmínkou vzdálenosti	37

Seznam tabulek

1.1	Kapacita stadionů extraligových klubů	13
2.1	Spearmanův korelační koeficient závislosti zaplnění stadionu na jednotlivých projevech úspěšnosti týmu	21
2.2	Spearmanův parciální korelační koeficient závislosti zaplnění stadionu na jednotlivých projevech úspěšnosti týmu	21
2.3	Spearmanův korelační koeficient vyjadřující míru závislosti mezi umístěním týmu a zaplněním jeho stadionu	22
2.4	Průměrná návštěva a zaplnění na jednotlivých stadionech	24
2.5	Dny, pro něž se významně liší zaplnění stadionů	26
2.6	Skupiny dní lišící se postavením vůči volnu, pro něž se významně liší zaplnění stadionů	26
2.7	Kontingenční tabulka vyprodání stadionu pro jednotlivé části play-off	30
2.8	Procentuální zaplnění stadionů pro jednotlivé soupeře	33
2.9	Spearmanův korelační koeficient závislosti zaplnění stadionu na době dojezdu	34
2.10	Doby dojezdu, pro něž se zaplnění stadionu významně liší	35
2.11	Zaplnění jednotlivých stadionů a nejvyšší počet odehraných zápasů se soupeři v jednom intervalu doby dojezdu	36

Poděkování

Děkuji Mgr. Ondřeji Vencálkovi Ph.D. za všechnen čas, rady a připomínky, kterými přispěl k tvorbě této práce.

Úvod

Tématem bakalářské práce je analýza návštěvnosti v české extralize za použití statistických metod a softwaru R. Záměrem je odhalit příčiny změn v návštěvnosti na zimních stadionech z dat sesbíraných za posledních sedm let.

Na začátku čtenáře stručně seznámím s českou hokejovou extraligou, konkrétně se systémem bodování a postupu do play-off, poté představím analyzovaná data a následně statistické testy, s nimiž jsem pracoval, a závěry, jež jsem vyvodil na základě jejich výsledků.

Kapitola 1

Česká hokejová extraliga

V České Republice se hraje řada hokejových soutěží, žádná však není tak známá a nemá tak vysokou návštěvnost jako Tipsport extraliga, jíž se účastní čtrnáct nejlepších českých týmů.

1.1 Základní část

Každý tým sehraje s třinácti ostatními týmy čtyři utkání (dvě na vlastním stadioně, dvě na soupeřově). Ta trvají šedesát minut, pokud je však skóre po této době nerozhodné, následuje pětiminutové prodloužení v němž vítězí tým, který jako první vstřelí branku, a pokud ani poté se nerozhodne, končí utkání samostatnými nájezdy. Po dvaapadeásti odehraných kolech postoupí deset nejlepších týmů do vyřazovacích bojů, (tzv. play-off), zatímco čtyři poslední týmy sehrají zápasy o udržení (tzv. play-out).

1.1.1 Bodování

O pozici týmu rozhoduje počet bodů, které nasbíral během již odehraných zápasů. Body za jednotlivá utkání se rozdělují následovně:

- Za vítězství v základní hrací době tým získá 3 body.
- Za remízu v základní hrací době a následné vítězství v prodloužení či samostatných nájezdech tým získá 2 body.

- Za remízu v základní hrací době při následné porážce v prodloužení či samostatných nájezdech tým získá 1 bod.
- Za prohru v základní hrací době tým nezíská žádné body.

1.2 Play-off

Vyřazovací boje začínají předkolem, kde se utká sedmý tým s desátým a osmý tým s devátým. Ta mužstva, jež vyhrají tři zápasy z nejvýše pěti, pak postoupí do čtvrtfinále, kterého už se účastní i zbylá šestice. Podobným systémem (tedy první proti poslednímu, druhý proti předposlednímu, ...) se utkají v sérii na čtyři vítězná utkání, vítězové se pak utkají podle stejného pravidla v semifinále a z něj vyjdou dva týmy, které se střetnou ve finále o Masarykův pohár a titul mistra České republiky. Zápasy se od základní části liší jen v prodloužení, které trvá v prvních čtyřech zápasech každého kola deset minut, v případném pátém až sedmém se prodlužuje dokud jedna strana nevstřelí branku.

1.3 Play-out

Nejhorší týmy Tipsport extraligy se spolu utkají v turnaji na šest (do roku 2012 dvanáct) zápasů, přičemž každý tým se utká se všemi ostatními jednou na vlastním ledě a jednou na ledě soupeře. Body z tohoto turnaje se přičítají k bodům ze sezony a dvojice týmů, která po tomto turnaji skončí na posledních místech, musí uhájit extraligovou příslušnost v baráži, kde změří síly s nejlepšími dvěma celky nižší soutěže.

1.4 Analyzovaná data

K testování hypotéz máme k dispozici data za posledních sedm kompletně odehraných sezon, konkrétně z ročníků 2009/2010 až 2015/2016. Kvůli rozdílným kapacitám jednotlivých zimních stadionů jsou u většiny testů využita procentuální zaplnění haly namísto počtu návštěvníků. Následuje ukázka analyzovaných dat:

rok	den	kolo	dom	hos	dgol	hgol	pozn	body	divaci	procent	umist	TV
2010	st	1	BUD	LIT	4	1		3	4800	74,75	1	NA
2010	ne	3	BUD	KLA	4	5	n	1	4429	68,98	2	NA
2010	pa	5	BUD	SLA	6	1		3	5003	77,92	6	NA
2010	ne	6	BUD	KOM	5	2		3	4728	73,63	4	NA
2010	pa	7	BUD	VIT	4	3		3	4935	76,86	3	NA
2010	ut	9	BUD	KVA	4	3		3	4478	69,74	2	NA
2010	ne	11	BUD	SPA	2	3	n	1	5131	79,91	3	NA
2010	ne	14	BUD	TRI	4	2		3	4409	68,67	7	NA
2010	ne	16	BUD	PLZ	1	3		0	5208	81,11	7	NA
2010	ne	19	BUD	PCE	1	4		0	5645	87,91	7	NA

Data zahrnují rok, jímž daná sezona končila, den v týdnu, v němž se utkání odehrálo, pořadí zápasu v sezoně, zkratky názvů obou týmů, výsledek, přičemž poznámka popisuje, jestli utkání skončilo v základní hrací době, v prodloužení (p) nebo na samostatné nájezdy (n), počet bodů, které získal zkoumaný tým, počet diváků na utkání, procentuální zaplnění stadionu, umístění týmu v tabulce a informaci, zda utkání přenášela Česká televize. V posledním sloupci jsou však data jen pro poslední rok. Pro úplnost doplníme ještě tabulku se změnami kapacity stadionů:

tým	rok						
	2010	2011	2012	2013	2014	2015	2016
BUD	6 421	6 421	6 421	6 421	NA	NA	NA
HKR	NA	NA	NA	NA	7 700	6 890	6 890
CHO	NA	NA	NA	5 250	5 250	NA	5 250
KLA	8 100	8 100	8 100	8 000	8 100	NA	NA
KOM	7 200	7 200	7 200	7 200	7 200	7 200	7 700
KVA	6 000	6 000	6 000	6 000	6 000	6 000	6 000
LIB	7 500	7 500	7 500	7 500	7 500	7 500	7 500
LIT	7 000	7 000	7 000	7 000	6 211	6 211	6 211
MBL	4 200	4 200	4 200	NA	NA	4 200	4 200
OLO	NA	NA	NA	NA	NA	5 500	5 500
PCE	10 194	10 194	10 194	10 194	10 194	10 194	10 194
PLZ	8 420	8 236	8 236	8 236	8 236	8 236	8 236
SLA	17 360	17 360	17 360	17 360	17 360	17 360	NA
SPA	13 238	13 238	13 238	13 238	13 238	13 238	17 360
TRI	5 200	5 200	5 200	5 200	5 200	5 200	5 200
VIT	10 004	10 004	10 004	10 004	10 004	10 004	10 004
ZLN	7 000	7 000	7 000	7 000	7 000	7 000	7 000

Tabulka 1.1: Kapacita stadionů extraligových klubů

System play-out je téměř totožný se základní částí sezony a data k němu jsou velmi podobná, play-off je však v mnohém specifické a v datech je proto zahrnuto, v které části se play-off nachází (4 – předkolo, ..., 1 – finále), kolikátý zápas série se hraje a jaký je její stav. Naopak ztratila význam poznámka o tom, zda tým vyhrál v základní hrací době či v prodloužení nebo na samostatné nájezdy. Také chybí informace o televizních utkáních, poněvadž v play-off Česká televize přenáší většinu zápasů. Ukázka datové množiny tedy vypadá takto:

rok	den	cast	zap	dom	hos	dgol	hgol	divaci	kapacita	stav	procent
2010	po	4	1	SLA	LIT	3	1	2464	17360	0x0	14,20
2010	ut	4	2	SLA	LIT	1	5	4872	17360	1x0	28,06
2010	ct	4	3	LIT	SLA	5	3	6001	7000	1x1	85,73
2010	pa	4	4	LIT	SLA	3	5	5883	7000	2x1	84,04
2010	ne	4	5	SLA	LIT	5	3	6128	17360	2x2	35,30
2010	po	4	1	LIB	BUD	2	1	3899	7500	0x0	51,99

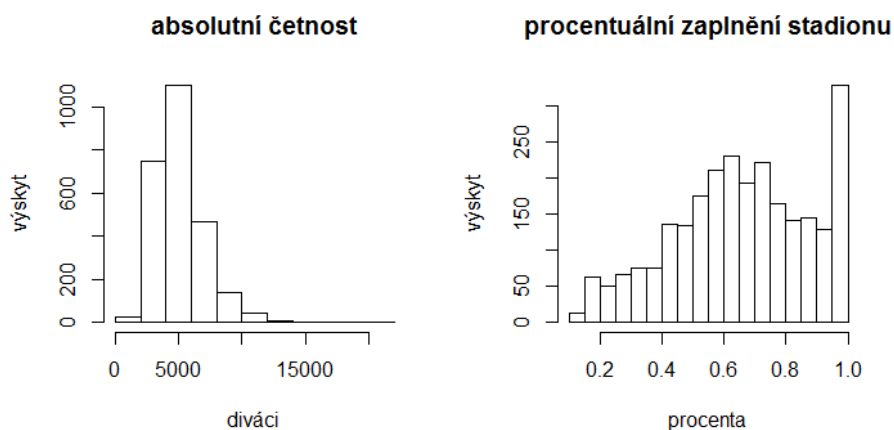
Mezi jednotlivými částmi Tipsport extraligy se dá očekávat poměrně výrazné rozdíly, proto pro většinu testů budeme uvažovat jen data ze základní části sezony, data z play-off a play-out využijeme jen v podkapitole věnované speciálně těmto pasážím.

Kapitola 2

Analýza

2.1 Vývoj návštěvnosti v čase

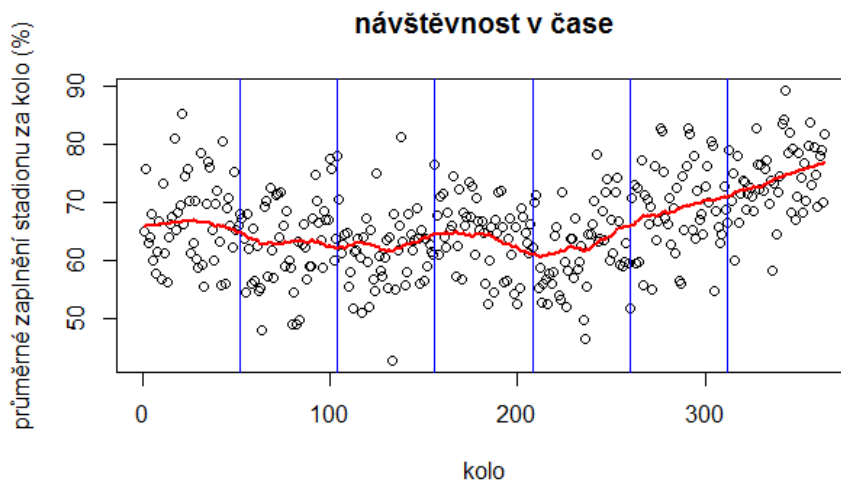
Nejprve si data zobrazíme graficky, abychom o nich získali základní přehled.



Obrázek 2.1: Histogram návštěvnosti – vlevo absolutní počet diváků, vpravo procentuální zaplnění stadiónu

Z dat je už na první pohled zřejmé, že jak absolutní četnosti, tak procentuální zaplnění zimních stadiónů nemají normální rozdělení a tedy k práci s nimi musíme využít neparametrické metody. Shapirův test normality tuto domněnku potvrzuje, pro absolutní počet diváků i procentuální zaplnění hal vychází p-hodnota menší než 0,0001.

Pro další práci by také bylo lepší mít přehled o vývoji návštěvnosti v posledních letech, proto si zobrazíme graf průměrného zaplnění stadionu v závislosti na čase, tedy spočítáme průměrné zaplnění stadionu pro každé kolo. Navíc využijeme tzv. prostých klouzavých průměrů, abychom zvýraznili vývoj v čase, který nebude zkreslený průběhem zaplnění stadionu v jednotlivých sezonách, tedy očistíme data od periodicky se opakujících vývoju návštěvnosti. To uděláme tak, že budeme uvažovat novou datovou množinu získanou jako průměry každých dvaapadesáti po sobě jdoucích průměrných zaplnění stadionů. Takto získaná data přiřadíme časovému okamžiku uprostřed 52 uvažovaných kol. To však neodpovídá žádnému z původních okamžiků, proto ještě zprůměrujeme každou dvojici takto získaných dat, opět ji přiřadíme středu těchto dvou bodů a teprve poté si vykreslíme časovou řadu¹. Prvních a posledních 26 pozorování doplníme na základě lineárního trendu určeného z prvních a posledních 52 průměrných zaplnění.



Obrázek 2.2: Průměrné zaplnění stadionů pro jednotlivá kola spolu s klouzavým průměrem

V grafu jsou modrými vertikálami odděleny jednotlivé sezony a červenou čarou vykreslený trend odhadnutý pomocí klouzavých průměrů. Z něj je pa-

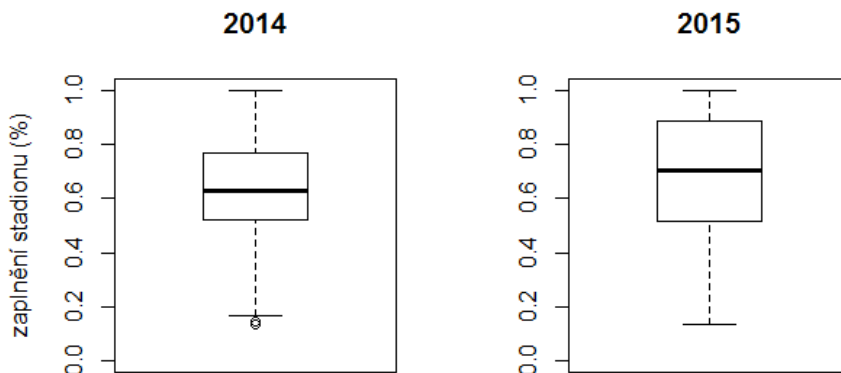
¹Vzhledem k neekvidantní povaze časové řady se ani po tomto kroku nevrátíme do původních časových okamžiků, budeme jim však blíže.

trné, že v závislosti na čase se návštěvnost opravdu měnila. Kromě jednoho výkyvu zaplnění stadionu klesalo zhruba v průběhu prvního dva a půl roku, poté došlo k výraznému nárůstu v první polovině sezony 2012/13, když se do Tipsport extraligy dočasně vrátil Jaromír Jágr, na jehož tým poté bylo velmi často vyprodáno. Po jeho odchodu návštěvnost klesala, ale už od následující sezony opět došlo k nárůstu zaplnění stadionů, který trval až do konce sledovaného období.

Data v závěru vybraných sedmi let nepatrně zkresluje snižující se kapacita zimních stadionů, její průměr od počátku sledovaného období klesl o více než osm set míst, graf absolutních návštěvností v čase však vykazuje podobný trend. Vzhledem k tomu, že pro vykreslení grafu a odhad trendu jsme data očišťovali od sezonnosti pomocí klouzavých průměrů, může být graf zkreslený bodem změny, tedy jediným okamžikem, v němž došlo k trvalému zvýšení či snížení procentuální návštěvnosti o nějakou konstantu, nebo ke změně trendu v časové řadě. Proto využijeme test pro jeho odhalení, k čemuž v softwaru R slouží příkaz `e.cp3o()` z knihovny `ecp`. Tento test je založený na měření vzdáleností mezi chronologicky uspořádanými průměrnými návštěvnostmi (v %), případně celými shluky těchto návštěvností. Kvůli očividné odlišnosti způsobené Jágrovou účastí v hokejové extralize však z testu vynecháme rok 2013.

Test prokázal existenci dvou bodů změny, konkrétně na začátku sezon 2011/12 a 2014/15. Vzhledem k tomu, že za bod změny bylo označeno vždy první kolo základní části, ke změně zřejmě došlo v období mezi dvěma sezonami. V ročníku 2014/15 se jako možná příčina nabízí odchod Chomutova a především Kladna, jehož návštěvnost v poslední sezoně nepřekročila 55 % kapacity zimního stadionu. Naopak místo těchto týmů přišly Mladá Boleslav a Olomouc, jež v následujícím roce nikdy nezaplňily méně než polovinu svých stadionů. Tento vliv lze snadno ověřit testem o shodě distribučních funkcí mezi příslušnými ročníky při vynechání týmů, které si vyměnily místa. K tomu poslouží Wilcoxonův test, který je neparametrickou obdobou dvouvýběrového t-testu, namísto procentuálního zaplnění stadionů však pracuje s pořadím těchto hodnot. Tento test však vyžaduje, aby data

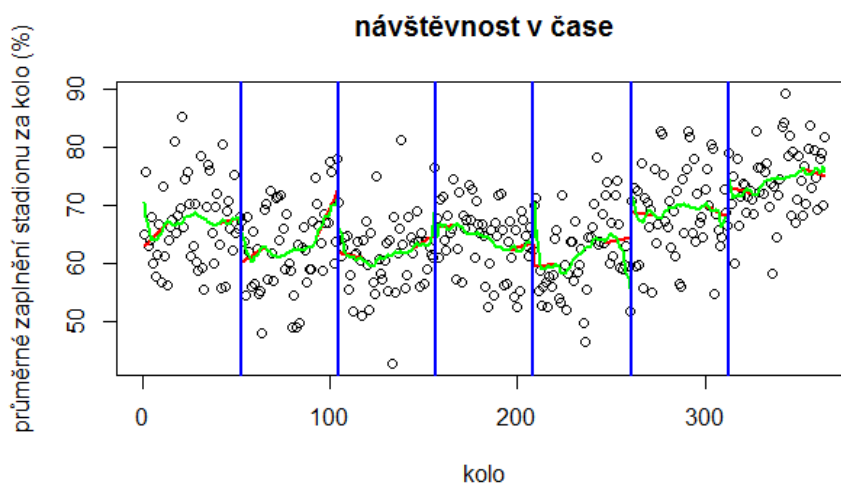
byla alespoň přibližně symetrická, nejprve si tedy vykreslíme jejich krabicové grafy a ověříme splnění této podmínky.



Obrázek 2.3: Krabicové grafy prokazující symetrii dat pro test vlivu příchodu Olomouce a Mladé Boleslavi na návštěvnost

Data jsou přibližně symetrická, můžeme proto přistoupit k Wilcoxonovu testu a ověřit hypotézu, že je návštěvnost stejná v obou letech. Tu ale na hladině významnosti 0,05 zamítáme, p-hodnota vychází menší než 0,0001 a mezi danými sezonami tak zřejmě došlo i k další změně, která se projevila v zájmu lidí o hokej.

Na závěr se s využitím klouzavých průměrů podíváme i na průběh jednotlivých ročníků. Průměrovat budeme vždy jen 26 sousedních pozorování, protože to je nejnižší počet kol, během nichž stihne hrát každý tým s každým na vlastním ledě i na ledě soupeře. Odstraníme tak vlivy jednotlivých týmů a stadionů, ale zachováme část informace o průběhu jedné sezony. Aby se ročníky vzájemně neovlivňovaly na svých koncích a začátcích, prozkoumáme data jako sedm samostatných množin a koncové a počáteční body doplníme tak, že budeme zmenšovat počet dat, která průměrujeme. První a poslední bod každé datové množiny tak bude roven přímo průměrnému zaplnění stadionu v daném kole. Protože se ale častěji k výpočtu počátečních a koncových klouzavých průměrů využívá odhad lineárního trendu, je v grafu zobrazen i ten, a to pomocí červené křivky.



Obrázek 2.4: Průběh jednotlivých sezon – zelenou čarou jsou vykresleny klouzavé průměry se zkracující se délkou okna na začátku a na konci, červenou odhady počátečních a koncových bodů pomocí lineárního trendu

Porovnáme-li počáteční a koncové body spočítané pomocí obou postupů, vidíme, že pomocí odhadu lineárního trendu téměř vždy dostaneme pro první zápasy sezony nižší hodnoty než pomocí klouzavých průměrů se zkracující se délkou okna, naopak ke konci sezony bychom pomocí lineárního trendu odhadovali hodnoty vyšší než ve skutečnosti jsou. To může být kromě rozdílné kapacity stadionů dáno také tím, že se fanoušci po dlouhé letní pauze více těší na začátek sezony a toto nadšení po prvních několika kolech opadne, což by při extrapolaci dat na základě lineárního trendu z prvních šestadvaceti pozorování nemuselo být tak patrné. Rozdíly v závěru sezony mohou být dány tím, jak některé týmy přicházejí o možnost postupu do play-off nebo mají jistou pozici v konečném umístění po základní části.

Z grafu jsou také vidět výrazné rozdíly mezi jednotlivými sezonami, přibližně v polovině případů zaplnění stadionu roste s počtem odehraných kol, v jednom případě naopak klesá a ve zbylých dvou nejprve došlo k poklesu, poté k nárůstu a ke konci sezony zaplnění opět klesalo. Rostoucí tendence zaplnění by mohla vycházet z toho, že s blížícím se koncem sezony se zdá každý bod důležitý a fanoušci očekávají napínavější zápasy, pokles v sezoně 2012/2013 zase může

vysvětlit, že extraligu hráli také slavní hráči ze zámoří, kteří se ale v průběhu ročníku postupně vraceli zpět do NHL.

2.2 Faktory ovlivňující návštěvnost

2.2.1 Úspěšnost týmu

První myšlenkou, kterou otestujeme, je mezi fanoušky rozšířené tvrzení, že někteří lidé chodí na hokej, jen když se právě jejich týmu daří. To je však příliš obecné tvrzení, proto budeme uvažovat jako projev úspěšnosti pět menší či větší měrou souvisejících faktorů:

1. počet branek, které tým zaznamenal v minulém utkání
2. počet bodů, které získal v uplynulé části sezony
3. počet bodů získaných od posledního utkání bez bodu
4. počet vítězství od poslední porážky
5. umístění týmu v tabulce v dané chvíli

Nejprve se podívejme, zda je zaplnění stadionu ovlivněno celou pěticí faktorů současně. K tomuto účelu si vytvoříme lineární regresní model závislosti zaplnění stadionu na těchto faktorech a podíváme se na korelaci hodnot odpovídajících modelu a nasbíraných dat. Kvůli absenci normálního rozdělení dat použijeme Spearmanův korelační koeficient, který počítá s pořadím hodnot a ne jejich velikostí. Výsledek je méně přesný než v případě, kdy můžeme použít Pearsonův koeficient, ale má tu výhodu, že odhalí každou monotónní závislost, ne jen lineární. Výsledný korelační koeficient $R_S = 0,19$ poukazuje na slabou závislost mezi zaplněním stadionu a pěticí sledovaných faktorů. Pro jistotu však ještě otestujeme, že tento koeficient je významně nenulový. Jelikož data nejsou normálně rozdělená a je jich více než třicet, použijeme neparametrický test, který má za platnosti nulové hypotézy následující podobu:

$$\frac{R_S - 0}{\sqrt{\frac{1}{n-1}}} = R_S \sqrt{n-1} \sim N(0, 1)$$

Výše spočítaný koeficient byl kladný, a pro test tedy budeme uvažovat pravostranou alternativu. Testová statistika vychází 9,678 a p-hodnota je menší než 0,0001 a test tedy potvrzuje, že tato závislost není pouze dílem náhody. Nicméně po prozkoumání Spearmanovy korelace s jednotlivými vlivy zvláště (odhady koeficientů a p-hodnoty testu jejich nulovosti jsou uvedeny v Tabulce 2.1) se zdá, že vliv ostatních faktorů lze zanedbat, protože koeficient mnohonásobné korelace je v absolutní hodnotě téměř stejný jako Spearmanův korelační koeficient pouze pro vliv umístění. Rozdíl ve znaménku je způsobený tím, že lepší umístění v tabulce je dáno nižším číslem, což při použití regresního modelu nehraje v korelaci roli.

faktor	Spearmanův korelační koeficient	p-hodnota
góly minule	0,053	0,0077
body	0,103	< 0,0001
forma (vítězství)	0,098	< 0,0001
forma (body)	0,096	< 0,0001
umístění	-0,178	< 0,0001

Tabulka 2.1: Spearmanův korelační koeficient závislosti zaplnění stadionu na jednotlivých projevech úspěšnosti týmu

Abychom si tuto teorii ověřili, spočítáme ještě parciální korelační koeficienty, tedy takové, které vyčíslí závislost mezi zaplněním stadionu a některým z pěti testovaných vlivů a přitom nebudou ovlivněny zbylými čtyřmi faktory. Odhalíme tak slabý, přesto však nenulový vliv získaných bodů.

faktor	Parciální korelační koeficient	p-hodnota
góly minule	-0,000	0,1371
body	0,047	0,0177
forma (vítězství)	0,034	0,0818
forma (body)	-0,029	0,1404
umístění	-0,137	< 0,0001

Tabulka 2.2: Spearmanův parciální korelační koeficient závislosti zaplnění stadionu na jednotlivých projevech úspěšnosti týmu

Protože se ale týmům často daří na vlastních ledových plochách více než na ledě jejich soupeřů, zopakujme testy i pro hodnoty zkoumaných vlivů spočítané jen z domácích utkání. Takto spočítané koeficienty se ovšem budou lišit jen o několik tisícín. Výjimkou je forma týmu v bodech, u níž parciální korelace zjištěná na datech jen z domácích stadiónů je v absolutní hodnotě desettisíckrát menší než parciální korelace počítaná na všech datech. Hypotézu její nulovosti nezamítneme ani v tomto případě. Můžeme tedy prohlásit, že není významný rozdíl mezi tím, zda zkoumáme úspěšnost týmu jen na jeho vlastním ledě nebo obecně v extralize.

Významnější byl vliv umístění a ten proto také blíže prozkoumáme. Především nás bude zajímat, jestli tato závislost není vyvolaná rozdílnými výsledky týmů, proto si Spearmanův korelační koeficient spočítáme pro každý tým zvlášť. Výsledkem je následující tabulka:

tým	Spearmanův korelační koeficient	p-hodnota
CHO	-0,625	< 0,0001
KLA	-0,622	< 0,0001
LIB	-0,452	< 0,0001
MBL	-0,452	< 0,0001
SPA	-0,337	< 0,0001
OLO	-0,314	0,0234
PLZ	-0,288	< 0,0001
KVA	-0,275	0,0002
LIT	-0,274	0,0002
BUD	-0,191	0,0522
PCE	-0,179	0,0158
KOM	-0,144	0,0525
ZLN	-0,139	0,0613
TRI	-0,117	0,1146
VIT	-0,069	0,3570
SLA	-0,064	0,4258
HKR	-0,028	0,8093

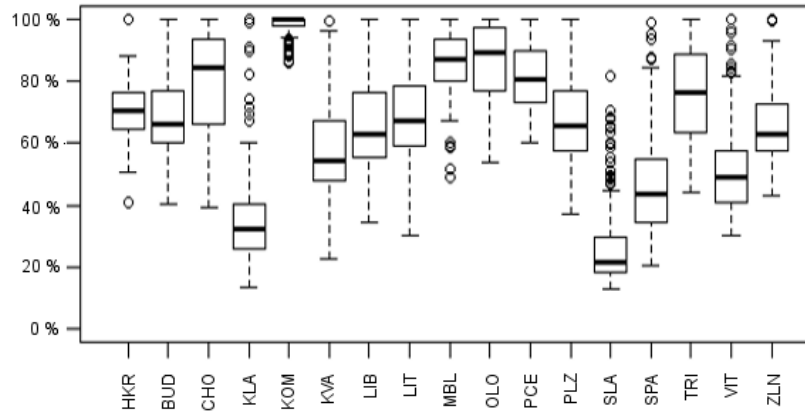
Tabulka 2.3: Spearmanův korelační koeficient vyjadřující míru závislosti mezi umístěním týmu a zaplněním jeho stadionu

Celkem u deseti týmů ze sedmnácti test objevil významnou korelaci, nejvíce je pak patrná u Chomutova a Kladna, kde absolutní hodnota koeficientu nabývá

hodnot vyšších než 0,6. Rozdíly v síle korelace budou patrně způsobeny dalším, neznámým faktorem jako je například odlišný přístup lidí k hokeji nebo další v daném městě nabízené sporty.

2.2.2 Město a velikost stadionu

Nyní se podívejme na rozdílnou návštěvnost mezi jednotlivými týmy, tedy který tým má v průměru největší fanouškovskou základnu a který má nejbližší k zaplnění celého stadionu.



Obrázek 2.5: Procentuální návštěvnost pro jednotlivé týmy

Z krabicového grafu můžeme vidět, že pro jednotlivé týmy se zaplněná část stadionu liší, suverénně nejplnější stadion má Kometa Brno, která vyprodala za sedm let neuvěřitelných 110 zápasů základní části (z celkem 182 možností). Odlišnosti mezi týmy lze snadno ověřit pomocí Kruskal-Wallisova testu, jež je neparametrickou obdobou analýzy rozptylu při jednoduchém třídění. Ten v tomto případě velmi přesvědčivě (p-hodnota vychází menší než 0,0001) zamítá nulovou hypotézu, že distribuční funkce návštěvnosti je na všech stadionech stejná. Pro srovnání jednotlivých týmů použijeme neparametrické mnohonásobné porovnávání, k němuž slouží nerovnost

$$|t_i - t_j| \geq \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right) h_{KW}(\alpha)}$$

kde t_i je průměrné pořadí pro i -tý výběr a $h_{KW}(\alpha)$ je kritická hodnota ze speciálních tabulek. Při dostatečném množství pozorování však můžeme kritickou hodnotu nahradit kvantilem χ^2 rozdělení, kde počet stupňů volnosti je o jedna nižší než počet skupin, do nichž pozorování třídíme. Platí-li tato nerovnost, pak zamítáme hypotézu o shodě distribučních funkcí obou skupin.

V případě testu o shodě distribučních funkcí mezi jednotlivými týmy je výsledkem 83 dvojic (z celkem 136 možností), pro něž hypotézu zamítáme, proto bude lepší uspořádat týmy do tabulky 2.4 spolu s jejich průměrnou návštěvou a zaplněním stadionů. Pro úplnost doplníme i velikost města podle údajů z cs.wikipedia.org.

#	tým	počet diváků	zaplnění haly v (%)	prům. kapacita	velikost města v tisících
1	KOM	7 284	98,26	7 423	378
2	OLO	4 750	86,38	5 500	99
3	MBL	3 590	85,48	4 200	44
4	PCE	8 371	81,57	10 275	90
5	CHO	4 213	80,25	5 250	49
6	TRI	3 962	76,19	5 200	36
7	BUD	4 522	70,43	6 421	94
8	HKR	4 924	69,05	7 160	93
9	LIT	4 552	68,71	6 661	25
10	PLZ	5 630	68,10	8 261	167
11	LIB	5 015	66,87	7 500	102
12	ZLN	4 607	65,82	7 000	75
13	KVA	3 481	58,02	6 000	49
14	VIT	5 229	52,27	10 004	297
15	SPA	6 468	47,03	13 780	1 247
16	KLA	3 118	36,24	8 276	68
17	SLA	4 738	27,30	17 360	1 247

Tabulka 2.4: Průměrná návštěva a zaplnění na jednotlivých stadionech

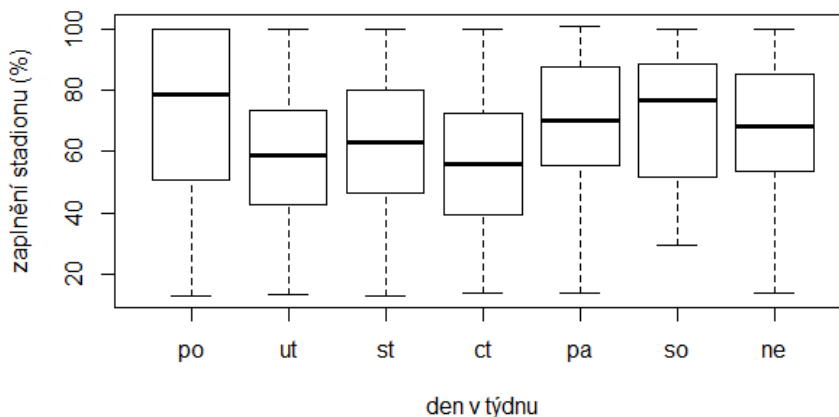
Kapacita některých stadionů se v průběhu let měnila, proto byla použita jejich průměrná velikost. Ta se pochopitelně podepíše i na zaplnění, neboť je snazší obsadit všechna místa na čtyřtisícovém stadionu Mladé Boleslavi než v sedmnáctitisícové aréně Slavie. Obdobně počet obyvatel města by mohl mít kladný vliv, zde je ale potřeba mít na paměti, že Sparta a Slavia hrají ve stejném městě a o diváky se

tedy dělí. V testech korelačního koeficientu proto budeme uvažovat jen polovinu obyvatel pro oba týmy.

Testy prokázaly, že se zvyšující se kapacitou stadionu klesá i jeho procentuální zaplnění, velikost města však tuto statistiku neovlivnila. Můžeme se podívat i na vliv velikosti města na průměrný počet diváků, kteří na stadion přišli. Zde vychází korelační koeficient vyšší než 0,7, po očištění od vlivu kapacity, která tentokrát působí kladně (představuje horní mez návštěvnosti na příslušném stadioně), vychází korelace přibližně 0,5.

2.2.3 Den v týdnu

Dalším důležitým faktorem by mohl být den v týdnu, kdy se utkání odehrálo. Zde využijeme skutečnosti, že se většinou hraje hned sedm zápasů v jediný den a vliv jednotlivých týmů tak bude ve výsledku zanedbatelný, tudíž můžeme testy provádět souhrnně.



Obrázek 2.6: Zaplnění stadionu podle dne v týdnu

Vykreslená data nám mnoho neřekla, rozdíl mezi jednotlivými dny může a nemusí být významný. Přejdeme proto ke Kruskal-Wallisovu testu, který uvádí p-hodnotu menší než 0,0001 a musíme proto zamítnout shodu distribučních funkcí.

Porovnáme-li jednotlivé dvojice dnů, dostaneme následující rozdíly:

rozdíly	
čtvrtek	neděle
čtvrtek	pátek
neděle	úterý
pátek	úterý

Tabulka 2.5: Dny, pro něž se významně liší zaplnění stadionů

Nás však budou více zajímat rozdíly mezi skupinami dnů lišících se jejich postavením vůči volným a pracovním dnům, proto si data rozdělíme do tří skupin podle toho, zda pocházejí z pátku či soboty (dnů, po nichž následuje volno), neděle či pondělí (volných dnů, po nichž následuje pracovní den)², nebo z trojice dnů úterý, středa a čtvrtek (pracovní dny, po nichž následuje další pracovní den). Pro Kruskal-Wallisův test vychází p-hodnota menší než jedna desetitisícina a přistoupíme proto k neparametrickému mnohonásobnému porovnávání, abychom objevili, které skupiny se od sebe liší. Z něj vzejde následující tabulka rozdílů:

rozdíly	
volno před pracovním dnem	pracovní den
pracovní den	den před volnem

Tabulka 2.6: Skupiny dní lišící se postavením vůči volnu, pro něž se významně liší zaplnění stadionů

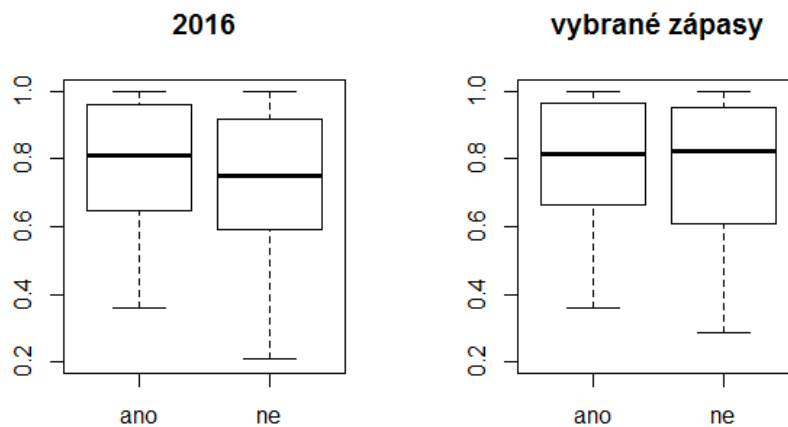
Vidíme tedy, že úterý až čtvrtek se významně liší od ostatních dnů, zatímco mezi dvojicemi pátek, sobota a neděle, pondělí test nenalezl na hladině 0,05 dostatečný rozdíl. Vrátime-li se ke krabicovému grafu na obrázku 2.6, je vidět, že od úterý do čtvrtka je návštěvnost nižší.

2.2.4 Televizní utkání

Podívejme se teď, jestli je rozdíl mezi utkáními, která v přímém přenosu vysílala Česká televize, a mezi těmi, která vysílána nebyla. Původní předpoklad

²V pondělí se hrává většinou jen pokud je příslušný den státním svátkem

byl, že by televizní přenosy návštěvnost snižovaly, ale už z boxplotu na obrázku 2.7 (vlevo) je vidět, že data budou tvrdit spíše opak. To je dáno tím, že Česká televize se snaží do přenosů zahrnout ta nejatraktivnější utkání, včetně většiny derby a především včetně utkání pod širým nebem, která jsou sice jen jedno nebo dvě do roka, ale vždy se těší vysoké návštěvnosti. Využijme nyní skutečnosti, že každý tým se utká s každým čtyřikrát do roka, ale nestává se, že by Česká televize odvysílala během jednoho ročníku všechna čtyři utkání mezi konkrétní dvojicí týmů. Omezíme tedy datovou množinu jen na zápasy dvojic týmů, které se v ročníku 2015/2016 střetly alespoň jednou před televizními kamerami. V takto vybraných datech porovnáme utkání, která Česká televize přenášela a která ne.

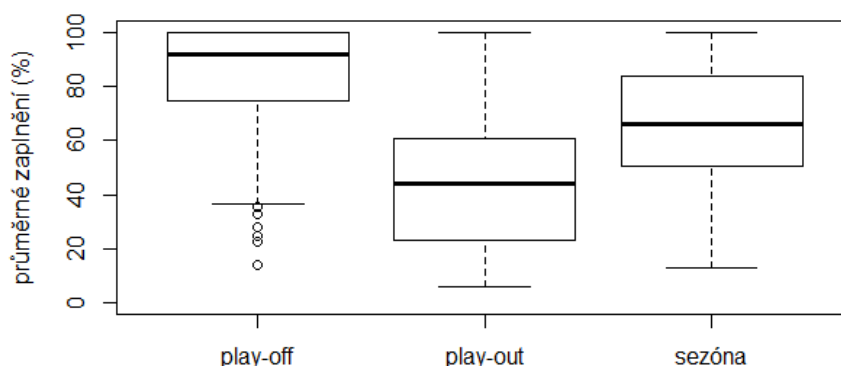


Obrázek 2.7: Graf návštěvnosti podle toho, zda utkání přenášela Česká televize – vlevo všechny zápasy sezony 2015/2016, vpravo zápasy vybraných dvojic týmů.

V grafu můžeme vidět, že omezení počtu zápasů skutečně srovnalo průměrnou návštěvnost pro přenášovaná i nepřenášovaná utkání, avšak nezměnilo ji natolik, aby data mluvila ve prospěch původní domněnky. To také potvrdí Wilcoxonův test, který na hladině významnosti 0,05 neumožňuje zamítnout nulovou hypotézu, že kamery nesnižují zaplnění stadionu, příslušná p-hodnota vychází větší než 0,606.

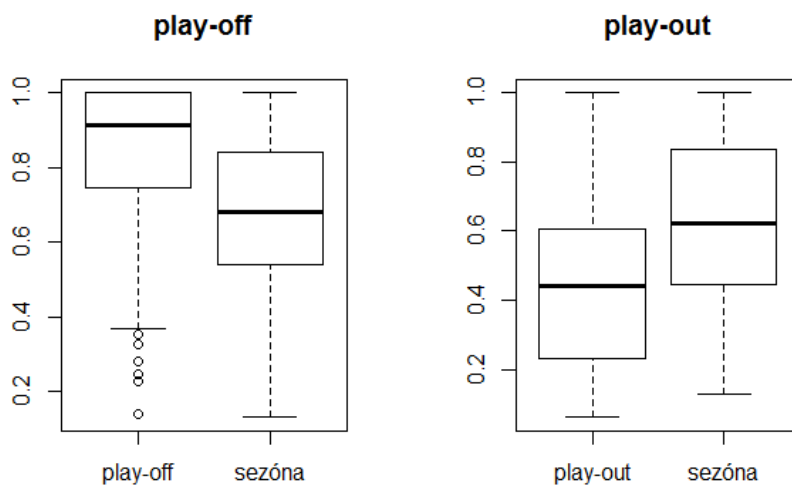
2.2.5 Play-off, play-out

Asi nejzřejmější vliv se dá očekávat mezi jednotlivými částmi české nejvyšší hokejové soutěže, jak byly představeny v 1. kapitole. Play-off se těší vysoké oblíbenosti, neboť už je pro každý tým důležitý každý zápas, pro fanoušky pak zřejmě bude nejvýznamnější rozhodující utkání, neboť budou chtít vidět na vlastní oči radost svého týmu z postupu, či jak jejich kapitán zvedá nad hlavu Masarykův pohár. V play-out se dá naopak očekávat pokles návštěvnosti, poněvadž výhrou je „jen“ udržení extraligové příslušnosti pro další rok a této části soutěže se účastní čtyři nejhorší týmy a z hlediska příležitostného diváka se tak může play-out jevit jako méně kvalitní soutěž. Nejprve se proto podíváme na rozdíly mezi částmi soutěže.



Obrázek 2.8: Krabicový graf zaplnění stadionu pro jednotlivé části soutěže

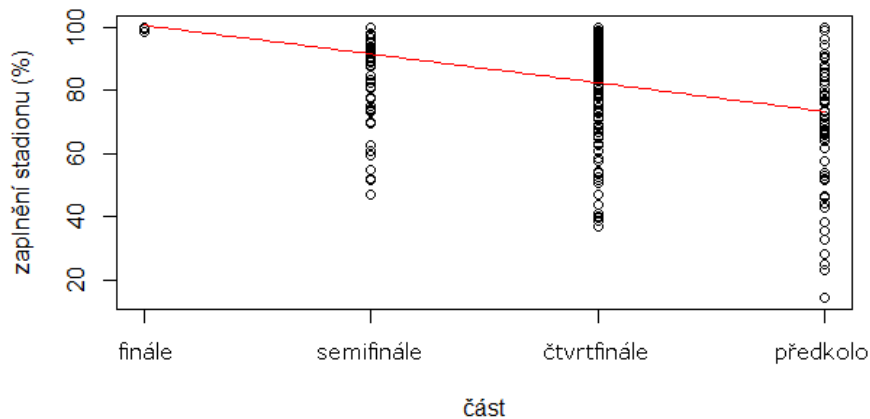
Grafické zpracování dat tuto domněnku na první pohled potvrzuje, ale je potřeba vzít v potaz i skutečnost, že do jednotlivých částí soutěže se dostanou jen některé týmy, což by mohlo celkovou statistiku ovlivnit. Proto data rozdělíme do dvou skupin, na zápasy týmů, které příslušný ročník zakončily bojem o Masarykův pohár, a na zápasy těch, kteří hráli o udržení, a porovnáme s dalšími částmi soutěže jen příslušné podskupiny.



Obrázek 2.9: Krabicový graf zaplnění stadionu pro jednotlivé části soutěže u vybraných týmů

Grafy se nezměnily nijak výrazně, ale k nepatrnému vyrovnání došlo, proto provedeme test pro obě části soutěže zvlášť. Data pro play-off postrádají symetrii a proto pro jejich porovnání zvolíme Kruskal-Wallisův test, který dá p-hodnotu menší než 0,0001, což znamená, že odhalil rozdíly mezi oběma skupinami. Play-out alespoň přibližně splňuje podmínku symetričnosti dat a k testu proto můžeme zvolit Wilcoxonův pořadový test s pravostranou alternativou (nepředpokládáme, že by play-out mělo vyšší návštěvnost než základní část). Také tentokrát vychází velmi nízká p-hodnota a tedy můžeme prohlásit, že i zde je rozdíl významný.

Na základě skutečnosti, že s blížícím se finále se tým, který stále hraje, jeví jako úspěšnější, můžeme otestovat ještě rozdíly návštěvnosti pro jednotlivé části play-off (od předkola do finále).



Obrázek 2.10: Zaplnění stadionu v závislosti na části play-off a odhadnutý lineární trend

Zde vidíme, že s postupem play-off se zvyšují šance týmu na vyprodání stadionu, ve finále pak nebývá vyprodáno jen zřídka. K otestování tohoto vlivu, využijeme lineární trend v kontingenční tabulce počtu zápasů, které si rozdělíme podle toho, v jaké části play-off se odehrály, a zda bylo nebo nebylo vyprodáno.

část	nebylo vyprodáno	vyprodáno
finále	3	37
semifinále	43	35
čtvrtfinále	124	36
předkolo	57	5

Tabulka 2.7: Kontingenční tabulka vyprodání stadionu pro jednotlivé části play-off

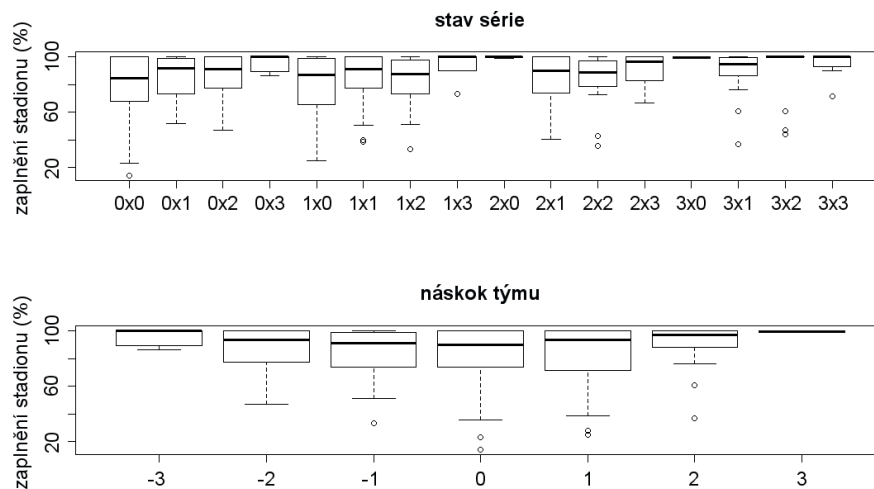
Z této tabulky spočítáme korelační koeficient jako

$$\hat{\rho} = \frac{\sum_i \sum_j (u_i - \bar{u})(v_j - \bar{v})p_{ij}}{\sqrt{\sum_i (u_i - \bar{u})^2 p_{i.} \sum_j (v_j - \bar{v})^2 p_{.j}}},$$

kde u_i jsou řádkové skóry (1 – finále, ..., 4 – předkolo), v_j sloupcové skóry (0 – ne, 1 – ano), $p_{ij} = \frac{n_{ij}}{n}$, $\bar{u} = \sum_i u_i p_{i.}$ a $\bar{v} = \sum_j v_j p_{.j}$. Výslednou hodnotu umocníme na druhou, vynásobíme číslem o jedna nižším než je počet pozorování

a porovnáme s 95% kvantilem χ^2 rozdělení o jednom stupni volnosti[3]. Koeficient $\hat{\rho}$ vyšel přibližně $-0,502$, po provedení výše uvedených výpočtů získáme hodnotu testové statistiky $85,581$, která je výrazně vyšší než 95% kvantil χ^2 rozdělení a zamítáme proto hypotézu o nezávislosti. Pravděpodobnost vyprodání stadionu roste s postupující fází play-off.

Poslední, co nás bude zajímat u play-off, je, jestli se na návštěvnosti podepíše i stav série, případně náskok či ztráta týmu ve vyhraných zápasech. Pro začátek si opět vykresleme krabicový graf pro obě dělení dat.



Obrázek 2.11: Krabicový graf návštěvnosti – nahoře rozdělená podle stavu série, dole podle náskoku či ztráty domácího týmu

Vidíme, že střední hodnoty se budou lišit v závislosti na stavu série, ale neobjevíme zřejmě monotónní závislost návštěvnosti na náskoku jednoho týmu nad druhým. Pozoruhodné také je, že i v okamžiku, kdy tým prohrává v sérii už 0:3, je vysoká šance, že bude vyprodáno. Podrobíme-li sesbíraná data Kruskal-Wallisovu testu, vyjde nám, že pro rozdílný stav série se návštěvnost liší, pro náskok jednoho týmu před druhým však ne. Z grafu je ale dobře vidět, že významný rozdíl bychom mohli objevit mezi rozptylem hodnot. K ověření této hypotézy nám poslouží Sieglův-Tukeyho test[24], který je založený na Wilcoxonovu dvouvýběrovém testu s odlišnou definicí pořadí $\tilde{t}_{(.)}$. Zatímco u předchozích testů jsme přiřadili pořadí

1 pro nejnižší hodnotu, 2 pro druhou nejnižší až n pro nejvyšší hodnotu, nyní budeme přiřazovat pořadí tak, že 1 bude patřit nejnižší hodnotě, 2 nejvyšší, 3 druhé nejvyšší, 4 druhé nejnižší, 5 třetí nejnižší, 6 třetí nejvyšší atd. až dosáhneme toho, že nejvyšší pořadí bude patřit jedné ze dvou prostředních hodnot uspořádaného výběru. Přiřazení pořadí lze znázornit následovně:

$$\tilde{t}_{(.)} = (1, 4, 5, \dots, 6, 3, 2)$$

Dále si označíme jako \tilde{S}_{kj} pořadí těch návštěvností, kde měl domácí tým nad soupeřem náskok k . Testovou statistikou je poté součet těchto pořadí:

$$ST = \sum_{j=1}^{n_k} \tilde{S}_{kj}$$

kde n_k je počet pozorování konkrétního výběru. Kritické hodnoty odpovídají kritickým hodnotám z Wilcoxonova testu. Tento test však neodhalil dostatečně velké rozdíly mezi rozptyly, získaná data nesvědčí o odlišnosti rozptylů pro jednotlivé situace.

2.2.6 Atraktivnost soupeře, derby

Všechny rozdíly jsme dosud hledali jen u domácího týmu, proto podrobíme testům i soupeře, jehož atraktivita jistě také přispívá k návštěvnosti. Začneme tedy testem shody distribučních funkcí pro jednotlivé soupeře, k čemuž nám opět poslouží Kruskal-Wallisův test. Ten prokázal, že mezi návštěvností na jednotlivé týmy skutečně existuje významný rozdíl i v případě, že přijedou na cizí stadion, po porovnání dvojic najdeme celkem sedm odlišností, ve všech figuruje Sparta Praha. Vytvoříme-li si tabulku průměrných zaplnění soupeřova stadionu, uvidíme, že na Spartu chodí nejvíce lidí.

#	tým	průměrné zaplnění
1	SPA	76,50 %
2	SLA	71,92 %
3	OLO	69,63 %
4	KOM	67,83 %
5	PCE	67,36 %
6	PLZ	67,31 %
7	KLA	67,09 %
8	HKR	64,90 %
9	LIT	64,66 %
10	LIB	63,81 %
11	TRI	63,65 %
12	KVA	63,14 %
13	ZLN	63,07 %
14	CHO	62,90 %
15	VIT	62,54 %
16	MBL	60,52 %
17	BUD	59,34 %

Tabulka 2.8: Procentuální zaplnění stadionů pro jednotlivé soupeře

To by mohlo být způsobeno i vysokým počtem fanoušků vyjíždějících se Spartou na venkovní utkání, přinejmenším na moravských stadionech jsem však zpravidla mnoho pražských fanoušků neviděl. Dá se proto předpokládat, že je to skutečně dáno spíše atraktivností soupeře, neboť utkání se Spartou bývají zpravidla vyhrocená bez ohledu na tým, proti němuž stojí.

Za zmínku stojí i vysoké umístění Brna a Olomouce, kterým patřila první dvě místa v domácí návštěvnosti. Ze zvědavosti jsem se proto podíval na vzájemná utkání těchto dvou mužstev, která na sebe narazila celkem osmkrát a hned sedm zápasů bylo vyprodáno. Plné nebylo jen brněnské Rondo ve vůbec prvním střetu těchto týmů po olomouckém návratu do Tipsport extraligy.

Na závěr se ještě můžeme pokusit popsat hokejová derby. Za derby jsou zpravidla považována utkání týmů ze stejného města (v případě střetu Sparty a Slavie) nebo měst relativně blízkých (např. Olomouc a Zlín). Ne všechny zápasy mezi blízkými týmy jsou v médiích označovány jako derby, k tomu, co je obecně označováno jako derby a co ne, se však vrátíme později. Nejprve se podíváme na

korelaci mezi návštěvností (resp. zaplněním stadionu) a vzdáleností dvou měst po silnicích, potažmo času nezbytnému k cestě z jednoho města do druhého. Informace o vzdálenostech a době dojezdu pocházejí z on-line map společnosti Google.

Souhrnná Spearmanova korelace poukazuje na to, že návštěvnost souvisí se vzdáleností i s časem potřebným k přepravě, přičemž s časem je koeficient vyšší (vychází přibližně $-0,109$ pro čas a $-0,103$ pro vzdálenost). Již dříve jsme si ale ukázali, že i přes nenulový koeficient u celé ligy se korelace u některých týmů nemusí objevit a proto jsem test o její nulovosti zopakoval pro každý tým zvlášť. Tentokrát jsem však uvažoval už jen závislost s časem, která v souhrnné korelaci vycházela významněji nenulová.

tým	Spearmanův korelační koeficient	p-hodnota
OLO	$-0,524$	$< 0,0001$
VIT	$-0,455$	$< 0,0001$
PCE	$-0,418$	$< 0,0001$
ZLN	$-0,384$	$< 0,0001$
KVA	$-0,365$	$< 0,0001$
CHO	$-0,333$	$0,0029$
SPA	$-0,311$	$< 0,0001$
HKR	$-0,311$	$0,0056$
SLA	$-0,304$	$0,0001$
MBL	$-0,283$	$0,0011$
LIB	$-0,276$	$0,0002$
PLZ	$-0,273$	$0,0002$
TRI	$-0,256$	$0,0005$
KLA	$-0,254$	$0,0036$
BUD	$-0,252$	$0,0098$
KOM	$-0,214$	$0,0038$
LIT	$-0,173$	$0,0194$

Tabulka 2.9: Spearmanův korelační koeficient závislosti zaplnění stadionu na době dojezdu

Tentokrát není korelace nulová ani u jednoho týmu a pohybuje se v rozmezí od $-0,52$ do $-0,17$. Můžeme tedy prohlásit, že vzdálenost od soupeřova stadionu hraje roli u každého týmu a zda zápas je derby nebo ne, je opravdu dáno

i vzdáleností. Nás však bude zajímat, jestli je rozdíl pouze mezi derby (která budeme uvažovat jako zápasy s délkou cesty do jedné hodiny) a ostatními zápasy, nebo se rozdíly vyskytují i mezi vzdálenějšími týmy. Rozdělíme proto týmy do čtyř kategorií podle toho, do kolika hodin je možné se do města dostat a porovnáme jednotlivé kategorie.

rozdíly	
do 1 hodiny	do 2 hodin
do 1 hodiny	do 4 hodin
do 2 hodin	do 3 hodin
do 2 hodin	do 4 hodin
do 3 hodin	do 4 hodin

Tabulka 2.10: Doby dojezdu, pro něž se zaplnění stadionu významně liší

Výsledek se jeví na první pohled poněkud překvapivý, jelikož test nenašel významný rozdíl mezi zápasy, kam soupeř dojížděl méně než hodinu a mezi zápasy, kam cesta trvala 2-3 hodiny. Mohli bychom ale dojít k závěru, že to je dáno polohou měst. Kupříkladu, kdyby do Brna (které má nejvyšší procentuální návštěvnost) dojížděli soupeři právě necelé tři hodiny, zatímco na Slavii (která se umístila naopak nejhůře) by se odehrálo nejvíce zápasů, jež jsme podle vzdálenosti klasifikovali jako derby, mohlo by to průměrné hodnoty vyrovnat.

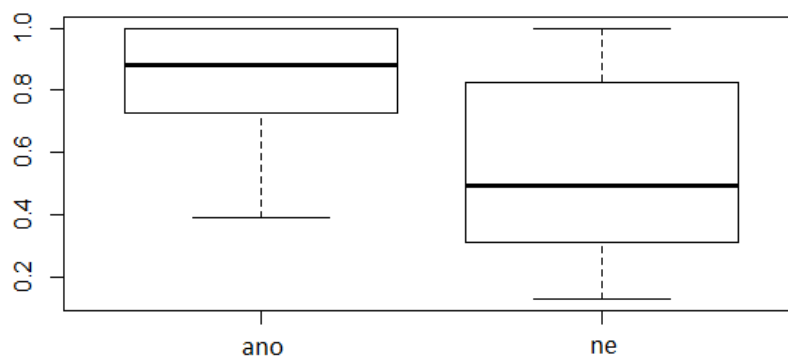
tým	doba dojezdu pro jednotlivé soupeře	zaplnění stadionu	počet zápasů
KOM	3	83,05 %	96
OLO	4	78,01 %	18
PCE	3	74,47 %	112
MBL	2	73,00 %	42
CHO	3	71,58 %	24
TRI	4	69,92 %	116
PLZ	2	67,71 %	87
HKR	3	66,98 %	46
LIT	2	66,68 %	74
LIB	3	65,34 %	70
BUD	3	64,89 %	48
ZLN	4	64,45 %	90
SPA	2	61,76 %	104
KVA	4	60,58 %	68
VIT	4	57,40 %	116
KLA	2	51,66 %	58
SLA	2	49,61 %	88

Tabulka 2.11: Zaplnění jednotlivých stadionů a nejvyšší počet odehraných zápasů se soupeři v jednom intervalu doby dojezdu

Z tabulky³ můžeme vidět, že hned dva ze tří týmů, které nejvíce zaplní haly i mimo derby, odehrají většinu zápasů se soupeři ve vzdálenosti 2 až 3 hodin jízdy autem, zatímco mezi šesti týmy s nejhorsí návštěvností se nenažde ani jeden, který by odehrál nejvíce zápasů s týmy v této vzdálenosti. Zřejmě proto pak dochází k vyrovnání průměrů se skupinou zápasů vzdálených do jedné hodiny cesty. Ani při vynechání zápasů Brna a Pardubic ale není rozdíl mezi první a třetí skupinou významný.

Zkusil jsem proto zápasy mezi týmy ve vzdálenosti nejvýše jedné hodiny od sebe rozdělit do skupin podle toho, jestli jsem se skutečně někdy dříve setkal s tím, že byly jako derby označeny v médiích, či nikoliv.

³Poslední sloupec udává počet zápasů, které tým odehrál na vzdálenost uvedenou v druhém sloupci.



Obrázek 2.12: Krabicový graf návštěvnosti – vlevo skutečná derby, vpravo ostatní zápasy zahrnuté podmínkou vzdálenosti

Zde je jasně vidět, že pouhá vzdálenost k definování hokejového derby nestačí. Například zápasy Sparty s Kladnem nejsou zahrnuty skutečnými derby, přestože k cestě z jednoho města do druhého stačí jen půl hodiny, naopak ze Zlína do Olomouce se jezdí téměř hodinu a přesto zápas do kategorie derby spadá. Kruskal-Wallisův test dává p-hodnotu menší než 0,0001 a potvrzuje, co jsme viděli už z grafu na obrázku 2.12. K popisu derby proto musíme kromě vzdálenosti uvažovat i nějaký další faktor, přinejmenším rivalitu mezi městy.

Závěr

V první kapitole se čtenář mohl obeznámit se základními informacemi o Tip-sport extralize, aby snáze pochopil uvažované vlivy na návštěvnost, a s daty, která posloužila k testování hypotéz v této práci.

V druhé kapitole jsme se pokusili odhalit faktory ovlivňující návštěvnost. Nejprve jsme se podívali na časovou řadu dat, abychom získali lepší přehled o vývoji návštěvnosti ve sledovaném období, čímž jsme odhalili, že návštěva v posledních letech výrazně vzrostla. Poté už jsme se zaměřili na test závislostí na možných prediktorech, k čemuž nám sloužily především Spearmanův korelační koeficient, Wilcoxonův pořadový test a Kruskal-Wallisův test, protože sesbíraná data neodpovídala výběru z normálního rozdělení.

Začali jsme testem souvislosti mezi úspěšností týmu a návštěvou v příštím zápase, na jejímž základě jsme byli schopni prokázat, že počet diváků je ovlivněn umístěním týmu v tabulce, následně jsme ale ukázali, že tato závislost se liší i podle toho, který tým právě sledujeme a u některých týmů je naprosto zanedbatelná. Pokračovali jsme testováním vlivu dne v týdnu, kde jsme objevili rozdíl mezi skupinami dnů úterý až čtvrtek a pátek až pondělí. Testům jsme podrobili i hypotézu, že televizní utkání snižují návštěvnost, k jejímuž testování jsme měli data jen z posledního roku, a na základě výsledku jsme dospěli k závěru, že především díky promyšlenému výběru České televize, její kamery neovlivní počet diváků. Mnohem patrnější vliv na návštěvu měla část, v níž se právě soutěž nacházela. Nejvíce diváků chodí na play-off a nejméně na play-out. Poslední vliv, který jsme zkusili prozkoumat, byla atraktivita soupeře, kde se prokázalo, že největší část stadionů se zaplní, když přijede pražská Sparta, což může být dáno

věrností jejích fanoušků, kteří by vyrazili na všechny výjezdy se svým týmem, ale daleko spíš její atraktivností a vyhrocenými utkáními.

V samotném závěru jsme se pak ještě pokusili prozkoumat, co určuje, který zápas je a který není derby. Zde bylo naší domněnkou, že toto přízvisko náleží zápasům dvou týmů, jejichž vzdálenost nepřesáhne nějakou hodnotu, ať už v ujetých kilometrech nebo v čase nezbytném k cestě z jednoho stadionu na druhý. Hypotéza, že návštěvnost roste se zkracující se vzdáleností měst, se ukázala správná, nestačila však k vysvětlení pojmu derby, do nějž bylo potřeba zahrnout i rivalitu mezi sousedními městy.

Na návštěvnost samozřejmě budou mít vliv i další faktory, které však z důvodu rozsahu textu byly vynechány. Dá se ale předpokládat, že výrazný vliv bude mít celková úspěšnost českého hokeje a dalších sportů pěstovaných na našem území, cena lístků či návrat slavných hráčů do české soutěže, kde často chtějí ukončit svou hráčskou kariéru.

Literatura

- [1] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky*. VUP, Olomouc, 2015.
- [2] RIZZO, M. L., SZÉKELY, G. J.: DISCO analysis: A nonparametric extension of analysis of variance. *The Annals of Applied Statistics*. Bethesda: Institute of Mathematical Statistics, 2010, (2), 1034-1055. DOI: 10.1214/09-AOAS245. ISBN 10.1214/09-AOAS245. Dostupné také z: <http://projecteuclid.org/euclid.aoas/1280842151>.
- [3] Agresti, A.: *Analysis of Ordinal Categorical Data* (2nd edition), Hoboken (NJ): Wiley, 2010.
- [4] Hokej.cz – web českého hokeje, <http://www.hokej.cz/> [cit. 14. 12. 2016].
- [5] HC Kometa Brno – oficiální internetové stránky, <http://www.hc-kometa.cz/>. [cit. 27. 11. 2016]
- [6] ČEZ Motor – oficiální stránky klubu, <http://www.hcmotor.cz/>. [cit. 27. 11. 2016]
- [7] Mountfield HK, <http://www.mountfieldhk.cz/>. [cit. 27. 11. 2016]
- [8] Piráti Chomutov – oficiální internetové stránky, <http://www.piratichomutov.cz/>. [cit. 28. 11. 2016]
- [9] Rytíři Kladno – Oficiální www stránky kladenského hokeje, <http://www.hc-kladno.cz/cz>. [cit. 28. 11. 2016]
- [10] HC Energie Karlovy Vary – oficiální internetové stránky, <http://www.hokejkv.cz/>. [cit. 1. 12. 2016]
- [11] HC Bílí Tygři Liberec – oficiální internetové stránky, <http://www.hcbilitygri.cz/>. [cit. 1. 12. 2016]
- [12] HC VERVA Litvínov – oficiální internetové stránky, <http://www.hcverva.cz/>. [cit. 20. 10. 2016]

- [13] BK Mladá Boleslav – oficiální internetové stránky, <http://www.bkboleslav.cz/>. [cit. 20. 10. 2016]
- [14] HC Olomouc – oficiální internetové stránky, <http://www.hc-olomouc.cz/>. [cit. 19. 11. 2016]
- [15] HC Dynamo Pardubice – oficiální internetové stránky, <http://www.hcdynamo.cz/>. [cit. 19. 11. 2016]
- [16] HC Škoda Plzeň – oficiální internetové stránky, <http://www.hcskoda.cz/>. [cit. 15. 12. 2016]
- [17] HC Slavia Praha – oficiální internetové stránky, <http://www.hc-slavia.cz/>. [cit. 15. 12. 2016]
- [18] HC Sparta Praha – oficiální internetové stránky, <http://www.hcsparta.cz/>. [cit. 8. 12. 2016]
- [19] HC Oceláři Třinec – oficiální internetové stránky, <http://www.hcocelari.cz/>. [cit. 8. 12. 2016]
- [20] HC Vítkovice Ridera, a.s., <http://www.hc-vitkovice.cz/>. [cit. 26. 10. 2016]
- [21] HC PSG Zlín – oficiální internetové stránky, <http://www.hokej.zlin.cz/>. [cit. 26. 10. 2016]
- [22] Česká televize, <http://www.ceskatelevize.cz> [cit. 30. 1. 2017]
- [23] Dubjaková, E.: *Metody mnohonásobného porovnávání pro jednoduché třídění* MUNI, Brno, 2009.
- [24] Lehmann, E.: *Nonparametrics statistical methods based on ranks*. Springer, New York, 2006
- [25] Mapy Google, <https://www.google.cz/maps/> [cit.23. 1. 2017]
- [26] Wikipedie, otevřená encyklopedie, <https://cs.wikipedia.org/> [cit. 7. 2. 2017]