

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY



MODELOVANIE A MERANIE VÝKONNOSTI
LUKOSTRELCOV

BAKALÁRSKA PRÁCA

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

**MODELOVANIE A MERANIE VÝKONNOSTI
LUKOSTRELCOV**

BAKALÁRSKA PRÁCA

Študijný program: Ekonomická a finančná matematika
Študijný odbor: 1114 Aplikovaná matematika
Školiace pracovisko: Katedra aplikovanej matematiky a štatistiky
Vedúci práce: Mgr. Martin Hurban



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

- Meno a priezvisko študenta:** Veronika Joščáková
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: aplikovaná matematika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický
- Názov:** Modelovanie a meranie výkonnosti lukostrelcov
Modeling and measurement of archers performance
- Cieľ:** Navrhnuť spôsob predikcie výkonu lukostrelcov. Zdokumentovať a porozumieť použitým metódam. Analyzovať a zhrnúť rôzne metódy porovnávania výkonnosti strelcov.
- Literatúra:** George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Time Series Analysis: Forecasting and Control, 4th Edition
- Vedúci:** Mgr. Martin Hurban
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.
- Dátum zadania:** 25.10.2016
- Dátum schválenia:** 19.11.2016 doc. RNDr. Margaréta Halická, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Podakovanie Touto cestou by som sa rada podakovala vedúcemu práce, Mgr. Martinovi Hurbanovi, za ochotu, cenné rady a čas, ktorý venoval pre úspešné spracovanie tejto práce.

Abstrakt

JOŠČÁKOVÁ, Veronika: Modelovanie a meranie výkonnosti lukostrelcov [Bakalárska práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; školiteľ: Mgr. Martin Hurban, Bratislava, 2017, 54 s.

V práci sa zaoberáme skúmaním časovej závislosti výstrelov počas preteku v športovej lukostrelbe. Vývoj výkonu lukostrelca modelujeme SARIMA procesmi. Ukazuje sa, že výkony lukostrelca možno popísať modelom $SARIMA(1, 0, 0) \times (1, 0, 0)_6$. Efekt po sebe vystrelených šípov vyšetríme pomocou Pearsonovho χ^2 testu závislosti. Analyzovali sme vplyv výkonu strelca na koeficient sezónneho AR člena. V druhej časti riešime problém nominácie vhodných strelcov na medzinárodné podujatia. Uvažujeme 4 rôzne prediktory a porovnáваме ich predikčné schopnosti so súčasnou metodikou SLZ.

Kľúčové slová: Lukostrelba, SARIMA procesy, Sezónnosť, Predikcie výkonu, Výber reprezentanta

Abstract

JOŠČÁKOVÁ, Veronika: Modeling and measurement of archers performance [Bachelor Thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: Mgr. Martin Hurban, Bratislava, 2017, 54 p.

This thesis deals with research of time dependence of shots during sport archery competition. The progress of archer's performance can be modeled by SARIMA processes. It turns out that performances of archers can be modeled by $SARIMA(1, 0, 0) \times (1, 0, 0)_6$. We examine dependence of two consecutive shots by Pearson's χ squared test. We analyzed impact of archer's performance into season AR coefficient. In the second part of the thesis is solved problem of nomination of suitable archers on international events. We consider 4 different predictors and compare their predictable abilities with current method of SLZ.

Keywords: Archery, SARIMA processes, Seasonality, Predictions of performance, Nomination of archer

Obsah

Úvod	9
1 Autoregresné procesy	11
1.1 Definície pojmov	11
1.2 Lineárne modely časových radov	13
1.3 Testy	16
2 Výkonnosť lukostrelca počas preteku	20
2.1 Formulácia problému	20
2.2 Popis dát	20
2.2.1 Výkon strelcov	22
2.3 Výber vhodného modelu	23
2.4 Výsledky	27
2.4.1 Závislosť hodnoty koeficientu pri sezónnom AR člene od výkonu	30
2.4.2 Závislosť výstrelu od predchádzajúceho výstrelu	32
2.4.3 Stredná hodnota v čase	34
3 Predikcie výkonu lukostrelcov	37
3.1 Formulácia problému	37
3.2 Popis dát	37
3.3 Metódy pre výpočet odhadov	38
3.3.1 Aritmetický priemer	38
3.3.2 Maximum	39
3.3.3 Medián	39
3.3.4 Vážený aritmetický priemer	40
3.4 Výber súťaží a výpočtový program	40
3.5 Výsledky	41
Záver	46
Zoznam použitej literatúry	48
Príloha A	50

Príloha B	51
Príloha C	52
Príloha D	54

Úvod

V súčasnosti je šport veľmi populárnou témou medzi ľuďmi, u mnohých aj ich súčasťou života. Avšak niektorí z nás mu zasvätili svoj život úplne a predstavuje tak spôsob, ktorým si zabezpečujú živobitie. Ale to platí len pre tých najlepších vo svojej športovej kategórii.

V každom športe sa usporadúvajú rôzne súťaže. Postupne od regionálnych tried cez rôzne postupové preteky a kvalifikácie, až po tie najväčšie súťaže v danom športe. U niektorých to však vrcholí najznámejšími olympiádami a majstrovstvami. A práve tu sa ukazuje forma jednotlivých súťažiacich prostredníctvom podaných výkonov. Tie sa medzi sebou následne porovnávajú a určuje sa víťaz danej športovej kategórie, respektíve víťazný tím v prípade kolektívnych športov.

V tejto práci sa budeme venovať športovej lukostrelbe. Práca je rozdelená do troch hlavných častí. V prvej kapitole je stručne zhrnutá teória časových radov, ktorú budeme využívať. V tejto časti vychádzame hlavne z prednášok [1], ktorých hlavným zdrojom je monografia [6].

Výkonnosť v závislosti od času bola zatiaľ málo skúmanou problematikou, pretože spôsob, akým si strelci zapisujú vystrelené šípy, nezachováva časovú postupnosť výstrelov. V súčasnosti sa pre zaznamenávanie výstrelov začínajú používať automatické prístroje, ktoré zachovávajú časovú závislosť. Takouto cestou boli zbierané aj dáta poskytnuté spoločnosťou Bowbook s.r.o., ktorá je prevádzkovateľom lukostreleckého portálu rcherz.com. V druhej kapitole práce sa preto pokúsime popísať výkonnosť lukostrelca počas preteku pomocou modelov časových radov. Správanie výkonu vybraných strelcov budeme modelovať konkrétne pomocou SARIMA modelov, ktoré popisujú závislosť výstrelov od predchádzajúcich striel. Cieľom je identifikovať vhodné modely, ktoré túto závislosť v čase popisujú čo najlepšie. Inou metódou merania závislosti v dátach sú kontingenčné tabuľky a testy popísané v práci [8].

V tretej časti budeme hľadať vhodný spôsob predikcie výkonu lukostrelcov pomocou historických dát. Tento problém je priamo motivovaný požiadavkami SLZ pre férovú a optimálnu voľbu lukostrelcov, ktorí budú nominovaní na vrcholné medzinárodné podujatia. Naším cieľom je skonštruovať a porovnať rôzne metódy predikcie výsledkov. Pre porovnanie nám poslúžia výpočty absolútnych a kvadratických chýb. Následne sa

pokúsime určiť, ktorá z metód je najlepším spôsobom pre predpoveď budúceho výkonu strelca. Súbor dát nám k tejto časti poskytol Slovenský lukostrelecký zväz.

1 Autoregresné procesy

1.1 Definície pojmov

Definícia 1.1. *Stochastický proces* je t - parametrický systém náhodných premenných $\{X_t, t \in I\}$, kde I je interval alebo diskrétna množina indexov. [2, str. 22]

Ako je uvedené v prednáškach [1], z ktorých budeme vychádzať pri zavádzaní ďalších pojmov, tento proces je úplne charakterizovaný T - rozmernou distribučnou funkciou. Zvyčajne sa zameriavame na prvé dva momenty:

- stredná hodnota $E[X_t]$, $t = 1, 2, 3, \dots$
- variancia $Var[X_t]$, $t = 1, 2, 3, \dots$
- kovariancie $Cov[X_t, X_s]$, $t, s = 1, 2, 3, \dots, t < s$, nazývané autokovariancie.

Definícia 1.2. *Realizácia stochastického procesu* X predstavuje časový rad, ktorý je vytvorený množinou pozorovaní v časoch $t = 1, 2, 3, \dots$.

Pre stochastický proces možno definovať tzv. silnú stacionaritu:

Definícia 1.3. *Stochastický proces je silno stacionárny*, ak sa združená distribučná funkcia pri posune v čase nemení.

Pre potreby našej práce budeme uvažovať slabú stacionaritu definovanú nasledujúcimi podmienkami:

- $E[X_t] = \mu, \forall t$
- $Cov[X_t, X_s] = \gamma(|t - s|), \forall t, s$, kde $t \neq s$
- $Var[X_t] = \gamma(0) = \text{konštanta}$, pre $\forall t$.

Pre procesy v našej práci budeme predpokladať vlastnosť ergodicity.

Definícia 1.4. *Ergodický proces* je proces, ktorého výberové momenty počítané z časového radu s T pozorovaniami konvergujú pre $T \rightarrow \infty$ k zodpovedajúcim momentom.

Definícia 1.5. *Lag operator* L je operátor posunu, ktorý vracia hodnotu procesu posunutú o jedno obdobie dozadu:

$$Lx_t = x_{t-1}. \quad (1)$$

Definícia 1.6. Nech $\hat{\theta}$ je odhadom parametra θ a n je počet pozorovaní. Potom parameter $\hat{\theta}$ je **konzistentným odhadom** θ , ak pre $\forall \epsilon > 0$ platí

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0.$$

Definícia 1.7. *Autokorelačná funkcia (ACF)* stacionárneho procesu je definovaná vzťahom:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}, \quad (2)$$

t. j.

$$\rho(\tau) = \text{Cor}(x_t, x_{t+\tau}) = \frac{\text{Cov}(x_t, x_{t+\tau})}{\sqrt{\text{Var}(x_t)\text{Var}(x_{t+\tau})}}. \quad (3)$$

Platí:

$$\rho(0) = 1, \rho(\tau) = \rho(-\tau)$$

a preto stačí počítať $\rho(\tau)$ pre $\tau = 1, 2, \dots$.

Keďže predpokladáme ergodicitu procesu, vieme strednú hodnotu, disperziu a kovariancie konzistentne odhadnúť z dát x_1, \dots, x_T , kde T je počet dát:

- $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T x_t$
- $\hat{\gamma}(0) = \frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})^2$
- $\hat{\gamma}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} (x_t - \hat{\mu})(x_{t+\tau} - \hat{\mu})$.

Potom konzistentný odhad autokorelačnej funkcie je

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}. \quad (4)$$

Definícia 1.8. *Parciálna autokorelačná funkcia (PACF)* je korelácia premenných x_t a $x_{t-\tau}$ očistená o vplyv premenných medzi nimi $x_{t-1}, \dots, x_{t-\tau+1}$.

Počítame ju ako podmienenú strednú hodnotu

$$\Phi_{k,k} = E[(x_t - \mu)(x_{t-k} - \mu) | x_{t-1}, \dots, x_{t-k+1}]. \quad (5)$$

Výberová parciálna autokorelačná funkcia $r_{k,k}$ s posunutím k odhaduje parciálny regresný koeficient $\Phi_{k,k}$ v autoregresii k -teho rádu

$$x_t = \Phi_{k,1}x_{t-1} + \Phi_{k,k}x_{t-k} + e_t, \quad (6)$$

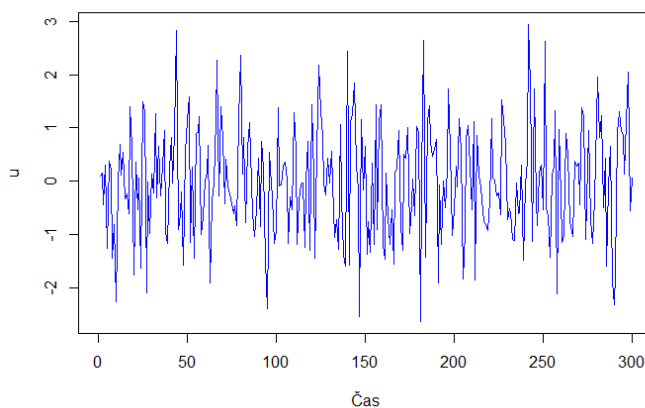
kde e_t je nekorelovaná premenná s náhodnými premennými x_{t-i} , $i \geq 1$. [3, str. 16]

1.2 Lineárne modely časových radov

Definícia 1.9. *Biely šum* $\{u_t\}$ je stacionárny proces definovaný prvými dvoma momentami:

- $E[u_t] = 0, \forall t$
- $Cov[u_t, u_s] = 0, \forall t \neq s$
- $Var[u_t] = \sigma^2, \forall t.$

Je to základný a dôležitý náhodný proces, pomocou ktorého sa definujú ďalšie procesy a modely pre dáta. Na obrázku 1 je znázornená realizácia bieleho šumu s parametrami $E[u_t] = 0$ a $Var[u_t] = 1$.



Obr. 1: Realizácia bieleho šumu

Definícia 1.10. *AR(p) je autoregresný proces rádu p daný:*

$$x_t = \delta + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + u_t, \quad (7)$$

kde δ a α_i sú konštanty a u_t je biely šum.

Pomocou operátora posunu L vieme AR proces zapísať v tvare:

$$\alpha(L)x_t = \delta + u_t, \quad (8)$$

kde

$$\alpha(L) = 1 - \alpha_1 L - \dots - \alpha_p L^p. \quad (9)$$

Podmienka stacionarity je splnená práve vtedy, keď korene rovnice (9) ležia mimo jednotkového kruhu, čo znamená, že sú v absolútnej hodnote väčšie ako jedna.

Charakteristické vlastnosti procesu:

- stredná hodnota: $E[X_t] = \mu = \frac{\delta}{1 - \alpha_1 - \dots - \alpha_p}$
- variancia: $Var[x_t] = \frac{\sigma^2}{1 - \alpha_1 \rho(1) - \dots - \alpha_p \rho(p)}$, kde σ^2 je variancia bieleho šumu
- autokorelačná funkcia ACF: $\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2) + \dots + \alpha_k \rho(k-p)$
- parciálna autokorelačná funkcia PACF: $\Phi_{kk} = 0, k > p$.

Pomocou PACF dokážeme zistiť rád AR procesu.

Definícia 1.11. *MA(q) je proces kľzavých priemerov q-teho rádu, ktorý reprezentuje lineárnu kombináciu bielych šumov v q predchádzajúcich okamihoch. Všeobecne je daný rovnicou:*

$$x_t = \mu + u_t - \beta_1 u_{t-1} - \beta_2 u_{t-2} - \dots - \beta_q u_{t-q}, \quad (10)$$

kde β_i a μ sú konštanty a u_t je biely šum.

Tento model možno pomocou lag operátora zapísať v tvare:

$$\beta(L)u_t = x_t + \mu, \quad (11)$$

kde

$$\beta(L) = 1 - \beta_1 L - \dots - \beta_q L^q. \quad (12)$$

MA(q) sú vždy stacionárne. Pre jednoznčnosť potrebujeme mať splnenú podmienku invertovateľnosti.

Definícia 1.12. *Proces $MA(q)$ je **invertovateľný** práve vtedy, ak je možné ho reprezentovať ako stacionárny $AR(\infty)$ proces a všetky korene rovnice (12) sú v absolútnej hodnote väčšie ako 1. [3]*

Charakteristické vlastnosti procesu $MA(q)$:

- stredná hodnota: $E[x_t] = \mu$
- variancia: $Var[x_t] = \sigma^2(1 + \beta_1^2 + \dots + \beta_q^2)$, kde σ^2 je variancia bieleho šumu
- autokorelačná funkcia ACF: $\rho(k) = 0$, pre $k > q$
- parciálna autokorelačná funkcia PACF: $\Phi_{kk} \neq 0$, pre $\forall k$.

Pomocou ACF dokážeme zistiť rád MA procesu.

Pri modelovaní však môže nastať situácia, že sa PACF ani ACF nevynuluje. Nevieme teda určiť, či ide o AR alebo MA proces. Takéto procesy vieme potom modelovať pomocou zmiešaných procesov.

Definícia 1.13. *Spojením AR a MA procesov vzniknú zmiešané **ARMA(p,q) procesy**, kde p je počet AR členov a q MA členov. Dané sú rovnicou:*

$$x_t = \delta + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + u_t - \beta_1 u_{t-1} - \dots - \beta_q u_{t-q}, \quad (13)$$

kde δ , α_i a β_i sú konštanty a u_t je biely šum.

Tento model takisto vyjadruje nasledujúca rovnica použitím lag operatora:

$$\alpha(L)x_t = \delta + \beta(L)u_t. \quad (14)$$

Od $\alpha(L)$ a $\beta(L)$ požadujeme, aby nemali spoločný koreň. Je dôležité, aby bol proces stacionárny. V prípade, že táto podmienka nie je splnená, je potrebné dáta transformovať. Nestacionaritu značí prítomnosť koreňa $L = 1$ vo vzťahu (14).

Ak sa v procese objavuje k -násobný jednotkový koreň, potom k -te diferencie procesu sú stacionárne. Tie sú modelované pomocou ARMA(p, q) modelov a pôvodný proces je potom ARIMA(p, k, q).

Definícia 1.14. *Proces, ktorý je potrebné k -krát diferencovať na to, aby bola splnená podmienka stacionarity, sa nazýva integrovaný ARMA proces rádu k **ARIMA** (p, k, q) , kde*

- p - počet AR členov
- k - počet diferencovaní
- q - počet MA členov.

Avšak proces môže mať aj sezónny charakter, čo značí, že štruktúra procesu sa periodicky opakuje. Na modelovanie sezónnosti existujú špeciálne modely nazývané SARIMA.

Definícia 1.15. ***SARIMA** $(p, k, q) \times (P, K, Q)_s$ je sezónny integrovaný ARMA proces daný rovnicou:*

$$\alpha(L)A_P(L^s)(1-L)^d(1-L^s)^D X_t = \beta(L)B_Q(L^s)u_t,$$

kde:

- P - počet sezónnych AR členov
- K - počet sezónnych diferencovaní
- Q - počet sezónnych MA členov
- s - perióda dát

a $\alpha(L)$ reprezentuje AR proces, $\beta(L)$ MA proces, člen $(1-L)^d$ predstavuje diferencie, $(1-L^s)^D$ sezónne diferencie, $A_P(L^s)$ sezónny AR proces a $B_Q(L^s)$ sezónny MA proces.

Cieľom po diferencovaní je časový rad, ktorý neobsahuje trend ani jednotkový koreň, sezónny charakter mať môže.

1.3 Testy

Definícia 1.16. ***Augmented Dickey-Fuller (ADF) test** slúži na testovanie prítomnosti jednotkového koreňa v časovom rade a teda na určenie stacionárnosti procesu.*

Testovacie hypotézy:

H_0 : časový rad obsahuje jednotkový koreň vs.

H_1 : časový rad neobsahuje jednotkový koreň,

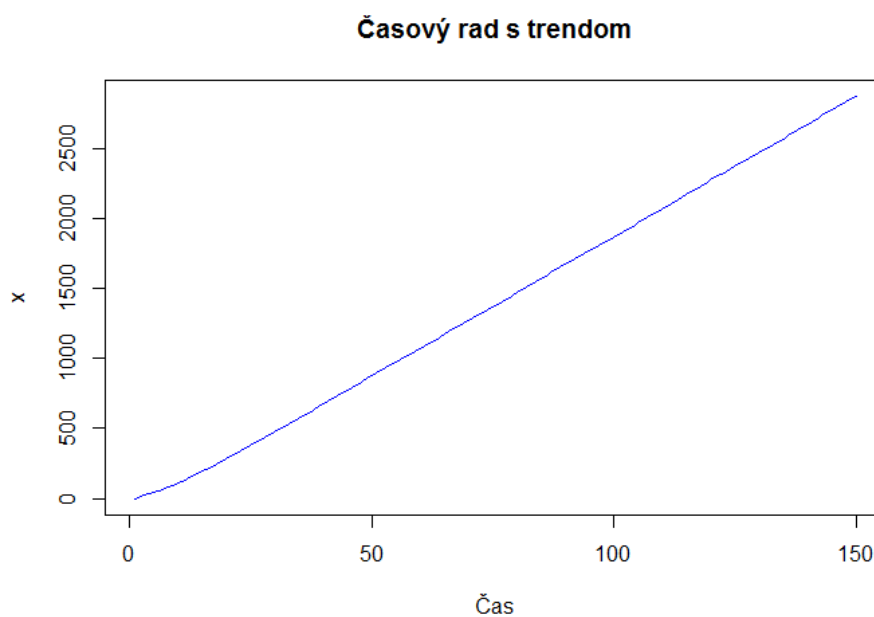
resp.

H_0 : proces nie je stacionárny vs.

H_1 : proces je stacionárny.

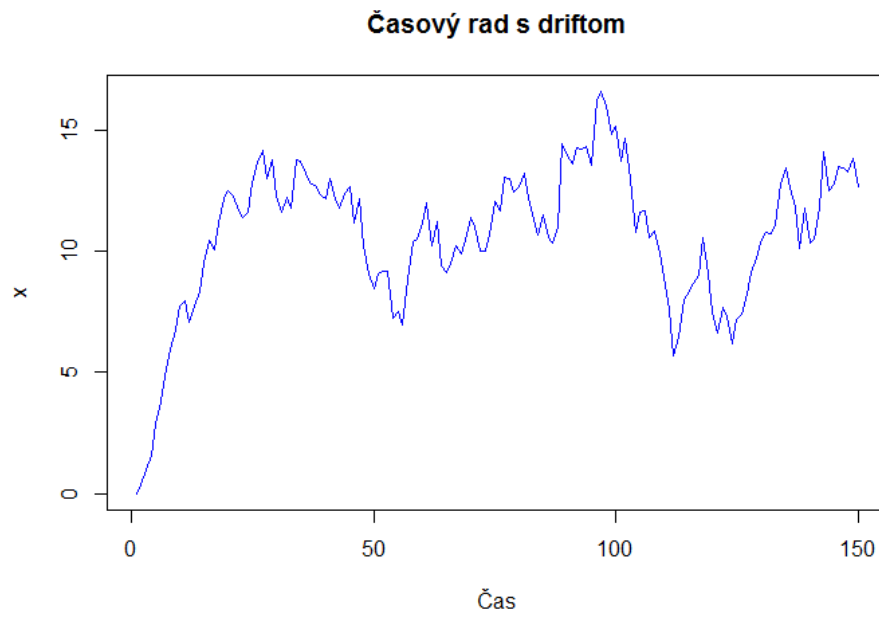
Rozlišujeme tri varianty testu podľa toho, ako sa časový rad správa:

- trend - obsahuje konštantu s lineárnym trendom. Na obrázku 2 je zobrazená realizácia procesu $x_t = 5 + 2t + 0.9x_{t-1} + u_t$.



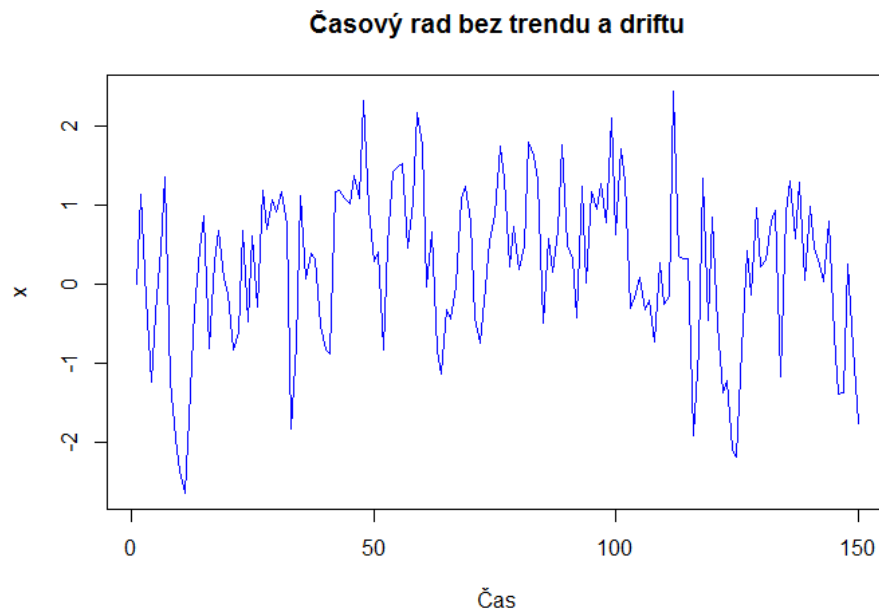
Obr. 2: Realizácia AR procesu s lineárnym trendom a konštantou

- drift - obsahuje konštantu bez lineárneho trendu. Obrázok 3 nám ponúka príklad realizácie procesu $x_t = 1 + 0.9x_{t-1} + u_t$.



Obr. 3: Realizácia AR procesu s konštantou

- none - neobsahuje konštantu ani lineárny trend. Príklad realizácie tohto procesu, $x_t = 0.9x_{t-1} + u_t$, môžeme vidieť na obrázku 4.



Obr. 4: Realizácia AR procesu bez konštanty a bez trendu

Definícia 1.17. *Ljung-Box test* je štatistický test, ktorý testuje hypotézu, či v postupnosti po sebe idúcich autokorelácií existuje nenulová autokorelácia.

Testovacie hypotézy:

H_0 : všetky¹ autokorelácie sú nulové vs.

H_1 : existuje nenulová autokorelácia,

čo vlastne znamená, že ak H_0 nezamietame, dáta možno považovať za nezávisle rozdelené a pozorované korelácie v dátach vyplývajú z náhodnosti procesu.

¹počet testovaných zodpovedá dĺžke časového radu

2 Výkonnosť lukostrelca počas preteku

2.1 Formulácia problému

Úlohou v tejto časti práce je zistiť, či sú nastrielané hodnoty lukostrelca počas súťaže ovplyvnené predchádzajúcimi výstrelmi alebo inými okolnosťami.

Predpokladáme, že medzi jednotlivými výstrelmi strelca existuje určitý vzťah. Tento súvis sa pokúsime zanalyzovať a otestovať pomocou teórie časových radov prezentovanej v 1. kapitole.

Výkon každého strelca pre nás predstavuje časový rad vytvorený množinou pozorovaní v časoch $t = 1, 2, \dots, 36$, čo predstavuje jednu realizáciu stochastického procesu.

Bodové hodnotenie sa pohybuje na stupnici 0 až 10, kde 10 značí maximálny počet dosiahnutých bodov jedného výstrelu. Nepredpokladáme preto výskyt trendu v testovaných procesoch. Súbor dát, ktorý máme k dispozícii uvádza súťažných strelcov. Vďaka skúsenostiam a schopnostiam súťažiacich predpokladáme, že sa jednotlivé výstrely budú pohybovať okolo nejakej konštanty.

V našom prípade každého strelca reprezentuje len jeden časový rad, preto pre štatistické závery potrebujeme dodatočné predpoklady pre odhady charakteristík stochastického procesu. Predpokladáme preto ergodicitu stacionárneho procesu, čo umožňuje výpočet výberových momentov z časového radu s T pozorovaniami, ktoré pre $T \rightarrow \infty$ konvergujú k zodpovedajúcim momentom.

Predpokladáme taktiež linearitu procesov, preto procesy vytvorené zo súboru dát, ktorý máme k dispozícii, chceme modelovať pomocou lineárnych ARMA modelov.

Počas súťaženia lukostrelec strieľa šípy v skupinách po šiesticiach. Preto predpokladáme možnú sezónnosť procesu a budeme testovať hypotézu o sezónnosti. Takéto procesy modelujeme prostredníctvom SARIMA modelov.

2.2 Popis dát

K dispozícii máme súbor dát, ktorý bol poskytnutý spoločnosťou Bowbook s.r.o., ktorá prevádzkuje lukostrelecký portál rcherz.com. Sú v ňom zaznamenané výsledky lukostrelcov z rôznych súťaží, ktorí vystupovali v rozličných súťažných kategóriách. To znamená, že sa líšia v streľbe na rôzne vzdialenosti a v počtoch vystrelených šípov.

Počty získaných bodov každého strelca sú zaznamenané po šesticiach, pretože počas preteku sa do terča vystrelí 6 šíпов za sebou, ktoré sa následne vyberú, aby nedošlo k prípadnému poškodeniu šíпов.

V tejto časti vymedzíme kritéria, podľa ktorých sme vyberali vhodné dáta pre naše testovanie.

Prvé kritérium pre vyfiltrovanie údajov bol výber tých strelcov, ktorí súťažili v kategórii, v ktorej sa strieľala sada 2x36 šíпов. Z celkového počtu 403 túto podmienku spĺňalo 169 strelcov.

Testovanie v kategórii s počtom výstrelov 36 na každú zo vzdialeností 30m, 50m, 70m a 90m, kde by sa vytvoril dlhší časový rad, sme zamietli. Dôvodom by bolo potrebné prepočítavanie vzdialeností, aby bol výkon strelca v podobe časového radu modelovaný za rovnakých podmienok. Preto sme zvolili už spomínanu sadu výstrelov, kde je konštantná vzdialenosť pre všetky vystrelené šípy a teda podmienky pre každý výstrel sú zhodné.

Druhým kritériom pre výber dát bude rovnaká vzdialenosť pre všetkých testovaných lukostrelcov. Pre testovanie sme si zvolili vzdialenosť dĺžky 70 metrov. Táto súťažná kategória 2x36 výstrelov so zvolenou vzdialenosťou sa nazýva malá fita. Zo 169 strelcov tejto podmienke vyhovuje 79.

Podrobnejšie delenia disciplín sú uvedené v prácach [4] a [5].

V súbore sa nachádzajú aj takí súťažiaci, ktorých bodové hodnotenie z celej súťaže alebo posledných pár šestic predstavovalo súčet 0, teda buď sa nezúčastnili preteku vôbec alebo ho nedokončili. Týchto strelcov sme takisto vylúčili z testovanej vzorky. Zo zvyšných 79 nám ostalo 64 lukostrelcov.

Tieto kritéria by nám za ideálnych podmienok postačovali, no narážame tu na problém. V niekoľkých prípadoch sú nastrieľané body zapísané v poradí vzostupne alebo zostupne po nastrieľaných šesticiach. Dôvodom pre takýto zápis je občasná porucha zariadenia zaznamenávajúca výstrely v čase. Ak dôjde k poruche, šípy zapíše človek usporiadane. Nie je teda zaznamenaný skutočný výkon strelca vyvíjajúci sa v čase. Náš súbor nie je výnimkou. Objavujú sa však aj strelci s neusporiadanými šesticami, čím predpokladáme, že daná šestica bola zapísaná v poradí, v akom sa šípy nastrieľali. A to je to, čo potrebujeme.

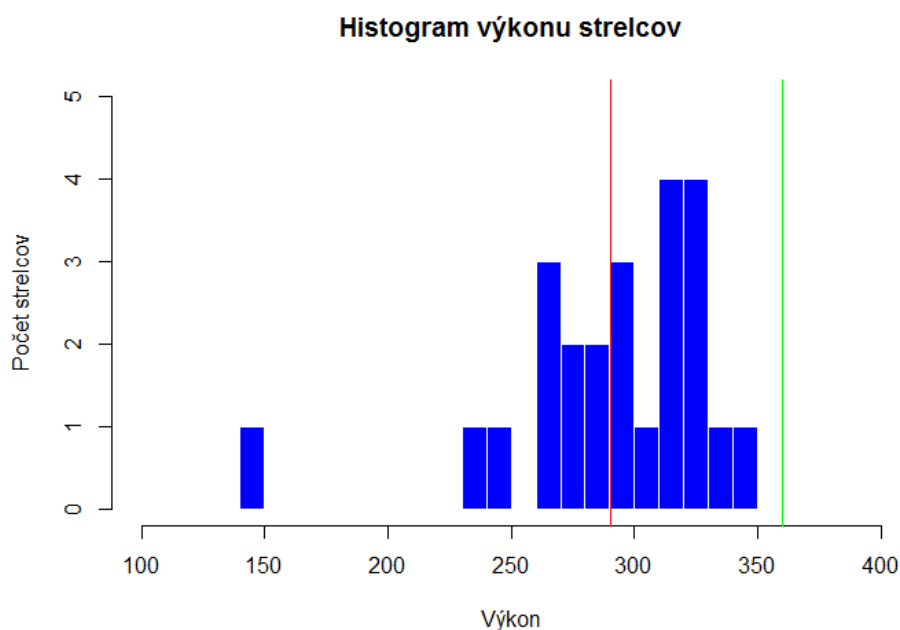
Vybrali sme preto strelcov, ktorý vo svojom preteku majú minimálne 3 neusporiadané šestice, keďže existuje určitá pravdepodobnosť, že strelec nastriela body aj usporiadane. V tejto časti sa nám naša vzorka zmenšila skoro o polovicu a ostalo nám 34 strelcov, ktorý spĺňajú doterajšie kritéria.

Narážame však na ďalší problém. Niektorí strelci sa vo zvyšnej vzorke nachádzajú viackrát, šiesti po 2-krát a jeden z nich dokonca 8-krát. Po vybratí jednej súťaže pre jedného strelca nám ostalo 24 vhodných vzoriek na testovanie. Následne sme pre testovanie zvolili tú polovicu strelca, v ktorej sa vyskytovali neusporiadané šestice častejšie. Pod polovicou rozumieme súbor prvých alebo druhých 36-ich výstrelov v poradí.

Ďalej pracujeme so súborom dát, ktorý obsahuje len tie informácie, ktoré vyhovujú našim kritériam.

2.2.1 Výkon strelcov

V súbore dát vybraných 24 strelcoch, s ktorým sa chystáme pracovať, je zaznamenaný priebežný výkon strelca, teda jednotlivé výstrely v čase. Ku každému lukostrelcovi sme si z dostupných dát vypočítali jeho celkový výkon, ktorý konkrétne uvádzame v prílohe B. Náhľad jeho rozloženia nám umožňuje histogram na obrázku 5.



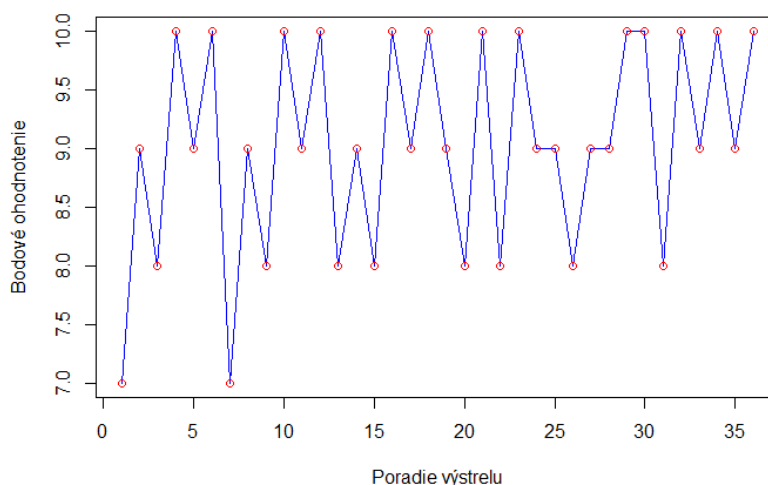
Obr. 5: Histogram celkového výkonu jednotlivých strelcov.

Lukostrelec počas preteku vo vybranej kategórii môže dosiahnuť maximálne 360 bodov, čo reprezentuje zelená čiara na obrázku 5. Môžeme si všimnúť, že výkony našich strelcov sa najčastejšie pohybujú v rozmedzí 260-330 bodov. V našom súbore dosiahol strelec 17 najmenší počet bodov 143 a naopak strelec 12 najväčší počet bodov 342. Priemer celkového výkonu je 290.3, čo je znázornené červenou čiarou v histograme.

2.3 Výber vhodného modelu

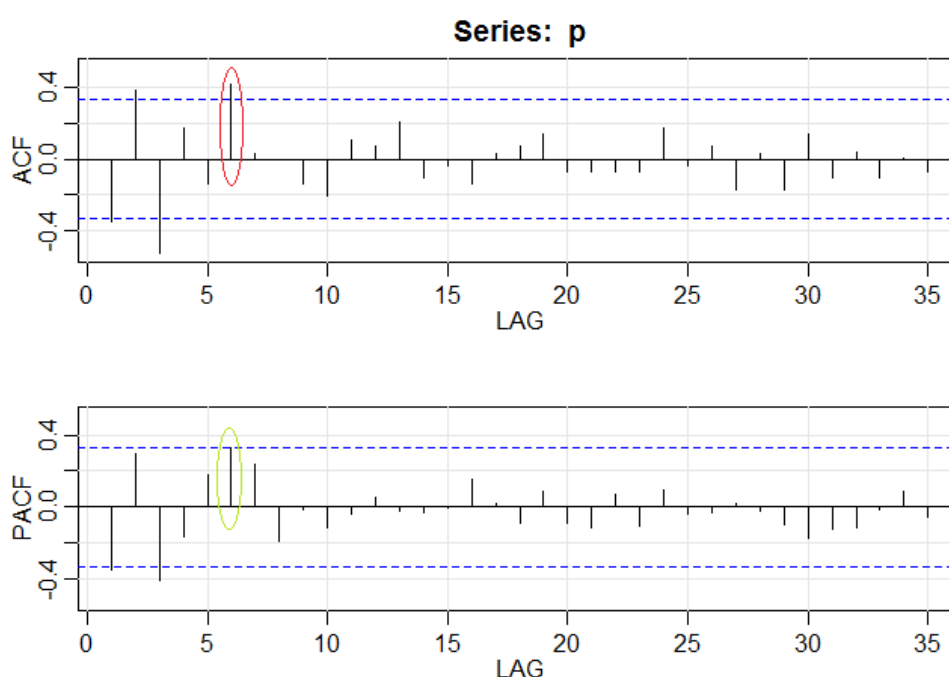
Teraz môžeme začať testovať dáta. Uvedieme si postup hľadania vhodného modelu, ktorý čo najlepšie popisuje výkon lukostrelca počas súťaže. Ako príklad použijeme strelca, ktorý je v súbore označený pod číslom 2. Testy vykonávame v prostredí R. Kód uvádzame v prílohe A. V tejto podkapitole opisujeme postup testovania v niekoľkých krokoch.

1. Po načítaní dát si vytvoríme maticu, v ktorej jeden riadok predstavuje jedného strelca a jeho výstrely v časovom poradí. Náhľad matice uvádzame v prílohe A.
2. Následne zvolíme číslo strelca, ktorého chceme testovať. V našom príklade je to strelec 2. Jeho výkon je zaznamenaný v druhom riadku matice.
3. Vytvoríme si časový rad výkonu strelca, ktorý je zobrazený na obrázku 6.



Obr. 6: Časový rad výkonu strelca.

4. Pozrieme sa na funkcie ACF a PACF procesu, ktoré nám ponúka obrázok 7. Vykreslíme ich príkazom `acf2()` z knižnice `astsa`. Pomocou týchto funkcií sa vieme orientovať, či je daný proces AR, MA alebo zmiešaný ARMA. V prípade, že by sa ACF vynulovala po q pozorovaniach, proces by sme modelovali pomocou $MA(q)$ modelu. Naopak, ak by sa to stalo vo funkcii PACF po p pozorovaniach, proces by sme modelovali $AR(p)$ modelom. Ak to neplatí, proces modelujeme pomocou zmiešaných $ARMA(p,q)$ modelov.



Obr. 7: Autokorelačná a parciálna autokorelačná funkcia procesu.

V našom prípade si na obrázku 7 môžeme všimnúť, že šiesta autokorelácia presahuje interval spoľahlivosti pre jej nulovosť. Ak sa pozrieme na parciálnu autokoreláciu, môžeme vidieť, že jej šieste pozorovanie je síce na hranici, ale je stále dosť veľká. To značí možnú sezónnosť procesu, ktorú predpokladáme.

5. Ďalším krokom je testovanie stacionarity procesu pomocou ADF testu, ktorý zisťuje prítomnosť jednotkového koreňa. V prostredí R na to slúži funkcia `ur.df()` z knižnice `urca` s parametrami

- `type`: možnosti drift, trend alebo none,

- lags: maximálny počet lagov,
- selectlags: AIC, BIC -kritérium, podľa ktorého sa vyberá počet lagov.

V našej práci uvažujeme proces s konštantou a preto na mieste parametra *type* uvádzame možnosť *drift* pre každého strelca.

Po spustení funkcie nám program ponúka výsledok vo forme, aký si môžeme všimnúť na obrázku 8. Pracujeme na kritickej úrovni 5 % a teda platí, že ak je testová štatistika menšia ako spomínaná úroveň, potom hypotézu H_0 o prítomnosti jednotkového koreňa zamietame, čo predstavuje aj naša konkrétna situácia. Znamená to, že proces je stacionárny.

```

value of test-statistic is: -3.227 5.2086

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.58 -2.93 -2.60
phi1  7.06  4.86  3.94

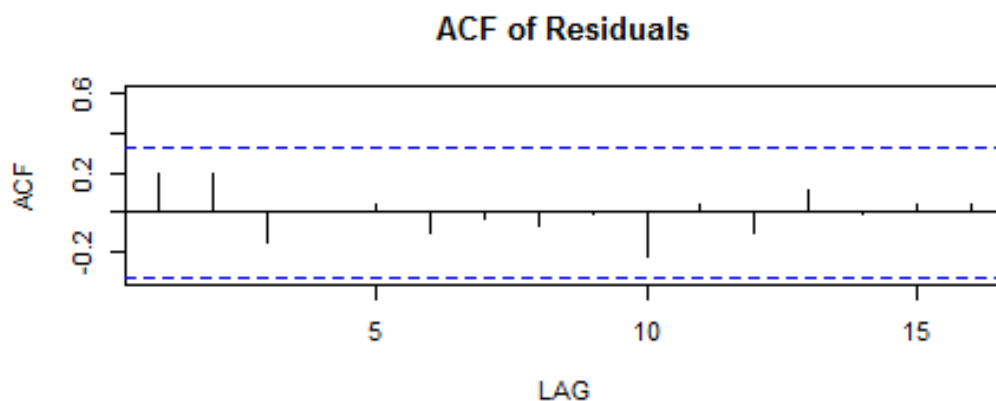
```

Obr. 8: Výsledok ADF testu.

V prípade, že sa stacionarita procesu potvrdí, pokračujeme posledným krokom 8 a to odhadom modelu. Inak sa vrátíme na krok 6.

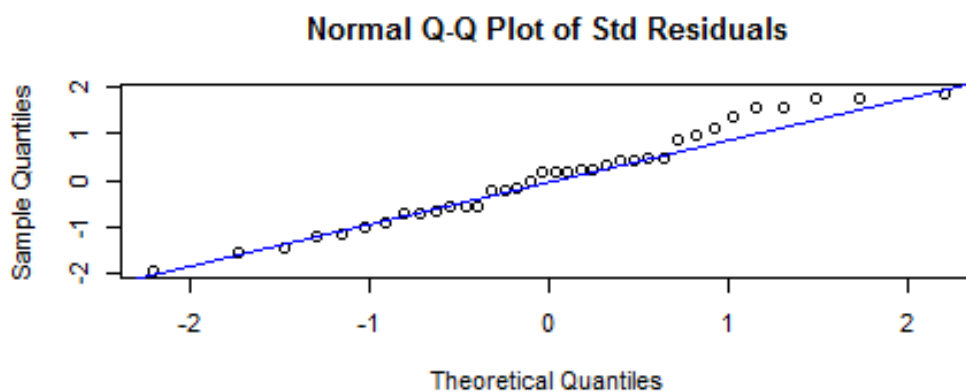
6. Ak testovaný proces obsahuje jednotkový koreň, je potrebné dáta diferencovať pomocou funkcie *dif*(*f*). Keďže predpokladáme, že naše procesy majú sezónny charakter, budeme používať sezónne diferencie. Následne sa vraciame na krok 5.
7. Krok 6 opakujeme dovtedy, kým nie je proces po niekoľko násobnom diferencovaní stacionárny.
8. Hľadáme SARIMA(p, k, q) \times (P, K, Q)_{*s*} model, ktorý čo najlepšie popisuje daný časový rad pomocou funkcie *sarima*(*f*), kde $s = 6$, pretože predpokladáme sezónnosť po nastrielaných šesticich.

Pre strelca č. 2 sme testovali rôzne modely. Jedným z nich bol SARIMA(1, 0, 0) \times (1, 0, 0)₆, ktorého výsledky testov uvádzame v obrázkoch 9 - 11.



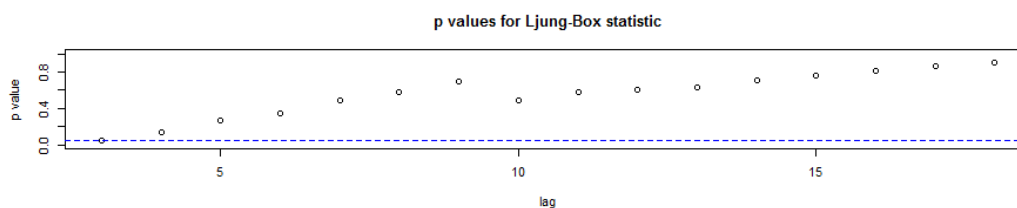
Obr. 9: ACF rezíduí modelu $SARIMA(1,0,0) \times (1,0,0)_6$

Na obrázku 9 je vykreslená autokorelačná funkcia rezíduí testovaného strelca pre model $SARIMA(1,0,0) \times (1,0,0)_6$. Tie sa nachádzajú v hranici intervalov spoľahlivosti, čo indikuje nulovú ACF funkciu. V prípade, že jednotlivé autokorelácie rezíduí sú nulové, znamená to, že sme vylúčili všetku deterministickú zložku a model potom popisuje dáta. Rezíduá sú teda biely šum a to chceme.



Obr. 10: Normalita rezíduí

Z obrázka 10 vieme vyčítať, či sú rezíduá testovaného modelu z normálneho rozdelenia. Riadia sa ním práve vtedy, keď čo najviac kopírujú zobrazenú priamku. Zdá sa, že to v našom príklade platí.



Obr. 11: Ljung-Boxova štatistika

Na obrázku 11 môžeme vidieť výsledky Ljung-Boxovej štatistiky pre jednotlivé pozorovania v čase. Vykreslené sú hodnoty p -value. Tie chceme mať nad hranicou 5%, pretože vyžadujeme nulovosť všetkých autokorelácií. Táto podmienka sa požaduje v hypotéze H_0 , ktorú nezamietame práve vtedy, ak je hodnota p -value $> 5\%$.

V našom konkrétnom príklade je výsledok prvého pozorovania na hrane, ale ostatné sú v pozícii, ktorú jednoznačne chceme. Výkon tohto strelca je teda vhodne popísaný testovaným modelom $SARIMA(1, 0, 0) \times (1, 0, 0)_6$.

Je však možné, že pre jeden proces nájdeme viac modelov, ktoré spĺňajú predchádzajúce kritéria. Rozhodnúť, ktorý z nich je lepší, kvalitnejší pri zadaných dátach vzhľadom na ostatné modely, nám pomáhajú Akaikeho (AIC) a Bayesovo (BIC) informačné kritéria. AIC je dané vzťahom

$$AIC = \log \hat{\sigma}^2 + 2 \frac{p + q + 1}{T}$$

a BIC

$$BIC = \log \hat{\sigma}^2 + \frac{p + q + 1}{T} \log T.$$

Po spustení funkcie `sarima()` pre konkrétny model nám prostredie R tieto hodnoty automaticky vypočíta. Vo všeobecnosti platí, že čím menšie sú tieto hodnoty, tým je daný model lepší.

2.4 Výsledky

Prvým testom, ktorý sme aplikovali na každého lukostrelca ešte pred hľadaním vhodného modelu, bol ADF test, test stacionarity. V teste sme si zvolili maximálny počet lagov 10, čo predstavuje maximálne oneskorenie AR členov. Hypotéza H_0 , ktorá hovorí

o prítomnosti jednotkového koreňa, bola týmto testom zamietnutá v 13 prípadoch, čo predstavuje 54%.

Každý časový rad je vytvorený z pomerne malého počtu pozorovaní (36), a preto test zvyčajne nie je dost silný na to, aby zamietol hypotézu H_0 aj v prípade, že proces stacionárny je. Ale ak sa naozaj vyskytne nestacionárny proces, nie je možné ho v takom stave modelovať, je potrebné ho diferencovať. Keďže predpokladáme sezónnosť procesov, aplikovali sme sezónne diferencovanie. Takto sa z nestacionárneho procesu stal stacionárny, ktorý je pripravený na modelovanie.

Následne sme pre jednotlivých strelcov hľadali vhodné modely, ktoré čo najlepšie popisujú výkon v čase tak, ako je uvedené v predchádzajúcej podkapitole. Z nich sme vybrali 3 najčastejšie sa vyskytujúce modely a to:

1. SARIMA(1, 0, 0) \times (1, 0, 0)₆,
2. SARIMA(0, 0, 0) \times (1, 0, 0)₆,
3. SARIMA(1, 0, 0) \times (0, 1, 1)₆.

Naším cieľom je nájsť nejaké spoločné znaky a porovnať jednotlivé modely medzi sebou. V prílohe D uvádzame presné výsledky testov a modelovania pre všetky tri uvedené modely, ktorými sme modelovali každého vhodného strelca. Ďalej budeme porovnávať a analyzovať len prvé dva modely, v ktorých sa na rozdiel od tretieho modelu nevyskytuje sezónna diferencia.

Podľa Ljung-Boxovej štatistiky sa až 19 procesov z 24 dalo modelovať pomocou modelu 1, čo predstavuje 79 %-nú úspešnosť. Na druhej strane model 2 bol oproti nemu horší a vyhovoval len 12-im procesom, čo je 50 %. Zaujímavé je, že tí strelci, ktorých výkon sa nedal modelovať pomocou modelu 1, nedal sa následne modelovať ani modelom 2.

Medzi informačnými kritériami AIC a BIC oboch modelov nie je významný rozdiel ani pri jednom z testovaných strelcov, preto nevieme jednoznačne rozhodnúť, ktorý z modelov lepšie popisuje tento výkon.

Naše modely majú spoločnú prítomnosť jedného sezónneho AR člena, čo značí závislosť výstrelu od strely uskutočnenej o sezónu pred ním, v našom prípade je to 6 časových krokov dozadu. Rozdiel medzi týmito modelmi činí prítomnosť jedného AR

člena v modeli 1. Ten poukazuje na závislosť výstrelu od strely uskutočnenej pred ním. Teda vhodnosť prvého modelu je skoro o 30 % väčšia ako v modeli 2. Z dát, ktoré sme analyzovali vyplýva, že prítomnosť AR člena a tým závislosť po sebe idúcich výstrelcov je významná.

Ďalej predpokladáme, že model SARIMA $(1, 0, 0) \times (1, 0, 0)_6$ by mohol byť vhodným pre modelovanie výkonnosti lukostrelcov počas preteku. Popísaný je rovnicou

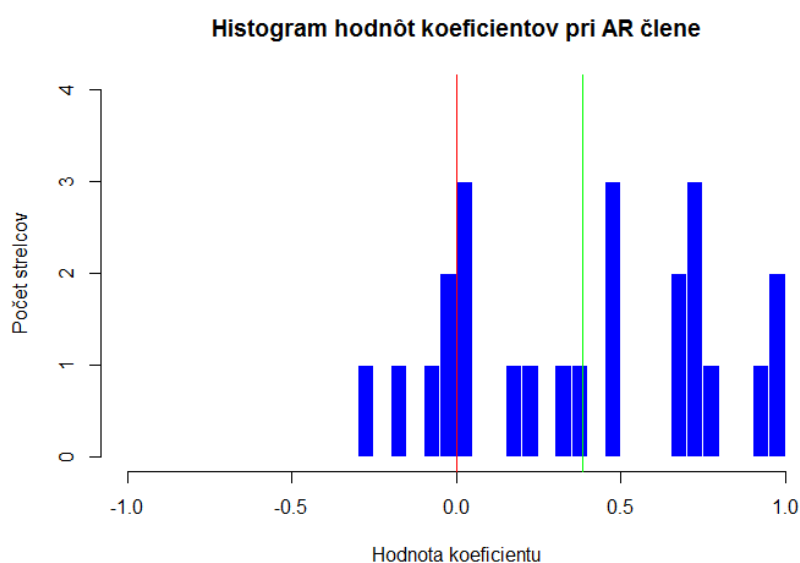
$$(1 - \alpha L)(1 - AL^6)x_t = u_t,$$

čo je ekvivalentné s

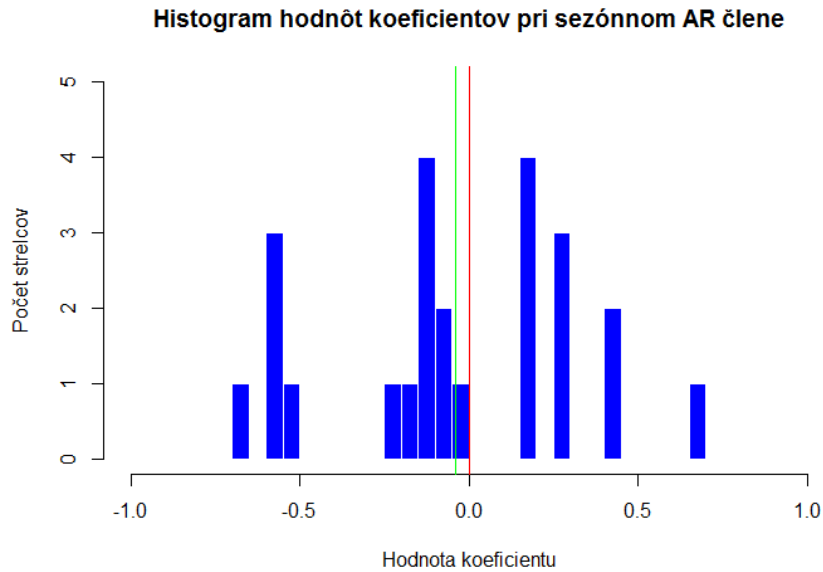
$$x_t = \alpha x_{t-1} + Ax_{t-6} - \alpha Ax_{t-7} + u_t,$$

kde α je koeficient pri AR člene a A je koeficient pri sezónnom AR člene.

V prílohe D sú uvedené hodnoty týchto koeficientov pre jednotlivé procesy, aj pre model 2. Na obrázku 12 a 13 uvádzame histogramy týchto hodnôt všetkých testovaných 24 strelcov pre model 1. Môžeme si všimnúť, že hodnoty α sú až v 19 prípadoch kladné. Priemerná hodnota je vyznačená zelenou čiarou a dosahuje hodnotu 0,3842792 a variácia nadobúda hodnotu 0,1489589. Kladná α značí pozitívnu závislosť, čo v našom prípade znamená, že lukostrellec v priemere strelí lepšie, ako strelil predchádzajúci výstrel.



Obr. 12: Histogram hodnôt koeficientov α modelu SARIMA $(1, 0, 0) \times (1, 0, 0)_6$



Obr. 13: Histogram hodnôt koeficientov A modelu SARIMA $(1, 0, 0) \times (1, 0, 0)_6$

Na obrázku 13 je zelenou farbou zobrazená priemerná hodnota sezónnych koeficientov rovná -0.04133333 . Záporný koeficient hovorí, že ak lukostrelec strelil v predchádzajúcej sade dobre, tak teraz strelí zle. Variancia dosahuje hodnotu 0.1336279 .

Naša testovaná vzorka, ktorú máme k dispozícii, je pomerne malá a nachádzajú sa v nej rôzni strelci. Nevieme z nej vyvodit' všeobecné jednoznačné závery o hodnotách koeficientov. Na to by sme potrebovali dostatok dát o lukostrelcoch, ktorých forma je porovnateľná. Na takejto sade dát by sa naše predpoklady dali testovať lepšie. Napriek tomu sa v nasledujúcich podkapitolách pokúsime otestovať možné závislosti.

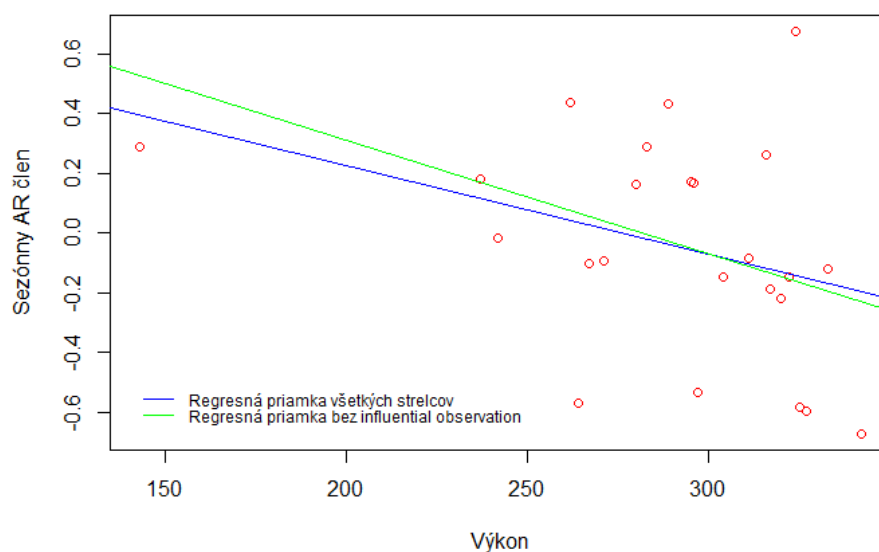
2.4.1 Závislosť hodnoty koeficientu pri sezónnom AR člene od výkonu

Predpoklad, že hodnota koeficientu pri sezónnom člene závisí od výkonu lukostrelca sme otestovali pomocou lineárnej regresie, ktorú vyjadruje rovnica

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

kde Y je vektor koeficientov pri sezónnom AR člene jednotlivých strelcov, ktorých modelujeme pomocou modelu SARIMA $(1, 0, 0) \times (1, 0, 0)_6$ a vektor X predstavuje celkový výkon jednotlivých strelcov.

Testovanie regresie vykonávame v prostredí R. Obrázok 14 nám ponúka grafické vy-



Obr. 14: Lineárna regresia

obrazenie testovanej lineárnej závislosti. Modrá priamka predstavuje závislosť, v ktorej sú zahrnutí všetci testovaní strelci. Avšak, ako si môžeme všimnúť, v našej sade sa nachádza jediný strelec s výrazne nižším výkonom, konkrétne 143 bodmi. Výkon ostatných lukostrelcov pochádza z intervalu (230,340), kde sú pomerne nahustení. Vieme teda skúmať správanie závislosti na tejto oblasti. Preto strelca, ktorý dosiahol výkon 143 bodov považujeme za tzv. *influential observation*, teda pozorovanie, ktoré výrazne ovplyvňuje výsledok regresie. Následne sme informácie o ňom vynechali a vytvorili nový model bez tohto pozorovania, ktorý je zobrazený zelenou priamkou.

Na obrázku 15 môžeme vidieť číselné výsledky testu o lineárnej regresii bez influential observation. Vyčítať vieme konkrétne hodnoty koeficientov nášho modelu a výsledky testov o ich nulovosti a ďalšie. Hodnota $p - value = 0.1569$ hovorí o význame našej regresie. Ak by sme brali do úvahy 5 %-nú kritickú hodnotu, musela by byť táto hodnota menšia ako 0.05, aby sme model považovali za významný a teda zamietali by sme hypotézu H_0 o tom, že regresia nie je významná. V našom prípade je táto hodnota väčšia o viac ako 0.1, teda test nepovažuje regresiu, ktorá predstavuje závislosť medzi hodnotami koeficientov pri sezónnom člene a celkovým výkonom strelca, za významnú.

```

Call:
lm(formula = SARM1[c(-17)] ~ 1 + vykon[c(-17)], x = TRUE)

Residuals:
    Min       1Q   Median       3Q      Max
-0.63953 -0.16548  0.00572  0.22292  0.83484

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.074964   0.773741   1.389   0.179
vykon[c(-17)] -0.003811   0.002596  -1.468   0.157

Residual standard error: 0.3574 on 21 degrees of freedom
Multiple R-squared: 0.0931,    Adjusted R-squared: 0.04991
F-statistic: 2.156 on 1 and 21 DF,  p-value: 0.1569

```

Obr. 15: Výsledok testovania lineárnej regresie

Najviac nás však zaujíma korelácia medzi testovanými veličinami. Tá je rovná odmocnine koeficientu determinácie, čo je rovné približne 0.305. Avšak test významnosti regresie nepotvrdil a potom aj koreláciu medzi testovanými veličinami nepovažujeme za význačnú. Pre všeobecnejšie závery a lepší model by bolo potrebné otestovať strelcov s porovnateľnými schopnosťami na väčšej vzorke.

2.4.2 Závislosť výstrelu od predchádzajúceho výstrelu

Rozhodli sme sa takisto testovať hypotézu, či sa medzi hodnotami dvoma po sebe idúcich výstrelov jednotlivých strelcov vyskytuje nejaká závislosť, iným spôsobom ako pomocou SARIMA modelovania. Výstrely sme rozdelili do dvoch skupín. V prvej skupine sa nachádzajú hodnoty výstrelov lepšie, ako je priemerná hodnota daného strelca a označíme ich ako *good*. V druhej sa nachádzajú naopak horšie hodnoty, ktoré označujeme ako *bad*. Kombinácie dvoch po sebe idúcich výstrelov, ktoré môžu nastať, sme si označili nasledujúcimi skratkami:

- *GG* označuje kombináciu dobrý-dobrý výstrel,
- *GB* dobrý-zlý,
- *BG* zlý-dobrý
- *BB* označuje kombináciu výstrelov zlý-zlý v poradí.

Následne sme označili všetky dvojice výstrelov príslušnou skratkou a zráтали počty jednotlivých kombinácií u každého strelca. Tieto počty sme následne zapísali do tabuľky, ktorú môžeme vidieť na obrázku 16. Celkový počet strelcov je 24 a ten zodpovedá počtu vytvorených tabuliek.

	G	B	
G	počet GG	počet GB	počet GG+BB
B	počet BG	počet BB	počet BG+BB
	počet GG+BG	počet GB+BB	počet GG+BB+BG+BB

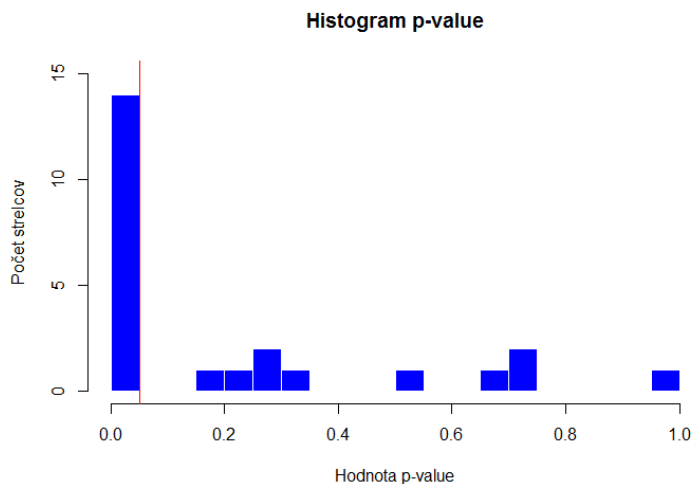
Obr. 16: Good/Bad výstrely

Pre testovanie hypotézy o závislosti dvoch po sebe idúcich výstrelov daného strelca sme použili test nezávislosti. Hypotézy tohto testu vyzerajú nasledovne:

H_0 : dva po sebe idúce výstrely sú nezávislé vs.

H_1 : dva po sebe idúce výstrely sú závislé.

Pre výpočet testovej štatistiky je potrebné vedieť súčty počtov daných pozorovaní v riadkoch a stĺpcoch tak, ako je to vyznačené v spomínanej tabuľke na obrázku 16. Túto testovú štatistiku porovnáваме v našom prípade s kritickou hodnotou rozdelenia χ_1^2 na hladine 5 %. Vieme si takisto dopočítať hodnotu p-value.



Obr. 17: Histogram p-value hodnôt

Platí, že ak je táto hodnota menšia ako 0.05, teda 5%, tak H_0 zamietame. Na obrázku 17 môžeme vidieť histogram týchto hodnôt testovaných strelcov. Hypotéza H_0 bola zamietnutá v 14 prípadoch z 24, čo predstavuje 58,3 %.

Tento test nezávislosti je oproti modelovaniu pomocou SARIMA procesov slabší, ale potvrdil predpoklad závislosti dvoch po sebe nasledujúcich výstrelov skoro v 60 %-ách strelcov, čo je celkom veľké číslo.

Avšak problémom aj v tomto prípade môže byť malý počet testovaných strelcov. Bolo by možno lepšie, ak by testovanie prebehlo na vzorke strelcov, ktorý sa nachádzajú v porovnateľnej forme.

2.4.3 Stredná hodnota v čase

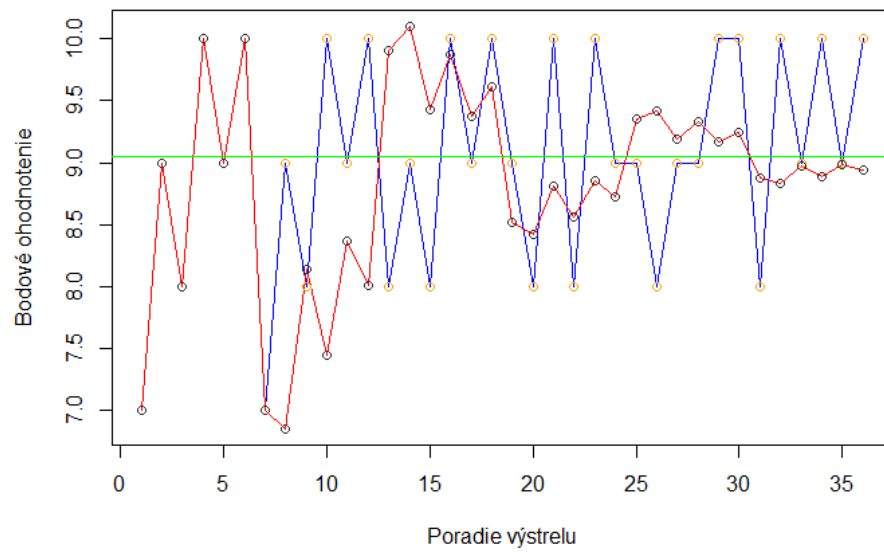
Na nasledujúcich troch obrázkoch sú zobrazené tri časové rady skutočných hodnôt výstrelov v čase pre troch strelcov. Sú na nich zachytené aj teoretické hodnoty výstrelov, kde prvých 7 hodnôt je pôvodných a následne z nich odhadujeme nasledujúce hodnoty, ktoré zodpovedajú modelu

$$x_t = \delta + \alpha x_{t-1} + Ax_{t-6} - \alpha Ax_{t-7},$$

kde rezíduá neuvažujeme, α a A sú hodnoty koeficientov procesu SARIMA $(1, 0, 0) \times (1, 0, 0)_6$ daného strelca a δ je konštanta, ktorú sme si dopočítali zo vzťahu

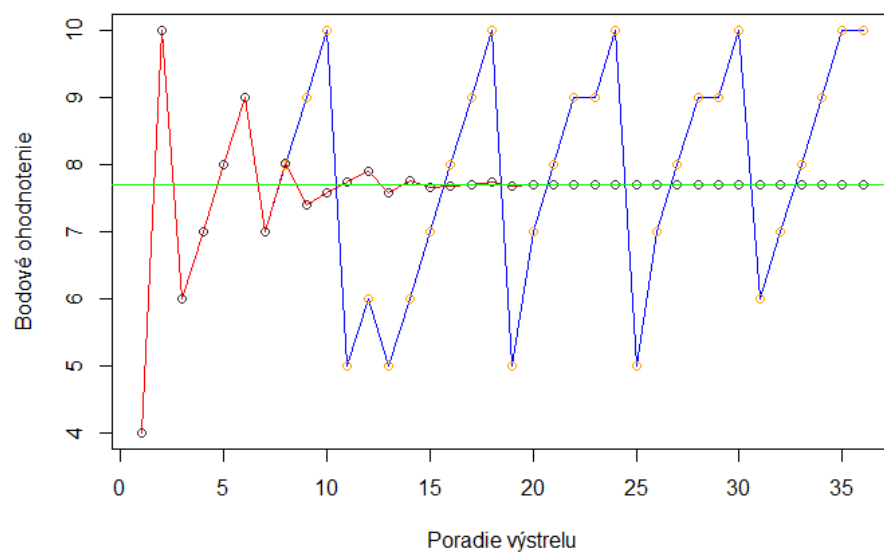
$$\delta = \hat{\mu}(1 - \alpha - A + \alpha A),$$

kde $\hat{\mu}$ je odhad strednej hodnoty, ktorý nám ponúkol program R pri modelovaní. Očakávame tak, že sa tieto teoretické hodnoty budú pohybovať okolo strednej hodnoty $E(x_t)$, ktorá je odhadnutá pomocou $\hat{\mu}$.



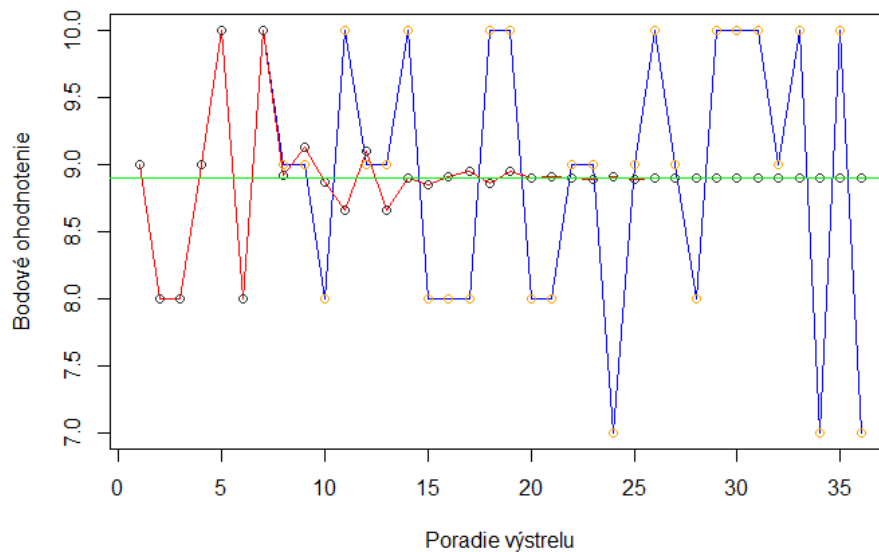
Obr. 18: Časový rad strelec 2

Na obrázku 18 je zobrazený proces strelca 2, ktorého hodnota α je kladná a koeficient A je záporný. Obrázok 19 ponúka časový rad strelca 9. Jeho hodnoty α a A sú kladné.



Obr. 19: Časový rad strelec 9

Posledný časový rad strelca 24 je zobrazený na obrázku 20. Koeficienty α a A pre tohto strelca dosahujú záporné hodnoty. Konkrétne čísla je možné vidieť v spomenu-
tej prílohe C. Na spomenutých obrázkoch je zelenou čiarou zobrazený odhad strednej
hodnoty.



Obr. 20: Časový rad strelec 24

Na všetkých troch obrázkoch si môžeme všimnúť, ako sa teoretické hodnoty pribli-
žujú k strednej hodnote $E(x_t)$ v čase.

3 Predikcie výkonu lukostrelcov

3.1 Formulácia problému

V tejto časti sa pokúsime navrhnúť čo najlepší spôsob, akým vieme odhadnúť formu lukostrelca na nasledujúcom preteku.

Ako v každom športe, aj v tomto sa usporadúvajú rôzne súťaže od najnižších tried až po tie najvyššie, ktoré vedú k reprezentácii krajiny. V slovenskej lukostrelbe sa tento výber reprezentanta na nominačný pretek určuje tak, že potenciálny strelec musel za posledných 9 mesiacov "splniť limit" a to streliť hodnotu vyššiu ako je stanovené číslo. V prípade, že takýchto kandidátov je viac, vyberie sa ten, ktorého medián z posledných troch súťaží je vyšší.

A to nás motivuje k testovaniu a porovnávaniu ďalších rôznych metód, ktorými môžeme predikovať výkon pomocou výsledkov z predchádzajúcich súťaží. Naším cieľom je nájsť spôsob, ktorý ponúka čo najkvalitnejší odhad pre predikovanie a následný výber lukostrelca, ktorého forma sa zdá byť momentálne najlepšia spomedzi ostatných. Preto vykonáme výpočet odhadov pomocou vybraných metód na historických dátach, ktoré máme k dispozícii. Tieto metódy sme zvolili podľa vlastného uváženia o kvalite ich novej výpovednej hodnoty. Odhady vypočítame aj pomocou súčasnej metódy, ktorou sa uskutočňuje výber strelca. Výpočet týchto hodnôt robíme len pomocou mediánu posledných troch súťaží rovnakej kategórie, akej je nominačná súťaž. Splnenie prvej podmienky tohto spôsobu neskúmame, pretože ju automaticky považujeme za splnenú z dôvodu, že sa skúmaní strelci reprezentačných súťaží naozaj zúčastnili.

Odhady budeme porovnávať so skutočnými hodnotami výkonov zo súťaží, na ktoré bola potrebná nominácia lukostrelca a následne budeme medzi sebou porovnávať kvalitu týchto odhadov získaných rôznymi metódami.

3.2 Popis dát

V tejto téme budeme na testovanie používať dáta, ktoré nám poskytol Slovenský lukostrelecký zväz (SLZ). Nachádzajú sa v ňom údaje o výkonoch slovenských lukostrelcov rôznych kategórií, ktorí sa zúčastňovali rozličných súťaží. Preteky, pre ktoré je nutná nominácia strelca, sme označili ako nominačné.

Výber zúžený na strelcov, ktorí strielali na nominačných pretekoch, sa týkal 18 lukostrelcov, 6 žien a 12 mužov. Z nich sa dvaja strelci a dve strelkyne zúčastnili niekoľkých zo spomínaných nominančných súťaží, ostatní boli prítomní na jednej, prípadne dvoch. Dokopy budeme odhadovať výkony 45 vybraných pretekov.

Vytvorili sme si nový súbor dát, ktorý obsahuje len údaje vybraných strelcov. Pred nominačnými súťažami mal každý strelec zaznamenaných niekoľko starších výkonov z rôznych pretekov.

Problémom však bolo, že počas zimného obdobia sa strieľa v hale za iných podmienok ako vonku, kde sa súťaži počas letného obdobia. Rozdiel je hlavne vo vzdialenosti a veľkosti terča, no aj v počasí. Preto bolo potrebné približné prepočítavanie týchto výkonov zo súťaží, ktoré sa líšili od nominačnej. Na to nám poslužil program v excelovskom prostredí, ktorý používa SLZ. Podrobný popis uvádza [4]. Prepočítané halové výkony boli u všetkých strelcov relatívne podobné terčovým výkonom. Následne sme už mali splnené porovnateľné podmienky pri všetkých súťažiach, ktoré boli pripravené pre ďalšie výpočty.

3.3 Metódy pre výpočet odhadov

Naším cieľom je odhadnúť výkon, ktorý podá daný strelec na preteku, z historických dát. Na tento odhad je možné použiť rôzne metódy, pomocou ktorých počítame odhady strednej hodnoty ďalšieho výkonu lukostrelca. V našej práci sme si vybrali štyri, ktoré v tejto kapitole podrobne opíšeme. Na odhad používame údaje zo starších súťaží, ktoré sme si časovo ohraničili. Zvolili sme interval, kde najstaršia súťaž, ktorú berieme do úvahy, nie je staršia ako 300 dní od nominovanej súťaže, čo je približne 10 mesiacov.

3.3.1 Aritmetický priemer

Prvou metódou pre výpočet odhadu ďalšieho výkonu je aritmetický priemer. Popísaný je rovnicou:

$$\hat{X}_1 = \frac{1}{n} \sum_{i=1}^n X_i,$$

kde n je počet súťaží, ktoré berieme do úvahy, X_i predstavujú výkony jednotlivých súťaží a i je poradie súťaže. Toto značenie budeme ďalej používať v ďalších metódach. Aritmetický priemer je vo všeobecnosti veľmi citlivý na zmenu pri výskyte extrémnych

hodnôt, avšak nesie v sebe veľa informácií. Ak predpokladáme, že sa nezávislé náhodné premenné X_i riadia rozdelením $N(\mu, \sigma^2)$, potom stredná hodnota priemeru je μ a disperzia $\frac{\sigma^2}{n}$. Aritmetický priemer je potom najlepší lineárny nevychýlený odhad strednej hodnoty.

3.3.2 Maximum

Druhou metódou je odhad pomocou maxima z množiny výkonov, ktoré boli dosiahnuté za vybrané časové obdobie. Tento spôsob bol kedysi využívaný v SLZ. Reprezentuje ho rovnica:

$$\hat{X}_2 = \max\{X_i\},$$

pre $i = 1, \dots, n$. Nie je najlepším odhadom formy streľca a toho, aký výkon podá, pretože vypustil všetky ostatné informácie o tom, ako sa správal výkon lukostrelca na iných súťažiach. Dá sa ukázať, že ak sú $X_i \sim N(\mu, \sigma^2)$, tak stredná hodnota maxima splňa nerovnosť $E(\max(X_i)) \leq \sigma\sqrt{2 \log n} + \mu$.

3.3.3 Medián

Treťou vybranou metódou je medián. V nej je najprv potrebné usporiadať výkony X_i vzostupne a následne nájsť stredný člen tohto radu. V prípade, že je počet súťaží nepárny, medián vypočítame pomocou vzorca:

$$\hat{X}_3 = X_{\frac{n+1}{2}}$$

a ak je počet párný:

$$\hat{X}_3 = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}.$$

Medián nám rozdelí množinu súťaží, ktoré berieme do úvahy, na dve skupiny v rovnakom pomere. Takže súťaže s nižším výkonom ako je medián tvoria 50 % z celkového počtu a to isté platí aj pre súťaže, kde bol dosiahnutý lepší výkon ako medián. Takisto ak predpokladáme $X_i \sim N(\mu, \sigma^2)$, stredná hodnota mediánu je μ s disperziou $\frac{\pi\sigma^2}{4k}$, kde k vypočítame zo vzťahu $n = 2k + 1$ pre n nepárne.

3.3.4 Vážený aritmetický priemer

Poslednou vybranou metódou je vážený priemer. Váhy sme si navolili v závislosti od času. Čím je staršia súťaž, ktorú berieme do úvahy, tým menšiu váhu jej pridáme. Keďže sme si ako časový interval zvolili obdobie 300 dní, postupne po 15 dňoch túto váhu meníme tak, ako je uvedené v nasledujúcej tabuľke na obrázku 21. Ich výpočet reprezentuje vzorec:

$$w_i(t_i) = \left\lceil \frac{300 - t_i}{15} \right\rceil,$$

kde t_i je počet dní medzi aktuálnym dátumom, ku ktorému chceme odhad vypočítať a dátumom súťaže, na ktorej strelec dosiahol výkon X_i a $i = 1, 2, \dots, n$, kde n je počet súťaží, ktoré berieme do úvahy.

Výpočet váženého priemeru charakterizuje rovnica:

$$\hat{X}_4 = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i},$$

kde w_i predstavuje váhu zodpovedajúcu danému výkonu zo súťaže i .

Počet dní	0-15	16-30	31-45	46-60	61-75	76-90	91-105	106-120	121-135	136-150
Váha	20	19	18	17	16	15	14	13	12	11
Počet dní	151-165	166-180	181-195	196-210	211-225	226-240	241-255	256-270	271-285	286-300
Váha	10	9	8	7	6	5	4	3	2	1

Obr. 21: Váhy

Za predpokladu $X_i \sim N(\mu, \sigma^2)$ je vážený priemer z normálneho rozdelenia so strednou hodnotou μ a disperziou $\sigma^2 \frac{\sum_{i=1}^n w_i^2}{(\sum_{i=1}^n w_i)^2}$.

Predpokladáme, že táto metóda by mohla byť najlepšia pre odhad ďalšieho výkonu strelca. A to preto, lebo ako sme spomínali, výkon sa v čase mení a čím je výkon starší, tým menšiu výpovednú hodnotu o momentálnej forme lukostrelca dostávame.

3.4 Výber súťaží a výpočtový program

Na výpočet odhadov pomocou vybraných metód, ktoré sú popísané v predchádzajúcej podkapitole, sme si vytvorili program v prostredí excel, ktorého fungovanie v krátkosti opíšeme.

Program obsahuje predvolenú tabuľku, do ktorej je potrebné nakopírovať údaje o výsledkoch súťaži tak, ako sú zapísané v pôvodnom súbore dát, ktorý máme k dispozícii. Na výpočet odhadu ďalšieho výkonu lukostrelca nám poslúžia jeho historické výkony. Avšak forma každého strelca sa v čase mení, preto volíme určité časové obmedzenie, ktoré by mali vhodné súťaže pre odhad výkonu spĺňať.

Najstaršia súťaž, ktorú budeme brať do úvahy, nesmie byť staršia ako 300 dní od nominačného preteku, čo je približne 10 mesiacov. Po vložení údajov do tabuľky a zadaní hraničného dátumu, program podľa spomenutého časového kritéria označí tie súťaže, ktoré ho spĺňajú. Potom sa už automaticky spočítajú odhady pomocou vybraných metód len z týchto pretekov a taktiež hodnoty odhadov získaných pomocou súčasnej metódy.

3.5 Výsledky

Vo vytvorenom programe sme vypočítali hodnoty výkonov pre jednotlivé nominačné súťaže. Výsledky výpočtov uvádzame v prílohe D aj s hodnotami reálnych výkonov na nominačnej súťaži.

Najprv sme sa pozreli na rozdiely jednotlivých odhadov a reálnych výkonov, čo je ekvivalentné absolútnym chybám. Tie až v 20-ich prípadoch boli najmenšie pri váženom priemere. V ostatných bol tento odhad vo väčšine prípadoch druhým najlepším, prípadne nastrelený reálny výkon bol buď nadpriemerný alebo podpriemerný. V 9-ich prípadoch zo zvyšných s najmenším rozdielom dominoval aritmetický priemer. Medián a maximum boli úspešné odhady v 8-ich prípadoch každý.

Pre meranie kvality odhadov sme uvažovali absolútnu priemernú chybu, ktorá uvažuje priemernú chybu predikcie preteku. Daná je vzorcom

$$d_m = \frac{1}{n} \sum_{j=1}^l \sum_{i=1}^{n_j} |x_{ji} - \hat{x}_{ji}|,$$

kde $m = 1, 2, \dots, 5$ je index pre metódu², pomocou ktorej bol odhad vypočítaný, n je celkový počet súťaží, $j = 1, 2, \dots, l$, kde l je počet strelcov, $i = 1, 2, \dots, n_j$ a n_j je počet nominačných súťaží daného strelca. Toto označenie budeme používať aj v ďalších vzťahoch.

²1-aritmetický priemer, 2-maximum, 3-medián, 4-vážený aritmetický priemer, 5-súčasná metóda

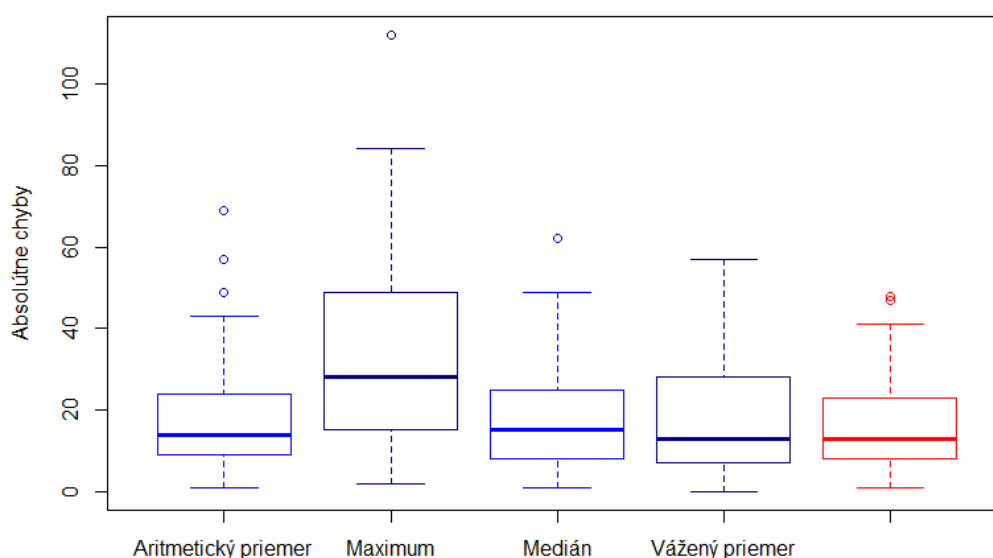
Vypočítali sme taktiež váženú absolútnu priemernú chybu pre všetky odhady. Tá uvažuje priemernú chybu predikcie strelca. Daná je vzťahom

$$d_m^v = \frac{1}{l} \sum_{j=1}^l \frac{1}{n_j} \sum_{i=1}^{n_j} |x_{ji} - \hat{x}_{ji}|.$$

	aritmetický priemer	maximum	medián	vážený priemer	súčasná metóda
priemerná chyba	19	33,6	18,4	18,2	17,8
vážená priemerná chyba	18,9	30,6	17,7	18,4	15,5

Obr. 22: Absolútne priemerné chyby

V tabuľke na obrázku 22 sa nachádzajú hodnoty týchto chýb pre vybrané metódy. Najmenšiu priemernú chybu z nami uvažovaných odhadov dosahuje vážený priemer. Aritmetický priemer, medián a vážený priemer dosahujú porovnateľné výsledky. Výrazne zaostáva maximum, ktoré bolo využívané ako odhad v starej metodike SLZ. Súčasný spôsob SLZ vykazuje lepšie predikčné schopnosti ako nami uvažované metódy.

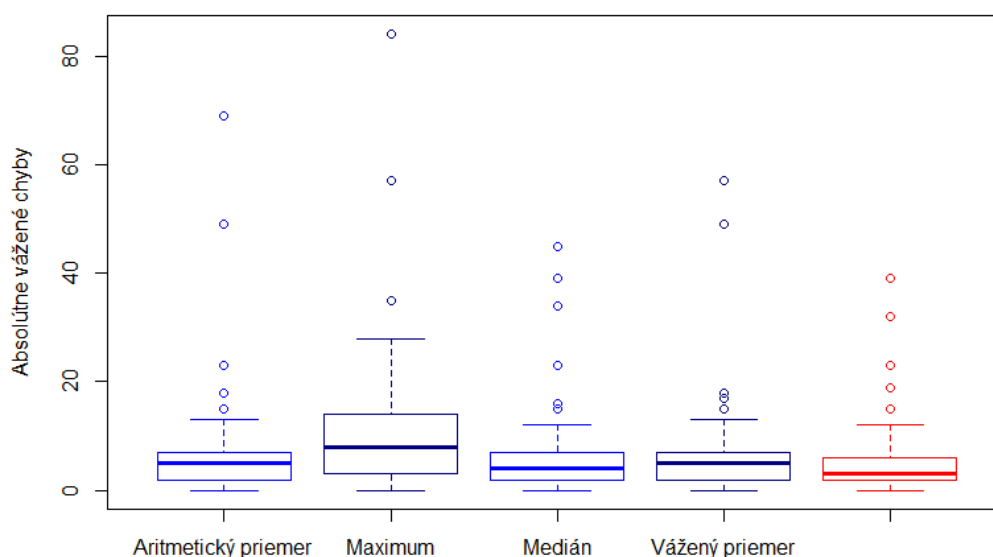


Obr. 23: Boxplot absolútnych chýb

Obrázok 23 nám ponúka náhľad na boxploty priemerných absolútnych chýb jednotlivých pozorovaní.

Je vidieť, že vážený priemer neobsahuje outlierov a medián chýb je pri všetkých metódach skoro rovnaký, okrem maxima. Jeho hodnota je trocha vyššia a takisto skupina prostredných 50% hodnôt je o niečo vyššia oproti ostatným, čo značí väčšie chyby odhadov od reálnych hodnôt a je teda menej spoľahlivým odhadom.

Podobne na obrázku 24 môžeme vidieť boxploty vážených absolútnych chýb. Týmito chybami sme sa snažili odstrániť možné ovplyvnenie výsledku rôznym počtom súťaží, ktoré prislúchajú jednotlivým strelcom. Tu sa už v testovaných metódach nachádza celkom veľa outlierov, no okrem metódy maxima sa 50% prostredných hodnôt chýb nachádza na rovnakej úrovni a na pomerne malom intervale. Je tak ťažšie rozhodnúť o tom, ktorá metóda je najlepšia, no vieme povedať, že maximum to nebude.



Obr. 24: Boxplot vážených absolútnych chýb

Rozumný spôsob porovnania kvality odhadov sú priemerné kvadratické chyby definované ako:

$$s^2 = \frac{1}{n} \sum_{j=1}^l \sum_{i=1}^{n_j} (x_{ji} - \hat{x}_{ji})^2,$$

a smerodajné odchýlky, ktoré sú ich odmocninou a sú rovné s . Podobne ako pri abso-

lútnych chybách sme si vypočítali aj vážené priemerné kvadratické chyby:

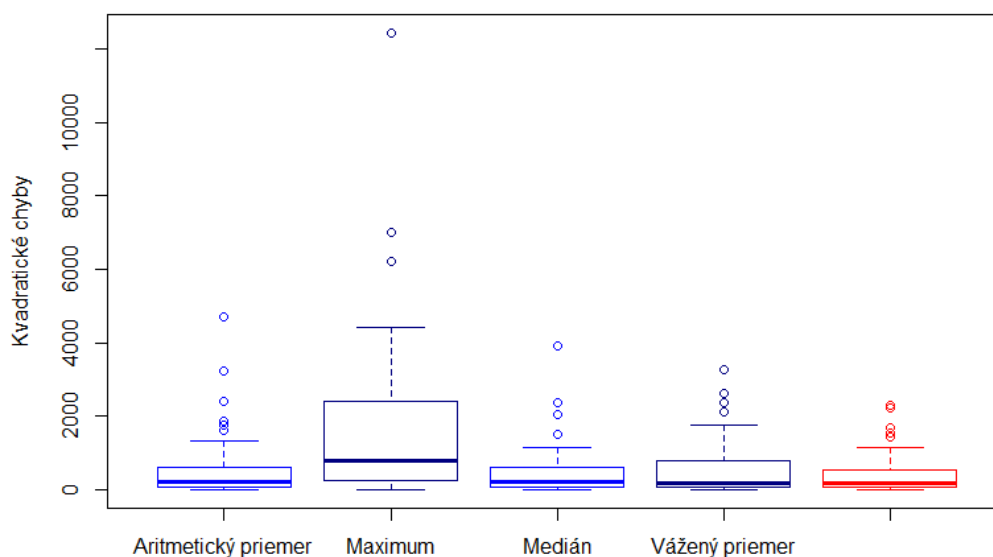
$$s^2 = \frac{1}{l} \sum_{j=1}^l \frac{1}{n_j} \sum_{i=1}^{n_j} (x_{ji} - \hat{x}_{ji})^2.$$

kde sú označenia ekvivalentné s označeniami z predchádzajúcich vzťahov.

	aritmetický priemer	maximum	medián	vážený priemer	súčasná metóda
priemerná kvadratická chyba	590,3	1695,5	542,3	548	479,6
smerodajná chyba PKO	24,3	41,2	23,3	23,4	21,9
vážená priemerná kvadratická chyba	655,8	1458,3	510,7	571,7	403,4
smerodajná chyba VPKO	25,6	38,2	22,6	23,9	20,1

Obr. 25: Priemerné kvadratické chyby

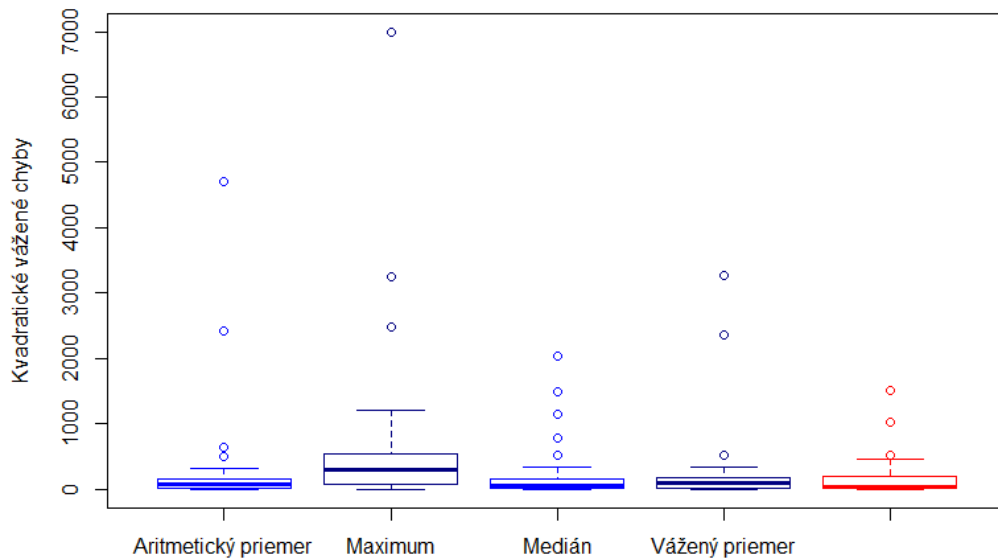
V tabuľke na obrázku 25 sú uvedené hodnoty priemerných kvadratických a smerodajných chýb. Môžeme si všimnúť, že medián je lepším odhadom ako priemer. Najhorším spôsobom odhadu je znova maximum a najlepšie predikčné schopnosti znova vykazuje súčasná metóda.



Obr. 26: Boxplot kvadratických chýb

Na obrázkoch 26 a 27 uvádzame boxploty aj pre kvadratické chyby. Podľa týchto boxplotov sa ostatné metódy zdajú byť porovnateľné a ťažko tak určiť jednoznačne

najlepšiu metódu pre odhad výkonu. Vieme však povedať, že najmenšiu výpovednú hodnotu predstavuje maximum.



Obr. 27: Boxplot vážených kvadratických chýb

Podľa výsledkov meraní kvality odhadov môžeme zhodnotiť, že najmenej presný je odhad pomocou maxima. Na druhej strane najlepšie predikcie poskytovala súčasná metóda a javí sa tak ako najkvalitnejšia spomedzi testovaných spôsobov.

Kvalitnejšie výsledky by sme dostali v prípade, že by sme mali k dispozícii väčší počet dát a zároveň viac rôznych strelcov, aby bola zabezpečená nezávislosť.

Záver

V práci sme rozoberali dva hlavné problémy. Prvý problém, ktorému sme sa venovali bolo testovanie existencie závislosti medzi výstrelmi v čase počas súťaže v športovej lukostrelbe. Analyzovanú vzorku tvorili výstrely od 24 lukostrelcov.

Závislosť sme modelovali pomocou autoregresných procesov. Konkrétne sme uvažovali SARIMA procesy s predpokladom sezónnosti po šiestich výstreloch. Následne sme dva modely, ktoré boli najčastejšie vyhovujúce pre testované dáta, medzi sebou porovnávali. Boli to $SARIMA(1, 0, 0) \times (1, 0, 0)_6$ (model 1) a $SARIMA(0, 0, 0) \times (1, 0, 0)_6$ (model 2), ktorých spoločným znakom bola prítomnosť jedného sezónneho AR člena, čo predstavuje závislosť výstrelu od výstrelu uskutočneného 6 časových krokov dozadu. Líšili sa len prítomnosťou jedného AR člena, čo určuje mieru závislosti po sebe idúcich výstrelov.

Model 1 bol vhodný v 79 %-ách prípadov. Oproti tomu bol model 2 vhodný v 50 %-ách. Závislosť dvoch po sebe idúcich výstrelov sme sa pokúsili otestovať aj slabším testom nezávislosti. Tá bola potvrdená u skoro 60 % strelcov. Ďalej sme závislosť strely od výstrelu o jeden krok a o 6 krokov dozadu považovali za významnú a pracovali sme s modelom 1.

Pomocou metód lineárnej regresie sme testovali, či hodnota koeficientu pri sezónnom AR člene závisí od celkového výkonu lukostrelca. V regresii sa vyskytovalo jedno pozorovanie, ktoré výrazne ovplyvňovalo regresiu, tzv. influential observation. Toto pozorovanie sme vynechali a vytvorili model so zvyšnými 23 strelcami. Avšak test významnosti regresie **nepotvrdil** významnosť nášho testovaného modelu. Korelácia medzi testovanými veličinami dosahovala hodnotu 0.305, čo však pri našom počte pozorovaní nie je štatisticky významné. Testovanie by bolo vhodné vykonať na vzorke strelcov, ktorí majú porovnateľné výkony a tak jednoznačne potvrdiť, resp. vyvrátiť predpokladané hypotézy.

Druhým problémom, ktorému sme sa venovali v tretej kapitole, je predikovanie výkonu strelca na súťaži. Motiváciou pre porovnávanie rôznych metód pre tento odhad bola skutočnosť, že na nominačné súťaže je potrebný výber reprezentanta, ktorý s najväčšou pravdepodobnosťou strelí najlepší výkon a je teda v lepšej forme oproti ostatným strelcom. Tento výber sa momentálne v slovenskej lukostrelbe uskutočňuje

tak, že možný reprezentant musel za uplynulých 9 mesiacov dosiahnuť výkon väčší, ako je vopred stanovené číslo. Ak je týchto uchádzačov viac, vyberie sa ten, ktorý má z posledných troch súťaží najväčší medián. Odhady pomocou tejto metodiky sme ďalej porovnávali s ďalšími nami zvolenými štyrmi metódami. Tými boli aritmetický priemer, maximum, medián a vážený priemer.

Analyzovaný súbor obsahoval 45 súťaží, ktorých sa zúčastnilo 18 strelcov.

Porovnávanie kvality metód sme vykonávali pomocou hodnôt priemerných absolútnych chýb a priemerných kvadratických chýb. Okrem nich sme vypočítali aj vážené verzie týchto meradiel kvality, pretože strelci v našom súbore mali rôzny počet pozorovaní.

Cieľom bolo nájsť metódu, ktorá bude dosahovať čo najmenšie hodnoty týchto chýb. Najhoršie obstálo maximum a preto ho označujeme jednoznačne za najhoršiu metódu pre predikovanie výkonu z testovaných spôsobov. Na druhej strane boli ostatné metódy porovnateľné, avšak **najmenšie hodnoty odchýlok dosahovala súčasná metóda**. Hodnoty priemerných chýb pre metódy medián a vážený priemer sa od tejto súčasnej metódy nelíšili oveľa. Bolo by možno vhodnejšie vykonať porovnanie na dátach, ktorých počet je väčší, no hlavne pri podmienke, že každá testovaná súťaž by bola nastriekaná iným strelcom.

Za hlavný prínos práce teda považujeme skutočnosť, že výber reprezentanta na nomináčne preteky nie je vhodné vykonávať odhadovaním pomocou metódy maxima. Na druhej strane ponúkame ďalšie možné varianty predikcie výkonu daného strelca okrem súčasnej metódy, ktoré sú kvalitou relatívne podobné.

Zoznam použitej literatúry

- [1] Stehlíková, B.: *Časové rady*, prednášky, FMFI UK Bratislava, 2016, dostupné na internete (28.3.2017): <http://www.iam.fmph.uniba.sk/institute/stehlikova/cr16.html>
- [2] Ševčovič D., Stehlíková B., Mikula K.: *Analytické a numerické metódy oceňovania finančných derivátov*, Slovenská technická univerzita, Bratislava, 2009
- [3] Holá, Z.: *Modelovanie ekonomických a finančných časových radov*, diplomová práca, FMFI UK Bratislava, 2013, dostupné na internete (28.3.2017): <http://www.iam.fmph.uniba.sk/studium/efm/diplomovky/2013/hola/diplomovka.pdf>
- [4] Kováč, L.: *Pravdepodobnosť v lukostrelbe*, bakalárska práca, FMFI UK Bratislava, 2016
- [5] Hurban V., Durný M.: *Stanovy a pravidlá FITA*, pravidlá a stanovy, Slovenský lukostrelecký zväz, Bratislava, 2010, dostupné na internete (28.3.2017): http://www.slz.sk/images/dokumenty/2010/FITARules2010_SVK_1.pdf
- [6] Box, G. E.P., Jenkins, G. M., Reinsel, G. C.: *Time Series Analysis: Forecasting and Control*, 4th Edition, John Wiley & Sons, New Jersey, 2008
- [7] Verbeek, M.: *A Guide to Modern Econometrics*, 2nd edition, John Wiley & Sons Ltd, Chichester, 2004
- [8] Anděl J.: *Statistické metódy*, 4. vydanie, vydavateľstvo Matematicko-fyzikálnej fakulty Karlovej Univerzity, Praha, 2007
- [9] Janková K.: *Pravdepodobnosť a štatistika*, 2. vydanie, vydavateľstvo UK Univerzity Komenského, Bratislava, 2013
- [10] Somorčík J.: *Prednášky zo štatistických metód*, učebné texty, FMFI UK Bratislava, 2016

- [11] Dasarathy G.: *A Simple Probability Trick for Bounding the Expected Maximum of n Random Variables*, dostupné na internete (5.5.2017): <http://www.cs.cmu.edu/~gautamd/Files/maxGaussians.pdf>
- [12] Merberg A. and Miller S. J.: *Course Notes for Math 162: Mathematical Statistics, The Sample Distribution of the Median*, dostupné na internete (5.5.2017): https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/162/Handouts/MedianThm04.pdf

Príloha A

V tejto prílohe je na obrázku A.28 uvedený kód, ktorý sme používali v prostredí R na modelovanie výkonu lukostrelcov v čase.

```
final <- read.csv(file="D:/data.csv", header = TRUE, sep = ";") #prečítanie dát

#vytvorenie matice
wantedShootMatrix <- matrix(c(1:(26*36)),nrow=26)
for(w in 1:26)
{
  wantedShootMatrix[w,1:36] <- final[final$strelec==w,9]
}

e =2 #poradové číslo strelca
p = wantedShootMatrix[e,] #vektor výstrelov vybraného strelca
p=ts(p) #vytvorenie časového radu z vektora

#grafické zobrazenie časového radu
plot(p,xlab="Poradie výstrelu",ylab="Bodové ohodnotenie",col="blue")
points(p, col="red")

#inštalácia potrebných knižníc
library(astsa)
library(urca)

acf2(p,35) #ACF a PACF časového radu

#ADF test o jednotkovom korení
summary(ur.df(p,type="drift",lags=11,selectlags="AIC"))

#diferencovanie v prípade, že je potrebné
pp=diff(p,6) #sezónne diferencovanie časového radu
plot(pp,.) #grafické zobrazenie nového časového radu

acf2(pp,20) #ACF a PACF diferencovaného časového radu
summary(ur.df(pp,type="drift",lags=10,selectlags="BIC")) #ADF test

sarima(p,1,0,0,1,0,0,6) #hľadanie modelu
```

Obr. A.28: Kód pre modelovanie výkonu strelcov

Obrázok A.29 nám ponúka náhľad na maticu, ktorá je v zdrojovom kóde označená ako *wantedShootMatrix*, kde riadky predstavujú výkon daného lukostrelca v čase.

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	2	5	8	8	9	9	3	6	8	8	9	10	6	8	8	4	9	10	6	8	9	9	4	10	!
2	7	9	8	10	9	10	7	9	8	10	9	10	8	9	8	10	9	10	9	8	10	8	10	9	!
3	7	8	9	1	4	5	2	4	6	7	8	9	3	5	6	8	9	9	3	5	10	8	9	6	10
4	9	7	10	8	10	9	9	7	9	8	9	10	9	8	9	9	10	6	10	9	9	8	8	6	10
5	8	10	9	10	9	10	8	10	9	10	9	10	8	10	9	10	9	10	8	10	9	10	9	10	!
6	7	9	8	10	9	10	8	9	8	10	10	9	9	8	10	8	10	9	9	8	10	8	10	9	10
7	9	10	10	8	9	5	10	9	10	8	9	6	10	9	10	8	9	6	10	9	10	8	9	7	10
8	5	7	7	8	9	9	10	9	8	7	7	5	5	7	7	8	9	10	6	7	8	10	9	8	!
9	4	10	6	7	8	9	7	8	9	10	5	6	5	6	7	8	9	10	5	7	8	9	9	10	!
10	4	8	6	8	7	10	4	8	6	8	7	10	5	8	6	8	7	10	5	8	6	8	7	10	!
11	9	7	10	8	10	9	10	7	10	8	9	9	6	9	8	10	9	10	10	10	9	9	8	6	10
12	10	9	8	10	9	10	9	10	9	10	9	10	10	9	10	9	10	9	10	9	10	9	10	9	10
13	8	4	8	5	9	7	8	4	8	6	9	7	8	4	8	6	10	8	8	4	8	6	8	10	!
14	8	7	9	8	10	10	10	6	9	7	10	8	9	6	10	7	10	8	9	5	10	7	10	8	!
15	10	10	10	9	9	8	10	9	8	8	7	6	10	10	7	7	7	6	8	8	8	8	6	7	!
16	9	8	8	10	8	7	8	10	9	7	8	9	10	10	9	6	7	6	7	10	9	6	7	10	!

Obr. A.29: Matica výstrelov v čase jednotlivých strelcov

Príloha B

Na obrázku B.30 uvádzame tabuľku, v ktorej sú zaznamenané presné hodnoty celkového výkonu jednotlivých strelcov v poradí.

Poradové číslo strelca	1	2	3	4	5	6
Celkový výkon	271	325	237	304	333	327
Poradové číslo strelca	7	8	9	10	11	12
Celkový výkon	317	283	280	267	311	342
Poradové číslo strelca	13	14	15	16	17	18
Celkový výkon	264	297	289	295	143	262
Poradové číslo strelca	19	20	21	22	23	24
Celkový výkon	242	322	316	296	324	320

Obr. B.30: Hodnoty celkových výkonov

Príloha C

Na obrázku C.31 je uvedená tabuľka výsledkov testov, v ktorej môžeme vidieť výsledky procesov o tom, či ADF test zamietol hypotézu H_0 a potom tento proces označujeme za stacionárny. Tabuľka obsahuje takisto počty náhodných šestic v testovanom procese.

strelec	šestice	test ADF	strelec	šestice	test ADF	strelec	šestice	test ADF
1	3	zamietá H_0	9	2	nezamietá H_0	17	2	nezamietá H_0
2	6	zamietá H_0	10	6	nezamietá H_0	18	2	zamietá H_0
3	4	zamietá H_0	11	3	nezamietá H_0	19	5	zamietá H_0
4	3	nezamietá H_0	12	6	zamietá H_0	20	6	zamietá H_0
5	6	nezamietá H_0	13	6	nezamietá H_0	21	6	zamietá H_0
6	6	zamietá H_0	14	5	nezamietá H_0	22	5	nezamietá H_0
7	5	nezamietá H_0	15	2	zamietá H_0	23	2	zamietá H_0
8	3	nezamietá H_0	16	6	zamietá H_0	24	6	zamietá H_0

Obr. C.31: Tabuľka výsledkov 1

strelec	AR1	SAR1	AR2	strelec	AR1	SAR1	AR2
1	0.4913	-0.0945	0.4643	13	0.7644	-0.5707	0.6509
2	0.6858	-0.5826	0.5235	14	0.7131	-0.5321	0.5597
3	0.0025	0.1815	0.0038	15	0.4692	0.433	0.4845
4	0.6791	-0.1458	0.6737	16	0.0346	0.1742	0.0614
5	0.9378	-0.1217	1	17	0.7461	0.2909	0.7462
6	0.7365	-0.5959	0.6725	18	0.3624	0.4386	0.2324
7	0.9506	-0.1864	0.9506	19	0.2159	-0.0183	0.2184
8	-0.2789	0.2911	-0.3297	20	-0.0023	-0.1465	0.0626
9	0.4575	0.1626	0.4311	21	0.1947	0.264	0.2094
10	0.9755	-0.1026	0.9782	22	-0.0373	0.1692	-0.0846
11	0.3373	-0.0834	0.3417	23	-0.0728	0.675	-0.2392
12	0.0194	-0.6713	0.1913	24	-0.1597	-0.2203	-0.1295

Obr. C.32: Tabuľka výsledkov 3

Na obrázku C.32 je tabuľka s hodnotami koeficientov modelov 1 a 2. $AR1$ je koeficient pri AR člene a $SAR1$ koeficient pri sezónnom AR člene v modeli 1. $AR2$ je hodnota koeficientu pri AR člene modelu 2. Hodnoty, ktoré sú zobrazené červenou farbou sú tie, kedy strelcovi podľa Ljung-Boxovej štatistiky daný model nevyhovoval.

V tabuľke na obrázku C.33 sú uvedené hodnoty AIC a BIC informačných kritérií pre jednotlivé modely všetkých strelcov. V stĺpci L-B je výsledok Ljung-Boxovej šta-

strelec	model 1			model 2			model 3		
	L-B	AIC	BIC	L-B	AIC	BIC	L-B	AIC	BIC
1		2,42784	1,55980		2,38565	1,47362	zle	2,62735	1,75930
2		0,18778	-0,68026	zle	0,58858	-0,32345		-0,09693	-0,96497
3		2,87963	2,01159		2,85909	1,94706	zle	2,81692	1,94888
4		0,96783	0,09979		0,93663	0,02461		1,09133	0,22329
5		-1,55534	-2,42338		-1,77928	-1,77928	zle	-1,61696	-2,48500
6		-0,12725	-0,99529	zle	0,25941	-0,65262		-0,15949	-1,02753
7		-0,49541	-1,36345	zle	-0,51812	-1,43014		-0,64303	-1,51107
8		1,73164	0,86360		1,74456	0,83254		1,81632	0,94828
9		2,04044	1,17240		2,01322	1,10120		1,83703	0,96899
10		-0,84915	-1,71719		-0,91389	-1,82591		-1,04028	-1,90832
11		1,62043	0,75239		1,57072	0,65870	zle	1,83149	0,96345
12		-0,62789	-1,49593	zle	-0,10725	-1,01928		-0,52280	-1,39084
13	zle	1,27546	0,40742	zle	1,67698	0,76495	zle	0,47395	-0,39409
14		1,41572	0,54768	zle	1,72393	0,81191		1,34567	0,47763
15		1,15504	0,28700	zle	1,28996	0,37793		0,75553	-0,11251
16	zle	1,56724	0,69920	zle	1,54232	0,63029		1,47723	0,60918
17	zle	2,55751	1,68947	zle	2,58290	1,67087	zle	2,19991	1,33187
18		2,19499	1,32695	zle	2,35664	1,44461		1,58735	0,71930
19		3,15036	2,28232		3,09496	2,18293	zle	3,50133	2,63329
20		0,64832	-0,21972		0,61010	-0,30192		0,71687	-0,15117
21	zle	1,15945	0,29141	zle	1,17527	0,26325	zle	1,04089	0,17285
22		1,83290	0,96486		1,79952	0,88750		1,74686	0,87882
23	zle	0,50316	-0,36488	zle	0,94520	0,03317	zle	0,33098	-0,53706
24		1,03255	0,16451		1,02559	0,11356		0,94882	0,08078

Obr. C.33: Tabuľka výsledkov 2

tistiky. Pre tých lukostrelcov, ktorí majú prázdnu bunku, dopadol tento test dobre a modelovanie daným modelom bolo prípustné. Naopak, ak sa v bunke nachádza slovo "zle", nebol daný model vhodný pre modelovanie procesu.

Príloha D

V tejto prílohe uvádzame výsledky odhadov výkonu pomocou vybraných metód pre súťaže, na ktoré bola potrebná nominácia strelca.

Meno	Reálny výkon	AP	MAX	MEDIÁN	VP	AKTUÁLNA M.
Duchoň	635	628	653	632	633	617
Duchoň	624	636	653	635	635	635
Duchoň	618	640	665	642	647	652
Duchoň	610	627	665	630	620	630
Duchoň	586	622	665	616	617	614
Duchoň	632	625	638	625	630	637
Hurban	655	615	651	616	615	634
Hurban	612	614	655	614	613	599
Hurban	628	616	642	618	625	640
Hurban	622	619	642	628	618	628
Hurban	610	619	642	625	618	622
Hurban	602	627	650	627	626	622
Lučaničová	618	612	645	623	618	631
Lučaničová	610	611	644	613	613	617
Lučaničová	532	589	644	594	583	564
Lučaničová	605	585	644	594	577	557
Lučaničová	581	604	634	604	603	623
Longová	591	623	644	623	627	632
Longová	584	626	648	629	630	631
Longová	582	625	648	631	624	620
Longová	634	619	645	624	627	636
Longová	642	621	645	629	617	614
Longová	647	623	645	634	622	639
Longová	631	632	655	630	632	630
Durná	670	652	676	655	655	674
Durný	672	664	686	664	666	673
Duchoň M.	622	615	641	624	609	624
Baloghová	588	519	585	549	531	556
Bužek	660	647	675	644	642	646
Tomčíková	502	551	586	547	551	541
Szegedi	569	584	597	592	584	574
Pavlík	673	679	698	682	677	687
Kondáš	674	651	690	640	657	686
Komarňanský	591	590	648	593	578	607
Harčarik	588	597	623	593	595	591
Dubašák	628	614	650	611	612	606
Dubašák	601	614	650	606	612	611
Kobza	596	582	629	585	580	517
Kobza	640	652	668	656	648	647
Baláž	623	636	674	639	632	635
Baláž	619	634	674	637	631	632
Baláž	657	644	662	642	645	642
Baláž	641	650	668	654	653	657
Baláž	622	651	668	656	651	660
Baláž	674	645	663	647	646	657

Obr. D.34: Tabuľka odhadov výkonov