UNIVERZITA PALACKÉHO V OLOMOUCI
PŘÍRODOVĚDECKÁ FAKULTA

BAKALÁŘSKÁ PRÁCE

Série vítězných utkání v hokeji - statistická analýza

Katedra matematické analýzy a aplikací matematiky
Vedoucí bakalářské práce: **Mgr. Ondřej Vencálek, Ph.D.**
Vypracoval(a): **Petra Sedláková**
Studijní program: B1103 Aplikovaná matematika
Studijní obor: Matematika–ekonomie se zaměřením na bankovnictví/pojišťovnictví
Forma studia: prezenční
Rok odevzdání: 2020

BIBLIOGRAFICKÁ IDENTIFIKACE

Autor: Petra Sedláková

Název práce: Série vítězných utkání v hokeji - statistická analýza

Typ práce: Bakalářská práce

Pracoviště: Katedra matematické analýzy a aplikací matematiky

Vedoucí práce: Mgr. Ondřej Vencálek, Ph.D.

Rok obhajoby práce: 2020

Abstrakt: Cílem bakalářské práce je zkoumat z různých hledisek závislost mezi výsledky po sobě jdoucích hokejových zápasů s ohledem na prostředí. Na datech pro sezóny 2016/2017 a 2017/2018 Tipsport Extraligy ledního hokeje jsme s využitím statistických metod (testy nezávislosti) provedli testování hypotéz nezávislosti. Na základě výsledků testů jsme pro některé týmy v určitých situacích zjistili vliv prostředí na jejich úspěšnost.

Klíčová slova: lední hokej, kontingenční tabulka, chí-kvadrát test nezávislosti, Fisherův faktoriálový test nezávislosti, poměr šancí

Počet stran: 48

Počet příloh: 1 CD

Jazyk: český

BIBLIOGRAPHICAL IDENTIFICATION

Author: Petra Sedláková

Title: Series of winning games in ice hockey - statistical analysis

Type of thesis: Bachelor's

Department: Department of Mathematical Analysis and Application of Mathematics

Supervisor: Mgr. Ondřej Vencálek, Ph.D.

The year of presentation: 2020

Abstract: The aim of this bachelor′s thesis is to examine the dependence between results of successive hockey matches from various points of view and with consideration to the environment. We tested independence hypotheses using statistical methods (independence tests) on the data for 2016/2017 and 2017/2018 Tipsport Extraliga ice hockey seasons. Based on the test results, we found out the influence of the environment on the success of some teams in certain situations.

Key words: ice hockey, contingency table, Pearson′s chi-squared test, Fisher′s exact test, odds ratio

Number of pages: 48

Number of appendices: 1 CD

Language: Czech

Prohlášení

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

V Brně dne ....

podpis

Obsah

Seznam tabulek

Poděkování

Ráda bych poděkovala vedoucímu práce Mgr. Ondřeji Vencálkovi, Ph.D., za odborné vedení, věcné připomínky, ochotu a trpělivost při vypracovávání mé bakalářské práce. Dále bych chtěla poděkovat své rodině za podporu při studiu.

Úvod

Hokej patří v České republice k jednomu z nejpopulárnějších sportů. Čeští fanoušci povzbuzují své oblíbené týmy buď přímo na zimních stadionech, nebo alespoň z pohodlí domova při sledování živých přenosů na televizních obrazovkách. Mnozí také podporují svoje děti v aktivním hraní a pomáhají tak vychovávat novou generaci hokejistů.

Hokejové týmy zažívají během sezóny lepší i horší období. Výhry a prohry se pravidelně střídají, nebo se vyskytují různě dlouhé série výher a proher. Také počet vstřelených gólů se v průběhu sezóny neustále mění. Mezi hokejovými fanoušky zní stále stejná otázka: Má prostředí nějaký vliv na úspěšnost týmu? Případně, jak velký tento vliv je?

Cílem této bakalářské práce je zkoumat z různých hledisek závislost mezi výsledky po sobě jdoucích utkání s ohledem na prostředí. Vycházíme z dat základní části Tipsport Extraligy pro hokejové týmy v rámci sezón 2016/2017 a 2017/2018. U týmu Karlových Varů uvažujeme kvůli sestupu pouze sezónu 2016/2017 a s nováčkem z Jihlavy v sezóně 2017/2018 vůbec pracovat nebudeme. S využitím statistického softwaru R a RStudio budeme testovat hypotézy nezávislosti výsledků po sobě následujících zápasů.

V první kapitole popíšeme datové soubory, se kterými budeme dále pracovat. V další kapitole se seznámíme s jednotlivými statistickými metodami, jež použijeme pro testování hypotéz. Ve třetí kapitole pak provedeme samotné testování hypotéz a rozbor získaných údajů.

Inspiraci, jak k problému závislosti přistoupit, jsme našli v deváté kapitole knihy Statistické úlohy, historky a paradoxy od Jiřího Anděla [1].

Kapitola 1

Datové soubory

V této kapitole se budeme zabývat popisem dat, s nimiž budeme pracovat a analyzovat je ve statistickém softwaru RStudio.

Tipsport Extraligu ledního hokeje (Tipsport ELH) hraje celkem 14 týmů. Pro naši práci jsou potřeba data ze základních částí sezón 2016/2017 a 2017/2018, která jsme získali z internetových stránek https://www.hokej.cz/. Protože byly Karlovy Vary v sezóně 2017/2018 nahrazeny Jihlavou, kterou do práce vůbec nezahrneme[1], tak v této sezóně uvažujeme pouze 13 týmů. Vytvořili jsme 14 datových souborů pro sezónu 2016/2017 (pro každý tým jeden soubor) a 13 datových souborů pro sezónu 2017/2018. Celkem tedy vzniklo 27 datových souborů.

Všechny datové soubory jsme rozdělili do sedmi sloupců a 52 řádků, které odpovídají celkovému počtu hraných zápasů neboli kol v základní části a jsou uspořádány podle data odehrání utkání. První sloupec **kolo** značí, kolikáté kolo se hrálo. Kvůli dřívějším předehrávkám a pozdějším dohrávkám v průběhu sezóny nejsou hodnoty ve sloupci uspořádány od 1 do 52. Další sloupec **domaci** označuje zkratku týmu, který hrál na domácí půdě. Vysvětlíme si jednotlivé zkratky:

- KOM – HC Kometa Brno
- MBL – BK Mladá Boleslav
- MHK – Mountfield HK (Hradec Králové)

[1] Ačkoli jsme tým HC Dukla Jihlava nezahrnuli do práce za účelem zkoumání závislosti, v datových souborech zbylých týmů pro sezónu 2017/2018 jsme zápasy těchto týmů proti jihlavské Dukle ponechali.

- CHM – Piráti Chomutov
- KVA – HC Energie Karlovy Vary
- LIB – Bílí Tygři Liberec
- LIT – HC VERVA Litvínov
- OLO – HC Olomouc
- PCE – HC Dynamo Pardubice
- PLZ – HC Škoda Plzeň
- SPA – HC Sparta Praha
- TRI – HC Oceláři Třinec
- VIT – HC VÍTKOVICE RIDERA
- ZLN – PSG Berani Zlín
- JIH – HC Dukla Jihlava

Třetí sloupec **goly_d** zobrazuje počet vstřelených gólů domácího týmu, následující sloupec **goly_h** naopak vstřelené góly hostujícím týmem. Další sloupec **hoste** udává zkratku týmu hrajícího na venkovním stadionu. Dále máme sloupec **goly_kom**, který v tomto případě značí počet vstřelených gólů týmem HC Kometa Brno, obecně však znamená vstřelené góly týmu s odpovídající zkratkou, jejichž význam jsme si uvedli výše. Sedmý a zároveň poslední sloupec **vyhra_kom** uvádí, zda tým HC Kometa Brno (obecně sledovaný tým s příslušnou zkratkou týmu) vyhrál (značeno 1) nebo prohrál (značeno 0). Následující tabulka (1.1) ukazuje část datového souboru pro tým HC Kometa Brno v sezóně 2016/2017.

kolo	domaci	goly_d	goly_h	hoste	goly_kom	vyhra_kom
1	ZLN	3	5	KOM	5	1
2	KOM	3	1	MHK	3	1
3	KOM	5	1	MBL	5	1
4	PLZ	3	6	KOM	6	1
5	KOM	0	1	VIT	0	0
6	KVA	2	4	KOM	4	1
7	KOM	7	1	LIT	7	1
8	PCE	3	4	KOM	4	1
9	KOM	6	2	SPA	6	1
10	TRI	4	3	KOM	3	0

Tabulka 1.1: Ukázka z dat týmu HC Kometa Brno pro sezónu 2016/2017

Kapitola 2

Statistické metody

V této kapitole uvedeme a popíšeme statistické metody, které v následující části využijeme pro testování našich hypotéz.

Zkoumání vztahů dvou kvalitativních znaků umožňují kontingenční tabulky. Základní testovanou hypotézou týkající se dvou kvalitativních znaků je hypotéza nezávislosti, kterou testujeme asymptotickým chí-kvadrát testem nezávislosti (pokud je dostatečně velký rozsah výběru), nebo s použitím exaktního Fisherova faktoriálového testu při malém rozsahu výběru.

U každého testu si ukážeme, jak testovat hypotézy pomocí příkazů v softwaru RStudio. Rozhodnutí o zamítnutí či nezamítnutí hypotézy provedeme na základě p-hodnoty. Jedná se o číslo z intervalu $\langle 0{,}1\rangle$, které udává nejmenší hladinu významnosti testu, při které bychom ještě H_0 zamítli. Nulovou hypotézu H_0 zamítáme v případech, kdy p-hodnota je menší nebo rovna číslu $\alpha \in \langle 0{,}1\rangle$, které určuje hladinu významnosti testu. Obvykle se používá testování na hladině α=0,05. Číslo $1-\alpha$ nazýváme koeficient spolehlivosti.

Jako zdroj informací nám poslouží publikace a práce [2], [3], [4], [5], [6], [7].

2.1. Kontingenční tabulky

Zkoumáme vztah dvou kvalitativních znaků, které si označíme jako X a Y. K dispozici máme dvourozměrný náhodný vektor $(X, Y)'$ s náhodnými složkami X a Y, které nabývají hodnot 1,…,r a 1,…, s. Symbolem n_{ij} označujeme četnost jevu $(X{=}i,\ Y{=}j)$ při provedení dvourozměrného náhodného výběru $(X_1, Y_1)', \dots,$ $(X_n, Y_n)'$, který přísluší náhodnému vektoru $(X, Y)'$. Marginální četnosti označíme

$$n_{i.} = \sum_{j=1}^{s} n_{ij}, \quad n_{.j} = \sum_{i=1}^{r} n_{ij}$$

a pro celkovou četnost platí

$$n = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{s} n_{.j} = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}.$$

Tabulka četností neboli kontingenční tabulka potom vypadá takto:

$X \backslash Y$	1	2	3	$\dots$	s	$\sum$
1	n_{11}	n_{12}	n_{13}	$\dots$	n_{1s}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	$\dots$	n_{2s}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
r	n_{r1}	n_{r2}	n_{r3}	$\dots$	n_{rs}	$n_{r.}$
$\sum$	$n_{.1}$	$n_{.2}$	$n_{.3}$	$\dots$	$n_{.s}$	n

2.2. Chí-kvadrát test nezávislosti

Nezávislost složek X a Y testujeme nejčastěji v kontingenční tabulce s využitím testovací statistiky

$$Z = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{\frac{n_{i.} n_{.j}}{n}}. \tag{2.1}$$

Pokud každý ze sledovaných znaků nabývá pouze dvou hodnot, kontingenční tabulku nazýváme jako *čtyřpolní* a vypadá následovně:

$X \backslash Y$	1	2	$\sum$
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
$\sum$	$n_{.1}$	$n_{.2}$	n

Po úpravě testovací statistiky (2.1) dostaneme tvar

$$Z = n\frac{\left(n_{11}n_{22} - n_{12}n_{21}\right)^2}{n_{1.}n_{2.}n_{.1}n_{.2}}.$$

Abychom mohli využít testovací statistiku Z, je potřeba splnění několika podmínek:

- Musíme mít velký soubor (dostatečně velké n).
- Pro četnosti musí platit: $n\widehat{p}_{ij} = \frac{n_{i.}n_{.j}}{n} \geq 5$, pro $\forall i, j$.

Test nulové hypotézy H_0 provádíme proti alternativě H_A

- H_0: veličiny X a Y jsou nezávislé.
- H_A: veličiny X a Y nejsou nezávislé.

Za platnosti nulové hypotézy H_0 má testovací statistika Z asymptoticky pro $n \to \infty$ rozdělení χ^2 s $(r-1)(s-1)$ stupni volnosti. Nulovou hypotézu zamítáme v případech, kdy platí: $z \geq \chi^2_{(r-1)(s-1),1-\alpha}$, tj. v situacích, v nichž se testová statistika Z realizuje v kritickém oboru $W = \langle \chi^2_{(r-1)(s-1),1-\alpha}, \infty)$.

Pro testování nezávislosti v programu RStudio použijeme příkaz **chisq.test()** a o výsledku testu rozhodneme na základě p-hodnoty.

Jestliže máme tabulku s menšími četnostmi a pokud má testová statistika pouze jeden stupeň volnosti (tj. kontingenční tabulka je čtyřpolní), pak ke zjištění hodnoty testové statistiky Z využijeme Yatesovu korekci na spojitost. Tuto korekci děláme z důvodu, že očekávané četnosti jsou většinou neceločíselné. V každém políčku tabulky odečteme hodnotu 0,5 z absolutní hodnoty rozdílu pozorované a očekávané četnosti a úpravou vztahu (2.1) vypočteme hodnotu testové statistiky Z s Yatesovou korekcí na spojitost ve čtyřpolní tabulce

$$Z = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(|n_{ij} - \frac{n_{i.}n_{.j}}{n}| - 0,5\right)^2}{\frac{n_{i.}n_{.j}}{n}} = n\frac{\left(|n_{11}n_{22} - n_{12}n_{21}| - \frac{n}{2}\right)^2}{n_{1.}n_{2.}n_{.1}n_{.2}}.$$

2.3. Fisherův faktoriálový test

Máme-li čtyřpolní kontingenční tabulku, ve které nejsou splněny podmínky pro četnosti $n\widehat{p}_{ij} = \frac{n_{i.}n_{.j}}{n} \geq 5$, pro $\forall i, j$, pak používáme pro testování nezávislosti znaků X a Y Fisherův faktoriálový test. Na rozdíl od chí-kvadrát (χ^2) testu nezávislosti jej můžeme použít pro výběry malého rozsahu.

Při testování vycházíme ze čtyřpolní tabulky

$X\backslash Y$	1	2	$\sum$
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
$\sum$	$n_{.1}$	$n_{.2}$	n

a z výpočtu podmíněné pravděpodobnosti p toho, že při daných marginálních četnostech $n_{1.}$, $n_{2.}$, $n_{.1}$, $n_{.2}$ a za platnosti hypotézy H_0 o nezávislosti znaků X a Y získáme tabulku s četnostmi n_{11}, n_{12}, n_{21}, n_{22}.

Nejprve zjistíme, kolika způsoby můžeme roztřídit soubor o n prvcích do čtyřpolních tabulek, aby všechny vzniklé tabulky měly marginální četnosti $n_{1.}$, $n_{2.}$, $n_{.1}$, $n_{.2}$. Soubor n prvků můžeme podle prvního znaku rozřadit na skupiny s četnostmi $n_{1.}$ a $n_{2.}$ a jejich počet vyjadřuje kombinační číslo

$$\binom{n}{n_{1.}} = \frac{n!}{n_{1.}!(n-n_{1.})!} = \frac{n!}{n_{1.}!n_{2.}!}.$$

Podle druhého znaku můžeme n-prvkový soubor rozlišit na skupiny o četnostech $n_{.1}$ a $n_{.2}$

$$\binom{n}{n_{.1}} = \frac{n!}{n_{.1}!(n-n_{.1})!} = \frac{n!}{n_{.1}!n_{.2}!}$$

možnými způsoby. Vzhledem k tomu, že vycházíme z předpokladu nezávislosti znaků, tak se každá kombinace prvního druhu může vyskytnout s každou kombinací druhého druhu. Celkem máme tedy

$$\binom{n}{n_{1.}}\binom{n}{n_{.1}} = \frac{n!n!}{n_{1.}!n_{2.}!n_{.1}!n_{.2}!} \tag{2.2}$$

možností, jak roztřídit n prvků do tabulky s marginálními četnostmi $n_{1.}$, $n_{2.}$, $n_{.1}$, $n_{.2}$.

Poté zjistíme, kolika možnými způsoby se v tabulce mohou realizovat četnosti n_{11}, n_{12}, n_{21}, n_{22}. Pro četnost n_{11} (první políčko) máme

$$\binom{n}{n_{11}} = \frac{n!}{n_{11}!(n-n_{11})!}$$

způsobů. Ve druhém políčku rozdělujeme zbylých $(n - n_{11})$ prvků, takže pro četnost n_{12} zbývá

$$\binom{n-n_{11}}{n_{12}} = \frac{(n-n_{11})!}{n_{12}!(n-n_{11}-n_{12})!}$$

možností. K rozdělení do posledních dvou políček máme k dispozici $(n-n_{11}-n_{12})$ prvků. Četnost n_{21} (třetí políčko) se může realizovat

$$\binom{n-n_{11}-n_{12}}{n_{21}} = \frac{(n-n_{11}-n_{12})!}{n_{21}!(n-n_{11}-n_{12}-n_{21})!} = \frac{(n-n_{11}-n_{12})!}{n_{21}!n_{22}!}$$

možnými způsoby. Pro poslední četnost n_{22} (čtvrté políčko) existuje už jen jedna možnost.

Celkový počet možností, jak můžeme n prvků rozdělit do všech políček čtyřpolní kontingenční tabulky s četnostmi n_{11}, n_{12}, n_{21}, n_{22}, je roven

$$\binom{n}{n_{11}}\binom{n-n_{11}}{n_{12}}\binom{n-n_{11}-n_{12}}{n_{21}} = \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}. \tag{2.3}$$

Hledanou pravděpodobnost p dostaneme jako podíl (2.2) a (2.3)

$$p = \frac{n_{1.}!n_{2.}!n_{.1}!n_{.2}!}{n!n_{11}!n_{12}!n_{21}!n_{22}!}.$$

Test nulové hypotézy H_0 provádíme proti alternativě H_A

- H_0: veličiny X a Y jsou nezávislé.
- H_A: veličiny X a Y nejsou nezávislé.

Fisherův test může mít podobu jednostranného nebo oboustranného testu. Při jednostranném Fisherově testu nejdříve spočítáme pravděpodobnost p pro danou tabulku. Následně snižujeme nejmenší četnost o jedničku při zachování marginálních četností a počítáme pravděpodobnosti p pro všechny nově vzniklé tabulky. Všechny vypočítané pravděpodobnosti sečteme a pokud je součet menší nebo roven α, tak zamítáme nulovou hypotézu na hladině (nejvýše) α. Nicméně tímto testem můžeme zkoumat možnou závislost jen v jednom směru.

Ve Fisherově oboustranném testu nejdříve postupujeme jako v testu jednostranném. Vypočítáme pravděpodobnost p pro danou tabulku a poté i pro všechny tabulky, které vzniknou snižováním nejmenší četnosti o jedničku při zachování marginálních četností. Následně v políčku s nejmenší četností porovnáme marginální četnosti sloupce a řádku, které se v tomto políčku protínají. Vybereme ten řádek nebo sloupec s nižší marginální četností, v něm prohodíme pozice původních četností, dopočítáme zbylé četnosti a vypočítáme pravděpodobnost pro nově vzniklou tabulku. Nejmenší četnost v tabulce snižujeme o jedničku při zachování marginálních četností a spočítáme pravděpodobnosti pro všechny tabulky. Jestliže je součet všech vypočítaných pravděpodobností menší nebo roven α, pak zamítáme hypotézu o nezávislosti na hladině (nejvýše) α.

Pro testování nezávislosti v programu RStudio použijeme příkaz **fisher.test()** a o výsledku testu rozhodneme na základě p-hodnoty.

2.4. Hypotéza nezávislosti a poměr šancí

Nulovou hypotézu můžeme také formulovat pomocí **poměru šancí (odds ratio, OR)**. Šanci definujeme jako podíl dvou pravděpodobností, a to pravděpodobnosti výskytu zkoumaného jevu A (P_A) a pravděpodobnosti jeho doplňku ($1 - P_A$). Šance (odds) výskytu zkoumaného jevu se vypočítá ze vztahu

$$šance = \frac{P_A}{1 - P_A}.$$

Uvažujme dvě náhodné veličiny X a Y (kvalitativní znaky), které nabývají dvou různých hodnot. Pro přehlednost vyjádříme hodnoty čísly jedna a dva a ukážeme, jak vypadá kontingenční tabulka:

	Y=1	Y=2
X=1		
X=2		

Chtěli bychom zjistit šanci, že X nabývá hodnoty jedna, když také Y je rovno jedné. Podle definice šance zjistíme tuto šanci z podílu (podmíněných) pravděpodobností zkoumaného jevu a jeho doplňku $\mathrm{P}(X = 1|Y = 1)/\mathrm{P}(X = 2|Y = 1)$. Zajímá nás také šance, že X opět nabývá hodnoty jedna, ale veličina Y se rovná dvěma. Stejně jako v předchozím případu vyjádříme šanci podílem pravděpodobností $\mathrm{P}(X = 1|Y = 2)/\mathrm{P}(X = 2|Y = 2)$. Podíl těchto dvou šancí nazýváme poměrem šancí a označujeme jej OR (odds ratio).

Bodový odhad poměru šancí dostaneme z odhadu pravděpodobností s využitím relativních četností. Po úpravě dostaneme následující vztah

$$\frac{šance(X = 1|Y = 1)}{šance(X = 1|Y = 2)} = \frac{\widehat{P}(X = 1|Y = 1)/\widehat{P}(X = 2|Y = 1)}{\widehat{P}(X = 1|Y = 2)/\widehat{P}(X = 2|Y = 2)} = \frac{n_{11} : n_{21}}{n_{12} : n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}.$$

Také by nás mohla zajímat šance, že veličina Y nabývá hodnoty jedna, když i X se rovná jedné, a šance, že Y je rovno jedné, když X nabývá hodnoty dva. S využitím předchozího případu platí pro bodový odhad poměru šancí vztah

$$\frac{šance(Y = 1|X = 1)}{šance(Y = 1|X = 2)} = \frac{\widehat{P}(Y = 1|X = 1)/\widehat{P}(Y = 2|X = 1)}{\widehat{P}(Y = 1|X = 2)/\widehat{P}(Y = 2|X = 2)} = \frac{n_{11} : n_{12}}{n_{21} : n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}},$$

který je úplně stejný s předchozím vyjádřením bodového odhadu poměru šancí.

Testujeme hypotézu H_0: OR=1 a alternativou je jedna ze tří možností:

- Poměr šancí je menší než jedna (OR<1).
- Poměr šancí je větší než jedna (OR>1).
- Poměr šancí není roven jedné (OR≠1).

Při použití jednostranného Fisherova faktoriálového testu s alternativní hypotézou OR<1 píšeme do RStudia příkaz *fisher.test(,alternative="less")*. K testování jednostranným Fisherovým faktoriálovým testem s alternativní hypotézou OR>1 používáme příkaz *fisher.test(,alternative="greater")* a pro oboustranný Fisherův test s alternativní hypotézou OR≠1 zadáme v programu RStudio příkaz *fisher.test(,alternative="two.sided")*.

Chí-kvadrát test nezávislosti je na rozdíl od Fisherova faktoriálového testu pouze oboustranný. Abychom mohli p-hodnoty obdržené Fisherovým testem s jednostrannou alternativou a chí-kvadrát testem správně porovnat a zhodnotit, musíme p-hodnoty získané použitím chí-kvadrát testu upravit. K tomu využijeme rezidua všech políček čtyřpolní kontingenční tabulky, která můžeme vyjádřit vztahem

$$e_{ij} = \frac{n_{ij} - \widehat{\mu}_{ij}}{\sqrt{\widehat{\mu}_{ij}}} = \frac{n_{ij} - \frac{n_{i.}n_{.j}}{n}}{\sqrt{\frac{n_{i.}n_{.j}}{n}}}, \quad i, j = 1, 2.$$

Ve čtyřpolních tabulkách jsou znaménka reziduí uspořádána jednou z následujících možností:

- Dostaneme-li při provedení chí-kvadrát testu kladná rezidua v políčkách n_{11} a n_{22}, pak polovina p-hodnoty tohoto testu přibližně odpovídá p-hodnotě jednostranného Fisherova faktoriálového testu s alternativou OR>1 (*alternative="greater"*).

+	-
-	+

- Naopak obdržíme-li při provedení chí-kvadrát testu v políčkách n_{11} a n_{22} záporná rezidua, pak polovina p-hodnoty tohoto testu přibližně odpovídá p-hodnotě jednostranného Fisherova faktoriálového testu s alternativou OR<1 (*alternative="less"*).

-	+
+	-

V práci budeme p-hodnoty při provedení chí-kvadrát testu značit následovně:

- Horní index ^{G} – původní p-hodnota je snížená na polovinu a odpovídá p-hodnotě jednostranného Fisherova testu s alternativou *"greater"*.
- Horní index ^{L} – původní p-hodnota je snížená na polovinu a odpovídá p-hodnotě jednostranného Fisherova testu s alternativou *"less"*.
- Horní index ^{T} – původní p-hodnota je ponechána a odpovídá p-hodnotě oboustranného Fisherova testu s alternativou *"two.sided"*.

Kapitola 3

Analýza hokejových utkání – testování různých závislostí

V této kapitole popíšeme naši práci s daty v programu RStudio a budeme se zabývat testováním hypotéz o nezávislosti výsledků dvou a více po sobě jdoucích utkání (kapitoly 3.1, 3.2 a 3.3). Zaměříme se také na testování hypotéz o nezávislosti výhry (resp. prohry) na počtu vstřelených gólů (kapitola 3.4) a budeme se věnovat i testování hypotéz o nezávislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (kapitoly 3.5 a 3.6).

K testování hypotéz o nezávislosti použijeme především Fisherův faktoriálový test (jednostranný i oboustranný) a při splnění podmínek četnosti také chí-kvadrát test nezávislosti. Testy provedeme zvlášť na datech pro sezónu 2016/2017 a pro sezónu 2017/2018. Ve všech případech hledáme p-hodnoty menší než 0,05, díky kterým zamítneme hypotézu nezávislosti, a zjištěnou závislost pak přesně zformulujeme.

3.1. Závislost mezi výsledky dvou po sobě jdoucích utkání

Zajímá nás, zda mezi výsledky dvou po sobě jdoucích utkání existuje závislost. Chceme zjistit, jestli týmy mají po výhře tendenci vyhrát a po prohře prohrát. Zaměříme se nejen na všechny zápasy základní části, ale také na domácí a venkovní utkání zvlášť. Pravděpodobnost vítězství je pro domácí a venkovní zápasy odlišná. Při domácích duelech hraje tým stále ve stejných podmínkách, tj. hraje na stejném kluzišti a mezi stejnými diváky, a má tedy vyšší pravděpodobnost vítězství než v duelech venkovních, kdy musí cestovat napříč Českou republikou a hrát na kluzištích různých rozměrů a při velmi rozdílných zvukových projevech diváků.

V programu RStudio se zabýváme sloupcem s výsledky utkání (např. **vyhra_kom** pro tým HC Kometa Brno) a pro dvojice po sobě jdoucích utkání zjišťujeme počet případů, kdy sledovaný tým vyhrál, resp. prohrál, a následně vyhrál či prohrál. V základní části TELH odehraje každý z týmů 52 zápasů (26 doma a 26 venku). Jelikož po posledním utkání už žádné nenásleduje, máme k dispozici 51 dvojic pro všechna utkání a po 25 dvojicích pro domácí a venkovní zápasy. Dvojice zápasů rozdělíme do čtyřpolní kontingenční tabulky následovně:

	výhra	prohra
výhra		
prohra		

Levé horní políčko značí počet dvojic zápasů, kdy tým vyhrál dvakrát po sobě, a v pravém horním políčku je počet dvojic zápasů, ve kterých tým nejdříve vyhrál a následně prohrál. Levé dolní políčko udává počet dvojic zápasů, v nichž tým nejdříve prohrál a pak vyhrál, a pravé dolní políčko zobrazuje počet dvojic zápasů, ve kterých tým prohrál dvakrát po sobě.

Vzhledem k tomu, že nás zajímá, zda týmy po výhře zase vyhrají, případně po prohře znovu prohrají (tzn. vyšší hodnoty v políčkách n_{11} a n_{22}), tak při použití Fisherova faktoriálového testu použijeme jednostrannou variantu s alternativní hypotézou OR>1 (*alternative="greater"*). Pro závislost mezi výsledky dvou po sobě jdoucích utkání svědčí p-hodnoty menší než 0,05.

Všechna utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	$0{,}197^G$	$0{,}365^L$	$0{,}420^L$	$0{,}210^L$	$0{,}232^L$	$0{,}445^L$	$0{,}428^L$
2017/2018	$0{,}102^L$	$0{,}466^L$	$0{,}371^G$	$0{,}315^G$	0,297	$0{,}262^G$	$0{,}500^L$
	VIT	**PLZ**	**MBL**	**ZLN**	**OLO**	**KVA**	**PCE**
2016/2017	$0{,}102^L$	$0{,}351^G$	$0{,}210^L$	$0{,}441^L$	$0{,}500^G$	0,989	0,657
2017/2018	$0{,}374^L$	0,537	$0{,}334^L$	$0{,}445^L$	$0{,}336^L$	–	$\mathbf{0{,}018}^L$

Tabulka 3.1: P-hodnoty testů nezávislosti (chí-kvadrát testu, pokud je v buňce index ^{L} nebo ^{G} – viz kapitola 2.4, nebo jednostranného Fisherova faktoriálového testu s alternativou "greater") k určení závislosti mezi výsledky dvou po sobě jdoucích utkání – všechna utkání základní části

Tabulka (3.1) ukazuje p-hodnoty obdržené testy nezávislosti pro všechny zápasy základní části sezón 2016/2017 a 2017/2018. Týmy jsou seřazené podle umístění po skončení základní části sezóny 2016/2017. Horní část tabulky zahrnuje sedm nejlepších týmů uspořádaných vzestupně podle pořadí (Bílí Tygři Liberec nejlepší), v dolní části je seřazeno sedm nejhorších týmů (HC Dynamo Pardubice nejhorší). U týmu HC Energie Karlovy Vary máme p-hodnotu pouze pro sezónu 2016/2017, neboť na konci této sezóny spadli Energetici do nižší ligy.

Ve většině případů jsme měli splněnou podmínku četnosti pro užití chí-kvadrát testu nezávislosti, pouze pro Karlovy Vary a Pardubice v sezóně 2016/2017 a pro Litvínov s Plzní v sezóně 2017/2018 jsme použili jednostrannou variantu Fisherova faktoriálového testu s alternativou "greater".

Hypotézu nezávislosti na hladině α=0,05 zamítáme pouze u týmu Pardubic v sezóně 2017/2018. Nicméně zjištěná p-hodnota 0,018 je rovna polovině p-hodnoty při provedení chí-kvadrát testu a odpovídá jednostrannému Fisherově faktoriálovému testu s alternativou OR<1. Našli jsme tedy opačnou závislost, než jsme hledali. Na základě tohoto zjištění můžeme říct, že Pardubice v sezóně 2017/2018 po výhře následující utkání často prohrály a po prohře v příštím utkání většinou zvítězily.

Ostatní p-hodnoty jsou vyšší než hodnota α=0,05, proto nemůžeme nulovou hypotézu zamítnout.

Domácí utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,895	0,847	0,736	0,289	0,343	0,834	0,834
2017/2018	1,000	0,222	0,582	0,847	0,138	0,895	0,582
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,988	0,987	0,919	0,287	0,430	0,769	1,000
2017/2018	0,893	1,000	0,919	0,418	0,942	–	0,659

Tabulka 3.2: P-hodnoty jednostranného Fisherova faktoriálového testu nezávislosti s alternativou "greater" k určení závislosti mezi výsledky dvou po sobě jdoucích utkání – domácí utkání

Podíváme se nyní pouze na domácí utkání v obou sezónách. Všechny p-hodnoty uvedené v tabulce (3.2) jsme získali provedením Fisherova faktoriálového testu při alternativě "greater". Je patrné, že každá p-hodnota je větší než hodnota α=0,05, takže nelze zamítnout nulovou hypotézu nezávislosti. Pouze u Litvínova v sezóně 2017/2018 se p-hodnota nachází celkem blízko k hodnotě α=0,05, avšak stále je příliš vysoká, abychom mohli hypotézu nezávislosti zamítnout.

Venkovní utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,163	0,723	0,659	0,769	0,987	0,393	0,893
2017/2018	1,000	0,862	0,988	0,989	0,066	0,418	**0,010**
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,984	0,769	0,824	1,000	0,138	1,000	0,934
2017/2018	0,418	0,081	0,834	0,999	0,296	–	0,407

Tabulka 3.3: P-hodnoty jednostranného Fisherova faktoriálového testu nezávislosti s alternativou "greater" k určení závislosti mezi výsledky dvou po sobě jdoucích utkání – venkovní utkání

Přesuneme se k venkovním utkáním. Stejně jako u domácích zápasů jsme k testování využili pouze Fisherův faktoriálový test při alternativě "greater" a výsledné p-hodnoty nalezneme v tabulce (3.3). Vidíme, že v sezóně 2017/2018 je p-hodnota u chomutovských Pirátů rovna číslu 0,01, a proto zamítáme nulovou hypotézu nezávislosti. Chomutovský tým tedy při venkovních zápasech po výhře v následujícím utkání většinou zvítězil a po prohře odcházel v dalším zápasu spíše poražen.

P-hodnota u týmu Litvínova (0,066) a Plzně (0,081) v sezóně 2017/2018 se blíží k hodnotě 0,05, nicméně nulovou hypotézu nezamítáme. Ostatní p-hodnoty přesahují hodnotu α=0,05 výrazněji, proto také v těchto případech nelze nulovou hypotézu zamítnout.

3.2. Závislost mezi sérií dvou výher/proher a výsledkem příštího utkání

Hokejové týmy v průběhu sezóny střídají různě dlouhé série vyhraných a prohraných utkání. Nejdříve se zaměříme na kratší série. Chtěli bychom zjistit, zda tým, který dvakrát po sobě vyhrál, znovu vyhraje, a jestli po dvou prohraných zápasech následuje další prohra. Kvůli malému množství údajů pro domácí a venkovní utkání se budeme zabývat pouze všemi zápasy základní části.

V RStudiu postupujeme podobně jako v předchozím případě. Ze sloupce výsledků utkání vybereme z celé základní části dvojice po sobě jdoucích zápasů, které skončily pro daný tým stejným výsledkem, tedy ty případy, kdy tým dvakrát po sobě vyhrál, respektive prohrál. Nejméně případů, kdy tým dvakrát po sobě vyhrál, nastalo u Karlových Varů (1 případ) v sezóně 2016/2017, nejvíce případů pak u Plzně (23 případů) v sezóně 2017/2018. Počet případů dvou proher v řade se pohybuje od pěti pro Litvínov v sezóně 2016/2017 až po 26 pro Karlovy Vary a Pardubice v sezóně 2016/2017. Následně zjistíme počet případů, v nichž po sérii následovala výhra či prohra. Vytvoříme si následující kontingenční tabulku:

	výhra	prohra
výhra 2×		
prohra 2×		

Levé horní políčko značí případy, kdy tým třikrát po sobě vyhrál, a pravé horní políčko udává, kolikrát tým po dvou výhrách prohrál. V levém dolním políčku je zaznamenáno, kolikrát po dvou prohrách následovala výhra, a v pravém dolním políčku jsou případy tří proher v řadě.

Naším cílem je zjistit, zda si týmy po dvou výhrách za sebou dokážou udržet vítěznou sérii, a naopak jestli při dvou prohraných zápasech v řadě prohrají i potřetí (tzn. vyšší hodnoty v políčkách n_{11} a n_{22}). Kvůli malým hodnotám v kontingenční tabulce nejsou splněny podmínky pro užití chí-kvadrát testu nezávislosti, proto použijeme pouze jednostranný Fisherův faktoriálový test s alternativní hypotézou OR>1 (*alternative="greater"*). Pro závislost mezi výsledky tří po sobě jdoucích utkání svědčí p-hodnoty menší než 0,05.

Všechna utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,217	0,455	0,680	0,905	0,717	0,950	0,986
2017/2018	0,975	0,851	0,803	0,786	0,413	0,385	0,325
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,995	0,637	0,968	0,962	0,769	1,000	0,888
2017/2018	0,235	0,625	0,966	0,893	0,878	–	0,998

Tabulka 3.4: P-hodnoty jednostranného Fisherova faktoriálového testu nezávislosti s alternativou "greater"k určení závislosti mezi sérií dvou výher/proher a výsledkem příštího utkání – všechna utkání základní části

V tabulce (3.4) nalezneme získané p-hodnoty pro všechny týmy v obou sezónách při provedení jednostranného Fisherova faktoriálového testu s alternativou "greater". Pro všechny p-hodnoty platí, že jsou vyšší než hodnota α=0,05, a více než polovina z nich dokonce převyšuje hodnotu 0,7. Jak je patrno ze získaných hodnot, nemůžeme pro žádný z týmů zamítnout nulovou hypotézu, tj. nemůžeme zamítnout hypotézu, že mezi sérií dvou výher (resp. dvou proher) a výhrou (resp. prohrou) v následujícím utkání není závislost.

3.3. Závislost mezi sérií tří výher/proher a výsledkem příštího utkání

Naším zájmem budou nyní delší série vítězství a proher. Oproti předchozímu případu, kdy jsme uvažovali dvouzápasové série výher, respektive proher, a zkoumali výsledek následujícího utkání, teď rozšíříme série na tři vítězné, resp. prohrané, zápasy za sebou. Vítězných sérií bylo nejméně pro Karlovy Vary (0) v sezóně 2016/2017 a nejvíce pro Liberec (17) v sezóně 2016/2017. Sérií třech proher pak bylo nejméně pro tým Chomutova (1) v sezóně 2016/2017 a nejvíce pro Litvínov (20) v sezóně 2017/2018. Zajímá nás, jestli tým po třech výhrách po sobě opět vyhraje, a po třech prohraných utkáních bude znovu poražen. Stejně jako u předešlé série nemáme dostatečné množství údajů, abychom se zabývali domácími a venkovními zápasy zvlášť, proto pracujeme se souborem všech zápasů.

Postup v RStudiu se od předchozího případu liší pouze prvním krokem, kdy ze sloupce výsledků vybereme trojice po sobě jdoucích zápasů se shodným výsledkem, tedy trojice tří výher, resp. tří proher po sobě. Pak určíme, kolikrát po této sérii následovala výhra a kolikrát prohra, a vytvoříme kontingenční tabulku:

	výhra	prohra
výhra 3×		
prohra 3×		

V levém horním políčku je zaznamenáno, kolikrát tým vyhrál čtyřikrát po sobě, a pravé horní políčko značí počet případů, ve kterých po třech výhrách následovala prohra. V levém dolním políčku jsou případy tří proher v řadě, po nichž tým vyhrál, a pravé dolní políčko obsahuje série čtyř proher.

Zajímá nás, jestli jsou týmy natolik dobré, že po třech vítězstvích v řadě jsou schopné znovu vyhrát, a zda když třikrát po sobě prohrají, tak v dalším zápasu opět odejdou poražené (tzn. vyšší hodnoty v políčkách n_{11} a n_{22}). Nemáme dostatečné četnosti pro splnění podmínek k použití chí-kvadrát testu nezávislosti, proto použijeme pouze jednostranný Fisherův faktoriálový test s alternativou

OR>1 (*alternative="greater"*). Pro závislost mezi výsledky čtyř po sobě jdoucích utkání svědčí p-hodnoty menší než 0,05.

Všechna utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	0,736	0,841	1,000	1,000	1,000	1,000
2017/2018	1,000	0,455	1,000	1,000	1,000	0,330	0,287
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	1,000	0,992	0,629	1,000	0,788	1,000	1,000
2017/2018	0,706	1,000	0,833	0,976	1,000	–	1,000

Tabulka 3.5: P-hodnoty jednostranného Fisherova faktoriálového testu nezávislosti s alternativou "greater"k určení závislosti mezi sérií tří výher/proher a výsledkem příštího utkání – všechna utkání základní části

Při zkoumání závislosti mezi sérií tří výher (resp. tří proher) a výhrou (resp. prohrou) v následujícím utkání dostaneme s využitím Fisherova jednostranného faktoriálového testu při alternativě "greater"p-hodnoty uvedené v tabulce (3.5). Stejně jako u dvouzápasové série výher (resp. proher) jsou p-hodnoty výrazně vyšší než hodnota α=0,05. Na základě získaných p-hodnot tak nemůžeme zamítnout hypotézu nezávislosti.

3.4. Závislost výhry/prohry na počtu vstřelených gólů

Další částí našeho testování (ne)závislosti je zkoumání závislosti mezi počtem vstřelených gólů a výsledkem utkání (tj. zda tým vyhrál nebo prohrál). Chtěli bychom zjistit, jestli při větším počtu vstřelených gólů tým vyhraje, a naopak při menším počtu vstřelených gólů je tým poražen. Rozebereme nejen všechna utkání základní části, ale také domácí a venkovní utkání zvlášť.

S použitím programu RStudio zjistíme počet zápasů, ve kterých tým vstřelil určitý počet gólů a vyhrál, respektive prohrál. Nejmenší počet vstřelených gólů je nula a nejvyšší počet je dán maximálním počtem vstřelených gólů. Vytvoříme tabulku, která má v řádcích počet vstřelených gólů, a sloupce označují výsledek utkání (výhra nebo prohra). Tato tabulka mnohokrát nesplňuje podmínky četností pro použití chí-kvadrát testu nezávislosti a navíc není ani čtyřpolní, takže nemůžeme použít Fisherův test nezávislosti. Z tohoto důvodu sloučíme řádky do dvou skupin a vytvoříme čtyřpolní kontingenční tabulku.

Hranici skupin vytvoříme mezi dvěma a třemi vstřelenými góly, abychom zamezili tomu, že některé políčko bude obsahovat nulu, a tedy p-hodnota testů nezávislosti by se rovnala jedné. Žádný tým nemůže vyhrát bez vstřelení gólu a jen málo týmů dokáže vyhrát při jednom vstřeleném gólu, ale při vstřelení dvou gólů už je schopný vyhrát každý. Naopak při vstřelení třech gólů všechny týmy prohrají alespoň jednou, kdežto po čtyřech vstřelených brankách týmy většinou neprohrají.

	výhra	prohra
≤ 2 góly		
≥ 3 góly		

Levé horní políčko udává počet zápasů, ve kterých tým vstřelil dva a méně gólů a vyhrál, pravé horní políčko pak počet zápasů, kdy tým prohrál po vstřelení maximálně dvou gólů. V levém dolním políčku je zaznamenán počet vyhraných zápasů po třech a více vstřelených gólech. Pravé dolní políčko zahrnuje počet prohraných zápasů, v nichž tým skóroval alespoň třikrát.

Zajímá nás, jestli týmy vyhrají, když vstřelí tři a více gólů, a prohrají, pokud vstřelí dva a méně gólů (tzn. vyšší hodnoty v políčkách n_{12} a n_{21}), proto při použití Fisherova faktoriálového testu pracujeme s jeho jednostrannou variantou s alternativou OR<1 (*alternative="less"*). Závislost mezi počtem vstřelených gólů a výsledkem utkání potvrdíme, pokud p-hodnoty jsou menší než 0,05.

	všechna utkání		domácí utkání		venkovní utkání	
	2016/17	2017/18	2016/17	2017/18	2016/17	2017/18
LIB	**0,0008**^{L}	**<0,0001**^{L}	0,1891	**0,0003**	**0,0006**	**0,0047**
TRI	**0,0224**^{L}	**<0,0001**^{L}	0,1119	**0,0048**	0,2142	**0,0089**
SPA	**<0,0001**^{L}	**<0,0001**^{L}	**0,0010**	**0,0027**	**0,0002**	**0,0006**
MHK	**<0,0001**^{L}	**<0,0001**^{L}	**0,0008**	**0,0053**	**0,0005**	**0,0016**
LIT	**0,0002**^{L}	**0,0026**^{L}	0,1826	**0,0425**	**0,0005**	**0,0209**
KOM	**<0,0001**^{L}	**0,0003**^{L}	**<0,0001**	**0,0034**	**0,0151**	0,0554
CHM	**<0,0001**^{L}	**<0,0001**^{L}	**0,0022**	**0,0089**	**0,0023**	**0,0056**
VIT	**<0,0001**^{L}	**<0,0001**^{L}	**0,0008**	**0,0022**	0,1077	**0,0085**^{L}
PLZ	**0,0003**^{L}	**<0,0001**^{L}	0,0554	**0,0278**	**0,0027**	**0,0024**
MBL	**<0,0001**^{L}	**<0,0001**^{L}	**0,0236**^{L}	**0,0001**	**0,0006**	**0,0033**
ZLN	**<0,0001**^{L}	**<0,0001**^{L}	**0,0016**	**0,0024**^{L}	**0,0001**	**0,0077**
OLO	**0,0001**^{L}	**<0,0001**^{L}	**0,0016**	**0,0001**	**0,0317**	**0,0001**
KVA	**<0,0001**	–	**0,0002**^{L}	–	**0,0092**	–
PCE	**<0,0001**^{L}	**0,0001**^{L}	**0,0074**	**0,0077**	**0,0002**	**0,0174**

Tabulka 3.6: P-hodnoty testů nezávislosti (chí-kvadrát testu, pokud je v buňce index ^{L} – viz kapitola 2.4, nebo jednostranného Fisherova faktoriálového testu s alternativou "less") k určení závislosti výhry/prohry na počtu vstřelených gólů

Tabulka (3.6) obsahuje p-hodnoty, které jsme získali z testů nezávislosti pro všechna utkání základní části a také pro domácí a venkovní utkání v sezónách 2016/2017 a 2017/2018. Týmy jsou uspořádané podle pořadí v tabulce po skončení základní sezóny 2016/2017. První dva sloupce tabulky se týkají všech utkání, další dva zahrnují domácí utkání a zbylé dva pouze zápasy venkovní. Pro testování závislosti ve skupině všech utkání byly vždy splněny podmínky četností k použití chí-kvadrát testu nezávislosti. Pouze pro tým Karlových Varů jsme využili Fisherův faktoriálový test. U domácích a venkovních utkání jsme pracovali především s Fisherovým faktoriálovým testem.

Pro všechny týmy jsme v rámci skupiny všech zápasů obdrželi p-hodnoty mnohem menší než je hodnota α=0,05, a to pro obě sezóny. Zamítáme tedy hypotézu nezávislosti a můžeme říct, že týmy při vstřelení nejvýše dvou gólů zápas spíše prohrály a při vstřelení alespoň tří gólů v utkání většinou zvítězily. Potvrdila se nám závislost, kterou jsme očekávali: Jestliže se týmům podaří vstřelit větší počet gólů, pak utkání vyhrají, a pokud soupeřovu brankáři dají méně gólů, skončí pro ně zápas porážkou. Výhra (resp. prohra) hokejových týmů tedy závisí na počtu vstřelených gólů.

V domácích zápasech sezóny 2016/2017 je p-hodnota u deseti týmů menší než hodnota α=0,05, proto v těchto případech zamítáme nulovou hypotézu. Tyto týmy v domácím prostředí vyhrávaly, pokud vstřelily alespoň tři góly, a prohrávaly, jestliže dokázaly dát nejvýše dva góly. U čtyř týmů (Liberec, Třinec, Litvínov a Plzeň) je v sezóně 2016/2017 p-hodnota testů nezávislosti vyšší než 0,05, takže není možné hypotézu nezávislosti zamítnout. V následující sezóně 2017/2018 hypotézu nezávislosti zamítáme u 13 týmů. Posledním týmem je Jihlava, která v sezóně 2017/2018 nahradila tým Karlových Varů, a v našich testech s ní nepracujeme.

U venkovních zápasů sezóny 2016/2017 pouze u Třince a Vítkovic dostaneme p-hodnoty přesahující hodnotu α=0,05, a tedy hypotézu nezávislosti nezamítáme. P-hodnoty u zbylých 12 týmů nepřekračují hodnotu 0,05, a proto zamítáme hypotézu nezávislosti. Tuto hypotézu zamítneme také pro všechny týmy v sezóně 2017/2018 s výjimkou brněnské Komety.

3.5. Závislost počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (0, nebo 1 a více gólů)

Týmy během sezóny střílí různý počet gólů. My se zaměříme na počet vstřelených gólů ve dvou po sobě jdoucích utkáních. Zajímá nás, zda existuje závislost mezi počtem vstřelených gólů v předchozím a následujícím zápasu. Nejdříve se zaměříme na situaci, kdy týmy v zápasech dají nula, nebo jeden a více gólů. Podíváme se jak na všechna utkání základní části, tak na domácí a venkovní utkání zvlášť.

Postup v Rku je následující: Pracujeme se sloupcem vstřelených gólů daného týmu (např. **goly_kom** pro tým HC Kometa Brno) a zjišťujeme, v kolika případech po vstřelení žádného, respektive jednoho a více gólů, tým nedá gól, nebo vstřelí jeden a více gólů. Pro domácí utkání využijeme sloupec **goly_d** a pro utkání venkovní sloupec **goly_h**. V základní části každý z týmů odehraje 52 zápasů (26 doma a 26 venku) a protože po posledním utkání už žádné nenásleduje, máme k dispozici 51 dvojic zápasů pro základní část a 25 dvojic pro domácí a venkovní utkání. Z těchto dvojic vytvoříme čtyřpolní kontingenční tabulku:

	0 gólů	≥ 1 gól
0 gólů		
≥ 1 gól		

Levé horní políčko tabulky značí počet dvojic zápasů, v nichž tým ani jednou nedal gól. Pravé horní políčko udává dvojice zápasů, kdy tým nejdříve neskóroval a v dalším zápase vstřelil jeden a více gólů. V levém dolním políčku je počet dvojic zápasů, ve kterých tým nejprve vstřelil alespoň jeden gól a následně se gólově neprosadil. V pravém dolním políčku jsou zaznamenány dvojice zápasů, kdy tým v obou zápasech vstřelil alespoň jeden gól.

Zajímá nás, zda mezi počtem vstřelených gólů ve dvou po sobě jdoucích utkáních existuje nějaká závislost. Kvůli nesplněným podmínkám četností pro použití chí-kvadrát testu pracujeme pouze s oboustranným Fisherovým faktoriálovým testem. Pro závislost mezi počtem vstřelených gólů ve dvou po sobě jdoucích utkáních svědčí p-hodnoty menší než 0,05.

Všechna utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	1,000	1,000	1,000	1,000	1,000	0,286
2017/2018	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,332	1,000	1,000	0,417	0,078	0,389	0,170
2017/2018	0,417	1,000	0,417	1,000	0,579	–	1,000

Tabulka 3.7: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (0, nebo 1 a více gólů) – všechna utkání základní části

V tabulce (3.7) můžeme vidět p-hodnoty, které jsme získali provedením testu nezávislosti pro soubor všech utkání základní části. Hypotézu nezávislosti nemůžeme zamítnout ani v jednom případu kvůli p-hodnotám převyšujícím hodnotu 0,05.

Domácí utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	1,000	1,000	1,000	1,000	1,000	1,000
2017/2018	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	1,000	0,330	1,000	1,000	1,000	0,330	1,000
2017/2018	1,000	1,000	1,000	0,080	1,000	–	1,000

Tabulka 3.8: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (0, nebo 1 a více gólů) – domácí utkání

Zaměříme-li se pouze na domácí zápasy, pak ani v jednom případu hypotézu nezávislosti nelze zamítnout. Pouze u zlínských Beranů dostáváme p-hodnotu blízkou hodnotě 0,05, avšak stále příliš vysokou pro zamítnutí nulové hypotézy. Všechny získané p-hodnoty jsou zaznamenány v tabulce (3.8).

Venkovní utkání

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	1,000	1,000	1,000	1,000	1,000	0,330
2017/2018	0,230	1,000	1,000	0,423	1,000	1,000	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,289	1,000	1,000	1,000	1,000	0,355	0,157
2017/2018	1,000	0,549	1,000	1,000	0,637	–	1,000

Tabulka 3.9: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (0, nebo 1 a více gólů) – venkovní utkání

Soubor venkovních utkání můžeme hodnotit stejně jako skupinu všech utkání i jako skupinu utkání domácích. Opět nám Fisherův test nezávislosti dal p-hodnoty vyšší než je hodnota pro zamítnutí nulové hypotézy, a proto na hladině α=0,05 nezamítáme hypotézu nezávislosti.

3.6. Závislost počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (ostatní případy)

Stejně jako v předešlém případu se zajímáme o počet vstřelených gólů ve dvou po sobě jdoucích zápasech a zkoumáme závislost mezi vstřelenými góly v předchozím a následujícím zápasu. Předchozí případ rozšíříme na další tři možné situace:

- 1 a méně, nebo 2 a více gólů
- 2 a méně, nebo 3 a více gólů
- 3 a méně, nebo 4 a více gólů

Postup v RStudiu je podobný jako u předchozího případu a znovu využijeme sloupec vstřelených gólů. Pro první uvedenou situaci (1 a méně, nebo 2 a více gólů) zjistíme, v kolika případech po vstřelení jednoho a méně gólů, respektive dvou a více gólů, tým následně dá jeden a méně gólů, nebo vsítí dva a více gólů. Pro druhou situaci (2 a méně, nebo 3 a více gólů) najdeme počet dvojic, ve kterých tým po vstřelení dvou a méně gólů, respektive tří a více gólů, v následujícím zápasu vstřelí dva a méně gólů, nebo dá tři a více gólů. Pro třetí situaci (3 a méně, nebo 4 a více gólů) zjistíme počet případů, kdy po vstřelení tří a méně gólů, respektive čtyř a více gólů, tým následně dá tři a méně gólů, nebo vsítí čtyři a více gólů. Označení sloupců a řádků v kontingenčních tabulkách se od předchozího případu liší pro každou z uvedených situací podle vstřelených gólů.

Zajímá nás, zda mezi počtem vstřelených gólů ve dvou po sobě jdoucích utkáních existuje nějaká závislost. Při použití Fisherova faktoriálového testu pracujeme s jeho oboustrannou variantou. Pro závislost mezi počtem vstřelených gólů ve dvou po sobě jdoucích utkáních svědčí p-hodnoty menší než 0,05.

Všechna utkání (1 a méně, nebo 2 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,185	1,000	0,573	1,000	0,730	1,000	0,309
2017/2018	0,503	0,178	0,418	0,322	0,746	0,651	1,000
	VIT	**PLZ**	**MBL**	**ZLN**	**OLO**	**KVA**	**PCE**
2016/2017	0,077	1,000	0,203	0,147^T	1,000	0,169^T	0,749
2017/2018	0,184	0,418	0,329	0,325	**0,041**	–	0,428

Tabulka 3.10: P-hodnoty testů nezávislosti (chí-kvadrát testu, pokud je v buňce index ^{T} – viz kapitola 2.4, nebo oboustranného Fisherova faktoriálového testu s alternativou "two.sided") k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (1 a méně, nebo 2 a více gólů) – všechna utkání základní části

Na základě získaných p-hodnot, které nalezneme v tabulce (3.10), zamítáme pro tým HC Olomouc v sezóně 2017/2018 hypotézu nezávislosti. Vzhledem k tomu, že při testování používáme oboustranný Fisherův faktoriálový test, tak nejsme schopni závislost přesně formulovat. V kontingenční tabulce olomouckého týmu provedeme také oba Fisherovy jednostranné testy.

	≤ 1 gól	≥ 2 góly
≤ 1 gól	1	13
≥ 2 góly	14	23

Pokud bychom testovali pomocí jednostranného Fisherova faktoriálového testu s alternativou OR>1, dostaneme p-hodnotu 0,997. Alternativa OR< 1 pak dává p-hodnotu přibližně 0,03.

Naše původní p-hodnota zhruba odpovídá jednostranné variantě Fisherova faktoriálového testu s alternativou OR< 1. Závislost lze formulovat následovně: Šance, že tým HC Olomouc vstřelí alespoň dva góly, je větší, pokud dal v předchozím utkání nejvýše jeden gól, než když vstřelil dva a více gólů. Také šance, že se olomoučtí hráči prosadí nejvýše jednou, je větší, jestliže v předchozím utkání skórovali alespoň dvakrát, než když dali jeden a méně gólů. Laicky bychom mohli říct, že gólová sklizeň přichází spíše po hubeném výsledku než po gólových hodech.

Zbylé p-hodnoty v tabulce (3.10) převyšují hodnotu 0,05, a proto hypotézu nezávislosti nezamítáme.

Domácí utkání (1 a méně, nebo 2 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	1,000	1,000	0,166	0,330	1,000	1,000
2017/2018	1,000	1,000	0,527	1,000	1,000	0,527	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	1,000	1,000	0,289	0,355	0,137	1,000	0,544
2017/2018	0,289	1,000	0,637	**0,032**	0,626	–	0,289

Tabulka 3.11: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (1 a méně, nebo 2 a více gólů) – domácí utkání

Závislost můžeme potvrdit i v domácích utkáních. Z p-hodnot uvedených v tabulce (3.11) zamítáme hypotézu nezávislosti pro tým PSG Berani Zlín v sezóně 2017/2018. Abychom zjistili, jak správně závislost interpretovat, provedeme také testování oběma alternativami Fisherova jednostranného testu v kontingenční tabulce zlínského týmu.

	≤ 1 gól	≥ 2 góly
≤ 1 gól	4	3
≥ 2 góly	2	16

Jednostranným Fisherovým testem s alternativou OR>1 dostaneme p-hodnotu 0,032, alternativou OR<1 získáme p-hodnotu 0,998.

P-hodnota, kterou jsem obdrželi oboustranným testem, je shodná s p-hodnotou, kterou dává jednostranný Fisherův test s alternativou OR>1. Pro domácí utkání zlínského týmu platí tato závislost: Šance, že tým vstřelí dva a více gólů, je větší, pokud v předešlém zápase skóroval alespoň dvakrát, než pokud se ze vstřeleného gólu radoval nejvýše jednou. Šance vstřelení jednoho a méně gólů je větší, jestliže i v přecházejícím utkání dali hráči nejvýše jeden gól, než pokud skórovali alespoň dvakrát. Laicky řečeno, na rozdíl od předchozího případu zde platí, že gólová úroda bude následovat spíše po gólově dobrém než po gólově špatném zápase.

Testy nezávislosti u zbylých týmů dávají p-hodnoty větší něž je hodnota α=0,05, tudíž hypotézu nezávislosti nelze zamítnout.

Venkovní utkání (1 a méně, nebo 2 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	0,549	1,000	1,000	**0,042**	0,278	0,661
2017/2018	0,111	1,000	0,133	0,637	0,671	0,544	0,330
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,397	1,000	1,000	1,000	1,000	1,000	0,241
2017/2018	0,394	0,362	1,000	1,000	1,000	–	1,000

Tabulka 3.12: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (1 a méně, nebo 2 a více gólů) – venkovní utkání

Ve venkovních zápasech také najdeme závislost. Ze všech obdržených p-hodnot, které jsou zaznamenány v tabulce (3.12), zamítáme hypotézu nezávislosti u týmu Litvínova v sezóně 2016/2017. K formulaci závislosti provedeme také jednostranné Fisherovy testy v kontingenční tabulce litvínovského týmu.

	≤ 1 gól	≥ 2 góly
≤ 1 gól	2	9
≥ 2 góly	9	5

Testováním pomocí jednostranného Fisherova testu s alternativou OR>1 získáme p-hodnotu 0,997, alternativa OR<1 pak dává p-hodnotu 0,027.

Původní p-hodnota uvedená v tabulce (3.12) je blízko k p-hodnotě, kterou získáme testováním jednostranného Fisherova testu s alternativou OR<1. Ve venkovních utkáních Litvínova v sezóně 2016/2017 pozorujeme tuto závislost: Šance, že hráči skórují alespoň dvakrát, je větší, pokud v předchozím zápase dali jeden a méně gólů, než když soupeřova brankáře překonali alespoň dvakrát. Také šance, že tým vstřelí nejvýše jeden gól, je větší, jestliže v předcházejícím venkovním utkání vsítil dva a více gólů, než když se z gólu radoval nejvýše jednou.

Ve zbylých případech hypotézu nezávislosti nezamítáme kvůli p-hodnotám převyšujícím hodnotu 0,05.

Všechna utkání (2 a méně, nebo 3 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,208^T	0,122	0,344^T	1,000^T	1,000^T	0,408^T	0,742^T
2017/2018	0,456^T	0,250	1,000^T	0,882^T	0,622^T	0,672^T	0,672^T
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,811^T	1,000^T	0,208^T	1,000^T	0,800^T	**0,018**^{T}	1,000^T
2017/2018	0,080^T	1,000^T	0,088^T	1,000^T	0,326^T	–	0,507^T

Tabulka 3.13: P-hodnoty testů nezávislosti (chí-kvadrát testu, pokud je v buňce index ^{T} – viz kapitola 2.4, nebo oboustranného Fisherova faktoriálového testu s alternativou "two.sided") k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (2 a méně, nebo 3 a více gólů) – všechna utkání základní části

Tabulka (3.13) obsahuje p-hodnoty obdržené testy nezávislosti. V jednom případu jsme dostali p-hodnotu menší než 0,05. Pro tým HC Energie Karlovy Vary v sezóně 2016/2017 na základě p-hodnoty 0,018 zamítáme hypotézu nezávislosti. Uvedenou p-hodnotu jsme získali z chí-kvadrát testu nezávislosti a tato hodnota by měla přibližně odpovídat p-hodnotě oboustranného Fisherova testu. Víme tedy, že mezi počtem vstřelených gólů je ve dvou po sobě jdoucích zápasech u karlovarského týmu nějaká závislost. Ke zjištění podoby závislosti provedeme v kontingenční tabulce všechny varianty Fisherova testu.

	≤ 2 góly	≥ 3 góly
≤ 2 góly	17	16
≥ 3 góly	16	2

S využitím Fisherova oboustranného testu získáme p-hodnotu 0,013, jednostranný test s alternativou OR>1 dává p-hodnotu 0,999 a pomocí jednostranného testu s alternativou OR<1 obdržíme p-hodnotu 0,007.

P-hodnota z tabulky (3.13) zjištěná chí-kvadrát testem je přibližně rovna jak p-hodnotě oboustranného Fisherova testu, tak p-hodnotě jednostranného Fisherova testu s alternativou OR<1. V zápasech základní části týmu HC Energie Karlovy Vary v sezóně 2016/2017 pozorujeme tuto závislost: Šance vstřelení alespoň tří gólů je větší, jestliže v předešlém utkání dali hokejisté dva a méně gólů, než když se prosadili alespoň třikrát. Stejně tak je šance vstřelení dvou a méně gólů větší, pokud v předchozím zápase hráči skórovali alespoň třikrát, než když dali nejvýše dva góly. Můžeme zde vidět příklad fungování regrese, tj. návratu k průměru. Více než tři vstřelené góly zřejmě znamená nadprůměrný výsledek, a proto lze očekávat, že po něm bude následovat už ne tak extrémně dobrý výsledek, tj. výsledek bude blíže k průměru.

Hypotézu nezávislosti ve zbylých případech nezamítáme, jelikož jsou p-hodnoty v tabulce (3.13) vyšší než hodnota α=0,05.

Domácí utkání (2 a méně, nebo 3 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,695	0,278	0,229	0,299	1,000	0,671	1,000
2017/2018	0,133	1,000	0,689	1,000	0,401	0,227	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,414	0,697	**0,015**	0,671	1,000	0,697	0,428
2017/2018	0,137	0,278	0,428	0,695	0,111	–	1,000

Tabulka 3.14: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (2 a méně, nebo 3 a více gólů) – domácí utkání

Tabulka (3.14) obsahuje všechny získané p-hodnoty pro domácí zápasy. Hypotézu nezávislosti zamítáme u týmu z Mladé Boleslavi v sezóně 2016/2017. Z oboustranného Fisherova faktoriálového testu nezjistíme, o jakou závislost se jedná, takže provedeme i oba jednostranné testy v kontingenční tabulce.

	≤ 2 góly	≥ 3 góly
≤ 2 góly	2	10
≥ 3 góly	9	4

Pro jednostranný test s alternativou OR>1 získáme p-hodnotu 0,999, test s alternativní hypotézou OR<1 dá p-hodnotu 0,011.

P-hodnota v tabulce (3.14) zhruba odpovídá jednostrannému Fisherově faktoriálovému testu s alternativou OR<1, a tedy šance, že se hráči Mladé Boleslavi v domácím prostředí gólově prosadí alespoň třikrát, je větší, jestliže v předešlém domácím zápase skórovali nejvýše dvakrát, než když vstřelili tři a více branek. Šance, že hokejisté doma překonají soupeřova brankáře nejvýše dvakrát, je větší, pokud v předchozím domácím utkání dali alespoň tři góly, než když se ze vstřeleného gólu radovali nejvýše dvakrát.

Ve zbylých případech nelze zamítnout hypotézu nezávislosti kvůli vysokým p-hodnotám.

Venkovní utkání (2 a méně, nebo 3 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,695	0,211	0,401	0,695	0,679	1,000	0,428
2017/2018	0,544	1,000	0,428	0,099	0,434	**0,049**	0,434
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	1,000	0,435	0,428	0,679	1,000	1,000	1,000
2017/2018	0,238	0,695	1,000	1,000	0,695	–	0,640

Tabulka 3.15: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (2 a méně, nebo 3 a více gólů) – venkovní utkání

Ve venkovních utkáních závislost také najdeme. Z tabulky (3.15) vyčteme, že u týmu HC Kometa Brno zamítneme v sezóně 2017/2018 hypotézu nezávislosti. Chtěli bychom zjistit, o jakou závislost se jedná, takže provedeme oba Fisherovy jednostranné testy v kontignenční tabulce brněnské Komety.

	≤ 2 góly	≥ 3 góly
≤ 2 góly	11	3
≥ 3 góly	4	7

P-hodnota jednostranného Fisherova testu s alternativou OR>1 je rovna číslu 0,042 a test s opačnou alternativou OR<1 dá p-hodnotu 0,995.

Výsledek v tabulce (3.15), který jsme získali s pomocí oboustranného Fisherova testu, je podobný tomu, který nám dal jednostranný test s alternativou OR>1. Závislost pro venkovní utkání u brněnského týmu v sezóně 2017/2018 popíšeme takto: Šance, že hráči vsítí tři a více branek, je větší, jestliže v předchozím venkovním utkání taky dali alespoň tři góly, než když skórovali nejvýše dvakrát. Šance vstřelení nejvýše dvou gólů je větší, pokud tým i v předcházejícím venkovním zápase překonal soupeře nejvýše dvakrát, než když dal soupeři tři a více gólů.

Nezávislost ve zbylých případech nemůžeme na základě p-hodnot zamítnout.

Všechna utkání (3 a méně, nebo 4 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	0,442	$0{,}901^T$	0,358	$0{,}465^T$	1,000	0,115	0,746
2017/2018	0,490	1,000	0,706	0,743	1,000	0,531	1,000
	VIT	**PLZ**	**MBL**	**ZLN**	**OLO**	**KVA**	**PCE**
2016/2017	0,684	0,329	0,082	**0,040**	1,000	1,000	1,000
2017/2018	$0{,}840^T$	$0{,}882^T$	1,000	1,000	0,297	–	1,000

Tabulka 3.16: P-hodnoty testů nezávislosti (chí-kvadrát testu, pokud je v buňce index ^{T} – viz kapitola 2.4, nebo oboustranného Fisherova faktoriálového testu s alternativou "two.sided") k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (3 a méně, nebo 4 a více gólů) – všechna utkání základní části

Z obdržených p-hodnot testů nezávislosti, které jsou zaznamenány v tabulce (3.16), zamítneme testovanou hypotézu nezávislosti pro tým PSG Berani Zlín v sezóně 2016/2017. Protože jsme k testování využili oboustranný Fisherův faktoriálový test, provedeme v kontingenční tabulce také jednostranné varianty Fisherova testu, abychom zjistili, o jakou závislost se jedná.

	≤ 3 góly	≥ 4 góly
≤ 3 góly	22	14
≥ 4 góly	14	1

Fisherův jednostranný test s alternativou OR>1 dává p-hodnotu 0,998. Využitím jednostranného testu s alternativou OR<1 obdržíme p-hodnotu 0,02.

P-hodnota testu nezávislosti pro zlínský tým v sezóně 2016/2017 přibližně odpovídá p-hodnotě, kterou získáme při testování jednostranným Fisherovým testem s alternativou OR<1. Závislost mezi počtem vstřelených gólů je následující: Šance, že hokejisté v utkání vstřelí čtyři a více gólů, je větší, pokud v předešlém zápase skórovali nejvýše třikrát, než když vsítili alespoň čtyři góly. Také šance vstřelení tří a méně gólů je větší, jestliže v předchozím utkání dali hráči alespoň čtyři branky, než když překonali soupeře nejvýše třikrát.

V ostatních případech nelze ze získaných p-hodnot hypotézu nezávislosti zamítnout.

Domácí utkání (3 a méně, nebo 4 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	1,000	1,000	1,000	0,527	0,116	0,695
2017/2018	1,000	1,000	0,527	0,227	0,544	1,000	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	0,278	0,394	1,000	0,211	0,330	1,000	0,137
2017/2018	0,227	0,697	0,544	0,394	0,661	–	1,000

Tabulka 3.17: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (3 a méně, nebo 4 a více gólů) – domácí utkání

Výsledné p-hodnoty testů nezávislosti pro domácí zápasy nalezneme v tabulce (3.17). Ve všech případech dostaneme hodnoty vyšší než je hodnota α=0,05 a hypotézu nezávislosti nezamítáme.

Venkovní utkání (3 a méně, nebo 4 a více gólů)

	LIB	TRI	SPA	MHK	LIT	KOM	CHM
2016/2017	1,000	0,401	0,122	1,000	1,000	0,070	1,000
2017/2018	1,000	0,562	1,000	1,000	1,000	0,087	1,000
	VIT	PLZ	MBL	ZLN	OLO	KVA	PCE
2016/2017	1,000	**0,032**	0,362	1,000	0,527	1,000	0,626
2017/2018	0,394	0,679	0,330	1,000	1,000	–	1,000

Tabulka 3.18: P-hodnoty oboustranného Fisherova faktoriálového testu nezávislosti s alternativou "two.sided"k určení závislosti počtu vstřelených gólů ve dvou po sobě jdoucích utkáních (3 a méně, nebo 4 a více gólů) – venkovní utkání

V domácích zápasech jsme závislost nenašli, avšak u zápasů venkovních můžeme v jednom případu hypotézu nezávislosti zamítnout. Z tabulky (3.18) p-hodnot zamítneme nulovou hypotézu u týmu HC Škoda Plzeň v sezóně 2016/2017. I nyní pro správnou formulaci závislosti provedeme Fisherovy jednostranné testy v kontingenční tabulce.

	≤ 3 góly	≥ 4 góly
≤ 3 góly	16	2
≥ 4 góly	3	4

Využitím Fisherova jednostranného testu s alternativní hypotézou OR>1 obdržíme p-hodnotu 0,032 a jednostranný test s alternativou OR<1 dá p-hodnotu 0,998.

P-hodnota uvedená v tabulce (3.18) je shodná s p-hodnotou, kterou získáme testováním pomocí jednostranného Fisherova faktoriálového testu s alternativní hypotézou OR>1. Na základě těchto výsledků zformulujeme závislost mezi počtem vstřelených gólů ve dvou po sobě jdoucích zápasech platnou pro venkovní utkání plzeňské Škodovky v sezóně 2016/2017: Šance, že hráči dají čtyři a více gólů, je větší, pokud v předcházejícím utkání také skórovali alespoň čtyřikrát, než když vstřelili tři a méně branek. Šance vstřelení nejvýše tří gólů je větší, jestliže tým v předešlém zápase překonal soupeřova brankáře nejvýše třikrát, než když se ze vstřelení gólu radoval alespoň čtyřikrát.

Ve všech zbylých případech p-hodnota přesáhla hodnotu α=0,05, a tak hypotézu nezávislosti nezamítáme.

Závěr

Hlavním úkolem bakalářské práce bylo zkoumat závislost mezi výsledky po sobě jdoucích hokejových utkání v rámci dvou sezón Tipsport Extraligy ledního hokeje. Ve statistickém programu RStudio jsme s využitím metod statistické analýzy testovali hypotézy nezávislosti mezi výsledky po sobě jdoucích zápasů a zjišťovali, jestli má prostředí nějaký vliv na úspěšnost týmu.

V první kapitole jsme popsali analyzované datové soubory. Na konci kapitoly jsme pro názornost uvedli ukázku z datového souboru jednoho týmu.

Druhá kapitola byla věnována statistickým metodám, jež nám posloužily pro testování našich hypotéz nezávislosti. Čtenář byl seznámen se způsobem značení výsledných p-hodnot, které jsme obdrželi provedením jednotlivých testů nezávislosti.

Třetí kapitola zahrnuje praktickou část práce. Věnujeme se v ní zkoumání závislosti mezi výsledky po sobě jdoucích utkání a tuto závislost rozebíráme z různých úhlů. V každé podkapitole obsahující jednu z možností rozboru závislosti uvádíme, jak jsme v RStudiu pracovali s datovými soubory.

Nejdříve jsme se zaměřili na zjišťování závislosti mezi výsledky dvou po sobě jdoucích utkání. V případě všech utkání základní části jsme u jednoho týmu v sezóně 2017/2018 našli opačnou závislost, než jsme čekali. Jednalo se o tým HC Dynamo Pardubice, jehož hráči po vítězném utkání následující zápas většinou prohráli a po prohraném zápasu převážně další zápas vyhráli. Když jsme zápasy základní části rozdělili na domácí a venkovní zvlášť, zjistili jsme závislost mezi výsledky dvou po sobě jdoucích venkovních utkáních pouze pro jeden tým v jedné sezóně. Tým Piráti Chomutov v sezóně 2017/2018 na venkovních kluzištích po výhře obvykle v příštím utkání opět vyhrál a po prohraném zápasu následující utkání většinou znovu prohrál.

Dále jsme rozbor závislosti rozšířili na dvouzápasové a třízápasové série výher (respektive proher). V rámci všech utkání základní části jsme závislost nikde nenašli a pro domácí a venkovní utkání jsme testování kvůli malým hodnotám nedělali. Laicky to znamená, že nic takového jako vítězné série neexistuje. Některé týmy jsou lepší, a tak zaznamenají více výher (a tím i více výher v řadě), ale výsledek každého utkání je nezávislý s výsledky předchozích utkání.

Následně jsme zkoumali závislost mezí výhrou (respektive prohrou) a počtem vstřelených gólů. Podle očekávání jsme v případě testování v rámci všech utkání základní části našli závislost u všech týmů v obou sezónách a překvapivě jen u většiny (nikoli u všech) týmů pro domácí a venkovní utkání zvlášť.

Poslední část byla zaměřena na zkoumání závislosti mezi různým počtem vstřelených gólů ve dvou po sobě jdoucích utkáních. Nejprve jsme se zabývali situací, kdy jsme pro předchozí a následující utkání uvažovali, že tým vstřelí žádný, nebo jeden a více gólů, tj. zkoumali jsme situaci, v níž tým neskóruje nebo skóruje. Pro takovou situaci jsme závislost nenašli ani ve všech utkáních základní části, ani při rozlišení zápasů na domácí a venkovní. V ostatních situacích (tj. 1 a méně, nebo 2 a více vstřelených gólů; 2 a méně, nebo 3 a více vstřelených gólů; 3 a méně, nebo 4 a více vstřelených gólů) jsme většinou našli závislost, ale vždy se jednalo nejvýše o jeden tým a jednu sezónu (jak pro všechna utkání základní části, tak i při rozdělení zápasů na domácí a venkovní).

Na základě provedených testů nezávislosti jsme pro některé týmy v určitých situacích (obvykle však jen pro jednu ze zkoumaných sezón) zjistili vliv prostředí na jejich úspěšnost. Především při hledání závislosti mezi výsledky po sobě jdoucích zápasů (výhra nebo prohra) a závislosti mezi různým počtem vstřelených gólů ve dvou po sobě jdoucích utkáních nebyla u žádného týmu nalezena závislost pro obě sezóny současně. Je třeba ale brát v úvahu, že se testování dělá na hladině α=0,05, takže v 5 % případů, v nichž platí nulová hypotéza, obdržíme chybný výsledek (tj. uděláme chybu prvního druhu, a tedy zamítneme pravdivou hypotézu).

Literatura

[1] Anděl, J.: *Statistické úlohy, historky a paradoxy.* 1. vydání. Praha: Matfyz-Press, 2018. ISBN 978-80-7378-360-0

[2] Hron, K., Kunderová, P.: *Základy počtu pravděpodobnosti a metod matematické statistiky.* 2. doplněné vydání. Olomouc: Univerzita Palackého v Olomouci, 2015. ISBN 978-80-244-4774-2

[3] Hron, K., Kunderová, P., Vencálek, O.: *Základy počtu pravděpodobnosti a metod matematické statistiky.* 3. přepracované vydání. Olomouc: Univerzita Palackého v Olomouci, 2018. ISBN 978-80-244-5398-9

[4] Sklenářová, Kristýna: *Bakalářská práce - Analýza finanční gramotnosti absolventů SŠ ve Znojmě* [online]. 2019, dostupné z: https://library.upol.cz/arl-upol/cs/csg/?repo=upolrepo&key=18790635131

[5] Agresti, A.: *Categorical Data Analysis.* 2. vydání. New Jersey: John Wiley & Sons, Inc., 2002. ISBN 0-471-36093-7

[6] Agresti, A.: *An Introduction to Categorical Data Analysis.* 2. vydání. New Jersey: John Wiley & Sons, Inc., 2007. ISBN 978-0-471-22618-5

[7] Osborn, Carol E.: *Statistical Applications for Health Information Management.* 2. vydání. Sudbury: Jones and Bartlett Publishers, 2006. ISBN 978-0-7637-2842-7