



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY

INSTITUTE OF MATHEMATICS

DEKOMPOZICE A PŘEDPOVĚDI V ČASOVÝCH ŘADÁCH

DECOMPOSITION AND FORECAST IN TIME SERIES

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Kristína Šramková

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Josef Bednář, Ph.D.

BRNO 2016

Zadání bakalářské práce

Ústav:	Ústav matematiky
Studentka:	Bc. Kristína Šramková
Studijní program:	Aplikované vědy v inženýrství
Studijní obor:	Matematické inženýrství
Vedoucí práce:	Ing. Josef Bednář, Ph.D.
Akademický rok:	2016/17

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma bakalářské práce:

Dekompozice a předpovědi v časových řadách

Stručná charakteristika problematiky úkolu:

V práci bude popsán lineární regresní model. Déle budou popsány dekompoziční modely v časových řadách a ty budou použity k modelování předpovědí na reálných nebo simulovaných datech.

Cíle bakalářské práce:

Popis lineárního regresního modelu.

Popis aditivních a multiplikativních dekompozičních modelů.

Využití modelů na reálných nebo simulovaných datech.

Seznam literatury:

ANDĚL, Jiří. Základy matematické statistiky. Vyd. 3. Praha: Matfyzpress, 2011, 358 s. : grafy, tab. ISBN 978-80-7378-162-0.

KARPÍŠEK, Zdeněk a Miloš DRDLA. Statistické metody. 4. vyd. Brno: PC-DIR Real, 2000, 103 s. ISBN 80-214-1534-7.

ZVÁRA, Karel. Regrese. Praha: Matfyzpress, 2008, 253 s. : il. ; 24 cm. ISBN 978-80-7378-041-8.

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2016/17

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Bakalárska práca sa zaoberá štatistickými metódami, ktoré sa používajú pri analýze a modelovaní časových radov. Konkrétne dekompozičnými metódami, ktoré rozkladajú časové rady na jednotlivé zložky a modelujú trend, čo umožní predikciu budúcich hodnôt. V prvej časti práce je súhrn potrebných poznatkov ohľadom regresnej analýzy, základné vlastnosti časových radov a ich dekompozičných modelov. Ďalej sú rozobrané modely trendových kriviek. Záver práce je venovaný aplikácii popísaných metód na reálnych dátach a ich modelovaniu v programe *Statistica*.

Abstract

Bachelor thesis is focused on statistic methods for analysis and modelling time series. In particular decomposition techniques, which decompose time series into particular components and model trend component to predict future values. First part of thesis is focused on summary of essential knowledge of regression, basic properties of time series and their decomposition models. Hereinafter are discussed models of trend curves. In conclusion the mentioned methods are applied on real data and its modeling in software *Statistica*.

Klíčové slová

časové rady, dekompozícia, trend, metóda klzavých priemerov, *Statistica*

Key words

time series, decomposition, trend, moving average method, *Statistica*

Bibliografická citácia

ŠRAMKOVÁ, K. *Dekompozice a předpovědi v časových řadách*. Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2017. 66 s. Vedoucí bakalářské práce Ing. Josef Bednář, Ph.D..

Čestné prehlásenie

Prehlasujem, že predložená bakalárska práca je pôvodná a spracovala som ju samostatne. Prehlasujem, že citácie použitých prameňov sú úplné, že som v svojej práci neporušila autorské práva (v zmysle Zákona č. 121/2000 Sb., o práve autorskom a o právach súvisiacich s právom autorským).

V Brne dňa 19.mája 2017

.....
Kristína Šramková

Pod'akovanie

Na tomto mieste by som rada pod'akovala vedúcemu bakalárskej práce Ing. Josefovi Bednářovi, Ph.D. za cenné rady a pripomienky pri vedení bakalárskej práce.

Obsah

ÚVOD	15
1 REGRESIA	16
1.1 Regresný model	16
1.2 Lineárny regresný model	17
2 ČASOVÉ RADY	22
2.1 Základné vlastnosti	22
2.2 Charakteristiky časových radov	22
2.3 Problémy časových radov	23
2.4 Analýza časových radov	24
3 MODELOVANIE ČASOVÝCH RADOV	27
3.1 Dekompozícia časových radov	27
3.2 Modelovanie trendu časového radu	28
3.2.1 Subjektívne metódy	28
3.2.2 Popis matematickými krivkami	29
3.2.3 Metóda kľzavých priemerov	36
4 APLIKÁCIA DEKOMPOZIČNÝCH MODELOV	40
4.1 Príklad 1: POPULÁCIA - regresná analýza	40
4.2 Príklad 2: CUDZINCI - regresná analýza	44
4.3 Príklad 3: PRELETENÉ MÍLE - metóda kľzavých priemerov	52
4.4 Príklad 4: NEHODY - metóda kľzavých priemerov	54
ZÁVER	57
ZOZNAM POUŽITÝCH ZDROJOV	58
ZOZNAM TABULIEK	59
ZOZNAM OBRÁZKOV	60
ZOZNAM PRÍLOH	61

A PRÍLOHY	i
A.1 Dáta k príkladu 1 - Populácia v Indii	i
A.2 Dáta k príkladu 2 - Cudzinci v ČR	ii
A.3 Dáta k príkladu 3 - Preletené míle	iii
A.4 Dáta k príkladu 4 - Dopravné nehody	v

ÚVOD

V súčasnej dobe je nevyhnutnou súčasťou rozhodovania znalosť vývoja sledovaného procesu, ako aj schopnosť predpovede jeho budúceho správania. Tieto aspekty je vzhľadom k ich nasledovnému spracovaniu potrebné odhadovať kvantitatívne. Práve v tomto kroku prichádza na rad matematicko-štatistický aparát nazývaný *analýza časových radov*.

Jej úlohou je skúmanie dynamiky pozorovaných údajov v čase, analýza príčin ich správania a predvídanie budúceho vývoja. Neodmysliteľnou súčasťou tohto procesu je zostavovanie vhodného matematického modelu, ktorý nám umožní nahliadnúť do procesu, pri ktorom časový rad vzniká. Jedná sa totiž o reálny proces a nie len postupnosť hodnôt, ktorú je potrebné analyzovať. Pre skúmanie je z tohto dôvodu kľúčová znalosť príčin a mechanizmu vzniku skúmaných hodnôt.

Ako už z názvu práce plynie, bude sa zaoberať najmä dekompozičnými technikami práce s časovými radmi. Cieľom bude predstaviť problematiku časových radov, od ich charakteristík až po metódy analýzy a hlavne ilustrovať popísané metódy na názorných príkladoch. Prvá časť práce bude zameraná na problematiku regresnej analýzy, pozornosť bude venovaná metóde najmenších štvorcov, odhadom parametrov, testovaniu štatistických hypotéz a určovaniu intervalov spoľahlivosti odhadov a predikcií.

Ďalej budú zavedené základné pojmy týkajúce sa časových radov a charakteristiky slúžiace k ich popisu. Bude nasledovať časť venovaná dekompozícii samotnej, spolu s popisom základných používaných trendových kriviek. Na záver teoretickej časti bude vysvetlená metóda kĺzavých priemerov, ktorá prinesie nový pohľad na výstavbu modelov časových radov svojou adaptivitou. Teda zoberie do úvahy možnosť časovej zmeny charakteru radu z dlhodobého hľadiska.

Posledná časť práce bude venovaná ilustráciám uvedených techník na príkladoch s využitím reálnych dát. Analýza, výpočty a modelovanie samotné bude realizované v programe *Statistica*.

1 REGRESIA

Regresná analýza je univerzálny štatistický nástroj, ktorý umožňuje hľadať závislosti medzi sledovanými veličinami v podobe rôznych typov funkcií. V ďalšom texte najskôr zavedieme všeobecný regresný model a ďalej sa zameriame na lineárny regresný model, ktorý budeme využívať neskôr pri dekompozícii časových radov.

1.1 Regresný model

Nech máme danú maticu čísel $\mathbf{X} = (x_{ij})$ typu $n \times k$, kde $k < n$ a náhodné veličiny Y_1, \dots, Y_n . Predpokladajme, že pre náhodný vektor $\mathbf{Y} = (Y_1, \dots, Y_n)'$ platí $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, kde vektor $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ je vektor neznámych parametrov a $\mathbf{e} = (e_1, \dots, e_n)'$ je náhodný vektor, ktorého stredná hodnota je nulová $E(\mathbf{e}) = \mathbf{0}$ a má konštantný rozptyl $\text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$ s neznámym parametrom $\sigma^2 > 0$. Takto zavedený model budeme nazývať regresný. Vzhľadom k tomu, že \mathbf{Y} závisí na vektore neznámych parametrov $\boldsymbol{\beta}$ lineárne, budeme hovoriť o *lineárnom regresnom modele* [1].

Pre takto zavedený regresný model sa požaduje, aby mala matica \mathbf{X} lineárne nezávislé stĺpce¹, máme teda $h(\mathbf{X}) = k$, z čoho vyplýva, že matica $\mathbf{X}'\mathbf{X}$ je regulárna. Túto maticu budeme značiť $\mathbf{H} = \mathbf{X}'\mathbf{X}$ [1].

Regresné parametre β_1, \dots, β_n sú odhadované metódou najmenších štvorcov, ktorá minimalizuje kvadrát rozdielu skutočných hodnôt a modelu. Vychádza teda z podmienky, že výraz $(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ má byť vzhľadom k $\boldsymbol{\beta}$ minimálny. Odhady parametrov β_1, \dots, β_n budeme označovať $\mathbf{b} = (b_1, \dots, b_n)'$ [1].

Veta 1.1. *Odhady metódou najmenších štvorcov sú $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ [1].*

Poznámka. V regresnej analýze sa často používa pojem *sústava normálnych rovníc*, je to sústava lineárnych rovníc $\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{Y}$, z ktorých sa vypočíta vektor odhadu regresných parametrov \mathbf{b} .

Poznámka. Pomocou odhadu regresných parametrov \mathbf{b} získavame tiež najlepšiu aproximáciu vektoru \mathbf{Y} , ktorá sa dá vytvoriť lineárnou kombináciou stĺpcov matice \mathbf{X} . Je ňou vektor $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

¹Táto požiadavka plynie z požiadavky na jednoduchosť modelu, chceme sa vyhnúť nadbytočným vysvetľujúcim premenným.

Veta 1.2. Platí $E(\mathbf{b}) = \beta$, $\text{var}(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ [1].

Poznámka. Z predchádzajúcej vety vyplýva, že odhad metódou najmenších štvorcov \mathbf{b} parametrov β je nestranným odhadom².

Veta 1.3. Pre reziduálny súčet štvorcov platí $S_e = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}$ [1].

Veta 1.4. Náhodná veličina $s^2 = \frac{S_e}{n-k}$ je nestranný odhad parametru σ^2 .

Poznámka. Reziduálny súčet štvorcov podmodelu³ je zdola ohraničený reziduálnym súčtom štvorcov modelu [2].

K popisu presnosti regresného modelu sa používa *koeficient determinácie* R^2 a *korigovaný koeficient determinácie* R_{adj}^2 . Model je tým presnejší (resp. tým lepšie vystihuje funkčnú závislosť medzi vysvetľujúcimi a vysvetľovanými premennými), čím sú tieto koeficienty bližšie k hodnote 1. Koeficienty sú dané vzťahmi [1]:

$$R^2 = 1 - \frac{S_e}{S_t}, \quad R_{adj}^2 = 1 - \frac{n-1}{n-r-1} \frac{S_e}{S_t}, \quad (1.1)$$

kde $r = k - 1$, pretože prvý stĺpec matice \mathbf{X} je väčšinou tvorený samými jedničkami, môžeme teda napísať $\mathbf{X} = (\mathbf{1}, \mathbf{X}_1)$, kde \mathbf{X}_1 je matica typu $n \times r$. S_e je reziduálny súčet štvorcov a S_t je dané vzťahom:

$$S_t = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2$$

Poznámka. Koeficient determinácie podmodelu nemôže byť väčší ako koeficient determinácie modelu, čo plynie z predchádzajúcej poznámky o reziduálnom súčte štvorcov podmodelu [2].

1.2 Lineárny regresný model

Lineárnym regresným modelom nazývame model

$$Y_i = \sum_{j=0}^m \beta_j f_j(x_i) + e_i, \quad i = 1, \dots, n, \quad (1.2)$$

jedná sa o model lineárny vzhľadom k regresným koeficientom β_0, \dots, β_m , funkcie $f_j(x)$ sú známe funkcie [1].

²Povieme že odhad \mathbf{T} parametru θ je *nestranný*, ak platí $E(\mathbf{T}) = \theta, \forall \theta \in \Omega$.

³Podmodelom rozumieme pôvodný model po vynechaní niektorých stĺpcov regresnej matice \mathbf{X} .

Ďalej sa budeme zaoberať špeciálnym prípadom lineárneho regresného modelu, *regresnou priamkou*, ktorá má tvar

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n,$$

kde $n \geq 3$. Pre názornú predstavu maticového zápisu z predchádzajúcej kapitoly ukážeme, ako vyzerajú matice v prípade lineárneho modelu

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \mathbf{H} = \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}, \quad \mathbf{X}'\mathbf{Y} = \begin{bmatrix} \sum Y_i \\ \sum x_i Y_i \end{bmatrix}$$

Odhady b_0, b_1 parametrov β_0, β_1 počítame metódou najmenších štvorcov, pre odhady poznáme vzorce [2]:

$$b_0 = \frac{\sum x_i^2 \sum Y_i - \sum x_i \sum Y_i x_i}{n \sum x_i^2 - (\sum x_i)^2}, \quad b_1 = \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - (\sum x_i)^2}. \quad (1.3)$$

Obvykle sa podľa uvedeného vzťahu počíta iba odhad b_1 a odhad b_0 sa dopočíta z rovnice $b_0 = \bar{Y} - b_1 \bar{x}$.

Dôkaz. Odvodíme vzťahy pre výpočet odhadov b_0, b_1 . Vychádzame z metódy najmenších štvorcov, minimalizujeme teda rozdiel medzi skutočnými hodnotami Y_i a modelovanými hodnotami $\beta_0 + \beta_1 x_i$, čo predstavuje reziduálny súčet štvorcov S_e , matematicky zapísané

$$\min S(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2.$$

Funkciu minimalizujeme, hľadáme teda extrém, to znamená, že S parciálne zderivujeme postupne podľa β_0 a β_1 a položíme rovné nule

$$\begin{aligned} \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} &= \sum_{i=1}^n -2(Y_i - \beta_0 - \beta_1 x_i) = 0 \\ \frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} &= \sum_{i=1}^n -2x_i(Y_i - \beta_0 - \beta_1 x_i) = 0. \end{aligned} \quad (1.4)$$

Sústavu (1.4) môžeme prepísať do tvar (pre prehľadnosť budeme v ďalšom texte vynechávať indexy pri sumačnom symbole, budeme predpokladať $\sum := \sum_{i=1}^n$)

$$\begin{aligned} \sum Y_i - \sum \beta_0 - \sum \beta_1 x_i &= 0 \\ \sum x_i Y_i - \sum \beta_0 x_i - \sum \beta_1 x_i^2 &= 0. \end{aligned}$$

Ďalej môžeme ešte sústavu upraviť do jednoduchšieho tvaru, ktorý sa nazýva *sústava normálnych rovníc*

$$\begin{aligned} n\beta_0 + \beta_1 \sum x_i &= \sum Y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 &= \sum x_i Y_i. \end{aligned} \quad (1.5)$$

Sústavu (1.5) vyriešime pre β_0, β_1 . V prvom kroku si z môžeme z oboch rovníc vyjadriť β_1

$$\begin{aligned} \beta_1 &= \frac{\sum Y_i - n\beta_0}{\sum x_i} \\ \beta_1 &= \frac{\sum x_i Y_i - \beta_0 \sum x_i}{\sum x_i^2}, \end{aligned} \quad (1.6)$$

tieto výrazy položíme do rovnosti a budeme ďalej upravovať, aby sme vyjadrili vzťah pre β_0

$$\begin{aligned} \frac{\sum Y_i - n\beta_0}{\sum x_i} &= \frac{\sum x_i Y_i - \beta_0 \sum x_i}{\sum x_i^2} \\ \sum x_i^2 (\sum Y_i - n\beta_0) &= \sum x_i (\sum x_i Y_i - \beta_0 \sum x_i) \\ \beta_0 (n \sum x_i^2 - (\sum x_i)^2) &= \sum x_i^2 \sum Y_i - \sum x_i \sum x_i Y_i. \end{aligned}$$

Dostávame výsledný vzťah (1.3) pre odhad b_0

$$b_0 = \frac{\sum x_i^2 \sum Y_i - \sum x_i \sum x_i Y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Odhad b_1 dostaneme dosadením do prvého vzťahu z (1.6) a postupnými úpravami

$$\begin{aligned} \beta_1 &= \frac{\sum Y_i - n \frac{\sum x_i^2 \sum Y_i - \sum x_i \sum x_i Y_i}{n \sum x_i^2 - (\sum x_i)^2}}{\sum x_i} \\ \beta_1 &= \frac{n \sum x_i^2 \sum Y_i - \sum Y_i (\sum x_i)^2 - n \sum x_i^2 \sum Y_i + n \sum x_i \sum x_i Y_i}{\sum x_i (n \sum x_i^2 - (\sum x_i)^2)} \\ \beta_1 &= \frac{\sum x_i [n \sum x_i Y_i - \sum x_i \sum Y_i \sum x_i]}{\sum x_i [n \sum x_i^2 - (\sum x_i)^2]}. \end{aligned}$$

Nakoniec dostávame vzťah (1.3) pre odhad b_1

$$b_1 = \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - (\sum x_i)^2}.$$

Dokázali sme vzťahy pre odhady b_0, b_1 parametrov β_0, β_1 . \square

Ďalej máme vzťah pre výpočet nestranného odhadu s^2 parametru σ^2 [1]:

$$s^2 = \frac{\sum Y_i^2 - b_0 \sum Y_i - b_1 \sum x_i Y_i}{n - 2} \quad (1.7)$$

Pre koeficient determinácie a korigovaný koeficient determinácie lineárneho regresného modelu platia všeobecné vzťahy (1.1), kde reziduálny súčet štvorcov vypočítame $S_e = \sum e_i^2 = \sum (Y_i - b_0 - b_1 x_i)^2$ a $S_t = \sum (Y_i - \bar{Y})^2$.

Ďalší spôsob analýzy, ktorý nás bude zaujímať, je testovanie hypotéz a s tým spojené intervaly spoľahlivosti [3]:

- **Bodový odhad regresnej priamky**

$$y = b_0 + b_1 x$$

- **Intervalový odhad regresného koeficientu β_j**

$$\left\langle b_j - t_{1-\frac{\alpha}{2}} s \sqrt{h^{jj}}; \quad b_j + t_{1-\frac{\alpha}{2}} s \sqrt{h^{jj}} \right\rangle,$$

kde h^{jj} je j -ty diagonálny prvok matice \mathbf{H}^{-1} . Pre regresnú priamku platí

$$h^{11} = \frac{\sum x_i^2}{\det \mathbf{H}}, \quad h^{22} = \frac{n}{\det \mathbf{H}},$$

kde $\det \mathbf{H} = n \sum x_i^2 - (\sum x_i)^2$.

Odhad je na hladine významnosti α a $t_{1-\frac{\alpha}{2}}$ je kvantil Studentovho rozdelenia s $n - k$ stupňami voľnosti.

- **Intervalový odhad strednej funkčnej hodnoty y**

$$\left\langle y - t_{1-\frac{\alpha}{2}} s \sqrt{h^*}; \quad y + t_{1-\frac{\alpha}{2}} s \sqrt{h^*} \right\rangle,$$

kde $h^* = \mathbf{f}(x)^T \mathbf{H}^{-1} \mathbf{f}(x)$. Pre regresnú priamku platí

$$h^* = \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum x_i^2 - n \bar{x}^2} = \frac{1}{n} + \frac{n(x - \bar{x})^2}{\det \mathbf{H}}.$$

Odhad je na hladine významnosti α a $t_{1-\frac{\alpha}{2}}$ je kvantil Studentovho rozdelenia s $n - k$ stupňami voľnosti.

- **Intervalový odhad individuálnej funkčnej hodnoty y**

Výpočet je analogický ako predchádzajúci intervalový odhad strednej funkčnej hodnoty s rozdielom, že h^* nahradíme $h^* + 1$.

Posledný spôsob analýzy modelov, ktorý si uvedieme, je **analýza rozptylu**, tzv. ANOVA⁴. Využíva sa pri skúmaní vzťahu medzi nezávislými a závislými premennými. Podstatou tejto analýzy je rozklad celkového rozptylu na objasnenú zložku a neobjasnenú (reziduálnu) zložku, o ktorej predpokladáme, že je náhodná. V jednorozmernom prípade, v ktorom sa pohybujeme, vyjadrujeme variabilitu ako súčet štvorcov. To znamená, že celkový súčet štvorcov S_A je rozložený na vysvetlený súčet štvorcov S_t a reziduálny súčet štvorcov S_e [4].

Testujeme hypotézu $H_0 : X$ neovplyvňuje Y , proti alternatíve $H_A : X$ ovplyvňuje Y .

Testové kritérium F je dané vzťahom

$$F = \frac{S_t/v_1}{S_e/v_2}, \quad (1.8)$$

ktorý odpovedá podielu priemerného vysvetľovaného súčtu štvorcov a priemerného reziduálneho súčtu štvorcov. Hodnoty v_1, v_2 sú počty voľností jednotlivých variabilít dané vzťahmi

$$v_1 = k - 1, \quad v_2 = n - k.$$

Štatistika F má pri platnosti nulovej hypotézy H_0 Fisher-Snedecorovo rozdelenie s v_1, v_2 stupňami voľnosti, zapisujeme $F \sim F(v_1, v_2)$. Hypotézu H_0 zamietame na hladine významnosti α , ak $F > F_{1-\alpha}(v_1, v_2)$.

⁴z anglického Analysis of Variance

2 ČASOVÉ RADY

2.1 Základné vlastnosti

Definícia 2.5. *Časovým radom* rozumieme rad hodnôt určitého vecne vymedzeného ukazovateľa, ktorý je v čase usporiadaný od minulosti do prítomnosti [5].

Časové rady môžeme klasifikovať podľa typu sledovaného ukazovateľa:

Definícia 2.6. *Intervalový časový rad* je rad ukazovateľov, ktoré sa vzťahujú k určitému obdobiu, sú zisťované počas celého časového intervalu (rok, mesiac, štvrťrok, týždeň). Je pre ne charakteristické, že závisia na dĺžke zvoleného intervalu a vyjadrujú množstvo. *Okamihový časový rad* je rad ukazovateľov, ktorých hodnoty sa vzťahujú k určitému časovému okamihu. Je pre ne charakteristické, že nezávisia na dĺžke časového intervalu a vyjadrujú úroveň skúmaného javu [6].

Poznámka. Údaje intervalového časového radu sú znázorňované väčšinou stĺpcovými alebo spojnicovými grafmi. Príkladom intervalového časového radu môže byť produkcia alebo spotreba surovín.

Poznámka. Údaje okamihového časového radu sú znázorňované výhradne spojnicovými grafmi. Príkladom okamihového časového radu môže byť počet zamestnancov k určitému dátumu.

Ďalej môžeme časové rady rozdeliť podľa periodicity, s akou sú hodnoty sledované:

Definícia 2.7. *Dlhodobý časový rad* je rad ukazovateľov, ktorých hodnoty sú sledované v rokoch alebo v dlhších časových úsekoch. *Krátkodobé časové rady* majú hodnoty sledované v časových úsekoch kratších ako je jeden rok (polročné, štvrťročné, mesačné, týždenné). *Vysokofrekvenčné časové rady* majú periodicitu sledovania kratšiu ako jeden týždeň [5].

Definícia 2.8. Analýza časového radu znamená použitie vhodných metód na zostavenie modelu popisujúceho správanie pozorovaných ukazovateľov a najmä umožňujúceho predikciu budúcich hodnôt.

2.2 Charakteristiky časových radov

Definícia 2.9. *Prvou diferenciou* časového radu rozumieme prírastky ${}_1d_t(y)$ definované vzťahom [7]:

$${}_1d_t(y) = y_{t+1} - y_t, \quad t = 1, \dots, n-1.$$

Poznámka. Prvá diferenciacia predstavuje prírastok hodnoty časového radu medzi susednými obdobiami, vyjadruje o koľko sa zmenila hodnota radu oproti predchádzajúcemu obdobiu. Ak je hodnota prvej diferencie približne konštantná, odporúčaný trend tohto radu je lineárny.

Definícia 2.10. Druhou diferenciou časového radu rozumieme prírastky ${}_2d_t(y)$ definované vzťahom [7]:

$${}_2d_t(y) = {}_1d_{t+1}(y) - {}_1d_t(y), \quad t = 1, \dots, n-2.$$

Poznámka. Druhá diferenciacia predstavuje prírastok hodnoty prvej diferencie časového radu medzi susednými obdobiami. Ak je hodnota druhej diferencie približne konštantná, odporúčaný trend tohto radu je kvadratický.

Definícia 2.11. Koeficientom rastu časového radu rozumieme podiel susedných hodnôt tohto radu, označujeme $k_t(y)$ a definujeme vzťahom [7]:

$$k_t(y) = \frac{y_{t+1}}{y_t}, \quad t = 1, \dots, n-1.$$

Poznámka. Koeficient rastu vyjadruje, koľkokrát sa zmenila hodnota časového radu oproti predchádzajúcemu obdobiu, charakterizuje rýchlosť rastu alebo poklesu tohto radu. Ak je hodnota koeficientu rastu približne konštantná, odporúčaný trend je exponenciálny.

2.3 Problémy časových radov

Pri modelovaní časových radov sa môžeme stretnúť s niektorými špecifickými problémami, ktoré sú zapríčinené štruktúrou a charakterom dát usporiadaných do časových radov. Najvýznamnejšie z týchto problémov zmienime v tejto podkapitole. Problémami s časovými radmi je potrebné sa zaoberať najmä pri ekonomických aplikáciách.

Problémy s voľbou časových bodov pozorovaní

Budeme pracovať s diskretnými časovými radmi, teda s radmi, ktorých pozorovania sú v určitých nespojitých časových bodoch. Tieto časové rady môžu vzniknúť tromi spôsobmi: časové rady *diskrétne svojou povahou* (napr. úroda plodiny za rok), ďalej *diskretizáciou spojitého časového radu* (napr. teplota na danom mieste v danom čase) a nakoniec *akumuláciou (agregáciou)* hodnôt za dané časové obdobie (napr. denné množstvo zrážok), niekedy sa tiež namiesto akumulácie používa priemerovanie [8].

Problémy s kalendárom

Niektoré problémy sú „zapríčinené“ prírodou, príkladom je neceločíselný počet dní v sl-

nečnom roku alebo časové posuny na zemeguli. Väčšinu problémom s kalendárom ale zapríčinil človek, vďaka čomu máme:

- rôznu dĺžku mesiacov, rôzny počet pracovných dní za mesiac
- štyri alebo päť víkendov za mesiac
- pohyblivé sviatky (napr. Veľká noc)
- zimný a letný čas

Tieto nezrovnalosti môžu mať nežiaduce účinky pri samotnej práci s časovými radmi. Ako príklad môžeme uviesť sviatok na začiatku mesiaca, v tomto prípade stúpnu obchodníkom tržby za predchádzajúci mesiac, keď si ľudia nakupujú tovar do zásoby.

Existujú metódy, ktoré dokážu očistiť časový rad od týchto vplyvov. Jednou z týchto metód je akumulácia dát, pri ktorej sú nezrovnalosti spôsobené kalendárom do istej miery potlačené, ďalej korekcia dát na "štandardný" mesiac s dĺžkou 30 dní. Pri ekonomických dátach sa často používa úprava cien pomocou cenových indexov [8].

Problémy s nezrovnalosťami jednotlivých meraní

Pri analýze dlhších časových radov môže vzniknúť problém s nezrovnalosťou na začiatku a na konci radu vzhľadom k technickému rozvoju, inflácii, a podobne. Jednotlivé hodnoty v časovom rade je preto potrebné upraviť pre násobením vhodným indexom [9].

Problémy s dĺžkou časových radov

Dĺžkou časového radu rozumieme počet pozorovaní, ktoré tvoria daný časový rad (nie časové rozpätie medzi prvým a posledným meraním). Problém s dĺžkou časového radu je spôsobený dvomi protichodnými teóriami, z ktorých jedna hovorí o tom, že dĺžka radu by mala byť čo najväčšia, aby sme mali dostatok informácií o generovanom modeli. Na druhej strane pri veľmi dlhých radoch vzniká nebezpečenstvo, že sa charakter modelu v priebehu času môže podstatne meniť [9].

2.4 Analýza časových radov

Existujú rôzne metódy analýzy časových radov, najzákladnejšie z nich popíšeme v tejto podkapitole. Výber metódy pri analýze danej časovej rady závisí na účele analýzy a type časového radu.

Dekompozícia časových radov

Analýzou časových radov pomocou metódy dekompozície sa budeme ďalej v práci zaoberať, podrobnejšie bude rozobratý v kapitole 3.

Box-Jenkinsova metodológia

Na rozdiel od dekompozičnej metódy, ktorej práca, ako uvidíme v kapitole 3, je zameraná najmä na systematické zložky (tj. trend, sezónna a cyklická zložka), základom prístupu Box-Jenkinsovej metódy je reziduálna zložka, ktorá môže byť tvorená korelovanými náhodnými veličinami. Postupy pri vyšetrení časových radov touto metódou sa zameriavajú práve na tieto závislosti medzi pozorovaniami, kladie sa teda dôraz na korelačnú analýzu [8].

Jedným z najjednoduchších typov modelu je tzv. model kľzavých súčtov prvého rádu, ktorý sa značí MA(1), ďalšie známe modely sú tzv. autoregresný model AR, zmiešaný model ARMA a integrovaný model ARIMA [8].

Analýza viacrozmerných časových radov

Úlohou analýzy viacrozmerných časových radov je modelovanie niekoľkých časových radov naraz, do modelu sú taktiež zahrnuté vzájomné väzby medzi jednotlivými časovými radmi a ich premennými [9].

Vo väčšine príkladov sú metódy pre viacrozmerné časové rady len zovšeobecnením metód pre jednorozmerné rady, kde namiesto skalárnych veličín pracujeme s vektormi. Avšak v istých prípadoch paralelný popis niekoľkých radov v čase prináša do analýzy nové prvky, ktorými sú napríklad vektorová autoregresia VAR alebo kointegrácia⁵ [9].

Spektrálna analýza časových radov

Predchádzajúce prístupy by sa dali spoločne charakterizovať pojmom analýza časových radov v časovej doméne. Spektrálna analýza časových radov patrí do inej kategórie, ktorú môžeme označiť ako analýza časových radov v spektrálnej doméne. Tento prístup považuje skúmaný rad za nekonečnú zmes sinusových a kosinusových kriviek s rôznymi amplitúdami a frekvenciami. Hlavnými prvkami analýzy sú periodogram a spektrálna hustota, ktoré umožňujú získať obraz o intenzite jednotlivých frekvencií a následne odhadnúť koeficienty periodických zložiek [8].

⁵Kointegrácia je prípad lineárnej kombinácie nestacionárnych radov, ktorej výsledkom je rad stacionárny

Špeciálne metódy analýzy časových radov

Existujú aj iné metódy, ktoré sa používajú pre špeciálne typy časových radov [9]:

- nelineárne modely časových radov
- rekurentné metódy v časových radoch
- metódy pre časové rady s chýbajúcimi alebo nepravidelnými pozorovaniami
- robustná analýza časových radov - identifikuje a eliminuje vplyv odľahlých pozorovaní, tzv. outliers, ktoré významne skresľujú výsledok
- intervenčná analýza časových radov - skúma jednorázové vonkajšie zásahy, ktoré majú vplyv na priebeh radu

3 MODELOVANIE ČASOVÝCH RADOV

3.1 Dekompozícia časových radov

Časové rady môžeme rozdeliť na jednotlivé zložky, tento prístup sa používa pri analýze časových radov pomocou tzv. dekompozície. Motívom dekompozície je snaha identifikovať pravidelné správanie radu po zložkách, čo by malo byť jednoduchšie ako pri pôvodnom nerozloženom rade. Časový rad rozkladáme na štyri základné zložky [5]:

1. *Trend* Tr_t - predstavuje výsledok dlhodobých procesov a zmien v priemernom správaní časového radu, inými slovami predstavuje všeobecnú tendenciu vývoja skúmaných dát za dlhé obdobie.
2. *Sezónna zložka* S_t - predstavuje periodické kolísanie v priebehu pozorovaného obdobia a v každom období sa opakuje. Tieto zmeny bývajú spojené so striedaním ročných období.
3. *Cyklická zložka* C_t - predstavuje fluktuáciu okolo trendu, pri ktorej sa striedajú fázy rastu a poklesu. Dĺžka a intenzita jednotlivých cyklov je premenlivá. Príčinu vzniku tejto zložky je väčšinou ťažké určiť.
4. *Reziduálna (náhodná) zložka* e_t - predstavuje náhodné vplyvy, ktoré fluktuujú okolo trendu, nemá systematický charakter. Reziduálna zložka zostáva v časovom rade po odstránení trendu, sezónnej a cyklickej zložky. Do tejto zložky časového radu patria tiež chyby v meraní údajov alebo aj zaokrúhľovacie chyby.

To znamená, že časový rad chápeme ako trend, okolo ktorého kolísajú periodické zložky (cyklická a sezónna zložka) a reziduálna zložka vytvára charakter bieleho šumu.

Charakter bieleho šumu znamená, že náhodné veličiny sú navzájom lineárne nezávislých, majú normálne rozdelenie s nulovou strednou hodnotou a konštantným rozptylom, matematicky formulované:

$$E(e_t) = 0, \quad D(e_t) = \sigma_e^2, \quad cov(e_{t_1}, e_{t_2}) = 0, \quad e_t \sim N(0, \sigma_e^2) \quad (3.1)$$

Ako bolo už spomenuté, časové rady delíme na uvedené zložky kvôli ich analýze pomocou dekompozície (rozkladu) časových radov. Rozlišujeme dva základné prístupy k dekompozícii časových radov [6]:

1. *Aditívny* - predpokladá, že hodnoty časového radu sú dané súčtom jednotlivých zložiek:

$$Y_t = Tr_t + S_t + C_t + e_t,$$

kde sú všetky zložky merané v rovnakých jednotkách ako Y_t .

2. *Multiplikatívny* - predpokladá, že hodnoty časového radu sú dané súčinom jednotlivých zložiek:

$$Y_t = Tr_t S_t C_t e_t,$$

kde je trendová zložka Tr_t meraná v rovnakých jednotkách ako Y_t a ostatné zložky sú bezrozmerné.

Poznámka. Aditívna dekompozícia je používaná v prípade časovo nemennej variability hodnôt časového radu, na druhej strane pri časovo premenlivej variabilite sa využíva multiplikatívna dekompozícia. Spomínané dva modely je možné navzájom prevádzať pomocou logaritmickej transformácie.

3.2 Modelovanie trendu časového radu

V tejto podkapitole sú uvedené základné prístupy k eliminácii trendu časového radu. Tento pojem sa často označuje tiež *vyrovnanie* alebo *vyhladenie* časového radu. Zmyslom tejto činnosti je odstránenie sezónnej, cyklickej a reziduálnej fluktuácie, takto získame popis trendu samotného, prostredníctvom ktorého môžeme ďalej predikovať správanie skúmaného javu.

3.2.1 Subjektívne metódy

Táto kapitola je uvádená najmä pre zaujímavosť a úplnosť textu, z matematického hľadiska nie sú tu spomenuté metódy pre analýzu relevantné.

Subjektívnymi metódami rozumieme konštrukciu trendu prevažne grafickým spôsobom. Výhodou je, že často týmto jednoduchým postupom získame postačujúce očistenie trendu, na ktorom vidíme základné vlastnosti skúmaného časového radu. Použitie metód je vhodné na začiatku analýzy, na ich základe je možné vybrať vhodnú „objektívnejšiu“ metódu analýzy. Nevýhodou metódy je, že po jej vykonaní nedostávame dostatočné informácie pre predikciu ďalšieho vývoja skúmaného javu [8].

Pre lepšiu predstavu popíšeme dve jednoduché subjektívne metódy. Prvá z nich spočíva

vo vyrovnání horných a dolných výkyvov okolo trendu. Pri tomto prístupe je často užitočné časový rad znázorniť v logaritmickom merítke, v takto znázornenom rade je jednoduchšie zachytiť rast a pokles [8].

Druhá metóda je postavená na *priemerovaní cyklov*, táto metóda je objektívnejšia ako predchádzajúca. Postupuje sa nasledovne, najskôr lomenou čiarou spojíme hodnoty horných bodov zvratu, potom rovnakým spôsobom spojíme hodnoty dolných bodov zvratu. Následne do grafu vynesieme stred vzdialenosti medzi dolnou a hornou lomenou čiarou pre každý časový okamih. Subjektivita tejto metódy spočíva v niektorých vhodných korekciách, ako napríklad odstránenie outlierov a iných prudkých výkyvov [8].

3.2.2 Popis matematickými krivkami

Ďalej sa budeme zaoberať exaktnejším popisom časových radov, kde pomocou rôznych metód budeme hľadať analytický popis trendu. Výhodou tohto prístupu je jednoduchá možnosť predikcie budúcich hodnôt, a teda predpoveď správania skúmaného javu. Spomínanú predikciu dostávame dosadením do rovnice odhadnutého trendu.

Predpokladom metód je skúmaný časový rad v tvare:

$$y_t = Tr_t + E_t. \quad (3.2)$$

Tento tvar môžeme docieľiť dvomi spôsobmi, buď plynie priamo z charakteru skúmaného javu alebo bol rad na tento tvar prevedený tzv. sezónnym očistením. Vďaka tomuto špeciálnemu tvaru môžeme pri jednoduchších krivkách k odhadu parametrov modelu použiť regresnú analýzu. V ďalšom texte budú uvedené základné typy trendových kriviek [8]:

Konštantný trend

Konštantným trendom nazývame trendovú krivku danú rovnicou:

$$Tr_t = \beta_0, \quad t = 1, \dots, n. \quad (3.3)$$

Odhad b_0 parametra β_0 má jednoduchý tvar, vypočítame ho ako aritmetický priemer daného časového radu:

$$b_0 = \bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (3.4)$$

Lineárny trend

Lineárny trend môžeme popísať rovnicou:

$$Tr_t = \beta_0 + \beta_1 t, \quad t = 1, \dots, n. \quad (3.5)$$

Odhady b_0, b_1 parametrov β_0, β_1 vypočítame pomocou sústavy normálnych rovníc:

$$\begin{aligned} b_0 n + b_1 \sum_{t=1}^n t &= \sum_{t=1}^n y_t, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 &= \sum_{t=1}^n t y_t. \end{aligned}$$

Vyriešením tejto sústavy dostávame odhady b_0, b_1 parametrov modelu:

$$b_1 = \frac{\sum_{t=1}^n t y_t - \bar{t} \sum_{t=1}^n y_t}{\sum_{t=1}^n t^2 - n \bar{t}}, \quad b_2 = \bar{y} - b_1 \bar{t}, \quad (3.6)$$

kde

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t, \quad \bar{t} = \frac{1}{n} \sum_{t=1}^n t = \frac{n+1}{2}$$

Kvadratický trend

Kvadratický trend môžeme popísať rovnicou:

$$Tr_t = \beta_0 + \beta_1 t + \beta_2 t^2, \quad t = 1, \dots, n. \quad (3.7)$$

Odhady b_0, b_1, b_2 parametrov $\beta_0, \beta_1, \beta_2$ vypočítame pomocou sústavy normálnych rovníc:

$$\begin{aligned} b_0 n + b_1 \sum_{t=1}^n t + b_2 \sum_{t=1}^n t^2 &= \sum_{t=1}^n y_t, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 + b_2 \sum_{t=1}^n t^3 &= \sum_{t=1}^n t y_t, \\ b_0 \sum_{t=1}^n t^2 + b_1 \sum_{t=1}^n t^3 + b_2 \sum_{t=1}^n t^4 &= \sum_{t=1}^n t^2 y_t \end{aligned}$$

Exponenciálny trend

Exponenciálny trend je dvojparametrický trend, ktorý môžeme popísať rovnicou:

$$Tr_t = \alpha \beta^t, \quad t = 1, \dots, n \quad (\beta > 0). \quad (3.8)$$

Odhady parametrov môžeme získať zlogaritmovaním trendu, čím transformujeme rovnicu na lineárnu funkciu, získavame takto lineárny trend:

$$\log Tr_t = \log \alpha + t \log \beta.$$

Pre tento trend odhadneme parametre $\log \alpha, \log \beta$ rovnakým spôsobom, aký bol popísaný vyššie, v časti o lineárnom trende. Tieto odhadnuté parametre následne spätne odlogaritmujeme a získame odhady parametrov pôvodného trendu (3.8).

Druhá metóda nazývaná *metóda najmenších vážených štvorcov* je pre modely nelineárnej regresie adekvátnejšia. Metóda najmenších vážených štvorcov spočíva v minimalizácii výrazu

$$\sum_{t=1}^n v_t (y_t - \alpha \beta^t)^2, \quad (3.9)$$

kde v_t sú vopred vhodne zvolené váhy. V tejto metóde však namiesto výrazu (3.9) minimalizujeme výraz

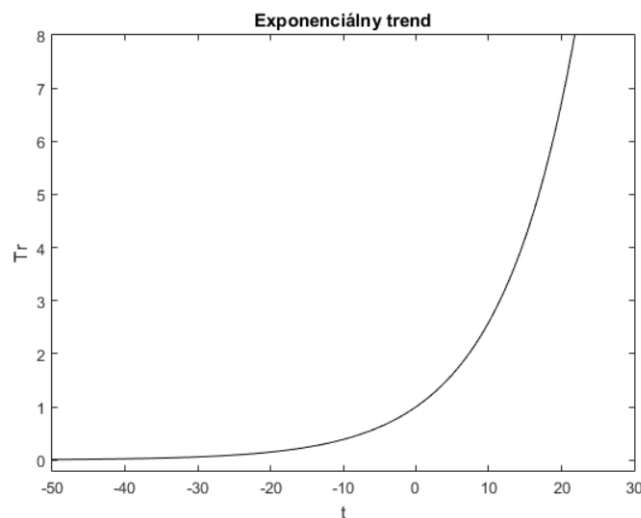
$$\sum_{t=1}^n w_t (\log y_t - \log \alpha - t \log \beta)^2, \quad (3.10)$$

kde w_t sú opäť vopred zvolené váhy podľa váh v_t , aby bol výsledný odhad parametrov približne rovnaký. Pre logaritmickú transformáciu je možné ukázať, že vhodná voľba týchto váh je $w_t = y_t^2 v_t$. Najčastejšou voľbou, keď nemáme žiadnu apriornú informáciu, je $v_t = 1$. Minimalizáciou výrazu (3.10) získavame odhady a, b parametrov α, β :

$$\log b = \frac{\sum_{t=1}^n (t - \bar{t}) y_t^2 \log y_t}{\sum_{t=1}^n (t - \bar{t})^2 y_t^2}, \quad \log a = \overline{\log y} - \bar{t} \log b,$$

kde

$$\overline{\log y} = \sum_{t=1}^n \frac{\log y_t}{n}, \quad \bar{t} = \sum_{t=1}^n \frac{t}{n} = \frac{n+1}{2}$$



Obr. 1: Exponenciálny trend, Zdroj: vlastný

Modifikovaný exponenciálny trend

Modifikovaný exponenciálny trend je trend s tromi parametrami α, β a γ , ktorý je zo-
všeobecnením predchádzajúceho prípadu, tento trend môžeme popísať rovnicou:

$$Tr_t = \gamma + \alpha\beta^t, \quad t = 1, \dots, n \quad (\beta > 0). \quad (3.11)$$

Tento trend vidíme na obrázku 2. Odhad parametrov tohto trendu sa väčšinou robí na-
sledujúcim spôsobom. Datový súbor rozdelíme na tri rovnako veľké časti s dĺžkou⁶ m .
Pozorovania v jednotlivých skupinách sčítame, získame tak čiastočné súčty:

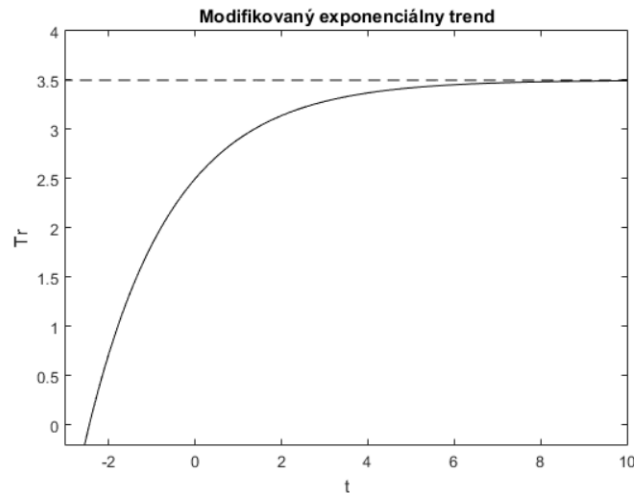
$$\begin{aligned} Sy_1 &= \sum_{t=1}^m y_t & \sim & STr_1 = m\gamma + \frac{\alpha\beta(\beta^m - 1)}{\beta - 1}, \\ Sy_2 &= \sum_{t=m+1}^{2m} y_t & \sim & STr_2 = m\gamma + \frac{\alpha\beta^{m+1}(\beta^m - 1)}{\beta - 1}, \\ Sy_3 &= \sum_{t=2m+1}^n y_t & \sim & STr_3 = m\gamma + \frac{\alpha\beta^{2m+1}(\beta^m - 1)}{\beta - 1}, \end{aligned}$$

kde Sy_1, Sy_2, Sy_3 sú čiastočné súčty pozorovaných hodnôt a STr_1, STr_2, STr_3 sú čiastočné
súčty hodnôt trendu.

Riešením tejto sústavy dostaneme odhady a, b, c parametrov α, β, γ :

$$\begin{aligned} b &= \left[\frac{S_3 - S_2}{S_2 - S_1} \right]^{\frac{1}{m}}, \\ a &= (S_2 - S_1) \frac{b - 1}{b(b^m - 1)^2}, \\ c &= \frac{S_1 - ab(b^m - 1)/(b - 1)}{m}. \end{aligned}$$

⁶Ak v danom súbore neplatí $n = 3m$, kde n je dĺžka celého súboru, potom niektoré pozorovania vynecháme,
väčšinou sa vynechávajú pozorovania na začiatku radu



Obr. 2: Modifikovaný exponenciálny trend, Zdroj: vlastný

Logistický trend

Logistický trend je trojparametrický trend s parametrami α, β, γ , ktorý môžeme popísať rovnicou:

$$Tr_t = \frac{\gamma}{1 + \alpha\beta^t}, \quad t = 1, \dots, n \quad (\beta > 0, \gamma > 0). \quad (3.12)$$

Ako je z obrázku 3 vidieť, tento trend radíme medzi tzv. *S-krivky*, presnejšie je logistický trend S-krivkou symetrickou okolo svojho inflexného⁷ bodu. Inflexia krivky sa nachádza v bode $t = -\log \alpha / \log \beta$. Ďalej môžeme rovnako vidieť, že krivka je asymptoticky obmedzená.

Pre odhady a, b, c parametrov α, β, γ existuje opäť niekoľko metód. Vzhľadom k tomu, že logistický trend môžeme považovať za inverziu k modifikovanému exponenciálnemu trendu, pre výpočet odhadov môžeme použiť metódu popísanú vyššie pri modifikovanom exponenciálnom trende pre rad s hodnotami $1/y_t$.

Druhou metódou je princíp tzv. *diferenčných odhadov parametrov*, v ktorom pracujeme s radom prvých diferencií $y_{t+1} - y_t$ namiesto hodnôt radu y_t . Pre potreby tejto metódy musíme uviesť deriváciu⁸ rovnice logistického trendu (3.12), ktorá má tvar:

$$\frac{dTr_t}{dt} = -\frac{\log \beta}{\gamma} Tr_t (\gamma - Tr_t). \quad (3.13)$$

⁷Inflexný bod je bod, v ktorom sa funkcia mení z konkávnej na konvexnú alebo naopak

⁸Derivácia logistického trendu je taktiež krivka symetrická okolo inflexného bodu, ktorý sa nachádza v bode $t = -\log \alpha / \log \beta$

Ďalej nahradením hodnôt trendu Tr_t hodnotami skutočných pozorovaní y_t v predchádzajúcej rovnici (3.13), dostaneme rovnicu:

$$\frac{dy_t}{dt} = -\frac{\log \beta}{\gamma} y_t (\gamma - y_t). \quad (3.14)$$

Použitím aproximácie radu derivácií skutočných pozorovaní radom prvých diferencií

$$\frac{dy_t}{dt} \sim \frac{y_{t+1} - y_t}{(t+1) - t} = y_{t+1} - y_t = {}_1d_t, \quad (3.15)$$

dostaneme po dosadení do rovnice (3.14) výraz

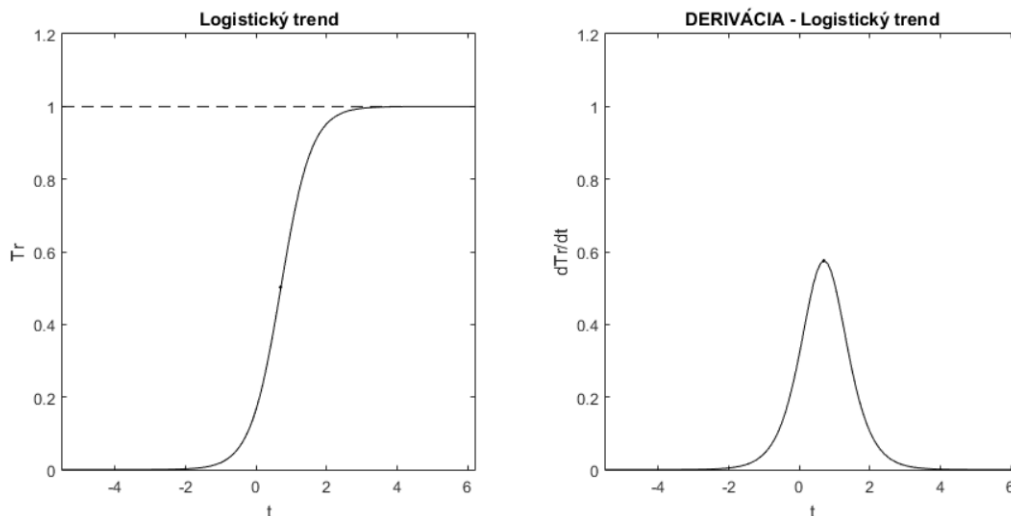
$$\frac{{}_1d_t}{y_t} = -\log \beta + \frac{\log \beta}{\gamma} y_t. \quad (3.16)$$

Odhady parametrov $-\log \gamma, (\log \beta)/\gamma$ lineárneho modelu (3.16) získame metódou najmenších štvorcov a z odhadov jednoducho môžeme vyjadriť odhady parametrov β, γ . Pre odhad parametru α musíme vo vzťahu (3.12) nahradiť hodnoty trendu Tr_t skutočnými hodnotami pozorovaní y_t a upraviť rovnicu do potrebného tvaru

$$y_t = \frac{\gamma}{1 + \alpha \beta^t} \quad \rightarrow \quad \alpha \beta^t = \frac{\gamma}{y_t} - 1.$$

Zlogaritmovaním a sčítaním cez $t = 1, \dots, n$ dostaneme tzv. *Rhodesov vzťah*, z ktorého už jednoducho vypočítame odhad parametru α

$$\log \alpha = -\frac{(n+1) \log \beta}{2} + \sum_{t=1}^n \frac{\log((\gamma/y_t) - 1)}{n}.$$



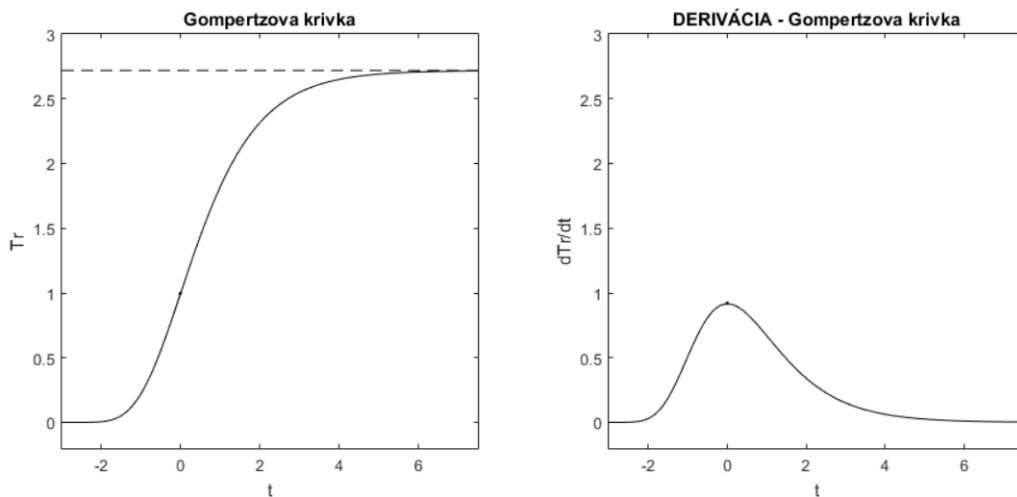
Obr. 3: Logistický trend, Zdroj: vlastný

Gompertzova krivka

Gompertzova⁹ krivka je trojparametrický trend, ktorý môžeme popísať rovnicou:

$$\log Tr_t = \gamma + \alpha\beta^t, \quad t = 1, \dots, n \quad (\beta > 0). \quad (3.17)$$

Podobne ako pri logistickom trende, aj Gompertzova krivka je transformáciou modifikovaného exponenciálneho trendu. Túto krivku radíme medzi tzv. *S-krivky nesymetrické okolo inflexného bodu*, pretože jej prvá derivácia nie je okolo bodu inflexie v bode $t = -\log(-\alpha)/\log\beta$ symetrická, ako môžeme vidieť na obrázku 4. Odhady a, b, c parametrov α, β, γ získame metódou používanou pre modifikovaný exponenciálny trend s rozdielom, že namiesto radu y_t budeme pracovať s radom $\log y_t$.



Obr. 4: Gompertzova krivka, Zdroj: vlastný

Vzhľadom k množstvu rôznych trendových kriviek je niekedy problematické zvoliť správny tvar pre daný časový rad. Na začiatku je vhodné si rad vykresliť a pokúsiť sa odhadnúť pravdepodobný tvar trendu. Ďalší spôsob je založený na určitých špecifických vlastnostiach daných kriviek, v Tabuľke 1. sú prehľadne uvedené testy na výber trendovej krivky.

⁹Benjamin Gompertz (1779 – 1865) bol britský matematik, ktorého meno je často spájané s demografickým modelom úmrtnosti v závislosti na veku

Tab. 1: Voľba trendovej krivky podľa orientačného testu, (Zdroj: [9])

Trend	Orientačný test	Matematický zápis
Lineárny	prvé diferencie sú približne konštantné	$y_{t+1} - y_t = konst.$
Kvadratický	druhé diferencie sú približne konštantné	$y_{t+2} - 2y_{t+1} + y_t = konst.$
Exponenciálny	prvé diferencie logaritmov sú približne konštantné (koeficient rastu je približne konštantný)	$\ln y_{t+1} - \ln y_t = konst.$ $(y_{t+1}/y_t = konst.)$
Modifikovaný exponenciálny	podiel susedných prvých diferencií sú približne rovnaké	$\frac{(y_{t+2}-y_{t+1})}{(y_{t+1}-y_t)} = konst.$
Logistický	histogram prvých diferencií má tvar hustoty normálneho rozdelenia s nulovou strednou hodnotou a rozptylom rovným jednej podiel susedných prvých diferencií prevrátených hodnôt sú približne konštantné	histogram $y_{t+1} - y_t$ má hustotu $\sim N(0, 1)$ $\frac{(1/y_{t+2}-1/y_{t+1})}{(1/y_{t+1}-1/y_t)} = konst.$
Gompertzov	podiel susedných prvých diferencií logaritmov sú približne konštantné	$\frac{(\ln y_{t+2}-\ln y_{t+1})}{(\ln y_{t+1}-\ln y_t)} = konst.$

3.2.3 Metóda kľzavých priemerov

Metóda kľzavých priemerov patrí medzi tzv. *adaptívne metódy*. Klasické metódy predpokladajú trend, ktorý má v čase konštantné parametre. Tento predpoklad je v dlhšom časovom období nereálny, riešením sú adaptívne metódy. Tieto metódy sú schopné pracovať s trendovou zložkou, ktorá v čase mení svoj charakter. Na druhej strane sa však predpokladá, že lokálne je vyrovnanie pomocou matematických kriviek možné na krátkych časových úsekoch [8].

Kĺzavými priemerami nazývame lineárne kombinácie členov pôvodného radu, tieto kombinácie sú potom ekvivalentné s vyrovnaním krátkych úsekov radu pomocou matematických kriviek [8].

Budeme predpokladať časový rad v tvare

$$y_t = Tr_t + E_t.$$

Konstruktia kĺzavých priemerov vychádza z predpokladu, že každá „rozumná“ funkcia môže byť aproximovaná polynómom. Časový rad y_t rozdelíme na kratšie časové úseky, na ktorých budeme odhadovať lokálne polynomicke trendy [8].

Postup vyrovnávania bude nasledovný, najskôr aproximujeme vhodným polynómom prvých $2m+1$ členov radu y_1, \dots, y_{2m+1} a týmto polynómom odhadneme hodnotu v strede uvažovaného úseku, t.j. v bode $t = m + 1$, ako vyrovnanú hodnotu \hat{y}_{m+1} . Ďalej postup opakujeme, teda pre získanie hodnoty \hat{y}_{m+2} vyrovnáme polynómom pozorovania y_2, \dots, y_{2m+2} . Tento postup je ekvivalentný s vytváraním lineárnych kombinácií hodnôt časového radu pre výpočet vyrovnaných hodnôt [10].

Kĺzavé priemery majú dva základné parametre. Prvým je *dĺžka kĺzavých priemerov*, ktorá udáva skutočnú dĺžku vyrovnaných úsekov. Predpokladá sa, že dĺžkou je nepárne číslo, t.j. $2m + 1$. Dĺžkou kĺzavých priemerov môžeme ovplyvniť mieru vyrovnania radu. Druhým parametrom je *rád kĺzavých priemerov*, ktorý určuje stupeň vyrovnávacieho polynómu [10].

Môžeme pozorovať *základné vlastnosti kĺzavých priemerov*:

1. Súčet váh kĺzavých priemerov je rovný jednej.
2. Váhy kĺzavých priemerov sú symetrické okolo prostrednej hodnoty.
3. Ak je rád r párne číslo, potom sú kĺzavé priemery rádu r a $r + 1$ s rovnakou dĺžkou $2m + 1$ totožné.

Je možné použiť rôzne druhy kĺzavých priemerov, dostávame ich voľbou váh jednotlivých členov vyrovnávaného úseku časového radu [8], [10], [11]:

- **Jednoduché kľzavé priemery**

Najjednoduchším typom kľzavých priemerov sú tzv. *jednoduché kľzavé priemery*, ktoré odpovedajú bežnému aritmetickému priemeru. Ako príklad uvidíme jednoduché kľzavé priemery dĺžky 5, ktoré majú tvar

$$\frac{1}{5}(1, 1, 1, 1, 1) = \frac{1}{5}(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}).$$

Vo všeobecnosti platí pre jednoduché kľzavé priemery dĺžky $2M + 1$ v čase $t = M+1, M+2, \dots, n - M$ nasledujúci vzťah

$$\hat{y}_t = \frac{1}{2M+1} \sum_{i=-M}^M y_{t+i} = \frac{y_{t-M} + y_{t-M+1} + \dots + y_{t+M}}{2M+1}.$$

Nie je dôvod, prečo by predpoveď časového radu nemohla byť konštruovaná kľzavými priermi párnej dĺžky, avšak toto vyrovnanie sa neodporúča. Dôvodom je, že vyrovnaná hodnota neprísluší k žiadnemu z uvažovaných okamihov meraní. V tejto situácii sa využívajú tzv. *centrované kľzavé priemery*.

- **Centrované kľzavé priemery**

Jedná sa o kľzavé priemery, ktorých váhy sú volené spôsobom, ktorý má za následok odstránenie sezónnej zložky časového radu. Vzhľadom k tomu, že kľzavé priemery potrebujú nepárny počet členov k tomu, aby odpovedali hodnote okamihu časového radu, namiesto prvého člena zoberieme priemer prvej a poslednej hodnoty. Pre $t = p + 1, \dots, n - p$ je tvar kľzavých priemerov nasledujúci

$$\hat{y}_t = \frac{\frac{y_{t-M} + y_{t+M}}{2} + y_{t-M+1} + \dots + y_{t+M-1}}{2M} = \frac{1}{4M}(y_{t-M} + 2y_{t-M+1} + \dots + 2y_{t+M-1} + y_{t+M}).$$

- **Vážené kľzavé priemery**

Vážené kľzavé priemery používame v prípade, že chceme časový rad popísať polynómom vyššieho stupňa. Koeficienty vyrovnávacieho polynómu sú odhadované metódou najmenších štvorcov.

Príklad 3.12. Princíp tejto metódy si vysvetlíme na kľzavých priemeroch 3. rádu a budeme vyrovnávať nepárny počet $2m + 1 = 5$ hodnôt časového radu, ktoré označíme [8]:

$$y_{t+\tau}, \quad \tau = -2, -1, 0, 1, 2.$$

Vyrovňavajúcim polynómom bude teda polynóm tretieho stupňa, ktorý určíme metódou najmenších štvorcov. Budeme minimalizovať výraz:

$$\sum_{\tau=-2}^2 (y_{t+\tau} - \beta_0 - \beta_1 \tau - \beta_2 \tau^2 - \beta_3 \tau^3)^2.$$

Tento výraz zderivujeme podľa všetkých štyroch koeficientov a získame sústavu 4 normálnych rovníc, všeobecne ju môžeme zapísať:

$$\sum_{\tau=-2}^2 y_{t+\tau} \tau^j - b_0 \sum_{\tau=-2}^2 \tau^j - b_1 \sum_{\tau=-2}^2 \tau^{j+1} - b_2 \sum_{\tau=-2}^2 \tau^{j+2} - b_3 \sum_{\tau=-2}^2 \tau^{j+3} = 0,$$

pre $j = 0, 1, 2, 3$.

Vzhľadom k tomu, že úseky radu volíme s nepárnym počtom členov $2m + 1$, sústava sa vďaka vlastnosti $\sum_{\tau=-2}^2 \tau^i = 0$ pre nepárne i zjednoduší na tvar:

$$\begin{aligned} 5b_0 + 10b_2 &= \sum y_{t+\tau}, \\ 10b_1 + 34b_3 &= \sum \tau y_{t+\tau}, \\ 10b_0 + 34b_2 &= \sum \tau^2 y_{t+\tau}, \\ 34b_1 + 130b_3 &= \sum \tau^3 y_{t+\tau} \end{aligned} \tag{3.18}$$

Pre výpočet vyrovnanej hodnoty v strede úseku radu (t.j. pre $\tau = 0$) nás budem zaujímať iba odhad b_0 ako vidíme z rovnice polynómu $b_0 + b_1 \tau + b_2 \tau^2 + b_3 \tau^3$. K výpočtu potrebujeme iba prvú a tretiu rovnicu sústavy. Jednoduchými úpravami dostaneme tvar:

$$b_0 = \frac{1}{35} (17 \sum y_{t+\tau} - 5 \sum \tau^2 y_{t+\tau}).$$

Odhadnutá vyrovnaná hodnota časového radu v čase t je:

$$\hat{y}_t = \frac{1}{35} (-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}). \tag{3.19}$$

Hodnota (3.19) je zároveň odhadom trendovej zložky v čase t . Symbolicky môžeme tento výraz zapísať pomocou tzv. *váh*, v našom prípade je to teda výraz:

$$\hat{y}_t = \frac{1}{35} (-3, 12, 17, 12, -3) y_t. \quad \square$$

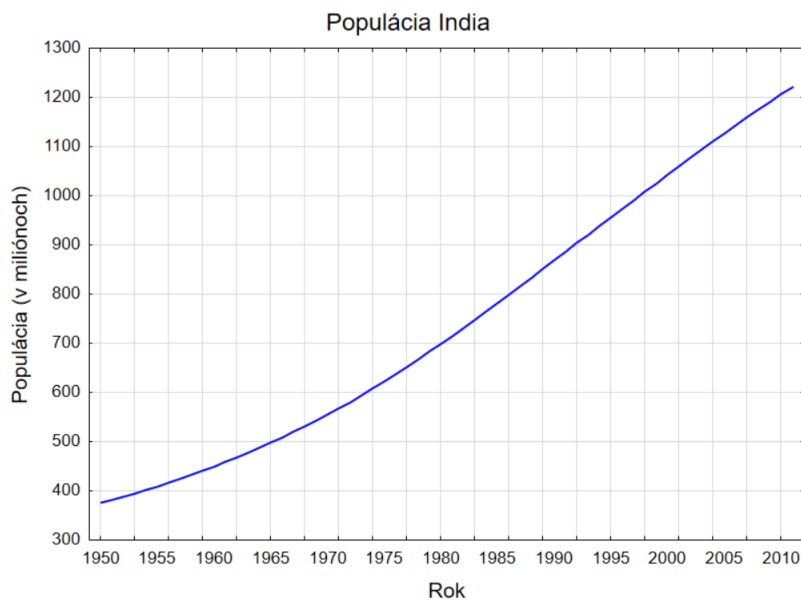
Metóda kľzavých priemerov má vplyv aj na ostatné komponenty časového radu. Sezónna zložka je použitím tejto metódy eliminovaná, na druhej strane, cyklické zložky v rade zostávajú. Biely šum stráca vlastnosť nekorelovanosti [10].

4 APLIKÁCIA DEKOMPOZIČNÝCH MODELOV

4.1 Príklad 1: POPULÁCIA - regresná analýza

Máme dáta¹⁰ vývoja populácie v Indii v ročných intervaloch od roku 1950 do roku 2011, viz. Príloha A.1. Máme teda časový rad s 62 pozorovaniami, dáta populácie sú uvedené v miliónoch. Pomocou softvéru *Statistica* budeme časový rad modelovať, analyzovať, ďalej vymodelujeme trend a spravíme predikciu ďalšej hodnoty na rok 2012. Skutočnú hodnotu pre tento rok máme tiež k dispozícii, preto na záver budeme môcť posúdiť vhodnosť zvoleného modelu na základe porovnania predikcie a skutočnej hodnoty. V tomto príklade sa zameriame na použitie regresnej analýzy.

Po načítaní dát do programu, si časový rad vykreslíme, použijeme spojnicový graf, ktorý môžeme vidieť na obrázku 5.



Obr. 5: Populácia - pôvodný časový rad, Zdroj: vlastný

Z grafu vidíme, že za trendovú krivku by sme mohli v najjednoduchšom prípade zvoliť priamku, prípadne polynóm druhého stupňa, t.j. kvadratickú funkciu. Obe možnosti zoberieme do úvahy, modely roanalyzujeme a na základe koeficientu determinácie, analýzy rozptylu a odchýlky predikcie ich porovnáme. Analýzu a výpočty budeme vykonávať v programe *Statistica*.

¹⁰zdroj: [12]

Ako prvú namodelujeme funkciu lineárnu a následne kvadratickú. Odhady parametrov spolu s 95% intervalmi spoľahlivosti sú uvedené v tabuľkách 2 a 3.

Tab. 2: Populácia - lineárny trend - odhady regresných parametrov, Zdroj: vlastný

LINEÁRNY TREND					
$Tr = \beta_0 + \beta_1 t$					
Parameter	Odhad	Dolná medz	Horná medz	t -hodnota	p -hodnota
b_0	282,333	266,226	298,439	35,063	0,000
b_1	14,586	14,141	15,031	65,626	0,000

Tab. 3: Populácia - kvadratický trend - odhady regresných parametrov, Zdroj: vlastný

KVADRATICKÝ TREND					
$Tr = \beta_0 + \beta_1 t + \beta_2 t^2$					
Parameter	Odhad	Dolná medz	Horná medz	t -hodnota	p -hodnota
b_0	350,387	341,981	358,793	83,406	0,000
b_1	8,206	7,590	8,822	26,669	0,000
b_2	0,101	0,092	0,111	21,394	0,000

Môžeme vidieť, že 0 nepatrí do žiadneho z uvedených intervalov spoľahlivosti pre odhad parametrov modelu, to znamená, že všetky parametre sú v modeloch významné na hladine významnosti $\alpha = 0,05$. Môžeme teda s oboma takto odhadnutými modelmi ďalej pracovať.

Ďalším krokom bude porovnanie modelov na základe koeficientu determinácie R^2 a tiež korigovaného koeficientu determinácie R_{adj}^2 , význam a výpočet týchto veličín je vysvetlený v kapitole 1. Porovnanie koeficientov je uvedené v tabuľke 4. Vidíme, že pomocou lineárneho trendu je vysvetlených 98,6% variability dát, zatiaľ čo pomocou kvadratického trendu až 99,8%.

Tab. 4: Populácia - porovnanie koeficientov determinácie, Zdroj: vlastný

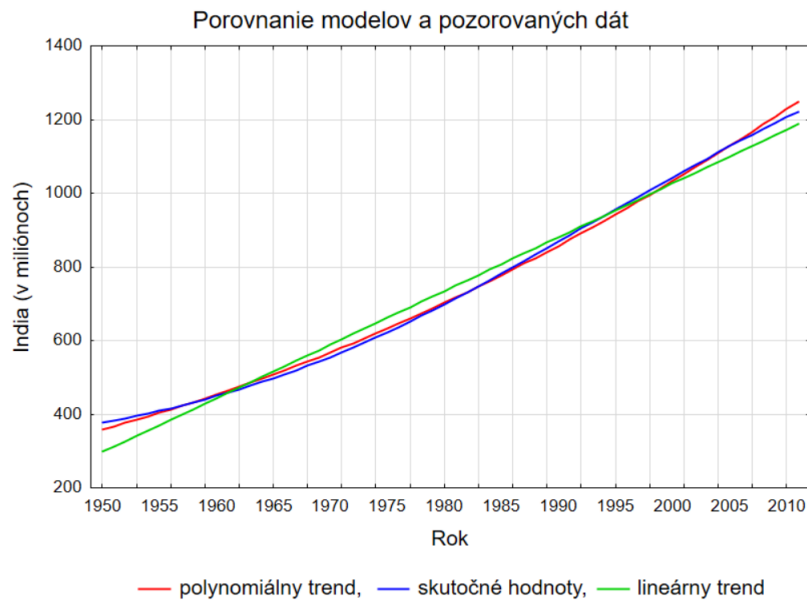
	Lineárny trend	Kvadratický trend
R^2	0,98626	0,99843
R^2_{adj}	0,98603	0,99838

Ďalej vidíme v tabuľke 5 porovnanie výsledkov analýzy rozptylu, ktorej hodnota vypovedá o vhodnosti zvoleného modelu. Pre oba modely vidíme, že hodnota F je podstatne vyššia ako 1 a p -hodnota menšia ako hladina významnosti 0,05, obidva modely sú teda konštruované vhodne.

Tab. 5: Populácia - analýza rozptylu, Zdroj: vlastný

	Lineárny trend	Kvadratický trend
Vysvetlený súčet štvorcov	4224314,91	2138222,87
Reziduálny súčet štvorcov	980,85	113,90
F	4306,80	18772,75
p -hodnota	0,000	0,000

Porovnaním koeficientov determinácie vidíme, že kvadratický trend lepšie vystihuje funkčnú závislosť medzi vysvetľujúcou a vysvetľovanou premennou. K rovnakému záveru dôjdeme aj pri pohľade na graf 6. Tiež sme k tomuto záveru mohli dôjsť z informácie, že koeficient pri kvadratickom člene je významne nenulový, už z tohto dôvodu má význam uprednostniť kvadratický trend pred lineárnym.



Obr. 6: Populácia - porovnanie modelov, Zdroj: vlastný

Posledným krokom analýzy bude spomínaná predikcia budúcej hodnoty pre rok 2013. Skutočná hodnota populácie v tomto roku je 1236,687. Predikovaná hodnota a reziduum oproti skutočnej hodnote sú uvedené v tabuľke 6. Opäť vidíme, že odhad pomocou kvadratického modelu je presnejší ako pri použití modelu lineárneho.

Tab. 6: Populácia - predikcia pre rok 2013, Zdroj: vlastný

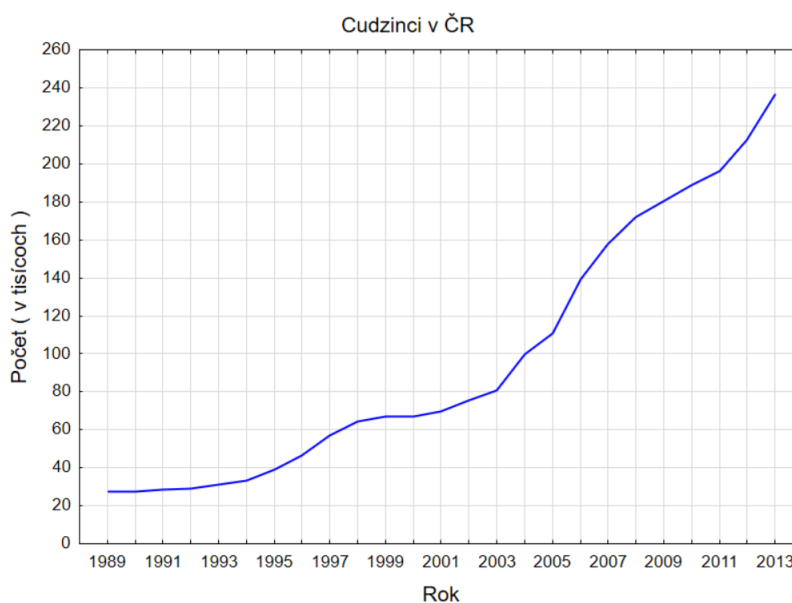
		Lineárny trend	Kvadratický trend
Predikcia		1201,251	1269,308
Reziduum		-35,436	32,621
Interval spoľahlivosti	Dolná medz	1185,147	1260,902
	Horná medz	1217,361	1277,715
Predikčný interval	Dolná medz	1136,570	1246,358
	Horná medz	1265,938	1292,259

Z analýzy časovej rady plynie, že najlepším zo skúmaných regresných modelov popisujúcich správanie dát je kvadratický trend: $Tr_t = 350,387 + 8,206t + 0,101t^2$.

4.2 Príklad 2: CUDZINCI - regresná analýza

Máme dáta¹¹ počtu cudzincov s trvalým pobytom v ČR v ročných intervaloch od roku 1989 do roku 2015, viz. Príloha A.2. Máme teda časový rad s 27 pozorovaniami, počet cudzincov je uvedený v tisícoch. Pomocou softvéru *Statistica* budeme analyzovať časový rad s prvými 25 pozorovaniami, namodelujeme rôzne typy trendov a spravíme predikciu d' ďalších dvoch hodnôt pre roky 2014 a 2015, ktorých skutočné hodnoty máme taktiež k dispozícii. Na záver zvolené trendy porovnáme a na základe rôznych kritérií vyberieme trend, ktorý najviac vyhovuje našim dátam. Tento príklad budeme analyzovať prostredníctvom nástrojov regresnej analýzy.

Po načítaní dát do programu, si časový rad najskôr vykreslíme, použijeme spojnicový graf, ktorý vidíme na obrázku 7.



Obr. 7: Cudzinci - pôvodný časový rad, Zdroj: vlastný

Z grafu vidíme, že trendová krivka by mohla byť aproximovaná polynómom druhého stupňa alebo exponenciálnou funkciou. Obidve možnosti budeme ďalej analyzovať a porovnávať. Grafy a výpočty budú realizované v programe *Statistica*. Ako prvú namodelujeme funkciu kvadratickú. V tabuľke 7 sú uvedené odhady parametrov, 95% intervaly spoľahlivosti parametrov a p -hodnota pre parametre modelu.

¹¹zdroj: [12]

Tab. 7: Cudzinci - kvadratický trend, Zdroj: vlastný

KVADRATICKÝ TREND				
$Tr = \beta_0 + \beta_1 t + \beta_2 t^2$				
Parameter	Odhad	Dolná medz	Horná medz	p-hodnota
b_0	28,725	17,876	39,575	0,000
b_1	-1,240	-3,162	0,683	0,195
b_2	0,384	0,312	0,456	0,000

Z tabuľky 7 vidíme, že koeficient pri lineárnom člene vychádza nevýznamný na hladine významnosti 0,05, t.j. $p > 0,05$, resp. 0 patrí do 95% intervalu spoľahlivosti. Preto namodelujeme ďalej ešte kvadratický trend bez lineárneho člena. Výsledky analýzy sú uvedené v tabuľke 8.

Tab. 8: Cudzinci - kvadratický trend bez lineárneho člena, Zdroj: vlastný

KVADRATICKÝ TREND BEZ LINEÁRNEHO ČLENA				
$Tr = \beta_0 + \beta_1 t^2$				
Parameter	Odhad	Dolná medz	Horná medz	p-hodnota
b_0	22,539	17,405	27,674	0,000
b_1	0,339	0,321	0,356	0,000

Druhým odhadovaným modelom je exponenciálna trendová funkcia, parametre tohto modelu môžeme vidieť v tabuľke 9.

Tab. 9: Cudzinci - exponenciálny trend, Zdroj: vlastný

EXPONENCIÁLNY TREND				
$Tr = \beta_0 + e^{\beta_1 + \beta_2 t}$				
Parameter	Odhad	Dolná medz	Horná medz	p-hodnota
b_0	-15,194	-41,041	10,652	0,236
b_1	3,467	2,910	4,023	0,000
b_2	0,083	0,064	0,103	0,000

Podobne ako pri kvadratickom trende, pre exponenciálny trend môžeme pozorovať nevýznamnosť absolútneho členu b_0 . Z tohto dôvodu ďalej namodelujeme tiež exponenciálny trend bez absolútneho členu, výsledky sú zaznamenané v tabuľke 10.

Tab. 10: Cudzinci - exponenciálny trend bez absolútneho členu, Zdroj: vlastný

EXPONENCIÁLNY TREND BEZ ABSOLÚTNEHO ČLENA				
$Tr = e^{\beta_1 + \beta_2 t}$				
Parameter	Odhad	Dolná medz	Horná medz	p-hodnota
b_0	3,114	2,969	3,260	0,000
b_1	0,096	0,089	0,103	0,000

Môžeme vidieť, že submodely majú všetky koeficienty významné na hladine významnosti $\alpha = 0,05$. V ďalšej analýze budeme brať do úvahy všetky štyri modely, vzájomne porovnáme ich vhodnosť. Aj napriek nevýznamným koeficientom v prvých dvoch trendoch, môžu tieto modely vykazovať inak dobré výsledky po iných stránkach.

Ďalej budeme porovnávať odhadnuté modely na základe koeficientu determinácie R^2 a tiež korigovaného koeficientu determinácie R_{adj}^2 , význam a výpočet týchto veličín je vysvetlený v kapitole 1. Porovnanie koeficientov je uvedené v tabuľke 11.

Tab. 11: Cudzinci - porovnanie koeficientov determinácie, Zdroj: vlastný

	Kvadratický trend	Kvadratický trend bez lineárneho člena	Exponenciálny trend	Exponenciálny trend bez absolútneho člena
R^2	0,9869	0,9859	0,9948	0,9943
R^2_{adj}	0,9858	0,9853	0,9943	0,9941

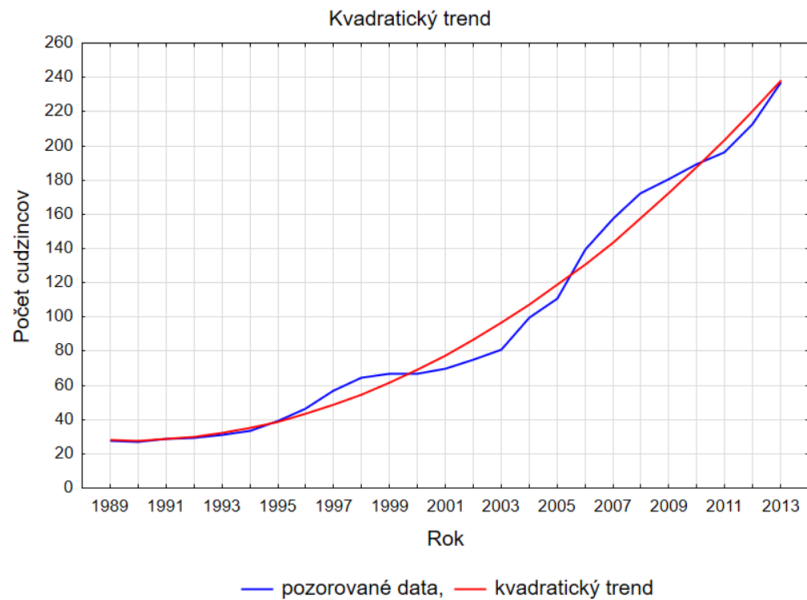
Z porovnania modelov môžeme pozorovať vlastnosť, ktorá bola spomenutá v prvej kapitole, a to, že koeficient determinácie podmodelu nemôže byť väčší ako koeficient determinácie modelu celkového. Z tohto dôvodu nemôžeme porovnávať dva kvadratické modely medzi sebou, rovnako ani exponenciálne. Porovnávať má teda zmysel iba exponenciálny a kvadratický trend, z tabuľky vidíme, že koeficient determinácie exponenciálneho trendu je o niečo väčší ako u kvadratického. Záverom analýzy na základe koeficientov determinácie môže byť, že exponenciálny trend vystihuje o niečo lepšie funkčnú závislosť pozorovaných dát.

Ďalej uvažujeme porovnanie modelov z hľadiska analýzy rozptylu. V tabuľke 12 sú zhrnuté výsledky tejto analýzy, ktorej hodnota vypovedá o vhodnosti zvoleného modelu. Pre všetky modely je hodnota F podstatne vyššia ako 1, p -hodnota je nulová u všetkých štyroch modeloch, modely sú teda zostavené vhodne.

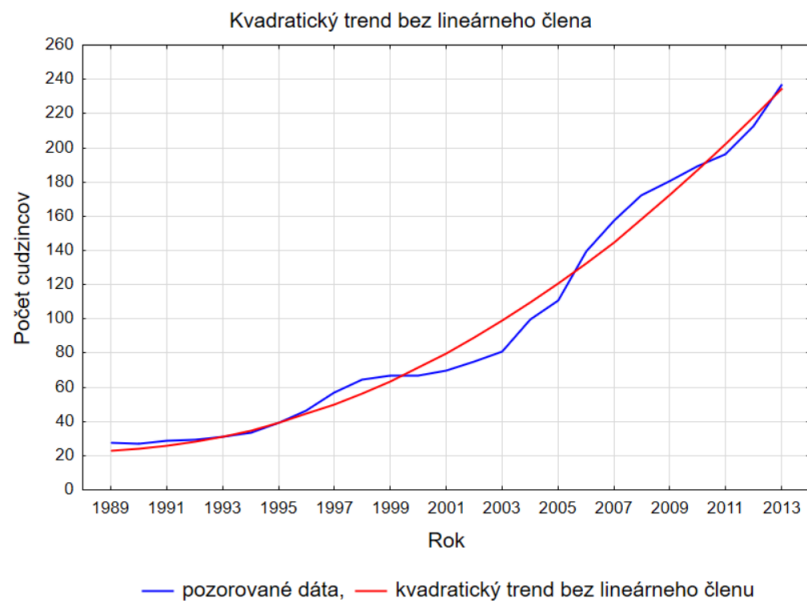
Tab. 12: Cudzinci - analýza rozptylu, Zdroj: vlastný

	Priemerný vysvetlený SŠ	Priemerný reziduálny SŠ	F	p -hodnota
Kvadratický trend	53640,094	64,496	831,683	0,000
Kvadratický trend bez lineárneho členu	107164,924	66,703	1606,595	0,000
Exponenciálny trend	114781,8	81,2	1414,385	0,000
Exponenciálny trend bez absolútneho členu	172089,127	84,9	2027,165	0,000

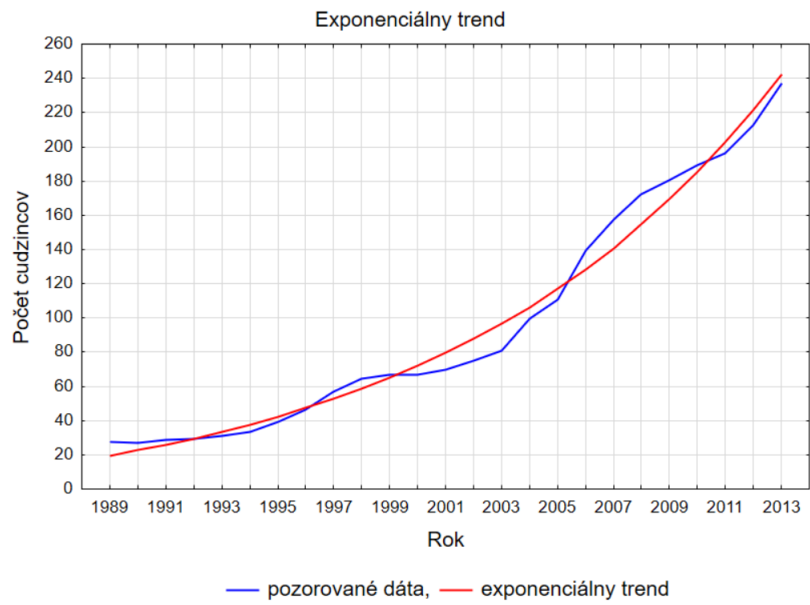
Nasleduje porovnanie grafické, na obrázkoch 8, 9, 10 a 11 vidíme porovnanie zvolenej trendovej krivky s pozorovanými dátami, ako bolo zrejmé z predchádzajúcej analýzy, aj v grafoch vidíme, že medzi zvolenými trendovými krivkami nie sú veľké rozdiely. Všetky modely z veľkej časti vystihujú charakter pozorovaných dát.



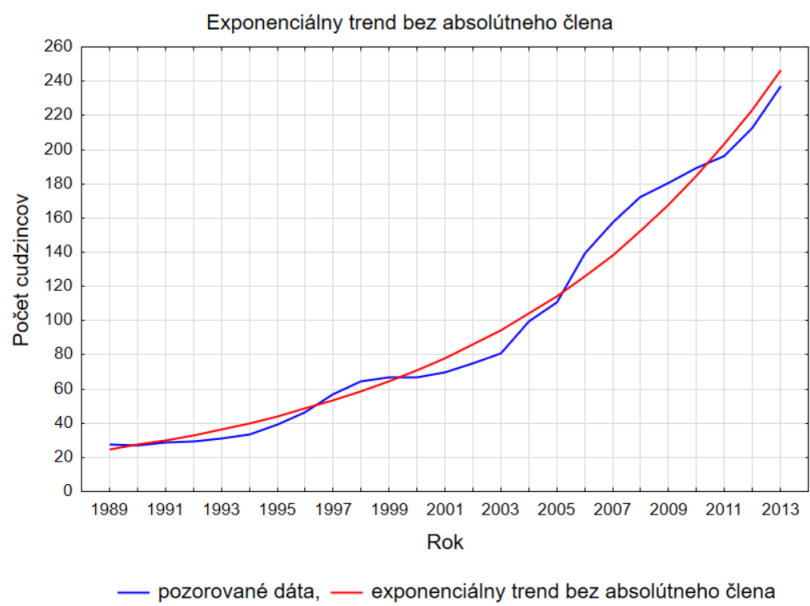
Obr. 8: Cudzinci - kvadratický trend, Zdroj: vlastný



Obr. 9: Cudzinci - kvadratický trend bez lineárneho členu, Zdroj: vlastný



Obr. 10: Cudzinci - exponenciálny trend, Zdroj: vlastný



Obr. 11: Cudzinci - exponenciálny trend bez absolútneho člena, Zdroj: vlastný

Posledným krokom analýzy bude spomínaná predikcia budúcej hodnoty pre roky 2014, 2015. Skutočné hodnoty počtu cudzincov v predikovaných rokoch je

2014: 249,856, 2015: 260,04.

Predikované hodnoty spolu s reziduami oproti skutočnej hodnote sú pre kvadratické trendy uvedené v tabuľke 13, tabuľka tiež obsahuje 95% intervaly spoľahlivosti a 95% predikčné intervaly. Pre pripomenutie, interval spoľahlivosti je intervalom okolo strednej hodnoty veličiny, zatiaľ čo predikčný interval je interval, v ktorom sa sledovaná veličina nachádza, nie len jej stredná hodnota.

Tab. 13: Cudzinci - predikcie - kvadratické trendy, Zdroj: vlastný

		Kvadratický trend	Kvadratický trend bez lineárneho člena
Predikcia (2014)		256,016	251,690
Reziduum		6,16	1,834
Interval spoľahlivosti	Dolná medz	245,166	243,042
	Horná medz	266,865	260,337
Predikčný interval	Dolná medz	236,139	232,710
	Horná medz	275,893	270,669
Predikcia (2015)		275,123	269,656
Reziduum		15,083	9,616
Interval spoľahlivosti	Dolná medz	262,482	260,148
	Horná medz	287,764	279,164
Predikčný interval	Dolná medz	254,214	250,269
	Horná medz	296,032	289,043

Z porovnania predikcií pomocou kvadratických trendov prichádzame k záveru, že kvadratický trend bez lineárneho člena lepšie predpovedá vývoj nami skúmaného časového radu.

Ďalej v tabuľke 14 môžeme vidieť porovnanie predikovaných hodnôt pomocou exponenciálnych modelov. Z dôvodu nelinearity modelov nemáme k dispozícii interval spoľahlivosti a predikčný interval vzhľadom k tomu, že túto možnosť neponúka používaný softvér.

Tab. 14: Cudzinci - predikcie - exponenciálne trendy, Zdroj: vlastný

	Exponenciálny trend	Exponenciálny trend bez absolútneho člena
Predikcia (2014)	262,078	273,144
Reziduum	12,222	23,288
Predikcia (2015)	286,074	300,666
Reziduum	26,034	40,626

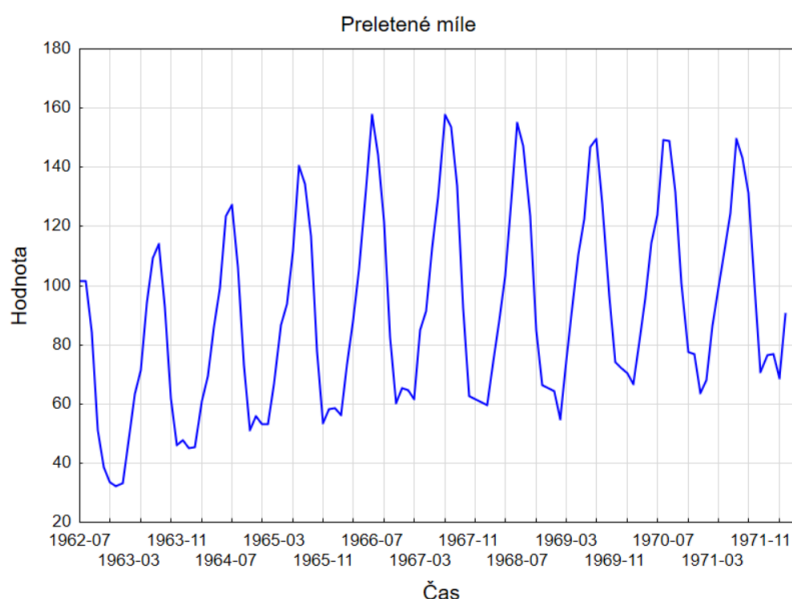
Z porovnania exponenciálnych modelov vidíme, že exponenciálny trend s absolútnym členom má presnejšiu predikciu ako exponenciálny trend bez absolútneho člena aj napriek tomu, že tento člen bol v modele nevýznamný. Avšak v porovnaní s kvadratickými trendmi vykazujú exponenciálne trendy globálne nepresnejšie predikcie budúceho vývoja. K opačnému záveru sme dospeli pri porovnávaní na základe koeficientu determinácie.

Môžeme teda povedať, že všetky zvolené modely vystihujú charakter skúmanej veličiny, čo vidíme aj z grafov 8, 9, 10, 11, a voľba najvhodnejšieho z modelov je podmienená zameraním analýzy.

4.3 Príklad 3: PRELETENÉ MÍLE - metóda kĺzavých priemerov

Máme mesačné dáta¹² preletených míľ pasažierov na vnútroštátnych letoch v Anglicku od júla 1962 do mája 1972, viz. Príloha A.3. Máme teda časový rad so 117 pozorovaniami. Pomocou softvéru *Statistica* budeme časový rad analyzovať, urobíme dekompozíciu a očistenie radu metódou kĺzavých priemerov.

Po načítaní dát do programu *Statistica*, si časový rad vykreslíme, použijeme spojnicový graf, ktorý môžeme vidieť na obrázku 12.

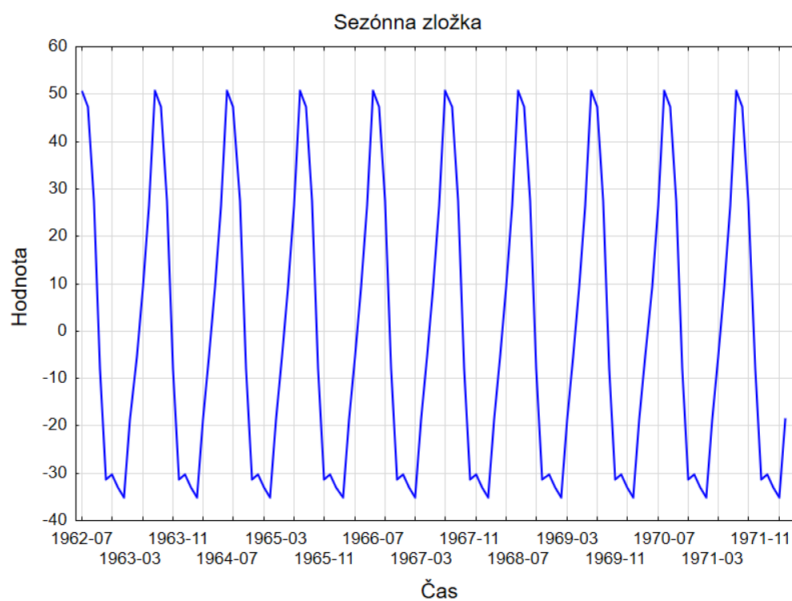


Obr. 12: Preletené míle - pôvodný časový rad, Zdroj: vlastný

Na prvý pohľad môžeme pozorovať silnú sezónnosť. Preto bude naším ďalším krokom sezónne očistenie časového radu, ako bolo v teórii uvedené, metóda kĺzavých priemerov nám umožňuje aj odstránenie sezónnosti. Ďalšiu analýzu budeme robiť v programe *Statistica*. Zvolíme sezónny rozklad, model zvolíme aditívny a sezónne posunutie nastavíme na hodnotu 12, keďže máme mesačné dáta. Aditívny model volíme z toho dôvodu, že z grafu môžeme pozorovať nezávislosť amplitúdy sezónnej zložky na úrovni trendu.

Po sezónnom očistení získame sezónnu zložku, vidíme na obrázku 13, ďalej získavame tiež hodnoty kĺzavých priemerov. Ich porovnanie s pôvodným časovým radom môžeme názorne vidieť v grafe 14.

¹²zdroj: [13]



Obr. 13: Preletené míle - sezónna zložka, Zdroj: vlastný

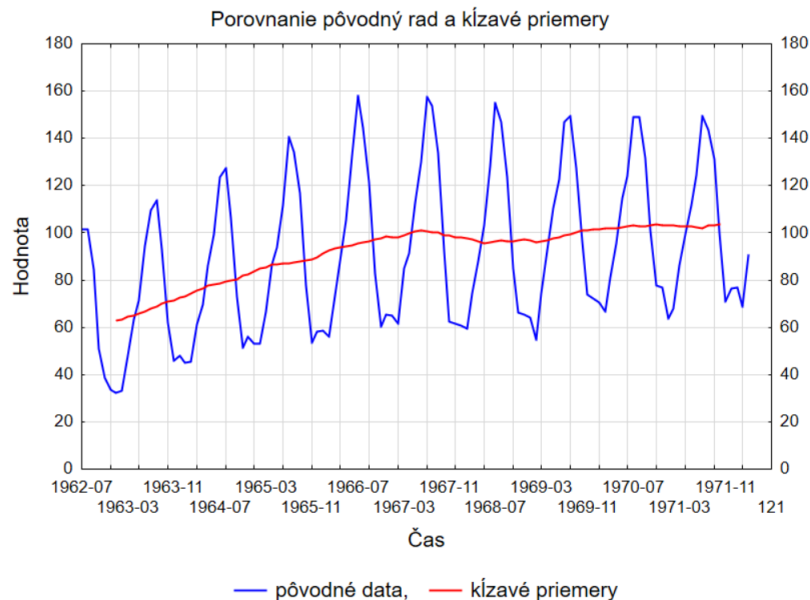
V tabuľke 15 sú zaznamenané vypočítané hodnoty sezónnej zložky v jednotlivých mesiacoch, tieto hodnoty sú v míľach, teda v rovnakých jednotkách ako pôvodný časový rad, vzhľadom k tomu, že sme použili aditívnu dekompozíciu.

Tab. 15: Preletené míle - odhad sezónnej zložky v jednotlivých mesiacoch, Zdroj: vlastný

Január	-33,0427
Február	-35,0427
Marec	-18,6297
Apríl	-5,2538
Máj	9,2666
Jún	26,4888
Júl	50,7935
August	47,3916
September	27,4055
Október	-7,9427
November	-31,2459
December	-30,1886

Môžeme overiť, sezónne vplyvy sú v priebehu roku vykompenzované, to znamená, že tzv. normalizačná podmienka je pre odhadnutú sezónnu zložku splnená, platí teda

$$\sum_{t=1}^{12} Sz_t = 0.$$



Obr. 14: Preletené míle - pôvodný časový rad a kľzavé priemery, Zdroj: vlastný

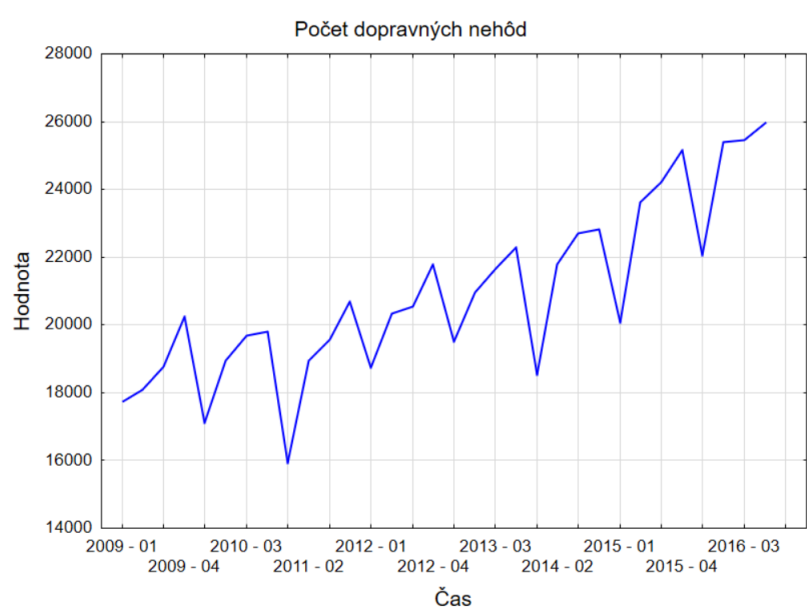
Záverom analýzy je vyrovnanie skúmaného časového radu preletených míľ, výsledkom nie je analytická funkcia, ako to bolo v predchádzajúcich príkladoch s regresnou analýzou. Dostávame opäť empirické vyjadrenie skúmaného vývoja, avšak vďaka očisteniu od sezónnosti sme získali trend, ktorý popisuje globálne správanie sledovanej veličiny.

4.4 Príklad 4: NEHODY - metóda kľzavých priemerov

Máme štvrťročné dáta¹³ počtu dopravných nehôd na cestách v Českej republike od roku 2009 do roku 2016, viz. Príloha A.4. Časový rad má 32 pozorovaní. Pomocou softvéru *Statistica* budeme realizovať analýzu tohto časového radu, ďalej urobíme dekompozíciu a očistenie radu metódou kľzavých priemerov.

Po načítaní dát časový rad vykreslíme, graf môžeme vidieť na obrázku 15.

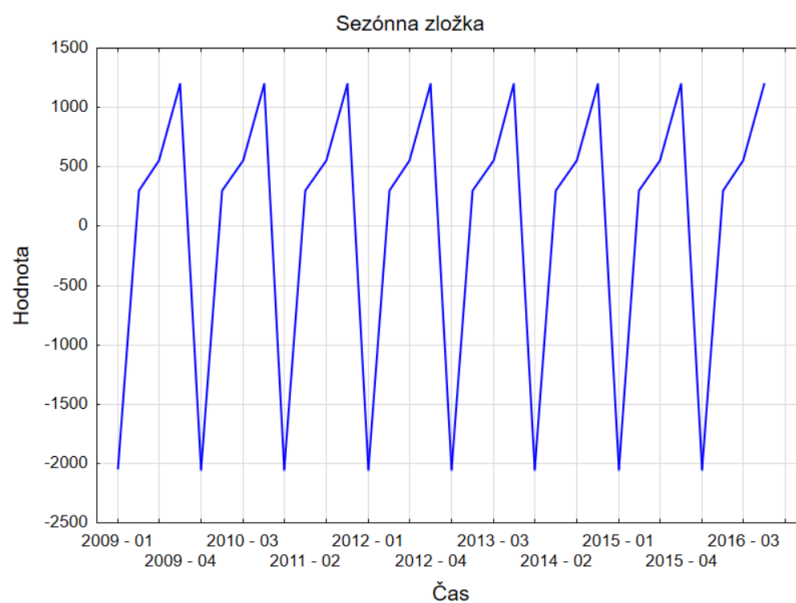
¹³zdroj: [12]



Obr. 15: Dopravné nehody - pôvodný časový rad, Zdroj: vlastný

Opäť vidíme sezónnosť, aj keď nie takú výraznú ako v predchádzajúcom príklade. Aj v tomto prípade zvolíme sezónny rozklad, model bude znova aditívny a sezónne posunutie nastavíme na hodnotu 4, vzhľadom k štvrtročným dátam.

Po sezónnom očistení získame dekomponovanú sezónnu zložku, ktorú vidíme na obrázku 16, ďalej získavame tiež hodnoty kľzavých priemerov. Ich porovnanie s pôvodným časovým radom môžeme názorne vidieť v grafe 17.



Obr. 16: Dopravné nehody - sezónna zložka, Zdroj: vlastný

Hodnoty sezónnej zložky v jednotlivých štvrťrokoch sú zaznamenané v tabuľke 16. Opäť môžeme overiť normalizačnú podmienku, teda platnosť rovnice

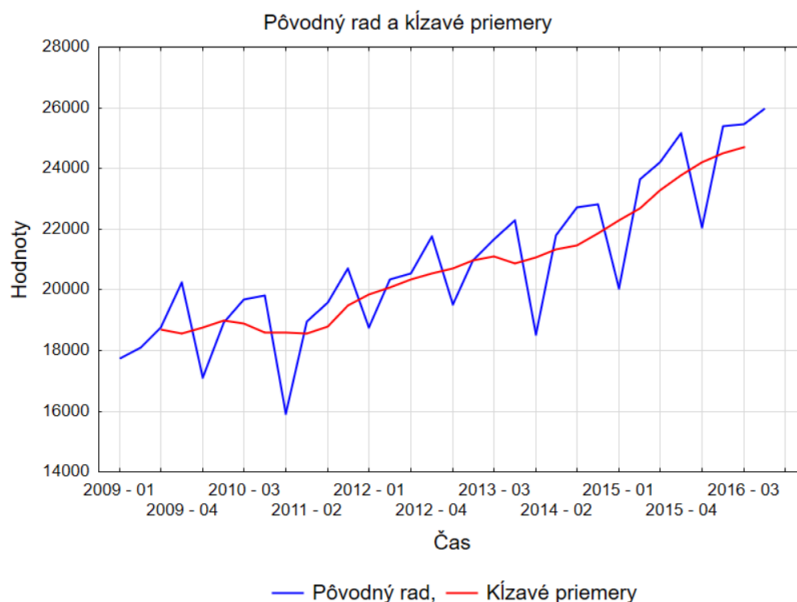
$$\sum_{t=1}^4 Sz_t = 0.$$

Sezónne vplyvy sú teda v priebehu roku vykompenzované.

Tab. 16: Dopravné nehody - odhad sezónnej zložky v jednotlivých štvrťrokoch , Zdroj: vlastný

1. štvrťrok	-2049,801
2. štvrťrok	301,484
3. štvrťrok	553,297
4. štvrťrok	1195,020

Výsledný model tohto príkladu môžeme vidieť na obrázku 17, kde je tiež porovnaný s pozorovanými hodnotami. Tento model sme získali metódou kľzavých priemerov, v rámci ktorej bol rad očistený aj o sezónnu zložku.



Obr. 17: Dopravné nehody - pôvodný časový rad a kľzavé priemery , Zdroj: vlastný

Ako v predchádzajúcom príklade, aj tu sme metódou kľzavých priemerov dostali vyrovnanie pôvodného časového radu, ktorým sme odhalili globálny charakter skúmaného javu.

ZÁVER

Práca sa zaoberala časovými radmi, ktoré už sami o sebe nesú množstvo informácií o správani a vývoji pozorovaného javu v danom období. Mojou úlohou bolo ich matematicko-štatistické spracovanie a nasledujúca analýza na základe teoretických informácií o ich vlastnostiach a metódach spracovania. Potrebné teda boli zrozumiteľné a prehľadné teoretické východiská, ich vysvetlenie a spracovanie bolo aj jedným z cieľov tejto práce. Hlavným cieľom bola následná ilustrácia metód na príkladoch s reálnymi dátami, analýza dát a predikcia budúceho vývoja.

Prvá kapitola bola zameraná na prácu s regresným modelom. Je to jeden z dôležitých prvkov štatistiky samotnej, ktorý je univerzálnym nástrojom v rôznych problematikách, nie len v oblasti časových radov. Medzi základné nástroje patrí metóda najmenších štvorcov, ďalej je významná teória odhadov a testovanie štatistických hypotéz, vďaka ktorým sme schopní stanoviť kvalifikované závery o tvrdeniach o predikčných modeloch aj predikciách samotných.

V druhej kapitole bol zavedený pre prácu esenciálny pojem časového radu a ďalej boli rozoberané jeho základné vlastnosti a charakteristiky. Z dôvodu častých aplikácií časových radov v oblasti ekonómie, sú spomenuté aj problémy, s ktorými sa pri časových radoch môžeme stretnúť. Z matematického hľadiska táto problematika nezohráva tak významnú rolu. Koniec druhej kapitoly bol venovaný spôsobom analýzy časových radov, ktoré sa v praxi používajú.

Z uvedených spôsobov bola práca zameraná práve na dekompozičné metódy, ktorými sa zaoberá tretia kapitola. V úvode je pozornosť venovaná jednotlivých zložkám a typom modelov. Nasleduje časť orientovaná na spôsoby modelovania trendu, kde je významný popis trendových kriviek, ktoré sa v praxi používajú. Dôležitou súčasťou je taktiež metóda kĺzavých priemerov, ktorá predstavuje adaptívny prístup k výstavbe trendu časového radu.

Posledná kapitola bola venovaná spomínanej aplikácii dekompozičných modelov na reálnych dátach. Dva príklady boli sústredené na použitie regresnej analýzy pri budovaní modelu. Ďalšie dva využívali adaptívny prístup, uvedenú metódu kĺzavých priemerov. Pri analýze a modelovaní časových radov bol využívaný program *Statistica*.

ZOZNAM POUŽITÝCH ZDROJOV

- [1] ANDĚL, Jiří. *Základy matematické statistiky*. Vyd. 3. Praha: Matfyzpress, 2011. ISBN 978-80-7378-162-0.
- [2] ZVÁRA, Karel. *Regrese*. Praha: Matfyzpress, 2008. ISBN 978-80-7378-041-8.
- [3] KARPÍŠEK, Zdeněk. *Matematika IV: statistika a pravděpodobnost*. Vyd. 2., dopl. Brno: Akademické nakladatelství CERM, 2003. ISBN 80-214-2522-9.
- [4] MELOUN, Milan a Jiří MILITKÝ. *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004. ISBN 80-200-1254-0.
- [5] ARTL, Josef a Markéta ARTLOVÁ. *Ekonomické časové řady*. Praha: Professional Publishing, 2009. ISBN 978-80-86946-85-6.
- [6] KARPÍŠEK, Zdeněk a Miloš DRDLA. *Statistické metody*. Vyd. 4. Brno: PC-DIR Real, 2000. ISBN 80-214-1534-7.
- [7] KROPÁČ, Jiří. *Statistika: náhodné jevy, náhodné veličiny, základy matematické statistiky, indexní analýza, regresní analýza, časové řady*. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská, 2010. ISBN 978-80-214-3866-8.
- [8] CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. Praha: SNTL, 1986.
- [9] CIPRA, Tomáš. *Finanční ekonometrie*. Praha: Ekopress, 2008. ISBN 978-80-86929-43-9.
- [10] KŘIVÝ, Ivan. *Analýza časových řad* [online]. [cit. 2015-09-18]. Dostupné z: <https://publi.cz/download/publication/20?online=1>
- [11] LITSCHMANNOVÁ, Martina. *Úvod do analýzy časových řad* [online]. 2010 [cit. 2017-05-19]. Dostupné z: http://homel.vsb.cz/lit40/SMAD/Casove_rady.pdf
- [12] *Český statistický úřad* [online]. [cit. 2017-05-10]. Dostupné z: www.czso.cz
- [13] *Data Market* [online]. [cit. 2017-04-10]. Dostupné z: www.datamarket.com

ZOZNAM TABULIEK

Tabuľka 1: Voľba trendovej krivky podľa orientačného testu	36
Tabuľka 2: Populácia - lineárny trend - odhady regresných parametrov	41
Tabuľka 3: Populácia - kvadratický trend - odhady regresných parametrov	41
Tabuľka 4: Populácia - porovnanie koeficientov determinácie	42
Tabuľka 5: Populácia - analýza rozptylu	42
Tabuľka 6: Populácia - predikcia pre rok 2013	43
Tabuľka 7: Cudzinci - kvadratický trend	45
Tabuľka 8: Cudzinci - kvadratický trend bez lineárneho člena	45
Tabuľka 9: Cudzinci - exponenciálny trend	46
Tabuľka 10: Cudzinci - exponenciálny trend bez absolútneho člena	46
Tabuľka 11: Cudzinci - porovnanie koeficientov determinácie	47
Tabuľka 12: Cudzinci - analýza rozptylu	47
Tabuľka 13: Cudzinci - predikcie - kvadratické trendy	50
Tabuľka 14: Cudzinci - predikcie - exponenciálne trendy	51
Tabuľka 15: Preletené míle - odhad sezónnej zložky v jednotlivých mesiacoch	53
Tabuľka 16: Dopravné nehody - odhad sezónnej zložky v jednotlivých mesiacoch	56

ZOZNAM OBRÁZKOV

Obrázok 1: Exponenciálny trend	31
Obrázok 2: Modifikovaný exponenciálny trend	33
Obrázok 3: Logistický trend	34
Obrázok 4: Gompertzova krivka	35
Obrázok 5: Populácia - pôvodný časový rad	40
Obrázok 6: Populácia - porovnanie modelov	43
Obrázok 7: Cudzinci - pôvodný časový rad	44
Obrázok 8: Cudzinci - kvadratický trend	48
Obrázok 9: Cudzinci - kvadratický trend bez lineárneho člena	48
Obrázok 10: Cudzinci - exponenciálny trend	49
Obrázok 11: Cudzinci - exponenciálny trend bez absolútneho člena	49
Obrázok 12: Preletené míle - pôvodný časový rad	52
Obrázok 13: Preletené míle - sezónna zložka	53
Obrázok 14: Preletené míle - pôvodný časový rad a kľzavé priemery	54
Obrázok 15: Dopravné nehody - pôvodný časový rad	55
Obrázok 16: Dopravné nehody - sezónna zložka	55
Obrázok 17: Dopravné nehody - pôvodný časový rad a kľzavé priemery	56

ZOZNAM PRÍLOH

Príloha 1: Dáta k príkladu 1 - Populácia v Indii

Príloha 2: Dáta k príkladu 2 - Cudzinci v ČR

Príloha 3: Dáta k príkladu 3 - Preletené míle

Príloha 4: Dáta k príkladu 4 - Dopravné nehody

A PRÍLOHY

A.1 Dáta k príkladu 1 - Populácia v Indii

ROK	POPULÁCIA	ROK	POPULÁCIA	ROK	POPULÁCIA
1950	376,3252	1971	567,8051	1992	903,7496
1951	382,2056	1972	580,7988	1993	921,1075
1952	388,4088	1973	594,193	1994	938,4525
1953	394,9376	1974	608,0026	1995	955,8044
1954	401,7929	1975	622,2324	1996	973,1476
1955	408,9736	1976	636,8837	1997	990,4602
1956	416,4765	1977	651,9352	1998	1007,747
1957	424,2967	1978	667,339	1999	1025,015
1958	432,4276	1979	683,0325	2000	1042,262
1959	440,8623	1980	698,9656	2001	1059,501
1960	449,5955	1981	715,1052	2002	1076,706
1961	458,6267	1982	731,4438	2003	1093,787
1962	467,9621	1983	747,9863	2004	1110,626
1963	477,616	1984	764,7494	2005	1127,144
1964	487,6074	1985	781,7365	2006	1143,289
1965	497,9523	1986	798,9418	2007	1159,095
1966	508,6563	1987	816,3288	2008	1174,662
1967	519,7222	1988	833,8335	2009	1190,138
1968	531,161	1989	851,3747	2010	1205,625
1969	542,9839	1990	868,8907	2011	1221,156
1970	555,1997	1991	886,3487	2012	1236,687

A.2 Dáta k príkladu 2 - Cudzinci v ČR

ROK	POČET CUDZINCOV	ROK	POČET CUDZINCOV
1989	27,325	2003	80,844
1990	27,204	2004	99,467
1991	28,457	2005	110,598
1992	29,145	2006	139,185
1993	31,072	2007	157,512
1994	33,164	2008	172,191
1995	39,242	2009	180,359
1996	46,388	2010	188,952
1997	56,797	2011	196,408
1998	64,352	2012	212,455
1999	66,754	2013	236,557
2000	66,855	2014	249,856
2001	69,816	2015	260,04
2002	75,249		

A.3 Dáta k príkladu 3 - Preletené míle

MESIAC	MÍLE	MESIAC	MÍLE	MESIAC	MÍLE
1962-07	101,6	1964-07	123,5	1966-07	157,8
1962-08	101,5	1964-08	127,4	1966-08	144,2
1962-09	84,3	1964-09	106,2	1966-09	121,6
1962-10	51	1964-10	73,2	1966-10	82,6
1962-11	38,5	1964-11	51,2	1966-11	60,3
1962-12	33,7	1964-12	55,9	1966-12	65,3
1963-01	32,2	1965-01	53,3	1967-01	64,8
1963-02	33,1	1965-02	53,2	1967-02	61,5
1963-03	48,1	1965-03	66,8	1967-03	84,8
1963-04	63,2	1965-04	86,6	1967-04	91,5
1963-05	71,5	1965-05	93,7	1967-05	112,7
1963-06	94,1	1965-06	111,7	1967-06	129,9
1963-07	109,4	1965-07	140,4	1967-07	157,7
1963-08	113,9	1965-08	134,4	1967-08	153,5
1963-09	92,9	1965-09	116,9	1967-09	133,8
1963-10	61,8	1965-10	77,8	1967-10	92,5
1963-11	46	1965-11	53,4	1967-11	62,5
1963-12	47,9	1965-12	58,3	1967-12	61,5
1964-01	44,9	1966-01	58,7	1968-01	60,6
1964-02	45,5	1966-02	56,1	1968-02	59,6
1964-03	61,1	1966-03	73,4	1968-03	74,2
1964-04	69,5	1966-04	88	1968-04	88
1964-05	86,1	1966-05	105,5	1968-05	103,1
1964-06	99,2	1966-06	132,6	1968-06	128,4

MESIAC	MÍLE	MESIAC	MÍLE	MESIAC	MÍLE
1968-07	154,9	1969-11	74,1	1971-03	86,2
1968-08	147,1	1969-12	72,2	1971-04	99,3
1968-09	123,4	1970-01	70,6	1971-05	111,6
1968-10	85,1	1970-02	66,8	1971-06	124,5
1968-11	66,4	1970-03	81	1971-07	149,6
1968-12	65,4	1970-04	95,7	1971-08	143,3
1969-01	64,2	1970-05	114,5	1971-09	131,3
1969-02	55	1970-06	124	1971-10	100,4
1969-03	74,5	1970-07	149,1	1971-11	70,9
1969-04	92,6	1970-08	149	1971-12	76,4
1969-05	110,5	1970-09	131,8	1972-01	76,8
1969-06	122,5	1970-10	100,9	1972-02	68,8
1969-07	146,8	1970-11	77,6	1972-03	90,5
1969-08	149,5	1970-12	76,8	1972-04	102,6
1969-09	127,5	1971-01	63,6	1972-05	128,3
1969-10	96,5	1971-02	68,1		

A.4 **Dáta k příkladu 4 - Dopravné nehody**

ŠTVRŤROK	POČET NEHÔD	ŠTVRŤROK	POČET NEHÔD
2009-01	17726	2013-01	19505
2009-02	18080	2013-02	20961
2009-03	18758	2013-03	21649
2009-04	20251	2013-04	22283
2010-01	17095	2014-01	18529
2010-02	18930	2014-02	21802
2010-03	19685	2014-03	22712
2010-04	19812	2014-04	22816
2011-01	15914	2015-01	20061
2011-02	18954	2015-02	23635
2011-03	19572	2015-03	24220
2011-04	20697	2015-04	25151
2012-01	18745	2016-01	22052
2012-02	20345	2016-02	25407
2012-03	20541	2016-03	25454
2012-04	21773	2016-04	25951