

**Univerzita Hradec Králové**  
**Fakulta informatiky a managementu**  
**Katedra informatiky a kvantitativních metod**

*Analýza časových řad*

**Návrh metod prognózy nezaměstnanosti v okrese Česká Lípa**

Diplomová práce

Autor: Tereza Zajícová  
Studijní obor: Informační management 5

Vedoucí práce: prof. RNDr. Hana Skalská, CSc.

**Prohlášení:**

Prohlašuji, že jsem diplomovou práci zpracovala samostatně a s použitím uvedené literatury.

V Hradci Králové dne 13. 11. 2015

Tereza Zajícová

**Poděkování:**

Děkuji vedoucí mé diplomové práce prof. RNDr. Haně Skalské, CSc. za metodické vedení. Mé poděkování je dále směřováno k Mgr. Barboře Kušnierikové za provedení korektur práce.

## **Anotace**

Diplomová práce se zabývá analýzou časové řady počtu nezaměstnaných v okrese Česká Lípa. Hlavním účelem analýzy je návrh alternativních metod prognózy budoucích hodnot. S použitím teorie časových řad je zkoumána konkrétní časová řada v období od ledna roku 2009 do května roku 2015. V návaznosti na provedenou analýzu je navrženo několik metod prognóz vhodných pro danou časovou řadu. Vhodné metody jsou vybrány podle doporučení z odborné literatury a podle průzkumu, který práci předcházel. Těmito metodami jsou ARIMA, Brownovo dvojitě exponenciální vyrovnání a klasický model dekompozice. Práce navíc řeší problematiku sezónnosti časových řad, jelikož ukazatel počtu nezaměstnaných je ovlivňován sezónní fluktuací.

Obsahem diplomové práce je uplatnění postupů analýzy časových řad a bližší seznámení se s možnostmi prognózy pomocí statistického software, konkrétně programů IBM SPSS Statistics a Statgraphics Centurion XVI. Práce se zabývá i návrhem rozšíření tabulkového editoru MS Excel, který je aktuálně využíván pro predikce hodnot Úřadem práce v České Lípě, o volně dostupné statistické doplňky.

**Klíčová slova:** ARIMA, Box-Jenkins, časové řady, dekompozice, sezónní index, SPSS, Statgraphics, vyrovnání časové řady.

## **Annotation**

**Title: Suggestion of forecasting methods for unemployment in Česká Lípa region.**

The diploma thesis is focused on the time series analysis of the number of unemployed in Česká Lípa region to propose alternative methods of forecast of future values. The time series is analyzed by using the theory of time series in the period of January 2009 to May 2015. Following on the analysis there is suggested several methods of forecasting that are appropriate for the time series. These appropriate methods are chosen by recommendations from specialized sources and by the survey that preceded the thesis. These methods are ARIMA, Brown's double exponential smoothing and classical model of decomposition. The thesis also deals with the issue of seasonality of the time series because the number of unemployed indicator is influenced by seasonal fluctuation.

The thesis contains implementation of the procedures of time series analysis and to learn about possible uses of the statistical software for forecast, namely IBM SPSS Statistics and Statgraphics Centurion XVI programs. The thesis also suggests an extension of Microsoft Excel by a free available statistical add-in. MS Excel is currently used for prediction of future values by Labor Office in Česká Lípa.

Key words: ARIMA, Box-Jenkins, decomposition, seasonal index, time series smoothing, SPSS, Statgraphics, time series.

## Obsah

1.	Úvod .....	1
2.	Teoretická část.....	2
2.1.	Časové řady .....	2
2.1.1.	Typy časových řad.....	4
2.1.2.	Elementární charakteristiky časových řad.....	6
2.2.	Předpověď v časových řadách.....	10
2.3.	Použitá metodika a software.....	14
2.3.1.	Klasický model dekompozice.....	14
2.3.2.	Brownovo exponenciální vyrovnání.....	18
2.3.3.	Box-Jenkinsova metodologie (ARIMA model) .....	20
2.3.4.	Statgraphics Centurion XVI .....	23
2.3.5.	IBM SPSS Statistics .....	28
3.	Praktická část.....	31
3.1.	Popis konkrétní časové řady .....	31
3.2.	Grafická analýza průběhu časové řady .....	35
3.3.	Elementární charakteristiky.....	38
3.4.	Předpověď .....	40
3.4.1.	Dosavadní forma předpovědi .....	41
3.4.2.	Klasický model dekompozice.....	42
3.4.3.	Brownovo dvojité exponenciální vyrovnání.....	44
3.4.4.	ARIMA model v programu Statgraphics.....	46
3.4.5.	ARIMA model v programu IBM SPSS Statistics .....	51
3.5.	Statistické doplňky pro MS Excel .....	56
3.5.1.	Real Statistics .....	56
3.5.2.	Statistician .....	59
3.5.3.	Unistat .....	62
4.	Shrnutí výsledků.....	65
5.	Závěr.....	68
6.	Seznam použité literatury .....	69
7.	Seznamy .....	72
8.	Přílohy .....	74

# 1. Úvod

Diplomová práce se zabývá analýzou časových řad se zaměřením na možnosti předpovědi budoucích hodnot. Práce je rozdělena do dvou hlavních kapitol (teoretická část a praktická část), které jsou dále členěny do několika podkapitol.

Oblastí zkoumání je nezaměstnanost, což je podstatná součást státní politiky a její hodnota ovlivňuje mnoho dalších sfér, jak ekonomických, tak sociálních. Nezaměstnanost v České republice nabývá v posledních měsících pozitivních hodnot. Klesající trend počtu nezaměstnaných je výbornou zprávou pro lidi, kteří hledají práci. Horší vliv má však tento faktor na zaměstnavatele, kterým se snižuje počet uchazečů vhodných na danou pozici. Nezaměstnanost je ovlivňována mnoha faktory a je proto velmi složité kvalitně předpovědět budoucí hodnoty tohoto ukazatele.

Teoretická část popisuje základní vlastnosti časových řad, přes definici, rozdělení časových řad podle podstatných parametrů, problém srovnatelnosti v časových řadách, až po elementární charakteristiky potřebné k analýze časových řad. Dále jsou v této části stručně vysvětleny principy předpovědi budoucích hodnot v časových řadách se zaměřením na základní rozdělení předpovědních metod. Následuje vysvětlení bodové předpovědi a předpovědního intervalu a představení výběrových kritérií pro hodnocení míry kvality předpovědi. Poslední kapitola se zabývá detailněji metodami a softwarem použitými v praktické části práce. Volba metod a statistických programů pro použití při analýze časových řad a predikci budoucích hodnot je výsledkem provedeného průzkumu zdrojů a doporučení týkajících se časových řad s ohledem na časovou řadu počtu uchazečů o zaměstnání.

V praktické části práce jsou řešeny možnosti předpovídání budoucích hodnot časové řady počtu nezaměstnaných v okrese Česká Lípa pro období leden 2009 až květen 2015. Tato časová řada je typická sezónní fluktuací, a proto je část odstavce věnována očištění časové řady od sezónních vlivů. Pro účel předpovědi budoucích hodnot řady počtu nezaměstnaných kapitola analyzuje využití předpovědních metod, především Box-Jenkinsovy metodologie, dále klasického modelu dekompozice a Brownova dvojitého exponenciálního vyrovnání. Z možností využití statistického softwaru pro analýzu a předpověď časových řad jsou prověřeny programy IBM SPSS Statistics a Statgraphics Centurion XVI, které nabízí možnost vygenerování hodnot predikce. Poslední kapitola obsahuje prověření možností využití doplňků pro Microsoft Excel.

Výsledky analýz a předpovědí časové řady jsou na závěr srovnány s aktuálními hodnotami předpovědi, kterou s půlročními intervaly vytváří Úřad práce v České Lípě.

## 2. Teoretická část

Teoretická část je členěna do tří částí: Časové řady, předpověď v časových řadách a použitá metodika a software. V odstavci o časových řadách jsou vymezeny základní pojmy nutné pro analýzu časových řad, jako jsou definice, jejich praktické využití a srovnatelnost v časových řadách. Dále jsou definovány typy časových řad a elementární charakteristiky používané při analýze časových řad.

Druhý odstavec teoretické části je věnován předpovědím v časových řadách. Pozornost je věnována především obecným metodám předpovědi, vysvětlení pojmů „předpovědní interval“ a „bodová předpověď“ a v neposlední řadě popisu kritérií hodnotících kvality předpovědi.

Třetí část obsahuje vysvětlení metodik použitých v praktické části, jimiž jsou „klasický model dekompozice“, „Brownovo exponenciální vyrovnání“ a „Box-Jenkinsova metodologie“. Dále jsou popsány použité software IBM SPSS Statistics a Statgraphics.

### 2.1. Časové řady

Časové řady jsou ve statistických zdrojích popsány různými definicemi. Podstata je však vždy stejná a záleží pouze na autorovi, jak definici vyjádří. Jednou z jednodušších definic je: „Data, která vytváří časovou řadu, vznikají jako pozorování, uspořádané chronologicky (v čase).“[10].

Upřesňující definicí je: „Časová řada je soubor pozorování v průběhu času uspořádaných chronologicky zpravidla od minulosti k přítomnosti.“[5].

Třetí definicí časové řady, tentokrát z pohledu pozorovaných dat, může být: „Data, která vytvářejí časovou řadu, vznikají jako chronologicky uspořádaná pozorování a podstatné pro ně je, že jsou v čase chronologicky uspořádána.“[7].

Vět definujících časovou řadu existuje nepřeborné množství, avšak všechny se shodují v jedné souhrnné charakteristice:

*Jedná se o chronologicky uspořádaný soubor pozorování v určitém čase.*

#### **Analýza časových řad**

„Statistická data popisující sociální nebo ekonomické jevy v daném čase, zapisujeme pomocí tzv. časových řad. Zápis těchto jevů pomocí časových řad umožňuje provádět nejen kvantitativní analýzu zákonitostí v jejich dosavadním průběhu, ale dává zároveň možnost prognózovat jejich vývoj v budoucích časových intervalech.“[5].

V analýze časových řad jde o zkoumání pravidelností a nepravidelností vývoje jevů pomocí modelovací techniky. Cílem analýzy časových řad je porozumět a popsat vlastnosti vývoje a změn hodnot proměnných v čase, analyzovat znalosti z minulého či přítomného



chování dat, sestavit kvalitní matematicko-statistický model časové řady a na jeho základě správně odhadnout hodnoty proměnných v budoucích obdobích.[11][19]

### **Vyjádření časových řad**

Je důležité na začátek vymežit pojem časové řady a pochopit jeho vlastnosti a charakteristiky. Jednou z matematických vyjádření časové řady může být tato: [5]

$$Y_1, Y_2, Y_3, \dots, Y_n,$$

kde  $n$  je celkový počet pozorování.

Jednotlivá pozorování jsou označována symbolem  $Y_t$ , který udává hodnotu pozorování v čase  $t$ , kde  $t = 1, 2, 3, \dots, n$

$t$  ... časová proměnná

### **Využití časových řad v praxi**

Příklady časových řad lze nalézt v různých oblastech. Existují časové řady: [21][22]

- politických věd (volební preference, důvěra v prezidenta)
- fyziky a techniky (seizmický záznam, průběh výstupního signálu určitého elektrického přístroje)
- ekonomické (HDP, HNP, nezaměstnanost, fluktuace měn, analýza poptávky po určitém výrobku, analýza vývoje kurzu akcií na burze)
- z oblasti sociologie (vývoj rozvodovosti, míra imigrace, míra kriminality, velikost populace)
- z psychologické oblasti (závislí na hazardu, počet alkoholiků, míra recidivy k drogám)
- z oblasti biomedicíny (tep, záznam EKG nebo EEG, krevní tlak)
- meteorologické monitorování teploty, tlaku, procenta oblačnosti

### **Srovnatelnost dat v časových řadách**

Podstatné pro správné závěry analýz je to, aby se údaje v časové řadě skládaly ze srovnatelných dat (pozorování, hodnot) určitého ukazatele.

Srovnatelnost je základním požadavkem pro smysluplnost časové řady. Srovnatelností dat se rozumí neměnnost metodických, obsahových, věcných, časových a prostorových vymezení použitých při jejich vytváření po delší časové období. Z pohledu uživatele časových řad je metodická, případně jiná nesrovnatelnost základní překážkou jejich využívání. I v případě uvedení popisu změn náplně ukazatele může být vysvětlen zdroj nestability dat, avšak to neumožňuje uživateli s daty pracovat.

Věcnou nesrovnatelností může být například legislativní změna, měnící obsah ukazatele (změna daní, účetních standardů, celních poplatků a podobně).

Prostorová srovnatelnost zahrnuje z hlediska územního například data ze stejného geografického území nebo ze stejné organizační struktury.

Časová nesrovnatelnost je častým problémem u intervalových ukazatelů, které nelze vždy srovnávat (měsíce nemají stejný počet dní). S těmito problémy se u okamžikových časových řad neseťkáváme, protože se vždy vztahují k předem zvoleným časovým okamžikům.

Metodickou nesrovnatelností může být například změna metodiky způsobu zjišťování ukazatele nebo použití různých metod výpočtu ukazatele. Například je tomu tak při výpočtu hrubého domácího produktu (HDP), který může být definován, respektive spočten třemi způsoby:

a) Produkční metodou:

$$\text{HDP} = \text{Produkce} - \text{Mezispotřeba} + \text{Daně z produktů} - \text{Dotace na produkty}$$

b) Výdajovou metodou:

$$\text{HDP} = \text{Výdaje na konečnou spotřebu} + \text{Tvorba hrubého kapitálu} + \text{Vývoz} - \text{Dovoz}$$

c) Důchodovou metodou:

$$\text{HDP} = \text{Náhrady zaměstnancům} + \text{Daně z výroby a z dovozu} - \text{Dotace} + \text{Čistý provozní přebytek} + \text{Čistý smíšený důchod} + \text{Spotřeba fixního kapitálu}$$

[16][21][25]

### **2.1.1. Typy časových řad**

Časové řady se dělí podle různých hledisek. Jednotlivá rozdělení slouží k jejich popsání a zjednodušení při analýze časových řad.

#### **Deterministické × stochastické**

Hlavní rozdělení časových řad je založeno na náhodné složce. Časové řady jsou proto rozděleny podle výstupu modelu na deterministické a stochastické. V deterministických modelech časových řad je výstup modelu plně určen hodnotami parametrů a počátečními podmínkami. Oproti tomu stochastické modely obsahují vlastní náhodnost, a tudíž stejný soubor hodnot parametrů a stejné počáteční podmínky povedou k souboru hodnot s různými výstupy.[1]

#### **Dlouhodobé × krátkodobé**

Podle délky sledování (takzvané periodicity) jsou časové řady děleny na dlouhodobé a krátkodobé. „Zatímco dlouhodobé časové řady jsou sledovány v roční frekvenci, časové řady krátkodobé jsou sledovány ve frekvencích kratších než jeden rok. Specifické postavení v rámci krátkodobých časových řad mají tzv. vysokofrekvenční časové řady, tj. časové řady sledované např. v denní či hodinové frekvenci.“[1].

### **Ekvidistantní × neekvidistantní**

Dále jsou časové řady děleny na ekvidistantní a neekvidistantní, které se liší délkou frekvence. Ekvidistantní mají stále stejnou frekvenci pozorování. Neekvidistantní jsou v tomto ohledu komplikovanější pro další analýzy, jelikož je nutné provést úpravy nekonstantní frekvence (tzv. očištění).

### **Okamžikové × intervalové**

Časové řady jsou nazývány podle charakteru pozorování okamžikové, kde hodnota pozorování je dána k určitému okamžiku, jednomu konkrétnímu dni nebo hodině. Okamžikové časové řady jsou definovány tím, že ukazatele časových řad charakterizují kolik jevů, věcí, událostí, apod. existuje v určitém časovém okamžiku.

Anebo jsou časové řady nazývány řadami intervalovými či úsekovými, kde je velikost sledovaného ukazatele dána délkou intervalu pozorování. To znamená, že ukazatele v časových řadách jsou nazývány intervalovými, pokud jsou popisem toho, kolik jevů, věcí nebo událostí vzniklo či zaniklo na daném časovém intervalu.

Zásadním rozdílem mezi okamžikovými a intervalovými typy časových řad je to, že údaje intervalových časových řad lze počítat, a tím lze vytvořit součty za více období. Naproti tomu počítání údajů okamžikových řad nemá reálnou interpretaci. Při zpracování intervalových časových řad je nutné přihlídnout k tomu, zda délka časových intervalů, v nichž se hodnoty časové řady měří, je stejná nebo rozdílná. Rozdílná délka intervalů totiž ovlivňuje hodnoty ukazatelů intervalových časových řad a tím zkresluje jejich vývoj. Například v měsících je různý počet dnů, takže při hodnocení ekonomických výsledků za jednotlivá období je nutno k tomu přihlídnout.[16]

### **Stacionární × nestacionární**

Další důležité rozdělení časových řad je podle takzvané stacionarity na stacionární a nestacionární. Stacionární časová řada se vyznačuje tím, že má konstantní průměr a rozptyl, a korelace dvou časově posunutých pozorování závisí na délce posunu. V podstatě se jedná o časovou řadu, kde je trendová složka konstantní ( $T_t = k$ ). „Úroveň a rozptyl stacionární řady jsou tedy konstantní v čase (trend, sezónnost či proměnný rozptyl jsou se stacionaritou neslučitelné a pro její dosažení musí být přednostně z řady odstraněny) a také kovarianční struktura takové řady musí být v čase neměnná (např. charakter závislosti mezi prvním a druhým kvartálem čtvrtletní stacionární časové řady musí zůstat ve všech rocích stejný).“[8].

Podrobnějším vysvětlením stacionarity časové řady se diplomová práce nezabývá. Pro hlubší pochopení stacionarity časových řad je vhodná kniha [1] ARLT, Josef; ARLTOVÁ, Markéta. *Finanční časové řady (Vlastnosti, metody modelování, příklady a aplikace)*, která na stranách 30-32 výborně popisuje tuto problematiku.

## Primární × sekundární

Druhem sledovaných ukazatelů je charakteristika dat, podle které mohou být časové řady rozděleny na řady s primárními (absolutními) daty (původní data jaká byla naměřena, např. odpracovaná doba) a řady se sekundárními daty (odvozená z dat primárních, např. zisk je odvozen z výnosů a nákladů). Sekundární data mohou být například součtová, průměrná nebo poměrová.[13]

### 2.1.2. Elementární charakteristiky časových řad

Tato kapitola uvádí základní metody, které pomáhají analyzovat časovou řadu a slouží tak k orientačnímu posouzení základních vlastností časových řad.

V dnešní době se také těžko obejdeme bez zpracování dat pomocí některého statistického softwaru, jako je např. Statistica, SPSS, případně statistických funkcí v běžném MS Excel nebo OpenOffice.[6]

#### Grafická analýza dat

Prvním obvykle používaným postupem, pomocí kterého mohou být snadno analyzovány základní charakteristiky časové řady, je grafická (vizuální) analýza. Jednoduše jde o vykreslení časové řady pomocí grafu. Intervalové časové řady lze graficky znázornovat třemi způsoby (sloupkovými, hůlkovými nebo spojnicovými grafy). Pro znázornění okamžikových časových řad se běžně používá *spojnicový graf (průběhový diagram)*. Jednotlivé hodnoty časové řady jsou vyneseny ve středech příslušných intervalů jako body, které jsou spojeny úsečkami. U spojnicových grafů vodorovná osa zaznamenává časovou proměnnou a na svislé ose se zobrazují hodnoty ukazatele časové řady  $y_t$ .

K posouzení rozložení dat podle velikosti pomocí kvartilů slouží krabicový diagram (Box plot), který poukazuje na případnou asymetrii. Velkou předností je, že vyznačuje odlehlé hodnoty. Pokud je v jednom obrázku více diagramů, pak se dá posoudit i variabilita. Odlehlá hodnota bývá označována kroužkem nebo bodem, jestliže je větší než  $x_{0,75} + 1,5q$  nebo menší než  $x_{0,25} - 1,5q$ , kde  $q = x_{0,75} - x_{0,25}$  je kvartilová odchylka. Některé programy rozlišují ještě extrémní hodnoty, které leží buď pod  $x_{0,25} - 3q$  nebo nad  $x_{0,75} + 3q$ .

Dalším praktickým znázorněním dat v časové řadě je histogram. Histogram představuje grafické zobrazení intervalového rozložení četností sledované veličiny. Je to graf skládající se z obdélníků sestrojených nad třídícími intervaly, přičemž obsah každého obdélníku je roven relativní četnosti příslušného třídícího intervalu. (Ve statistických programových systémech se na vodorovnou osu vynášejí třídící intervaly, avšak na svislou osu se vynášejí absolutní nebo relativní četnosti třídících intervalů.) Pomocí histogramu se získává názor na rozptýlení hodnot sledované veličiny či odhad tvaru rozložení. [7][13][16]

A nyní popis výpočtů elementárních statistických charakteristik, které jsou použity v praktické části diplomové práce.

### **Průměr**

Průměr intervalové časové řady se počítá jako aritmetický průměr hodnot časové řady v jednotlivých intervalech. Je dán vzorcem:

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$$

Průměrná úroveň okamžikové časové řady  $Y_t$ , kde  $t = 1, 2, \dots, t$  za delší sledované období, se zjišťuje pomocí chronologického průměru. Podle délky časového intervalu mezi rozhodujícími okamžiky, v kterých jsou hodnoty okamžikového ukazatele zjištěny, se počítá jednoduchý nebo vážený chronologický průměr.

Při shodných, stejně dlouhých intervalech mezi všemi okamžiky zjišťovaných údajů časové řady se počítá jednoduchý chronologický průměr podle vztahu:

$$\bar{Y} = \frac{\sum_{t=1}^n \frac{(Y_{t-1} + Y_t)}{2}}{t - 1}$$

Jakmile jsou časové intervaly mezi rozhodujícími okamžiky různé délky (značeno  $d_t$ ), je pro výpočet průměru vhodnější použít vážený chronologický průměr, který je daný vztahem:

$$\bar{Y} = \frac{\sum_{t=1}^n \frac{(Y_{t-1} + Y_t)}{2} * d_t}{\sum d_t}$$

Průměrování se často používá při transformaci časových řad s vyšší frekvencí sledování na řady s frekvencí kratší (např. měsíční na čtvrtletní).

[16][19]

### **Diference (průměrný přírůstek)**

Nejjednodušší charakteristikou popisu vývoje časové řady jsou první diference (někdy absolutní přírůstky), které se počítají jako rozdíl dvou po sobě jdoucích hodnot časové řady. První diference vyjadřují přírůstek hodnoty časové řady, tedy o kolik se změnila její hodnota v určitém okamžiku oproti určitému okamžiku bezprostředně předcházejícímu. Je-li zjištěno, že první diference kolísají kolem konstanty, lze říci, že sledovaná časová řada má lineární trend, tedy její vývoj lze popsat přímkou. Pokud uvedeme, že  $n$  je celkový počet pozorování, pak je možno výpočet prvních diferencí vyjádřit vzorcem:

$$\Delta Y_t = Y_t - Y_{t-1}, \text{ kde } t = 2, 3, \dots, n$$

Jedná se o jednoduchou míru dynamiky poklesu (růstu), která se určí rozdílem hodnot sledované proměnné ve dvou po sobě jdoucích obdobích. Porovnává se hodnota  $Y_t$  s hodnotou v předcházejícím časovém období  $Y_{t-1}$  časové řady  $Y_1, Y_2, \dots, Y_t$ . S postupným posuvem o jedno časové období pro všechna  $t = 2, 3, \dots, t$  je získána nová časová řada příslušné charakteristiky, avšak o první hodnotu kratší.

Průměrný absolutní přírůstek (úbytek) je počítán jako aritmetický průměr  $t-1$  absolutních přírůstků (úbytků) podle vztahu:

$$\overline{\Delta Y}_t = \frac{\sum_{t=1}^n \Delta Y_t}{t-1}$$

[16][19]

### **Tempo růstu (průměrné tempo růstu)**

Tempo růstu neboli relativní přírůstek vyjádřený většinou v procentech udává, o kolik procent vzrostla (poklesla) hodnota proměnné  $Y$  v čase  $t$  oproti hodnotě z předcházejícího období. Zjednodušeně, jde o podíl sousedních hodnot. Pokud tento podíl vynásobíme 100, určuje, na kolik procent hodnoty v čase  $t-1$  se změnila hodnota v čase  $t$ .

Průměrný koeficient růstu se vypočítá podle vzorce geometrického průměru  $t-1$  koeficientů růstu: [19]

$$\bar{k}_t = \sqrt[t-1]{k_1 k_2 k_3 \dots k_t}$$

### **Koeficient růstu (průměrný koeficient růstu)**

Prostý koeficient růstu se vypočítá podle vzorce

$$k_t = \frac{Y_t}{Y_{t-1}}; \text{ kde } t = 2, \dots, n.$$

Vzorec meziročního koeficientu růstu s měsíčním krokem je

$$k_{t(12)} = \frac{Y_t}{Y_{t-12}}; t = 13, \dots, n.$$

Zde se předpokládá časová řada, která má za časový krok měsíc. Pak  $k_{t(12)}$  ukazuje na změnu za 1 rok pro každý měsíc vzhledem k minulému roku.

Průměrný koeficient růstu je pak počítán podle vzorce

$$\bar{k} = \sqrt[n-a]{\frac{Y_n}{Y_1}}$$

kde u prostého koeficientu je  $a=1$ , u meziročního s měsíčním krokem je  $a = 12$  a u meziročního se čtvrtletním krokem je  $a = 4$ . [7]

## Elementární charakteristiky v nabídce statistických programů

„Analýza časových řad se v současnosti provádí výhradně na počítačích pomocí vhodného softwaru (například SPSS, Statistica, S+, Statgraphics,...). Velká většina statistických a ekonometrických softwarů má algoritmy těchto analýz zabudované ve svých standardních nabídkách.“[13].

### a) Statgraphics

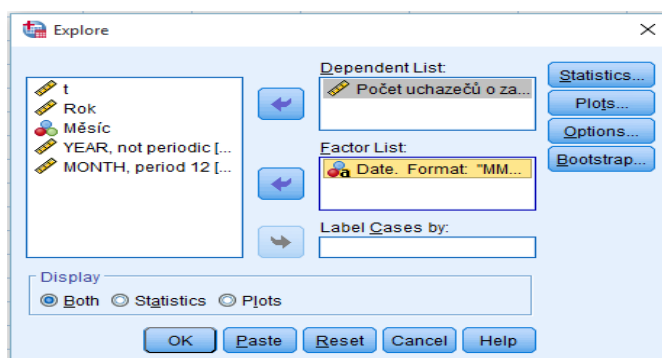
Výše uvedené elementární charakteristiky lze vypočítat pomocí programu STATGRAPHICS v panelu Generate Data. Výsledky se zobrazí ve zvoleném sloupci tabulkového editoru po vepsání příslušného příkazu do řádku Expression.

Pro výpočet absolutních přírůstků (1. diferencí) proměnné  $Y_t$  lze v řádku Expression použít příkaz  $\text{diff}(Y_t)$ . Pro výpočet průměrného absolutního přírůstku použijeme příkaz  $\text{avg}(\text{diff}(Y_t))$ . Koeficienty růstu časové řady  $Y_t$  se počítají pomocí kombinace operátorů příkazem  $\text{exp}(\text{diff}(\log(Y_t)))$ . A průměrný koeficient růstu lze získat příkazem  $\text{geomean}(\text{exp}(\text{diff}(\log(Y_t))))$ . [2]

### b) IBM SPSS Statistics

Pro průzkum spojitě proměnné v programu IBM SPSS Statistics slouží funkce Explore, která je vyvolána pomocí příkazu Analyze/Descriptive Statistics/Explore. Obrázek 1 zobrazuje úvodní dialog této funkce, ve kterém se volí proměnné. Do okna Dependent list patří spojitě proměnné, které chceme analyzovat. Okno Factor list umožňuje vložit kategoriální proměnnou, která rozdělí soubor na několik částí (pro každou kategorii jednu) a ty je možné mezi sebou porovnat. Pro časové řady se sezónními vlivy je vhodné vyplnit okno Label Cases by, kam se vkládá proměnná, která identifikuje jednotlivé případy. Pomocí tlačítka Statistics se volí možnosti Descriptives a Outliers. Pomocí tlačítka Plots jsou v části Descriptives zvoleny grafy pro výstup (Histogram, Stem-and-Leaf). Analýza se spouští pomocí tlačítka OK. Výstupem funkce je několik grafů a tabulek. Důležitá je tabulka Descriptives, která shrnuje základní popisnou statistiku. [6]

Obrázek 1 - Úvodní dialog funkce Explore



(zdroj: statistický program IBM SPSS Statistics)

## 2.2. Předpověď v časových řadách

Účelem modelování časových řad není jen popis minulého vývoje, ale rovněž odhad vývoje budoucího. Předpovídání budoucího vývoje je v podstatě založeno na extrapolaci, která může být krátkodobá či dlouhodobá. Předpokladem úspěšné předpovědi je, že nedojde k podstatným změnám podmínek, které by mohly způsobit změnu ve vývoji jevu.

Prognózou neboli předpovědí se rozumí předpovídání vývoje budoucnosti založené na vědeckém postupu, s použitím statistických metod. Při statistickém odhadu je usuzováno z části na celek, což znamená, že na základě dosažených výsledků ve výběrovém souboru (například charakteristik, funkcí, ...) jsou činěny odhady platné pro základní soubor.[11]

### Předpovědní metody

Analýza základních charakteristik a vlastností časové řady je podstatnou součástí práce před samotnou tvorbou předpovědí. „Znalost modelu umožňuje předpovídat budoucí vývoj systému. Konstrukce modelu umožňuje řídit a optimalizovat činnost příslušného systému vhodnou volbou vstupních parametrů a počátečních podmínek.“[10].

Technika předpovědí v časových řadách je spjata se dvěma hlavními metodami.

- a) Kvantitativní metody (jsou k dispozici dostatečné kvantitativní informace)

Jedná se o objektivní přístup, který je založen na naměřených statistických údajích. Předpovídají pokračování historických vzorů, jako je například růst tržeb nebo hrubého národního produktu.

- b) Kvalitativní metody (nejsou k dispozici žádné nebo jen nedostatečné kvantitativní informace, ale existují dostatečné kvalitativní znalosti)

Tato metoda má naopak subjektivní charakter, odráží se od názorů odborníků. Jedná se například o předpověď vlivu cen oleje na jeho spotřebu. Příkladem může být metoda Delfi.

„Výběr předpovědní techniky závisí na celé řadě faktorů, zejména na požadované formě předpovědi, časovém horizontu předpovědi, požadované přesnosti, charakteru vstupních dat a jejich dostupnosti.“[16]. Pro aplikaci kvantitativní předpovědi je nutné, aby soubor splňoval tři podmínky. V první řadě musí být dostupné informace a data z minulosti. Další podmínkou je, že tyto informace je možné kvantifikovat ve formě numerických dat. A v neposlední řadě musíme mít jistotu, že lze předpokládat budoucí pokračování alespoň některých aspektů z minulosti.



## **Bodová předpověď a předpovědní interval**

V období analýzy nebo interpolace řady  $t = 1, 2, \dots, t$  se zjišťuje, zda řada  $Y_t$  má trend a hledá se jeho model. Z hodnot časové řady  $Y_t$  jsou potom odhadovány parametry modelu. Je-li odhad trendu statisticky významný, je možnost jej využít jako prognostický model pro výpočet předpovědí neboli extrapolací. Extrapolacemi jsou nazývány kvantitativní odhady budoucích hodnot časové řady, které vznikají prodloužením vývoje z minulosti a přítomnosti do budoucnosti s horizontem  $h$  za předpokladu, že se tento vývoj nezmění. Extrapoláční předpovědi se obecně dělí na bodové a intervalové.[2]

„Bodovým odhadem se rozumí odhad na základě jednoho čísla. Odhadem charakteristiky či parametru základního souboru je výběrová charakteristika či parametr. Obvykle je volen výběrový protějšek, jako je tomu například u směrodatné odchylky  $s^2$  a výběrové směrodatné odchylky  $\sigma^2$ .“[11]. Bodová předpověď tedy udává bodový odhad hodnoty časové řady v určitém budoucím okamžiku. Jedná se o numerickou hodnotu v čase  $t$ , která je spíše orientační. Avšak skutečná hodnota se se spolehlivostí 95% bude pohybovat v 95%ním předpovědním intervalu.

„Intervalovým odhadem charakteristiky či parametru základního souboru se rozumí stanovení intervalu, v němž se odhadovaná charakteristika či parametr s danou pravděpodobností nachází. Mluví se o  $100(1-\alpha)\%$  intervalu charakteristiky  $\Theta$ .  $P(T' \leq \Theta \leq T'') = 1 - \alpha$ ; kde  $T'$  je dolní hranice intervalu a  $T''$  je horní hranice intervalu. Za  $\alpha$  se zpravidla volí  $\alpha=0,05$  a  $\alpha=0,01$ , takže s 95% nebo s 99% pravděpodobností v sobě interval zahrnuje odhadovanou charakteristiku. Hodnoty  $\alpha$  jsou rizikem odhadu.“[11]. Intervalová předpověď (předpovědní interval) je v podstatě analogií intervalu spolehlivosti a vyjadřuje dolní a horní hranici hodnot, ve kterých se bude pohybovat skutečná budoucí hodnota. Další výhodou předpovědních intervalů je to, že uživatelé nabízí nejlepší a nejhorší možný případ budoucí hodnoty.

### **Základní kroky předpovědi**

Pro jakýkoliv předpovědní úkol, pro který jsou dostupná kvantitativní data, existuje pět základních kroků.

- a) Definice problému
- b) Shromáždění informací
- c) Předběžná (průzkumná) analýza
- d) Výběr modelu a jeho aplikace
- e) Využití a vyhodnocení prognostického modelu

[18]

## Jednoduché míry kvality předpovědi

Používané postupy při předpovědi budoucího chování jsou založeny na principu, že „historie se opakuje“. Tento předpoklad bývá v praxi splněn s různou přesností, a proto je vhodné u vyhlazování a předpovědi v časových řadách uvádět i spolehlivost získaných výsledků a hodnotit úspěšnost predikce. Chyba předpovědi  $r_t$  v čase  $t$  je definována vztahem  $r_t = y_t - \hat{y}_t(t-1)$ ,  $t = 1, 2, \dots, n$ , kde  $y_t$  značí skutečně naměřenou hodnotu ukazatele v čase  $t$  a  $\hat{y}_t(t-1)$  značí předpověď této hodnoty pořízenou v časovém okamžiku bezprostředně předcházejícím.

[13][16]

Při posuzování kvality předpovědi v dané časové řadě je nutno vzít v úvahu všechny zkonstruované předpovědi. „Při analýze časové řady někdy můžeme dojít k závěru, že existuje několik akceptovatelných modelů. Někdy je poměrně jednoduché z této množiny vybrat ten nejlepší. Jsou však situace, kdy tuto jistotu nemáme a výběr nejlepšího modelu je velice obtížný. Pro řešení této úlohy bylo proto navrženo několik dodatečných kritérií. Tato kritéria jsou založena na porovnání reziduí jednotlivých modelů prostřednictvím souhrnných statistik.“[1]. V praxi se nejčastěji používají následující míry kvality předpovědi:

- a) Součet čtvercových chyb SSE (Sum of Squared Errors)

$$SSE = \sum_{t=1}^n r_t^2$$

- b) Průměrná čtvercová chyba MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{t=1}^n r_t^2 = \frac{1}{n} SSE$$

- c) Průměrná absolutní odchylka MAD (Mean Absolute Deviation)

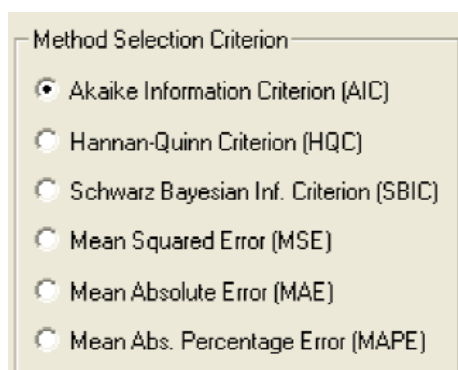
$$MAD = \frac{1}{n} \sum_{t=1}^n |r_t|$$

Jsou-li srovnány všechny uvedené míry, zjistíme, že míry MSE a SSE (ve srovnání s MAD) posuzují mnohem přísněji velké chyby předpovědi než chyby malé.

## Míry kvality předpovědi používané ve specializovaných programech Statgraphics Centurion XVI a IBM SPSS Statistics

V programu Statgraphics jsou užívané metody výběrového kritéria pro výběr nejlepšího modelu. Procedura otestuje každý z uvedených modelů a vybere ten, u kterého vyjde nejnižší hodnota výběrového kritéria. Statgraphics nabízí šest kritérií k výběru.

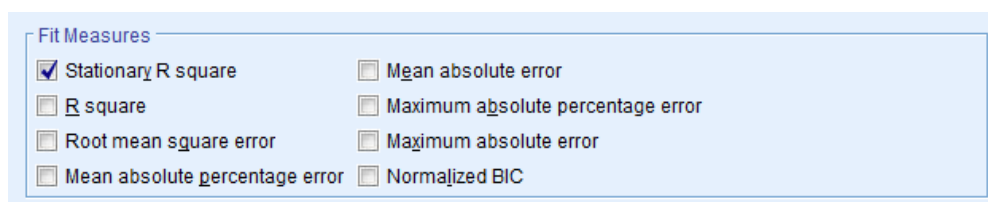
Obrázek 2 - Metody výběrových kritérií používaných pro výběr vhodného modelu



(zdroj: Statgraphics)

Program IBM SPSS Statistics nabízí volbu z osmi výběrových kritérií.

Obrázek 3 - Metody výběrových kritérií používaných pro výběr vhodného modelu



(zdroj: IBM SPSS Statgraphics)

Popis metod, které znázorňuje Obrázek 3 a Obrázek 2:

a) Akaikeho informační kritérium (AIC)

Akaike (1974) navrhl kritérium ve formě:

$$AIC = 2 \ln(RMSE) + \frac{2c}{n},$$

kde RMSE je průměr čtverců odchylek vyjádřený vzorcem

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2};$$

kde  $c$  je počet odhadovaných koeficientů ve zkoumaném modelu; a  $n$  je počet pozorování v časové řadě. Všimněte si, že AIC je vlastně funkcí rozptylu z reziduí modelu ovlivněné počtem odhadovaných koeficientů. Obecně, kritérium vybírá model, který minimalizuje průměrnou odchylku čtverců s použitím minimálního počtu koeficientů (v souvislosti s počtem dostupných dat).

b) Hannan-Quinn kritérium (HQC) - je vyjádřen vzorcem:

$$HQC = 2 \ln(RMSE) + \frac{2p * \ln(\ln(n))}{n}$$

Toto kritérium, na rozdíl od AIC, využívá jiný parametr pro počet odhadnutých koeficientů.

c) Schwarz-Bayesian informační kritérium (SBIC) – je vyjádřen vzorcem:

$$SBIC = 2 \ln(RMSE) + \frac{p * \ln(n)}{n}.$$

Opět je parametr pro počet odhadnutých koeficientů rozdílný od předešlých dvou kritérií.

d) Průměrná absolutní odchylka (Mean Absolute Error neboli MAE)

Vybraný model je ten s celkově nejmenší průměrnou absolutní odchylkou.

e) Průměrná absolutní procentuální odchylka (Mean Absolute Percentage Error neboli MAPE)

Vybraný model je ten s nejmenší průměrnou absolutní procentuální odchylkou.

[34]

## 2.3. Použitá metodika a software

První tři podkapitoly popisují metodiky pro analýzu a predikci časových řad. Pro krátké časové řady je vhodnější použít dekompozici časových řad, naopak pro dlouhé časové řady je vhodnější použít Box – Jenkinsovu metodologii. Pro předpovědi na krátká období dopředu je zase vhodné použít metodu Brownova dvojitého exponenciálního vyrovnání.

V posledních dvou podkapitolách je popsána práce se specializovanými statistickými programy. Programy Statgraphics Centurion XVI a IBM SPSS Statistics jsou zvoleny s ohledem na možné budoucí uživatele, jimiž budou pracovníci úřadu práce v České Lípě. Oba programy nabízejí analýzu trendu a vývoj předpovědí časové řady rychle a jednoduše i pro uživatele se základními analytickými schopnosti.

### 2.3.1. Klasický model dekompozice

Hodnoty časové řady mohou být, zejména z ekonomické praxe, rozloženy na několik složek. Časovou řadu si lze představit jako trend, na který jsou „nabaleny“ ostatní složky. Rozklad neboli dekompozice časové řady na tyto složky je motivován tím, že v jednotlivých složkách se snadněji podaří zjistit zákonitosti v chování řady než v původní nerozložené řadě. U některých časových řad mohou v jejich dekompozici některé složky chybět.[16]

#### Multiplikativní a aditivní forma dekompozice

Klasická analýza časových řad vychází z předpokladu, že časovou řadu  $Y_t$  pro  $t = 1, 2, \dots, T$  je možné rozložit na čtyři složky: trendovou, cyklickou, sezónní a nesystematickou.

Modely dekompozice mohou být dvojího typu:

a) Multiplikativní model

Hodnoty časové řady se dají určit jako součin hodnot jednotlivých složek ve tvaru

$$Y_t = T_t * C_t * S_t * I_t$$

Po multiplikativní dekompozici je trendová složka časové řady ve stejných měrných jednotkách jako původní časová řada, ale ostatní složky časové řady (cyklická, sezónní a nesystematická) jsou v relativním vyjádření. Multiplikativní dekompozice se používá v případě, že variabilita časové řady roste v čase, nebo se v čase mění.

b) Aditivní model

Hodnoty časové řady se dají určit jako součet hodnot jednotlivých složek ve tvaru

$$Y_t = T_t + C_t + S_t + I_t$$

Po aditivní dekompozici jsou jednotlivé složky časové řady ve stejných měrných jednotkách jako původní časová řada. Aditivní dekompozice se používá v případě, že variabilita hodnot časové řady je přibližně konstantní v čase.

[2]

V některých výjimečných případech může být použit ještě model semimultiplikativního typu, který je vhodný tehdy, existuje-li mezi složkami časové řady logický vztah funkčního či korelačního charakteru. Proto je třeba při volbě promyslet logické důvody a důsledky. Například je-li sezónní kolísání proporcionální k trendu (Spotřeba plynu ve městě se zvyšuje, pak je o dovolených větší absolutní pokles).[11]

### **Trendová složka**

Trendová složka vyjadřuje dlouhodobou tendenci vývoje zkoumaného jevu. Je výsledkem faktorů, které dlouhodobě působí stejným směrem např. technologie výroby, demografické podmínky, podmínky na trhu. Je-li ukazatel dané časové řady v průběhu celého sledovaného období v podstatě na stejné úrovni, a kolem této úrovně pouze kolísá, pak mluvíme o časové řadě bez trendu. Nejpoužívanější metodou popisu vývoje časové řady je regresní analýza, neboť umožňuje nejen vyrovnání pozorovaných dat časové řady, ale také prognózu jejího dalšího vývoje. Při regresní analýze se předpokládá, že analyzovanou časovou řadu, jejíž hodnoty jsou  $Y_1, Y_2, \dots, Y_t$ , lze rozložit na složky trendovou a reziduální, tj.

$$Y_t = T_t + e_t, \text{ kde } t = 1, 2, \dots, n$$

[2][16]

Základním problémem je volba vhodného typu regresní funkce. Ten je určován z grafického záznamu průběhu časové řady nebo na základě předpokládaných vlastností trendové složky, vyplývajících z ekonomických úvah.

Klasická analýza časových řad předpokládá, že trendová funkce má v čase konstantní parametry. V delším časovém období je tento předpoklad nereálný, proto je vhodné využívat adaptivní techniky, jako je například metoda klouzavých průměrů. Modelování trendu pomocí klouzavých průměrů nebo pomocí klouzavých mediánů se používá, je-li vývoj řady v důsledku silného vlivu nesystematické složky nerovnoměrný nebo má extrémní hodnoty.[2]

### **Cyklická složka**

Cyklická, někdy označována jako periodická složka vyjadřuje kolísání okolo trendu, ve kterém se střídají fáze růstu a poklesu. Jednotlivé cykly (periody) se vytvářejí za období delší než jeden rok a mají nepravidelný charakter.

Cyklická složka bývá považována za nejspornější složku časové řady. Délka jednotlivých cyklů časové řady, která je rovna vzdálenostem mezi dvěma sousedními horními respektive dolními body zvratu, a také intenzita jednotlivých fází cyklického průběhu se mohou měnit. Cyklická složka může být důsledkem evidentních vnějších vlivů, někdy je ale určení jejich příčin velmi obtížné. Eliminace cyklické složky je obtížná jak z věcných důvodů, neboť je obtížné nalézt příčiny vedoucí k jejímu vzniku, tak i z výpočetních důvodů, protože charakter této složky se může v čase měnit.[2][16]

### **Sezónní složka**

Periodické modely obsahují některou z periodických složek charakterizující kolísání časové řady. S ohledem na výchozí data může jít o kolísání denní, měsíční, roční, několikaleté, anebo i o kolísání, které nemusí odpovídat kalendářním obdobím. Podmínkou však je, aby toto kolísání bylo pravidelné. Sezónní složka ( $S_t$ ) tedy vyjadřuje pravidelné kolísání okolo trendu v rámci kalendářního roku. Sezónní výkyvy se opakují každoročně ve stejných obdobích (délka periody je jeden rok) a vznikají v důsledku střídání ročních období nebo vlivem různých institucionalizovaných zvyků, jako jsou např. svátky, dovolené a podobně. Při klasických metodách analýzy časových řad může dojít k závěrům, které neodpovídají realitě. Vzhledem k tomu, že jsou trend a cykličnost zkoumány samostatně, avšak ve skutečnosti působí současně, mohou se chyby projevit ve zkreslení výsledků.[2][11]

Sezónní očišťování časové řady:

Jednou z běžných operací časových řad je takzvané sezónní očišťování časových řad. Sezónním očišťováním se rozumí odstranění sezónní složky, takže časová řada potom obsahuje pouze trendovou, cyklickou a reziduální složku. Důvody pro sezónní očišťování jsou následující:

- a) možnost konstrukce předpovědí krátkodobých časových řad
- b) použití očištěných časových řad při konstrukci ekonometrických modelů
- c) dosažení srovnatelnosti hodnot za jednotlivé měsíce případně čtvrtletí.

Sezónně očištěné řady se obecně používají proto, že umožňují snadno srovnávat hodnoty v měsících či ve čtvrtletích, neboť nejsou zatíženy specifickými podmínkami těchto časových úseků. Máme-li například časovou řadu spotřeby elektrické energie v jednotlivých měsících, je patrně velmi obtížné srovnávat hodnoty neočištěné řady např. v červenci a v lednu vzhledem k tomu, že spotřeba elektrické energie je v zimních měsících vždy větší než v měsících letních.[3]

K sezónnímu očištění se obecně používají takzvané centrované klouzavé průměry. Při vhodně zvolené délce klouzavých průměrů dojde k odstranění sezónní složky z vyrovnané řady.

Vyhazení časové řady pomocí klouzavých průměrů:

Jedná se o jeden z postupů dekompoziční metody, který pomocí průměrů dat vyrovná křivku časové řady. Metoda klouzavých průměrů patří k takzvaným adaptivním metodám. Vyhlazenou hodnotu časové řady v okamžiku  $t$  získáme určitým zprůměrováním původních pozorování z jistého okolí pozorování  $Y_t$ . Můžeme si představit, že podél dané časové řady klouže okénko, v jehož rámci se průměruje. Pak se očištění časové řady řídí následujícími definicemi.

- a) Délka klouzavého okénka je liché číslo

Nechť toto okénko zahrnuje  $d$  členů nalevo od bodu  $t$  a  $d$  členů napravo od bodu  $t$ . Hovoříme pak o vyhlazovacím okénku šířky  $h = 2d + 1$  a jedná se o prostý (jednoduchý) klouzavý průměr. Prvních a posledních  $d$  hodnot časové řady nevyhlazujeme, protože pro  $t \in \{1, \dots, d\} \cup \{n-d+1, \dots, n\}$  není vyhlazovací okénko symetrické. Vyhlazená hodnota ve středu vyhlazovacího okénka je dána vztahem

$$\hat{Y}_t = \frac{1}{2d+1} (Y_{t-d} + Y_{t-d+1} + \dots + Y_{t+d}); t = d+1, \dots, n-d$$

b) Délka klouzavého okénka je sudé číslo

V případě, že šířka vyhlazovacího okénka je sudé číslo, to znamená  $h = 2d$ , klouzavý průměr se počítá jako průměr dvou sousedních jednoduchých klouzavých průměrů. Říká se mu centrováný klouzavý průměr a je získáván ze vztahu:

$$\hat{Y}_t = \frac{1}{4d} \left( \sum_{i=0}^{2d-1} Y_{t+i-d} + \sum_{i=1}^{2d} Y_{t+i-d} \right); t = d + 1, \dots, n - d$$

[7]

V další fázi očišťování časové řady se pracuje s hodnotami centrováných klouzavých průměrů, kterými podělíme původní hodnoty časové řady tedy  $Y_t/CMA$ . Podíl pak vyjadřuje dvě složky časové řady, a to sezónní a nesystematickou ( $S_t * I_t$ ). Samotnou sezónní složku vyjádříme zprůměrováním těchto hodnot v jednotlivých stejných obdobích. V případě měsíčních časových řad získáme dvanáct průměrů, které jsou číselným vyjádřením sezónních indexů. Pokud chceme původní časovou řadu očistit od sezónních vlivů, podělíme ji vypočítanými sezónními indexy.

#### **Nesystematická (reziduální) složka**

Tato složka vyjadřuje nahodilé a jiné nesystematické výkyvy, ale také chyby měření. Reziduální složka zbývá v časové řadě po odstranění trendu a sezónní i cyklické složky. Je tvořena náhodnými fluktuacemi v průběhu časové řady, které nemají rozpoznatelný systematický charakter. Proto se také nepočítá mezi systematické složky časové řady. Reziduální složka pokrývá také chyby v měření údajů časové řady a některé chyby (například chyby v zaokrouhlování), kterých se dopouštíme při jejím zpracování. Při zkoumání dlouhodobé vývojové tendence ukazatele časové řady neboli trendu v časové řadě, je nutné „očistit“ zadané údaje od ostatních vlivů, které tuto vývojovou tendenci zastírají. Postup, kterým se toho dosahuje, se nazývá vyrovnávání časových řad.[2][16]

### **2.3.2. Brownovo exponenciální vyrovnání**

Modely jednoduchého exponenciálního vyrovnání, Brownova lineárního exponenciálního vyrovnání a kvadratického exponenciálního vyrovnání jsou založeny na odhadu trendu podobně jako model průměru, lineárního trendu a kvadratického trendu. Avšak modely exponenciálního vyrovnání odhadují trend tím, že nedávna pozorování vnímají jako více důležitá než pozorování starší.[35]

Při použití této metody mají starší data při výpočtu menší váhu než data čerstvější. Váha údajů je tedy při výpočtu nepřímo úměrná jejich „věku“.



Pro generování předpovědí můžeme využít vyrovnání prvního, druhého až třetího stupně. Postupy pro vyrovnání časové řady jsou vyjádřeny vzorci:

$$S1_t = \alpha Y_t + (1-\alpha)S1_{t-1},$$

kde  $S1_t$  je řada hodnot vyrovnaných pomocí jednoduchého Brownova exponenciálního vyrovnání.

$$S2_t = \alpha S1_t + (1-\alpha)S2_{t-1},$$

kde  $S2_t$  je řada exponenciálních průměrů druhého stupně.

$$S3_t = S2_t + (1-\alpha)S3_{t-1},$$

kde  $S3_t$  je řada exponenciálních průměrů třetího stupně.

[11]

### Postup při exponenciálním vyrovnání druhého stupně

Brownovo dvojitě exponenciální vyrovnání s vyrovnávací konstantou  $\alpha$  se používá pro předpovídání časové řady  $Y_t$  bez sezónní složky.

Při exponenciálním vyrovnání druhého stupně postupujeme následujícím způsobem:

- Stanovení libovolné velikosti  $\alpha$  v intervalu (0;1).
- První fáze obsahuje výpočet první vyrovnávací statistiky  $S1_t$  podle vztahu:

$$S1_t = (1-\alpha) * y_t + \alpha * S1_{t-1}$$

- Druhá fáze obsahuje výpočet vyrovnávací statistiky druhého stupně  $S2_t$

$$S2_t = (1-\alpha) * S1_t + \alpha * S2_{t-1}$$

- Pro vyrovnanou časovou řadu pak platí vztah:

$$\hat{y}_t = \left(2 + \frac{1-\alpha}{\alpha}\right) * S1_t - \left(1 + \frac{1-\alpha}{\alpha}\right) * S2_t,$$

kdy hodnoty  $S1$  a  $S2$  v čase  $t=1$  aproximujeme hodnotou  $y_1$ .

- Následuje vypočtení míry odlišnosti  $E$  (nejlépe  $E2$ ) vyrovnaných hodnot od skutečných a výpočet průměru z hodnot (tedy MSE).
- Provedení optimalizace  $\alpha$  pomocí Řešitele tak, že je minimalizována hodnota MSE.
- Následně je potřeba provést výpočet odhadnuté konstanty pomocí vyrovnávacích statistik dle vzorců:

$$\beta_0(t) = 2 * S1_t - S2_t \text{ a } \beta_1(t) = \frac{1-\alpha}{\alpha} * (S1_t - S2_t)$$

- A nakonec je nutné provést bodovou predikci pomocí odhadnutých konstant dle následujícího vztahu:

$$\hat{y}_t = \beta_0(n) + \beta_1(n)(t - n + 1), \text{ kde } n \text{ je počet naměřených hodnot.}$$

[13]

### 2.3.3. Box-Jenkinsova metodologie (ARIMA model)

Autoregressive/Integrated/Moving Average neboli ARIMA modely byly zpopularizovány díky George Box a Gwilym Jenkins v roce 1970. Pro popsání této metody jsou v následujících odstavcích použity především citace dle MAKRIDAKIS, Spyros, WHEELWRIGHT, C., Steven, HYNDMAN, J., Rob z knihy *Forecasting: Methods and Applications* (1998).

„Box-Jenkinsova metodologie dekompozice časové řady bere za základní prvek konstrukce modelu časové řady reziduální (náhodnou) složku, která může být tvořena korelovanými (závislými) náhodnými veličinami. Těžiště spočívá právě ve vyšetřování těchto závislostí. Box-Jenkinsovy modely jsou daleko flexibilnější než klasické dekompoziční modely (flexibilitou modelu rozumíme schopnost modelu adaptovat se rychle na změněný charakter časové řady).“[10]

ARIMA(p,d,q) je složeným modelem obsahujícím část autoregresní AR(p), modely klouzavých průměrů MA(q) a diferenci I(d). Je tedy vhodné nejprve vysvětlit jednotlivé součásti modelu.

#### AR (= řád autoregresního procesu)

Autoregresní model řádu p (označení AR(p)) je možné zapsat ve tvaru:

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + e_t.$$

Autoregresní model časové řady je založen na poznatku, že každá hodnota v časové řadě je v relaci (závislosti) s předchozími hodnotami této řady.[10][18]

#### MA (= řád procesu klouzavých průměrů)

Proces klouzavých průměrů řádu q (označení MA(q)) je možné zapsat ve formě:

$$Y_t = b_0 + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_q e_{t-q} + e_t,$$

kde  $e_{t-1}, e_{t-2}, \dots, e_t$  představují složky bílého šumu.[18]

#### I (= stupeň zapojení první diference)

Proces náhodné procházky I(d) je vyjádřen jako:  $X_t = X_{t-1} + a_t$ . Jedná se o zvláštní případ procesu AR(1), kdy  $b_1 = 1$ . Za předpokladu počátku v čase  $t = 0$ , lze proces náhodné procházky přepsat do formy  $X_t = X_0 + a_t + a_{t-1} + a_{t-2} + \dots$ , kde  $X_0$  je počáteční podmínka. Je tedy zřejmé, že proces náhodné procházky je tvořen kumulováním náhodných veličin tvořících proces bílého šumu. Protože jeho první diference je proces bílého šumu, nazývá se integrovaný proces řádu jedna a označuje I(1).[5]

Stanovení hodnoty parametru  $d$  je možné třemi přístupy:

- Vizuální posouzení stacionarity výchozí řady  $\{y_t\}$  a diferencovaných řad  $\Delta^d \{y_t\}$   $y_t, d = 1, 2, \dots$
- Studium odhadů autokorelační funkce  $r_k$  pro řady  $\{y_t\}$  a  $\Delta^d \{y_t\}$   $y_t, d = 1, 2, \dots$ . Jestliže hodnoty  $r_k$  klesají pomalu (přibližně lineárně), pak je nutno provést další diferenciaci.
- Studium odhadů rozptylu pro řady  $y_t$  a  $\Delta^d \{y_t\}$   $y_t, d = 1, 2, \dots$ . Za optimální se volí taková hodnota parametru  $d$ , která poskytuje nejmenší odhad rozptylu.

[16]

### **Integrovaný smíšený model ARIMA**

Složitější model ARIMA( $p, d, q$ ) obsahuje autoregresní složku AR( $p$ ), složku klouzavých průměrů MA( $q$ ) a náhodnou složku I( $d$ ).

Tyto modely umožňují popis procesů, u kterých nejenže dochází ke změnám úrovně, ale tyto změny mohou mít nesystematický, náhodný charakter (což je pro časové řady z praxe běžné). ARIMA modeluje stochasticky vedle náhodných fluktuací i trendovou složku. Smíšený integrovaný model (model ARIMA) je určen pro popis časových řad s náhodnými změnami trendu (úrovně a sklonu). Výchozí časová řada nemusí být stacionární, je však nutné, aby byla převoditelná na stacionární. Tento převod se uskutečňuje diferencováním výchozí časové řady.[10][16]

Výše zmíněný model je například v programu Statgraphics uváděn ve formě ARIMA( $p, d, q$ )x( $P, D, Q$ )s model. Tento model je složen z několika termínů, které v první řadě udávají jakého řadu je termín ARIMA modelu:

Tabulka 1 - Význam termínů ARIMA modelu

označení	Význam
p	Nesezónní autoregrese
d	Nesezónní diference
q	Nesezónní klouzavý průměr
P	Sezónní autoregrese
D	Sezónní diference
Q	Sezónní klouzavý průměr
s	Počet opakujících se sezón v jedné periodě

(zdroj: program Statgraphics)

[34]

Konstrukce smíšeného integrovaného modelu se realizuje ve dvou krocích:

- a) Nejprve se výchozí nestacionární časová řada  $\{y_t\}$  převede diferencováním, respektive vhodnou transformací, na stacionární řadu  $\{w_t\}$ . Přitom je třeba si uvědomit, že diferencováním se výchozí časová řada zkracuje.
- b) Na stacionární časovou řadu se aplikuje smíšený model ARMA(p,q). Modely ARIMA zřejmě představují “rozumný” kompromis při zeslabování předpokladu o stacionaritě ARMA modelů.

[16]

### Identifikace modelu

Základním úkolem identifikace je výběr typu modelu (AR, MA, ARMA, ARIMA) a určení řádu modelu. Při identifikaci modelu se doporučuje dodržovat následující postup:

Nejprve se pořídí grafický záznam časové řady a alespoň vizuálně se ověří, zda je řada stacionární.

V případě, že sledovaná řada není stacionární, je nutno ji stacionarizovat s využitím diferenciací. Pokud je střední hodnota stacionárního procesu (časové řady) nenulová, je třeba řadu vycentrovat. Princip centrování je jednoduchý – stačí nahradit řadu  $\{y_t\}$  řadou  $\{y_t - \bar{y}\}$ , kde  $\bar{y}$  představuje aritmetický průměr všech pozorování. Jestliže existují pochybnosti o nenulovosti střední hodnoty, doporučuje se použít běžných statistických testů (např. srovnání aritmetického průměru  $y$  s dvojnásobkem směrodatné odchylky tohoto průměru).

Vlastní identifikace je založena na analýze odhadů autokorelační funkce  $r_k$  a parciální autokorelační funkce  $r_{kk}$ . Ve většině případů stačí spočítat prvních 20 hodnot obou funkcí. Cílem vlastní identifikace je určit identifikační bod  $k_0$  těchto funkcí, pokud existuje, s využitím Bartlettovy, resp. Quenouilleovy, aproximace.

- a) Bartlettova aproximace. Je-li  $\rho_k = 0$  pro  $k > k_0$ , pak platí

$$\sigma(r_k) \approx \sqrt{\frac{1}{n} (1 + 2 \sum_{j=1}^{k_0} r_j^2)}, \text{ přičemž } n \text{ je délka analyzované časové řady.}$$

- b) Quenouilleova aproximace. Je-li  $\rho_{kk} = 0$  pro  $k > k_0$ , potom platí  $\sigma(r_{kk}) \approx \sqrt{\frac{1}{n}}$ .

Odhady  $r_k$  i  $r_{kk}$  mohou být navzájem silně korelované, takže se doporučuje netrvat na jednoznačném určení řádu modelu, ale vyzkoušet více modelů.

Na závěr identifikačních procesů je vhodné spočítat počáteční odhady parametrů. Tyto odhady jsou zpravidla velmi hrubé, využívají se jako výchozí hodnoty při odhadování parametrů modelu.[16]

### **Výhody a nevýhody Box-Jenkinsova přístupu**

Box-Jenkinsova metodologie má ve srovnání s metodami založenými na dekompozici následující výhody:

- a) Modely se rychle adaptují na změny v průběhu časové řady, proto je Box-Jenkinsova metodologie úspěšná i v těch případech, kde dekompozice selhává.
- b) Box-Jenkinsův přístup vykazuje nejlepší výsledky při analýze ekonomických časových řad.
- c) Box-Jenkinsův přístup je systematický, proto může být plně automatizován.

Nevýhody Box-Jenkinsova přístupu mohou být spatřovány v tom, že:

- a) je vhodný jen pro časové řady o délce nejméně 50 pozorování,
- b) jeho praktické aplikace jsou mnohem náročnější než aplikace dekompozičních metod,
- c) výsledné modely, zejména modely s větším počtem parametrů, se obtížně interpretují.

[16]

### **2.3.4. Statgraphics Centurion XVI**

V následující kapitole bude představen statistický program Statgraphics, který je v praktické části použit pro práci s daty a následnou analýzu časové řady. Informace k odstavci jsou čerpány z internetových stránek od softwarového vývojáře a hlavního architekta programu Statgraphics (Dr. Neil W. Plochému), z oficiálních stránek produktu (Dostupné z: <http://www.statgraphics.com>), knižních zdrojů (Altová, Matušů, Kozák; 2004 a Cuřínová, Valentová; 2012) a z vlastních zkušeností a poznatků při práci s programem.

Veškeré výstupy z programu Statgraphics (výsledky, tabulky a grafy), použité v praktické části, jsou získány pomocí 30denní zkušební verze programu, která je dostupná ke stažení z oficiálních stránek programu Statgraphics.

#### **Obecný popis**

Statgraphics je statistický program od společnosti Saponit, Inc., který je velmi oblíbený na vysokých školách. Produkt Statgraphics je jedním ze starších, avšak existuje v nových rozšířených verzích, které obsahují vše potřebné pro analýzu časových řad. Následující odstavce se věnují významnému upgradu programu, verzi nazývané se Statgraphics Centurion XVI.

„Statgraphics Centurion byl vytvořen především pro lidi, kteří chtějí provádět seriózní analýzu dat bez zdlouhavého učení, jak používat statistický balíček. Balíček Statgraphics Centurion XVI je aplikace vytvořená pro Windows, která obsahuje přes 170 statistických procedur (obsahujících vše od souhrnných statistik po návrhy experimentů). Přesto uživatel nemusí být statistik, aby mohl program používat. Program je velice intuitivní a uživatele ke všemu zcela navede orientační menu, a navíc Statgraphics obsahuje pomocné nástroje jako je StatWizard a StatAdvisor (okno statistického rádce), které uživatelům pomohou program používat efektivně.“[30]

Uživatelé mohou data zadávat přímo do programu, importovat je z textových či tabulkových aplikací, jako je například Microsoft Excel nebo extrahovat z databází pomocí ODBC.

### **Obsah**

Statistické procedury ve Statgraphics jsou rozděleny do několika hlavních sekcí:

*Plot* – Grafy zobrazující základní data (graf časových řad, 3D, bodové, maticové, vícerozměrné grafy).

*Describe* – Popisná statistika včetně intervalů spolehlivosti, testů hypotéz, vyhlazování časových řad a sezónní dekompozice.

*Compare* – Srovnávání více než jednoho vzorku.

*Relate* – Regresní a klasifikační metody.

*Forecast* – Analýza časových řad a předpovědi včetně automatické volby modelu. Exponenciální vyrovnání, metoda klouzavých průměrů, metoda náhodné procházky, lineární a nelineární trendy, ARIMA modely. „Metoda Forecast umožňuje provádět předpovědi budoucího vývoje časových řad za použití celé řady modelů, včetně ověřování jejich vhodnosti pomocí testů.“[12].

*SPC* – Metody pro statistické řízení procesů.

*DOE* – Design experimentů

### **Automatizace statistických analýz**

Možnost automatizace je vhodná pro analýzu počtu nezaměstnaných, jelikož data každý měsíc narůstají a je tak možno analýzu provádět opakovaně bez zdlouhavého nastavování parametrů. „Provádíme-li statistickou analýzu dat, často se v praxi stává, že je potřeba stejnou analýzu několikrát zopakovat pro různá data. V tomto případě můžeme využít nabízenou automatizaci statistických analýz. Podstatou této automatizace je StatFolio obsahující všechna vytvořená okna analýz. Po požadované datové změně (ve vstupním panelu) dojde k přepočítání výsledků a překreslení grafů. Pokud načteme jiný datový soubor, opět lze celou analýzu provést pro aktuální data pouhým vyvoláním příslušných oken analýzy (před zobrazením musí být zmenšena do ikony).“[12].

## Analýza časových řad

Jelikož je práce s programem velice intuitivní a uživatelské rozhraní se podobá jiným programům vytvořeným pro Windows, nejsou zde uvedeny postupy pro základní ovládání. Statgraphics nabízí více možností práce se statistickými daty, avšak následující řádky jsou zaměřeny na analýzu časových řad. Je v nich popsáno, co vše Statgraphics pro tuto analýzu nabízí.

„Metody analýzy časových řad jsou dostupné v nabídkách Describe a Forecast. Nabídku Describe, položku Time Series volíme, jestliže chceme provést vyrovnání časových řad ukazatelů (Smoothing) či očistit časovou řadu od sezónní složky (Seasonal Decomposition). Položka Forecast je využita při popisu trendové složky časové řady a předpovídání budoucího vývoje analyzovaného ukazatele.“[12].

Pro analýzu časových řad a prognostické metody nabízí Statgraphics tyto speciální statistické techniky: [31]

V první řadě jsou to **Run Charts**, což je grafické zobrazení dat v jednom číselném sloupci. Předpokládá se, že data jsou přirozeně sekvenční a skládají se buď z jednotlivců (v každém časovém období bylo provedeno jedno měření) nebo podskupin (v každém časovém období byla provedena skupina měření). Tuto možnost nalezneme v hlavní nabídce cestou Plot/Time Sequence Plots/Run Charts. Metoda Run Charts se používá pro zjištění trendu nebo jiných nenáhodných vzorů vyskytujících se v časové řadě. Přehledné schéma zobrazuje Tabulka 2.

Tabulka 2 - Zjišťování nenáhodných vzorů v časové řadě pomocí metody Run Charts

	Hodnoty nad nebo pod mediánem	Body zvratu (růst nebo pokles)
Směšování	příliš	-
Shlukování	málo	-
Oscilace	-	příliš
Trend	-	málo

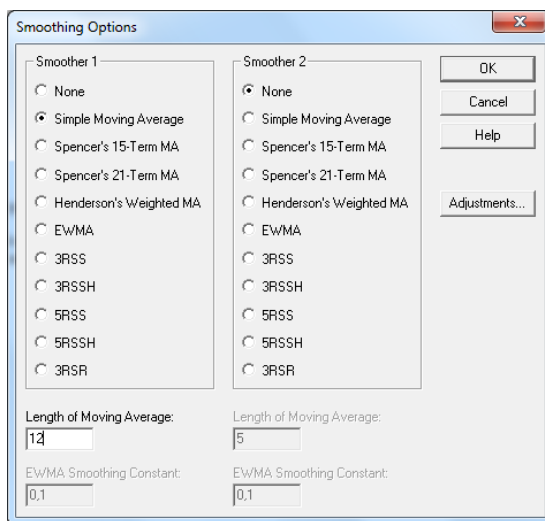
(zdroj: vlastní tvorba)

Další metoda pro použití při analýze časových řad je **Descriptive methods**, kterou lze v programu nalézt cestou Describe/Time Series/Descriptive methods. Deskriptivní metody analyzují časové řady pomocí základních charakteristik (střední hodnota, odchylka, korelace).

Třetí uvedenou metodou vhodnou pro analýzu časových řad je **Smoothing**. Pokud časové řady obsahují velké množství vychýlených hodnot (tzv. hluků), může být obtížné představit si linii trendu. Různé lineární a nelineární metody vyhlazení mohou být použity pro oddělení trendu od hluků. Nejznámější metoda je jednoduchá metoda klouzavých průměrů (simple moving average). Statgraphics umožňuje provést dvojí vyrovnání.

K volbě, kterou zobrazuje Obrázek 4, vede cesta Describe/Time Series/Smoothing po vyplnění dat do úvodní tabulky.

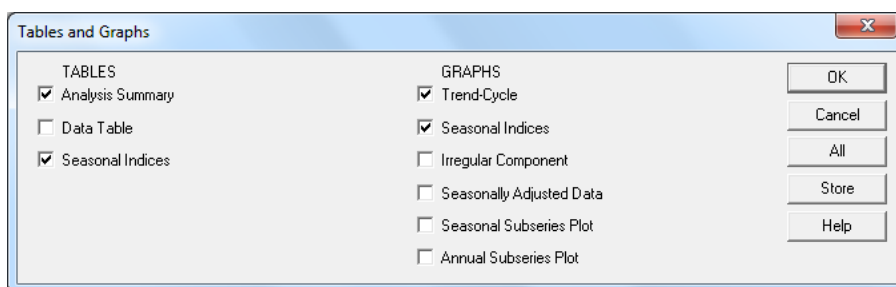
Obrázek 4 - Rozhraní při volbě metody vyrovnaní časové řady



(zdroj: Statgraphics)

**Seasonal Decomposition** se používá, pokud data obsahují silný sezónní efekt a je v tom případě užitečné oddělit sezónnost od ostatních složek časové řady. Tato metoda napomáhá odhalit sezónní vlivy a vygeneruje sezónně očištěná data. Navíc lze vše v možnostech znázornit graficky. V nabídce Tables and Graphs programu Statgraphics (viz obrázek) je možnost zvolit, jaké tabulky a grafy (volby uvádí Obrázek 5) budou zobrazeny ve výstupu. Sezónní dekompozici provedeme opět pomocí hlavního menu cestou Describe/Time Series/Seasonal Decomposition.

Obrázek 5 - Nabídka Tables and Graphs pro volbu výstupů



(zdroj: Statgraphics)

„Při volbě položky Forecast Plot z nabídky Graphs se zobrazí graf předpovídaných hodnot na stanovený počet období do budoucna, kdy je opět možné nastavit spolehlivost odhadu v % pomocí položky Pane Options, která umožní zobrazit doplňkový panel Forecast Limit Options.“[12].

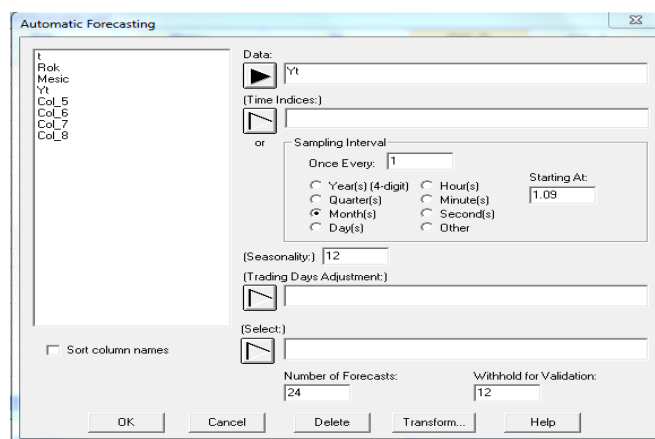


Pro předpovědi časových řad nabízí Statgraphics dvě základní metody dostupné z hlavního menu cestou Forecast/User-Specified Model nebo Forecast/Automatic Model Selection. Jejich cílem je extrapolace minulého chování do budoucnosti. Předpovědi ve Statgraphics zahrnují náhodné procházky, klouzavé průměry, modely trendů, jednoduché, lineární, kvadratické a sezónní exponenciální vyrovnání a ARIMA modely. Uživatelé mohou srovnávat různé modely pomocí vzorků z konce časových řad odebraných pro účely ověření daného modelu (=“withholding samples“). „V poli Withhold for Validation je uveden počet pozorování na konci časové řady, která nebudou při konstrukci předpovědi brána v úvahu, budou tedy považována za neznámá a předpovědi budou konstruovány pouze na základě předchozích údajů. Tato položka se využívá v případech, kdy chceme posoudit, do jaké míry je vybraná trendová funkce vhodná pro předvídaní budoucího vývoje časové řady.“[12]

První z nabídky předpovědí v časových řadách je **Forecasting (User Specified Model)**, který je dán úvodní volbou uživatele. Na rozdíl od něj druhá nabídka pod názvem **Automatic Forecasting** zkouší různé typy modelů a automaticky vybere nejvhodnější model. Pokud je to žádoucí, mohou uživatelé zvolit model prognózy pomocí automatického výběru. V této nabídce Statgraphics porovnává více modelů, přičemž maximalizuje specifikovaná kritéria.

U všech skupin metod je třeba začít vstupním panelem charakterizujícím druh popisovaných dat. Zvolená časová řada se vkládá do políčka Data. Dále je třeba specifikovat časové souřadnice analyzované časové řady. Zda jde o roční data (Year(s)), čtvrtletní data (Quarter(s)), měsíční data (Month(s)), denní data (Day(s)), data za minuty (Minute(s)), data za vteřiny (Second(s)) či data v jiném průřezu (Others). Do políčka Starting At: se uvádí pořadové číslo prvního z popisovaných pozorování, do políčka (Seasonality) se uvádí délka sezóny v případě, že hodnoty časové řady jsou sezónní (například číslice „12“ značí měsíční pozorování).[4]

Obrázek 6 - Vstupní panel při volbě Automatic Forecasting



(zdroj: statistický program Statgraphics)

### **2.3.5. IBM SPSS Statistics**

Následující kapitola popisuje program IBM SPSS Statistics ve verzi 20 a přídavný modul SPSS Forecasting. Informace vychází ze zdrojů dostupných na oficiálních stránkách firmy IBM, z knižních zdrojů (YAFFEE, MCGEE, 2000, Borůvková, Horáčková, Hanáček, 2014) a z vlastní tvorby v programu IBM SPSS Statistics.

#### **Obecný popis**

SPSS (=Statistical Package for the Social Sciences) je americká softwarová firma sídlící v Chicagu, od roku 2009 vlastněná společností IBM. Software je dodáván od roku 1968 a patří tak k jednomu z nejstarších a nejrozšířenějších paketů.

IBM SPSS Statistics je dodáván v několika plně integrovaných modulech, které nabízí speciální funkcionality. Základní modul SPSS Statistics Base je dostupný ke stažení z oficiálních stránek společnosti IBM a nabízí základní funkce standardního statistického programu, jako je import dat z datové tabulky připravené v MS Excel nebo v Access, manipulace s daty, deskriptivní statistiky, tvorbu grafů, ukládání a export dat. SPSS Statistics umožňuje rychlý a snadný přístup ke všem typům souboru dat, jejich rychlou organizaci a analýzu.

SPSS Statistics Base a ostatní produkty SPSS Statistics mohou být používány samostatně nebo jako součást jedné ze tří speciálních edic: IBM SPSS Statistics Standard/Professional/Premium. Software je uživatelsky velice přívětivý, většina ikon je dobře známá z programu MS Excel – mají podobný vzhled i funkce.

#### **Modul IBM SPSS Forecasting**

Modul, který vylepšuje metodu prognózy širší nabídkou statistických procesů s kompletní analýzou časových řad, včetně několika modelů stanovení vyhovující křivky trendu, modely vyrovnání časových řad, metody pro odhadnutí autoregresních funkcí. Procedura Expert Modeler pomáhá automaticky určovat, které ARIMA procesy nebo exponenciální vyrovnávací modely nejlépe vyhovují pozorované časové řadě a nezávislým proměnným. Procedura je založena na eliminaci výběru prostřednictvím pokusů a omylů.

Modul nabízí techniky vytvořené pro analýzu časových řad. Mezi hlavní funkce modulu patří tvorba spolehlivé předpovědi, redukce chyb v předpovědi pomocí automatického výběru vhodného modelu a jeho parametrů a aktualizace nebo organizace předpovědních modelů. Konkrétními procedurami jsou TSMODEL (pomocí Expert Modeler modeluje proměnné časových řad, buď pomocí ARIMA nebo exponenciálního vyrovnání), TSAPPLY (umožňuje aplikovat uložené modely při práci s novými nebo aktualizovanými daty), SEASON (generuje sezónní faktory pro periodické časové řady pomocí multiplikativní nebo aditivní metody), SPECTRA (rozkládá časovou řadu do základních čtyř komponentů).

## Postup zadávání dat a analýza

Tradiční cestou pro import dat je File/Open/Data, kde je potřeba v okénku Files of type změnit typ souboru, pomocí kterého budou data otevřena. Na importovaná data je možné pohlížet dvěma způsoby: vlastní pohled na data umožňuje záložka Data view a bližší informace o proměnných jsou v záložce Variable view, kde je možné proměnné přidávat, mazat nebo měnit jejich pořadí.

Dále program umožňuje provést kontrolu správnosti dat pomocí frekvenčních tabulek a popisné statistiky. Ke kontrole spojité proměnné slouží funkce pro popisnou statistiku, pomocí které lze zkontrolovat minimum, maximum, aritmetický průměr a směrodatnou odchylku. Tyto údaje mají vypovídací hodnotu o správnosti dat a dají se získat cestou Analyze/Descriptive Statistics/Descriptives, kde budou po zobrazení dialogového okna do pole Variable(s) zadány proměnné, které jsou podrobeny kontrole.

Tabulka 3 - Výpočet popisné statistiky pro Počet uchazečů o zaměstnání

### Descriptive Statistics

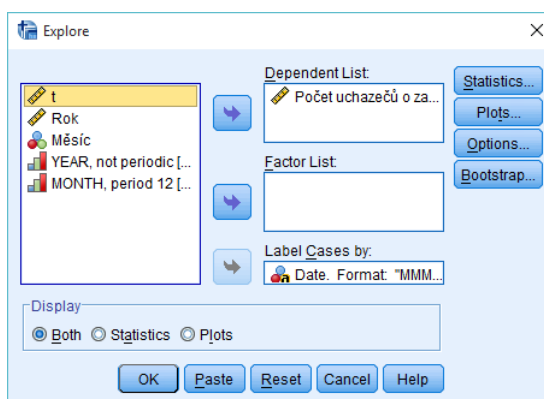
	N	Min	Max	Mean	Std. Deviation
Počet uchazečů o zaměstnání	77	4358	7460	6271,38	623,287
Valid N (listwise)	77				

(zdroj: výstup z programu SPSS Statistics)

Veškeré výstupy analýz jsou shromážděny v samostatném okně Output s názvem dokumentu. Grafy a tabulky jsou zaznamenávány postupně podle času zadání procedury a obsahují vždy popis a vložené proměnné. K jedné z dalších metod využitelných pro popisnou statistiku se lze dostat příkazem Analyze/Descriptive/Statistics/Explore, kde po výběru proměnné, která má být analyzována a proměnné, která identifikuje jednotlivá pozorování, program provede analýzu popisné statistiky. Zadání dat ve vstupním panelu metody Explore znázorňuje

Obrázek 7

Obrázek 7 - Vstupní panel k metodě Explore (popisná statistika)



*(zdroj: výstup programu SPSS Statistics)*

## **Grafická prezentace dat**

Program SPSS Statistics nabízí několik možností zobrazení dat, například výsečové grafy, sloupcové grafy nebo kategorizované grafy. Pro spojitou proměnnou, která nabývá mnoha hodnotami, je vhodným typem grafu histogram, krabicový graf nebo spojnicový graf. Všechny typy grafů jsou k nalezení na hlavním panelu pod příkazem Graphs/Legacy Dialogues/...

### a) Histogram

V hlavní nabídce programu je uveden pod názvem Histogram. Proměnná se vepíše do položky Variable. Zaškrtneme-li Display normal curve, je možné graf proložit Gaussovou křivkou normálního rozdělení. Díky tomu můžeme získat lepší grafický přehled o možnosti normálního rozdělení naměřených dat. Pro změnu počtu intervalů se v programu musí dvakrát kliknout na graf, poté opět dvojklikem na libovolný sloupec histogramu se zobrazí okno vlastností histogramu, kde lze na kartě Binning upravit šířku intervalů.

### b) Krabicový graf

V hlavní nabídce je k nalezení pod názvem Boxplot. Krabicové grafy jsou v programu SPSS tvořené mediánem (čára uvnitř krabice), kvartily (dolní a horní víko krabice), nejmenší a největší hodnotou a odlehlými hodnotami (kolečka). Program Microsoft Excel tento typ grafu nenabízí, je proto pro jeho tvorbu výhodné používat specializované statistické programy.

### c) Spojnicový graf

Vykreslení toho grafu je iniciováno v nabídce funkcí pod názvem Line. Spojnicový graf se nejčastěji používá k prezentování časových řad. Chceme-li zobrazit jednu časovou řadu, používá se označení Simple. Naopak pro více časových řad v jednom grafu se využívá označení Multiple. Dále je nutné označit možnost Values of individual cases a v následujícím okně vložit proměnnou do pole Lines Represent. V nabídce Category Labels je možno změnit popis vodorovné osy.

### **3. Praktická část**

V následujících kapitolách je v první řadě představena konkrétní časová řada a její podstatné specifikace, které slouží k hlubšímu poznání dat. Specifikace časové řady je zaměřena na geografický průzkum oblasti související s časovou řadou. V další fázi se věnuje představení metodiky ukazatele a určení typu časové řady.

Zvolená časová řada je následně analyticky rozebrána podle teorie analýzy časových řad. Důkladné provedení analýzy je podstatnou součástí předpovědi v časových řadách. Analýza časové řady sestává ze dvou částí. První část se zabývá grafickou analýzou průběhu časové řady, tedy vizuální stránkou. Druhá část se zaměřuje na výpočty elementárních charakteristik časové řady, jejichž postup je popsán v teoretické části práce.

Další samostatná kapitola praktické části je věnována předpovědím pro danou časovou řadu. Nejprve je představena forma předpovědi, kterou doposud používá Úřad práce v České Lípě. V dalších odstavcích jsou zkonstruovány návrhy alternativních metod předpovědi budoucích hodnot časové řady, které by pracovníci úřadu práce mohli v budoucnu využívat.

Poslední část se věnuje volně dostupným statistickým doplňkům pro Microsoft Excel a jejich možnostem při analýze časových řad.

#### **3.1. Popis konkrétní časové řady**

Následující odstavce představují časovou řadu, která je použita v praktické části práce k analýze a výpočtům předpovědi budoucího vývoje. Jsou zde popsána data obsažená v časové řadě, definován typ časové řady, metodika získávání jednotlivých pozorování a další charakteristiky. Hlavním zdrojem informací je roční zpráva kontaktního pracoviště úřadu práce v České Lípě z roku 2014 (viz zdroj [20]).

Zdrojem pro praktickou část diplomové práce je časová řada z oblasti demografické. Při studiu změn populačních ukazatelů se vyskytují různé časové řady (příklady zahrnují porodnost, počet obyvatel nebo nezaměstnanost). Demografové obvykle požadují předpovědi změn v populaci pro 10 až 20 let dopředu.

### **Specifikace okresu:**

Okres Česká Lípa je rozlohou 1073 km<sup>2</sup> největším ze čtyř okresů Libereckého kraje, naopak hustotou osídlení 95,8 obyvateli na km<sup>2</sup> se řadí k nejnižším v kraji. Na území okresu se nachází 57 obcí, z toho 11 měst a 1 městys. Počet obyvatel v celém okrese byl (k datu 31. 12. 2014) 102 816 obyvatel, z toho 69 977 osob ve věku 15-64 let tvoří 68,1 %.

Obrázek 8 - Poloha města Česká Lípa v rámci okresu



(Zdroj: wikipedia.org)

V celorepublikovém srovnání nezaměstnanosti byl okres Česká Lípa k datu 30. 6. 2015 na 34. místě z celkového počtu 77 okresů.[20]. Podíl nezaměstnaných osob v celé ČR v tuto dobu činil 6,2%, což je o 0,3 % více než v okrese Česká Lípa. Českolipský okres je tedy v rámci nezaměstnanosti přibližně na hodnotě celorepublikového průměru.

### **Přehled hlavních zaměstnavatelů:**

Okres Česká Lípa patří mezi okresy s tradičně rozšířeným průmyslem. Mezi nejvýznamnější průmyslové obory okresu patří výroba zaměřená na automobilový průmysl a sklářská výroba.

Sklářská výroba v okrese, mající v některých technologiích mnohaletou tradici a jedinečnost, je soustředěna zejména ve městech Nový Bor a Kamenický Šenov a mezi nejvýznamnější zaměstnavatele v tomto oboru patří:

- EGERMANN, s. r. o.,
- Crystalex CZ, s. r. o.,
- PRECIOSA-LUSTRY, a. s.,
- Lasvit s.r.o.

Firmy zaměřené na výrobu automobilových dílů a doplňků našly své sídlo především v okresním městě Česká Lípa a ve Stráži pod Ralskem. Mezi hlavní zaměstnavatele v automobilovém odvětví patří tyto firmy:

- Johnson controls automobilové součástky, k. s.,
- Fehrer Bohemia, s. r. o.
- Johnson Controls Autobaterie spol. s r. o.,
- Grammer Automotive CZ s.r.o.

K dalším významným firmám v okrese Česká Lípa patří:

- Bombardier Transportation Czech Republic, a. s.,
- DIAMO, státní podnik,
- Festool s.r.o.,
- MODUS, spol. s r. o.,
- International Automotive Components Group s.r.o.,
- MEGA a.s.,
- Nemocnice s poliklinikou Česká Lípa, a. s.,
- Jizerské pekárny, spol. s.r.o.

Mimo jiné, v okrese působí obchodní řetězce OBI, Kaufland, Albert, Penny Market, LIDL, které částečně přispívají k řešení nezaměstnanosti v regionu a jsou významnými zaměstnavateli.

#### **Metodika ukazatele:**

Analyzovaná časová řada obsahuje data, která jsou prezentována podle následující definice metodiky:

*Počet dosažitelných, neumístěných uchazečů o zaměstnání ve věku 15-64 let, občanů ČR a občanů EU (EHP), vedených úřady práce podle bydliště uchazeče ke konci sledovaného měsíce.*

Zkráceně: *Počet uchazečů o zaměstnání.*

Pro výpočty pod označením:  $Y_t$ .

Doplnění k definici ukazatele: „Jedná se o evidované nezaměstnané, kteří nemají žádnou objektivní překážku pro přijetí do zaměstnání a při nabídce vhodného pracovního místa mohou do něj bezprostředně nastoupit (nejsou ve vazbě, ve výkonu trestu, nepobírají peněžitou pomoc v mateřství, hmotné zabezpečení po dobu mateřské dovolené, nejsou v pracovní neschopnosti, nejsou zařazeni do rekvalifikačních kurzů a nevykonávají krátkodobé zaměstnání). Neumístění uchazeči o práci jsou občané bydlící na příslušném území, kteří nejsou v pracovním nebo obdobném vztahu k zaměstnavateli,

ani nevykonávají samostatnou výdělečnou činnost a požádali úřad práce o zprostředkování vhodného zaměstnání.“[24]

Počet uchazečů o zaměstnání je primárním ukazatelem. Veškerá data jsou získána z měsíčních statistik nezaměstnanosti uváděných na integrovaném portálu Ministerstva práce a sociálních věcí (portal.mpsv.cz). Analýza zahrnuje data z období leden/2009 až květen/2015 (celkem 77 pozorování) zahrnující oblast okresu Česká Lípa. Tabulka s konkrétními seskupenými daty je k nahlédnutí v Příloze 1.

### **Změna metodiky:**

Před začátkem práce nastal problém nesrovnatelnosti dat v časové řadě, a to z hlediska změny metodiky ukazatele registrované nezaměstnanosti. „Ministerstvo práce a sociálních věcí počínaje lednem 2013 přechází na nový ukazatel registrované nezaměstnanosti v ČR s názvem Podíl nezaměstnaných osob, který vyjadřuje podíl dosažitelných uchazečů o zaměstnání ve věku 15 – 64 let ze všech obyvatel ve stejném věku. Tento ukazatel nahrazuje doposud zveřejňovanou míru registrované nezaměstnanosti, která poměřuje všechny dosažitelné uchazeče o zaměstnání pouze k ekonomicky aktivním osobám. Nový ukazatel má kvůli odlišné definici jinou úroveň a je tudíž s původním ukazatelem nesrovnatelný.“[29].

Zmíněná změna metodiky uměle snižuje procento nezaměstnaných, a proto jsou hodnoty ukazatele registrované nezaměstnanosti před rokem 2012 nesrovnatelné s hodnotami po roce 2012. Z tohoto důvodu je práce zaměřena na analýzu časové řady počtu uchazečů o zaměstnání, která obsahuje srovnatelné hodnoty a vhodně vyjadřuje měnící se nezaměstnanost v okrese.

### **Typ časové řady:**

Časová řada počtu uchazečů o zaměstnání v okrese Česká Lípa má charakter okamžikového ukazatele (okamžiková časová řada viz odstavec 2.1.1 Typy časových řad). Veškeré hodnoty dat se vztahují k poslednímu dni měsíce, tedy hodnota s názvem „2013/6“ je naměřena dne 30. června 2013.

Jelikož jsou hodnoty zapisovány s měsíční frekvencí, jedná se o časovou řadu krátkodobou.

Počet dní mezi jednotlivými pozorováními se liší, proto mluvíme o neekvidistantní časové řadě (pro analýzu dat je potřeba neekvidistantní časovou řadu upravit, jak je naznačeno v odstavci 2.1.1 Typy časových řad).

Data, ze kterých časová řada vychází, jsou primární. Jedná se tedy o data, která byla přímo naměřena a nebyla dále upravována.



### 3.2. Grafická analýza průběhu časové řady

První částí grafické analýzy časové řady je vytvoření spojnicového grafu, který podává souhrnné informace o průběhu dat v celé časové řadě. V grafu mohou být pozorovány odchylky, nesrovnalosti nebo sezónní výkyvy. Pakliže je v grafu patrná sezónní složka, je v dalším kroku vhodné provést sezónní dekompozici časové řady. Metoda sezónní dekompozice očistí časovou řadu od sezónních vlivů a číselně vyjádří sezónní indexy v jednotlivých měsících. Z průběhu očištěné časové řady je pak zřetelný trend.

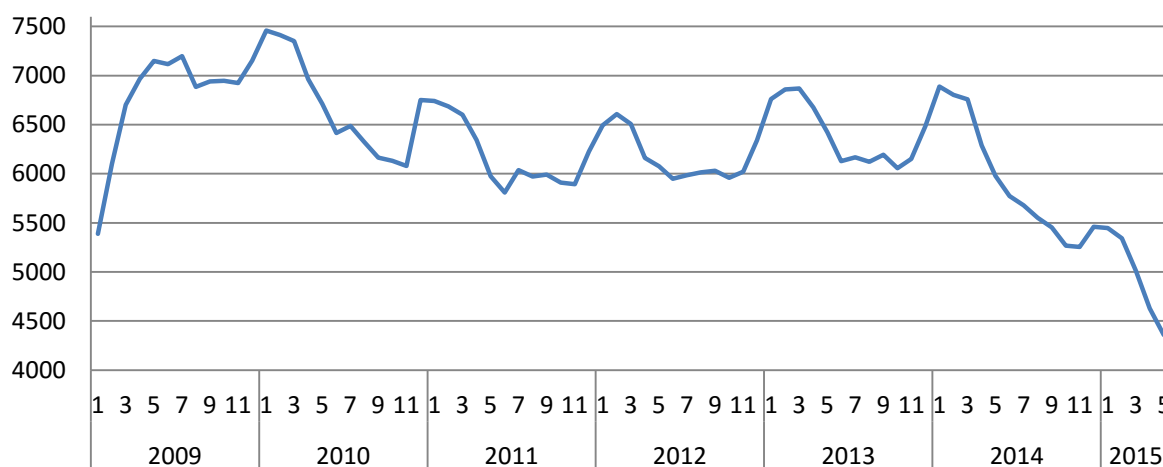
Pro grafickou analýzu průběhu časové řady je použit tabulkový procesor Excel, který nabízí základní potřebné typy grafického zobrazení. Dále byly pro srovnání použity statistické software Statgraphics a IBM SPSS, které nabízí více možností grafické analýzy časové řady.

#### Zobrazení vývoje časové řady:

Vývoj počtu uchazečů o zaměstnání v čase nejlépe znázorňuje spojnicový graf časové řady, jehož tvar ukazuje Obrázek 9. Graf je vyobrazen pomocí funkce programu MS Excel. Pro srovnání Obrázek 23 zobrazuje spojnicový graf stejné časové řady vyobrazený pomocí funkce ve specializovaném statistickém programu IBM SPSS Statistics.

Pomocí vizuálního rozboru průběhu časové řady je možno rozpoznat periodicky se opakující výkyvy a odhadnout trend. V grafu je také možné pozorovat zvyšující se počet uchazečů o zaměstnání v zimních měsících. Tyto výkyvy jsou vyjadřovány sezónní složkou. V konkrétním případě časové řady počtu uchazečů o zaměstnání jde o vliv sezónní nezaměstnanosti, která se projevuje sníženou zaměstnaností především v zimních měsících, kdy jsou omezené počty pracovních míst (především v oborech stavebnictví a zemědělství) Z grafu je také patrný mírný klesající trend za posledních 5 let.

Obrázek 9 - Spojnicový graf časové řady počtu uchazečů o zaměstnání

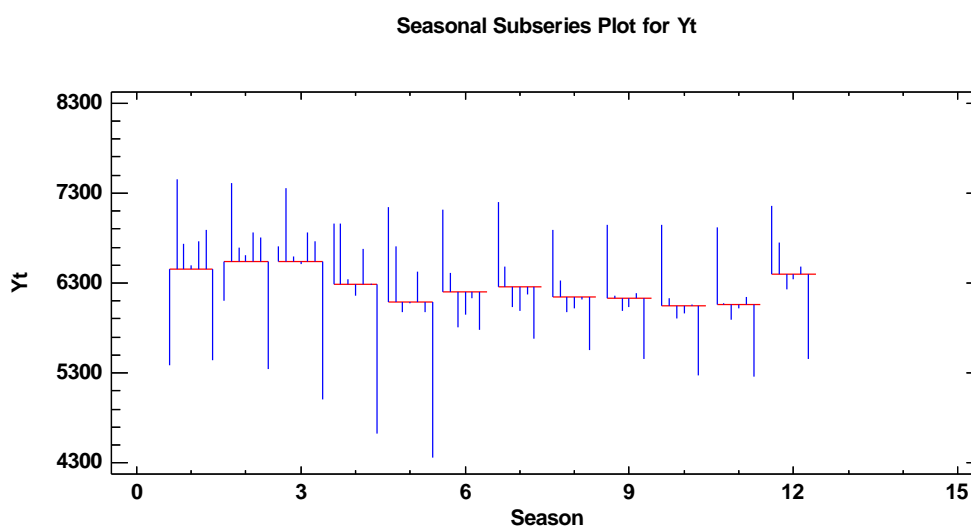


(zdroj: vlastní tvorba)

### Graf časové řady podle jednotlivých sezón:

Graf, který zobrazuje Obrázek 10, se používá při analýze sezónních časových řad. „Zobrazuje hodnoty časové řady uspořádané podle jednotlivých sezón. Vodorovné čáry představují průměrnou úroveň hodnot v sezónách za všechny zkoumané roky. Svislé čáry znázorňují odchylky skutečných hodnot od průměru pro každou sezónu.“[2]. V grafu jsou vidět značné odchylky od sezónních průměrů vždy v posledním pozorovaném roce (2015), což znázorňují modré svislé čáry umístěné úplně vpravo na červených vodorovných čarách. Tvar grafu sezónních hodnot udává přesnější podobu sezónních indexů. Například pro dvanáctou sezónu (prosinec) jsou hodnoty v prvních dvou letech vyšší než průměrná úroveň hodnot v dané sezóně a naopak hodnota v posledním roce je hluboko pod průměrem.

Obrázek 10 - Graf sezónních hodnot měsíční časové řady počtu uchazečů o zaměstnání



(zdroj: výstup z programu Statgraphics)

### Sezónní dekompozice časové řady:

Metoda klasické sezónní dekompozice časové řady je popsána v odstavci 2.3.1 Klasický model dekompozice.

Pro krátkodobou časovou řadu jsou typické sezónní výkyvy v datech. Tabulka 4 ukazuje přehled sezónních indexů pro každý měsíc v roce a porovnání jejich hodnot vypočtených programy Statgraphics a IBM SPSS Statistics a vlastního výpočtu v programu MS Excel. Indexy jsou získány pomocí multiplikativní metody sezónní dekompozice. Rozptyl indexů se pohybuje od nejnižšího v 6. měsíci po nejvyšší ve 2. měsíci. To znamená, že se v časové řadě vyskytuje vliv sezónní složky v průběhu jednoho cyklu od 95,31% z průměru do 107,13% z průměru (průměr je 100%). Únorový počet uchazečů o zaměstnání převyšuje až o 7% celkový průměrný počet.

Tabulka 4 – Srovnání sezónních indexů pro Yt

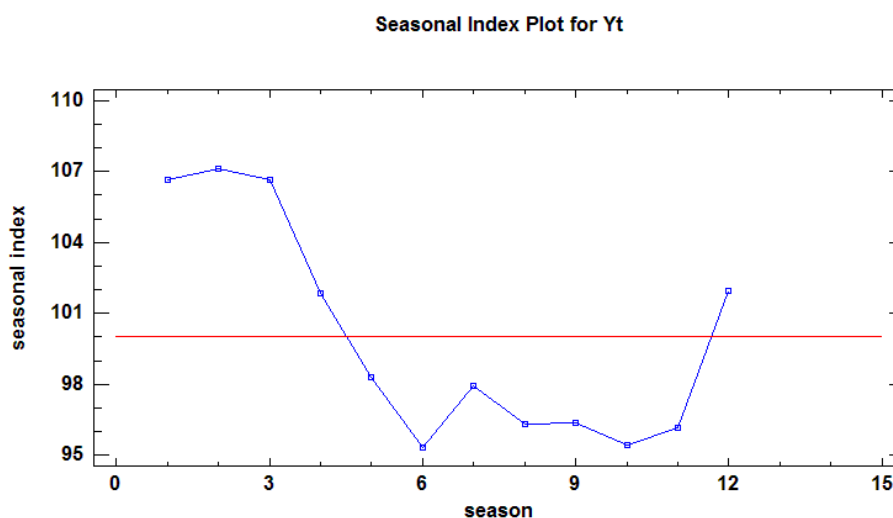
Měsíc	sezónní index (%)		
	Statgraphics	Excel	SPSS
1	106,635	106,979	106,400
2	107,131	107,476	107,300
3	106,628	106,971	106,700
4	101,851	102,179	101,700
5	98,291	98,608	98,600
6	95,311	95,618	95,800
7	97,938	98,253	96,900
8	96,316	96,627	96,200
9	96,357	96,668	96,700
10	95,437	95,745	95,300
11	96,170	96,480	96,400
12	101,937	102,265	102,100

(zdroj: výstup z programu Statgraphics a IBM SPSS Statistics, vlastní tvorba a úprava)

Výsledné hodnoty sezónních indexů z jednotlivých programů jsou srovnatelné, liší se pouze v desetínách. Vlastní výpočet těchto indexů je však časově náročnější a vyžaduje znalosti klasického modelu dekompozice časových řad. Podle výsledků, které uvádí Tabulka 4, může být tento způsob výpočtu plně nahrazen některým z nabízených statistických programů.

Obrázek 11 znázorňuje sezónní indexy, získané pomocí statistického programu Statgraphics z předchozí tabulky pomocí spojnicového grafu. Můžeme tak názorněji pozorovat změny v průměrných sezónních vlivech v průběhu roku (periody).

Obrázek 11 - Graf sezónních indexů pro Yt získaných pomocí multiplikativní metody dekompozice



(zdroj: výstup programu Statgraphics)

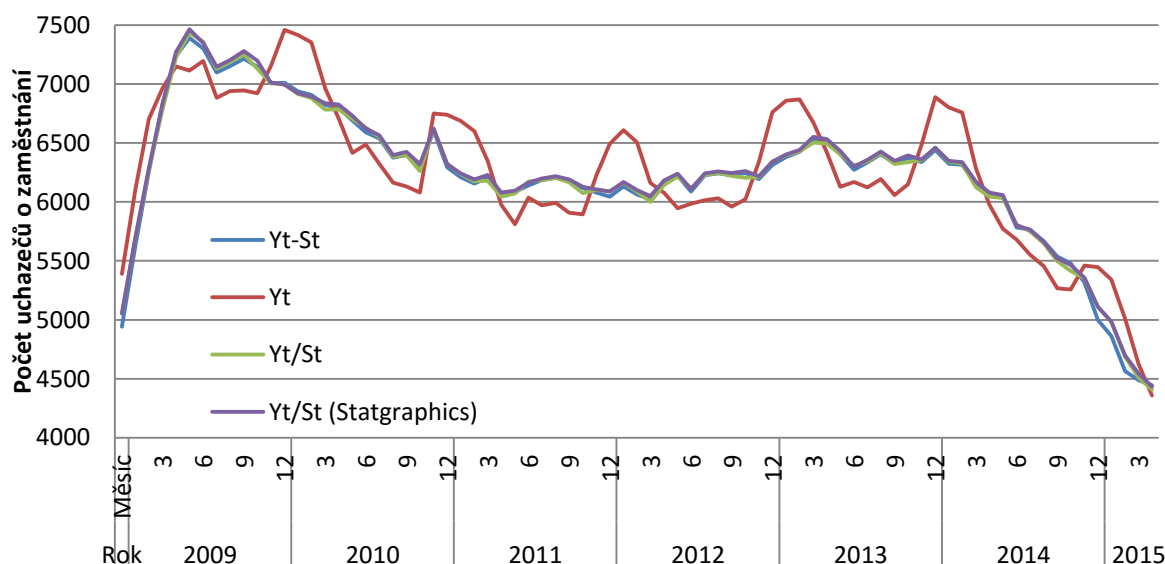
### Očištění časové řady od sezónní složky:

Pro očištění časové řady je využita metoda centrovaných klouzavých průměrů s délkou dvanácti hodnot. Metoda klasické sezónní dekompozice časové řady je popsána v odstavci 2.3.1 Klasický model dekompozice.

Z grafického srovnání metod sezónní dekompozice (viz Obrázek 12) je patrné, že aditivní ( $Y_t - S_t$ ) a multiplikativní metody ( $Y_t / S_t$ ) jsou srovnatelné. Hodnoty očištěných dat v jednotlivých metodách se liší minimálně, a proto záleží pouze na nás, kterou metodu dekompozice zvolíme. Tabulka se sezónně očištěnými daty (srovnání multiplikativní a aditivní metody pomocí tabulkového procesoru Excel) je v Příloze 2.

Pro úplnost následující graf obsahuje sezónní dekompozici vytvořenou pomocí programu Statgraphics, který automaticky volí multiplikativní metodu. Hodnoty jsou opět srovnatelné s výpočty v MS Excel, avšak časová náročnost je mnohem nižší.

Obrázek 12 - Grafické srovnání metod sezónní dekompozice



(zdroj: vlastní tvorba)

### 3.3. Elementární charakteristiky

a) Vážený chronologický průměr

$$\bar{Y} = \frac{\sum \frac{(Y_{t-1} + Y_t)}{2} * d_t}{\sum d_t}$$

$$\bar{Y} = \frac{14532963}{2311} = 6288,6$$

Volba váženého chronologického průměru je zdůvodněna v odstavci 2.1.2 Elementární charakteristiky časových řad. Pomocí průměru se dají porovnávat hodnoty v jednotlivých časových obdobích a můžeme tyto hodnoty dělit na optimistické (nižší než průměr) a pesimistické (vyšší než průměr). Výsledný průměr v celém zkoumaném období je 6288,6 a roční průměry (zaokrouhлено na celá čísla) v období 2009 – 2014 zobrazuje Tabulka 5. Z hodnot ročních průměrů z tabulky je pozorovatelný mírný klesající trend ve sledovaném období.

Tabulka 5 - Roční průměrné počty uchazečů o zaměstnání v okrese Česká Lípa

Rok	Průměr
2009	6844
2010	6701
2011	6202
2012	6173
2013	6399
2014	5967

(zdroj: vlastní výpočty)

b) Průměrný absolutní přírůstek

$$\overline{\Delta Y}_t = \frac{\sum \Delta Y_t}{T - 1}$$

$$\overline{\Delta Y}_t = \frac{-1031}{76} = -13,6$$

Hodnota průměrného absolutního přírůstku ve sledované časové řadě je -13,6, což značí celkový mírný pokles. Zajímavostí je, že v obdobích růstu má průměrný absolutní přírůstek hodnotu 209,3 a v období poklesu má tato charakteristika hodnotu -158,9. Lze tedy konstatovat, že období růstu jsou delší a rychlost růstu časové řady v těchto obdobích je vyšší než rychlost poklesu v ostatních obdobích. Charakteristiky přírůstků jsou vypočítány mechanicky pomocí průměrů z kladných hodnot (přírůstky) a ze záporných hodnot (poklesy).

c) Průměrné tempo růstu (poklesu)

$$\bar{k}_t = \sqrt[t-1]{k_1 k_2 k_3 \dots k_t}$$

$$\bar{k}_t = \sqrt[76]{0,808684} = 0,99721$$

V případě časové řady počtu nezaměstnaných jde spíše o průměrné tempo poklesu. Výsledné číslo ze vzorce vynásobené stem udává, na kolik procent průměrně poklesla hodnota proměnné Y v měsíci t oproti hodnotě z měsíce t-1. Podle této charakteristiky bychom mohli v období 2015/06 předpokládat 4346 uchazečů o zaměstnání. Avšak tato hodnota je velice přibližná, jelikož se jedná o průměr z celého období 2009-2015.

#### d) Koeficienty růstu

Pro časovou řadu, která je ovlivněna sezónními faktory, je další vhodnou elementární charakteristikou výpočet meziročního koeficientu růstu s měsíčním krokem. Hodnoty vypovídají o procentuální změně mezi stejnými měsíci dvou po sobě jdoucích let.

$$k_{i(12)} = \frac{y_i}{y_{i-12}}; i = 13, \dots, n.$$

Tabulka 6 obsahuje meziroční koeficienty růstu pro měsíc květen, kde je například vidět značný pokles počtu nezaměstnaných v roce 2015 oproti roku 2014 až o 20,1%.

Tabulka 6 - Meziroční koeficienty růstu s měsíčním krokem pro měsíc květen

t	Y <sub>t</sub>	k <sub>i(12)</sub>
2010/5	6711	<b>0,938733</b>
2011/5	5976	<b>0,890478</b>
2012/5	6077	<b>1,016901</b>
2013/5	6423	<b>1,056936</b>
2014/5	5977	<b>0,930562</b>
2015/5	4358	<b>0,729128</b>

(zdroj: vlastní tvorba)

V Příloze 5 je tabulka výpočtů absolutních přírůstků a koeficientů růstu v jednotlivých obdobích provedena pomocí statistického programu Statgraphics.

### 3.4. Předpověď

V následujících odstavcích je zhodnocena dosavadní forma předpovědi, kterou provádí úřad práce České republiky, kontaktní pracoviště Česká Lípa. V dalších krocích jsou navrženy alternativní formy předpovědi počtu uchazečů a jejich porovnání s hodnotami dostupnými z pololetních a ročních zpráv Úřadu práce v České Lípě. Alternativními možnostmi jsou (po důkladném průzkumu tedy vyhledání ve zdrojích a konzultacích s vedoucí práce) vybrány metody Brownova dvojitého exponenciálního vyrovnání, klasického modelu dekompozice a ARIMA. Zmíněné metody jsou pro předpověď časové řady počtu uchazečů nejvhodnější a jako návrh alternativní metody predikce pro Úřad práce Česká Lípa vyhovují jejich požadavkům a možnostem.

Metody Brownovo dvojité exponenciální vyrovnání a klasický model dekompozice jsou představeny za pomoci tabulkového procesoru Microsoft Office Excel, který má kontaktní pracoviště Úřadu práce v České Lípě k dispozici. Pro komplikovanější model ARIMA je vhodné využít sofistikovanější statistické programy (v případě diplomové práce jsou v této souvislosti představeny programy Statgraphics a IBM SPSS Statistics).

### 3.4.1. Dosavadní forma předpovědi

Kontaktní pracoviště Česká Lípa úřadu práce České republiky vydává každé pololetí detailní Analýzu vývoje na trhu práce v okrese Česká Lípa. Tato zpráva, kromě shrnutí nezaměstnanosti, aktivní a pasivní politiky, obsahuje také odstavec zabývající se prognózou vývoje nezaměstnanosti na trhu práce v dalším období. Jedná se o předpokládaný vývoj trhu práce v okrese, který je tvořen zaměstnanci úřadu práce kvalitativní metodou (viz odstavec 2.2 Předpověď v časových řadách). Budoucí hodnoty jsou odhadovány za pomoci předchozích dat, informací získaných od zaměstnavatelů a především zkušeností zaměstnanců a jejich přehledu v okrese.

Pracoviště uvádí vždy dvě možnosti budoucího vývoje (mírnější a pesimističtější varianta). Predikce jsou uváděny vždy pro dva ukazatele, podíl nezaměstnaných osob a počet evidovaných uchazečů o zaměstnání. Z důvodu změny metodiky podílu nezaměstnaných osob (viz odstavec 3.1 Popis konkrétní časové řady) je práce zaměřena na práci s hodnotami počtu uchazečů o zaměstnání.

Odhady pro rok 2015 (zveřejněny v lednu roku 2015) v mírnější i pesimističtější variantě zobrazuje Tabulka 7. Podobné tabulky jsou zveřejňovány vždy v půlročních intervalech a obsahují předpovědi k první polovině roku a ke konci roku. Pro červen 2015 byl počet uchazečů o zaměstnání odhadován mezi hodnotami 5 300 a 6 000. Již známá hodnota pro červen 2015 činila 4 115 uchazečů. V tomto případě byla kvalitativní metoda predikce velmi nepřesná (rozdíl 1185 osob od skutečné hodnoty).

Tabulka 7 - Předpokládaný vývoj nezaměstnanosti v dalším období

Očekávaný stav k	Mírnější varianta		Pesimističtější varianta	
	počet evidovaných uchazečů celkem	podíl nezaměstnaných osob (PNO) v %	počet evidovaných uchazečů celkem	podíl nezaměstnaných osob (PNO) v %
30. 6. 2015	5 300	7,2	6 000	8,1
31. 12. 2015	5 500	7,7	6 300	8,6

(zdroj: kontaktní pracoviště Česká Lípa)

### 3.4.2. Klasický model dekompozice

Jako první forma předpovědi budoucích hodnot časové řady je představena metoda klasické dekompozice, která je založena na rozdělení časové řady na 4 složky (trendovou, sezónní, cyklickou a náhodnou). Více informací o této metodě je rozepsáno v teoretické části práce v odstavci 2.3.1 Klasický model dekompozice. Tato metoda se opírá především o tvar křivky dlouhodobého trendu, která je extrapolována v čase  $t$  podle zvolené trendové funkce. Výhodou metody dekompozice je, že pracuje se sezónními vlivy a v předpovědi jsou zahrnuty sezónní výkyvy, hodnoty předpovědi jsou tak bližší skutečným hodnotám.

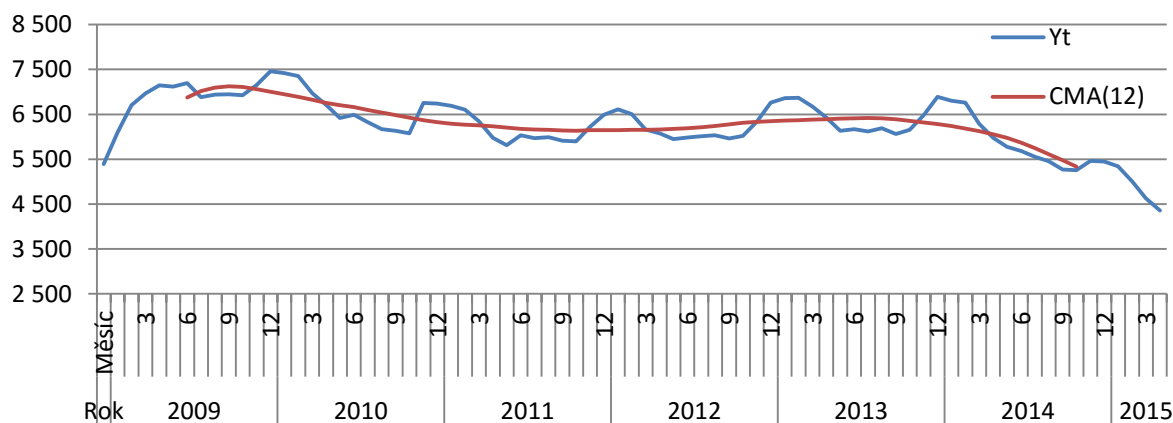
#### Centrované klouzavé průměry

Pro vyhlazení časové řady je použita metoda centrovaných klouzavých průměrů neboli centered moving average (CMA). O této metodě je vše popsáno v odstavci 2.3.1 Klasický model dekompozice.

Práce se opírá o hodnoty z tabulky uvedené v Příloze 1, kde jsou hodnoty časové řady seřazeny podle měsíců a roků a navíc označeny časovou proměnnou  $t$ , kde  $t = (1;77)$ . Nejprve jsou hodnoty  $Y_t$  upraveny pomocí metody centrovaných klouzavých průměrů řádu 12 neboli CMA(12), řád je zvolen z důvodu opakující se 12ti měsíční periody. „Je-li délka sezónnosti liché číslo, odhaduje se trend jednoduchými klouzavými průměry, je-li délka sezónnosti sudé číslo, odhaduje se váženými (centrovanými) klouzavými průměry.“[2]. Metoda klouzavých průměrů modeluje trendovou složku časové řady, jde o vyrovnaní časové řady v krátkých úsecích matematickou křivkou.

Z grafu (viz Obrázek 13) je vidět rostoucí trend na začátku sledované časové řady, který se na přelomu let 2009 a 2010 vyrovnal a osciloval kolem konstantního trendu, až do poloviny roku 2013. V posledních dvou letech sledované časové řady je trend mírně klesající.

Obrázek 13 - Graf vyrovnaní časové řady hodnot  $Y_t$  metodou klouzavých průměrů (CMA)



(zdroj: vlastní tvorba)

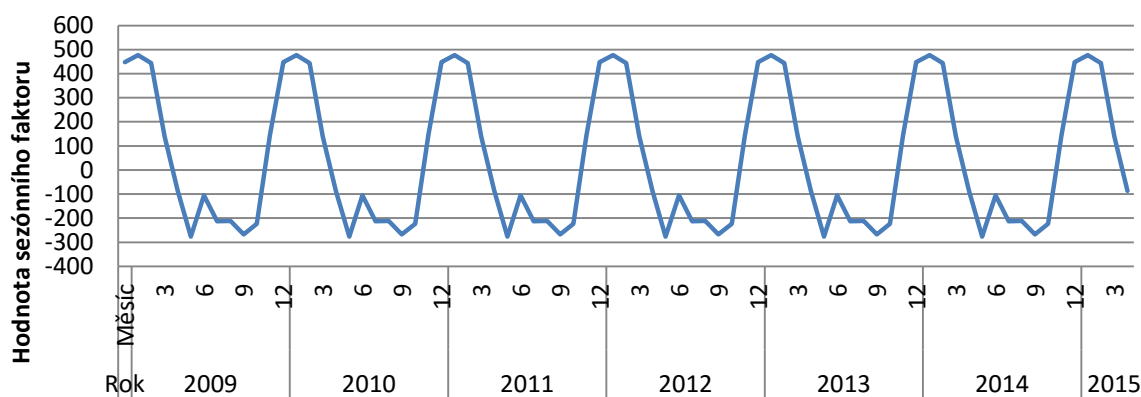


## Výpočet sezónního faktoru

V následujícím kroku jsou pomocí původních a vyrovnaných dat extrahovány sezónní a náhodná složka. V případě analyzované časové řady, kdy je zvolen multiplikativní model, tyto složky spočítáme vzorcem:  $S_t * I_t = Y_t / CMA(12)$ .

Z výsledných hodnot pak jsou snadno spočitatelné průměrné sezónní složky v jednotlivých měsících, a to zprůměrováním hodnot ve stejných měsících za celé období časové řady. V odstavci 4.2. Grafická analýza průběhu časové řady jsou k porovnání hodnoty výpočtu sezónních indexů v programech IBM SPSS Statistics, Statgraphics a MS Excel. Obrázek 14 znázorňuje graf sezónních indexů pro celé sledované období. Graf byl vytvořen v MS Excel a obsahuje hodnoty vlastnoručně vypočtené podle metody centrovaných klouzavých průměrů.

Obrázek 14 - Hodnoty sezónního faktoru pro  $Y_t$  získaných pomocí aditivní metody



(zdroj: vlastní tvorba)

## Křivka trendu

V posledním kroku dochází k formulaci křivky trendu pro výpočet předpovědí. K tomu je použita jednoduchá lineární regrese pomocí dat očištěných od sezónních vlivů a časová jednotka  $t$ . Z regrese jsou vygenerovány koeficienty rovnice, pomocí které je možné vyjádřit trendovou funkci.

Tabulka 8 - Koeficienty získané pomocí jednoduché lineární regrese

	Koeficienty
Hranice	6930,80556
$t$	-17,71680016

(zdroj: výstup funkce regrese programu Excel)

Trendová funkce je podle výstupu, který uvádí Tabulka 8, ve tvaru  $T_t = 6930,80556 - 17,71680016 * t$ . Tvar trendové funkce značí mírný klesající trend, a to díky hodnotě -17,71680016.

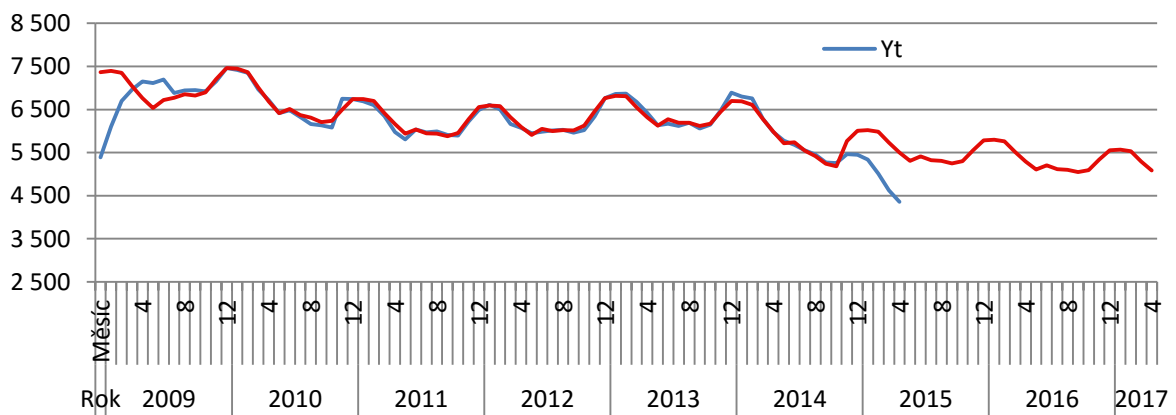
### Bodová předpověď pro další dva roky

Nakonec už je jen potřeba k funkci opětovně připojit sezónní složku, a to se provede opět multiplikační formou. Finální předpověď tedy získáme pomocí vzorce  $\hat{Y}_t = T_t * S_t$ .

Obrázek 15 zobrazuje porovnání původních hodnot s předpověďmi a bodové předpovědi pro následujících dvacet čtyři období.

Výkyv v hodnotách  $Y_t$  a hodnotách předpovědi na začátku zkoumaného období je způsoben prudkým nárůstem počtu nezaměstnaných v předchozích obdobích. Vlivem vyrovnání časové řady a zakomponováním sezónního faktoru je v grafu vidět nesrovnalost, která se však s postupem času eliminuje.

Obrázek 15 - Předpovědi klasického modelu dekompozice a původní hodnoty



(zdroj: vlastní tvorba)

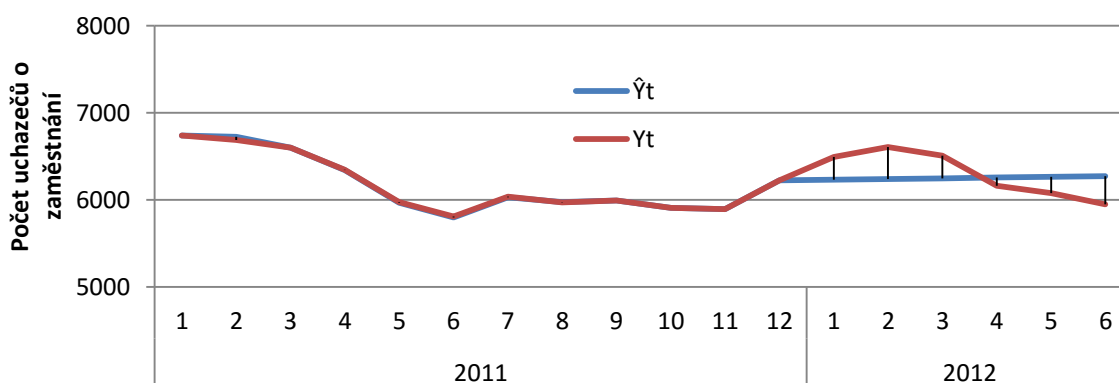
Jedná se pouze o bodovou predikci, intervalové predikce jsou na výpočet v Excelu komplikované. Intervaly je možné zkonstruovat podle Cipry (1986, str. 70). Numerické hodnoty předpovědi jsou obsaženy v Příloze 6.

### 3.4.3. Brownovo dvojité exponenciální vyrovnání

Pokud bychom chtěli vytvářet predikci maximálně na půl roku dopředu, je vhodné využít metodu dvojitého exponenciálního vyrovnání podle Browna, která je přesnější pro krátkodobé predikce (1 až 6 období dopředu).

Z dostupných dat jsem vytvořila nejprve kontrolní predikce tak, že jsem z ročních dat předpověděla následujících 6 měsíců (postup podle odstavce 2.3.2 Brownovo exponenciální vyrovnání). Předpovědní hodnoty jsem následně porovnála se skutečnými hodnotami. V grafu níže jsou znázorněna data pro rok 2011, předpovědi ( $\hat{Y}_t$ ) a skutečné hodnoty ( $Y_t$ ) pro první polovinu roku 2012.

Obrázek 16 - Graf kontrolních predikcí Brownova dvojitého exponenciálního vyrovnání

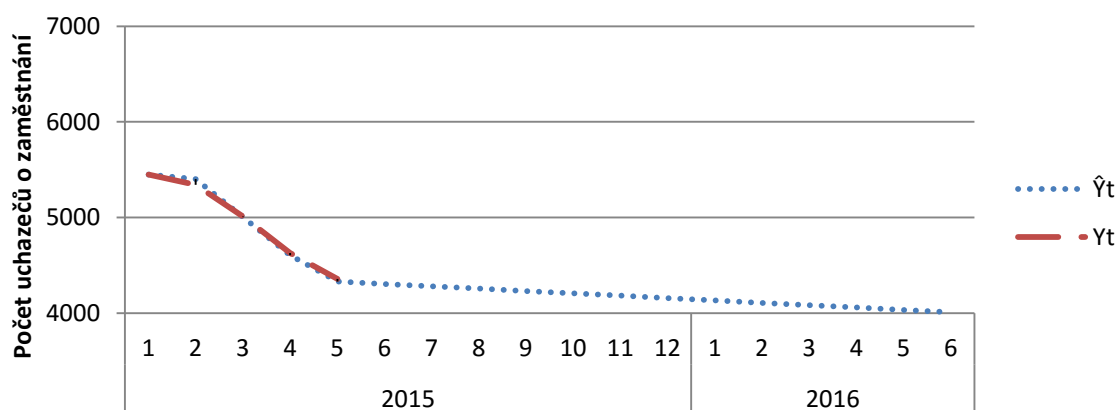


(zdroj: vlastní tvorba)

Touto metodou se vyplatí počítat pouze nejbližší predikce, jelikož s rostoucím  $t$  rostou i odchylky od skutečných hodnot. Predikce a jejich grafické porovnání se skutečnými hodnotami pro všechna sledovaná období jsou v Příloze 3.

Po zhodnocení kontrolních predikcí v každém roce, jsem vytvořila předpověď bez kontrolních hodnot. Predikce pro rok 2015/2016 vypadá následovně. Jedná se pouze o bodovou predikci, intervalové predikce jsou na výpočet v Excelu komplikované. Intervaly je možné zkonstruovat podle Cipry (1986, str. 70).

Obrázek 17 - Predikce pomocí metody Brownova dvojitého exponenciálního vyrovnání



(zdroj: vlastní tvorba)

Přesné hodnoty bodových odhadů počtu uchazečů o zaměstnání pro další rok, které jsou zobrazeny v grafu, uvádí Tabulka 9.

Tabulka 9 - Bodové odhady počtu uchazečů o zaměstnání pro období 6/2015 - 6/2016

Rok	Měsíc	Ÿt
2015	6	<b>4305</b>
	7	<b>4280</b>
	8	<b>4256</b>
	9	<b>4231</b>
	10	<b>4206</b>
	11	<b>4182</b>
	12	<b>4157</b>
2016	1	<b>4132</b>
	2	<b>4107</b>
	3	<b>4083</b>
	4	<b>4058</b>
	5	<b>4033</b>
	6	<b>4009</b>

(zdroj: vlastní tvorba)

### 3.4.4. ARIMA model v programu Statgraphics

Statistický program Statgraphics, který je vhodný pro analýzu časových řad, nabízí volbu automatické předpovědi.

#### Postup zadávání dat do programu Statgraphics

Do vstupního panelu Automatic Forecasting se vloží časová řada pro  $Y_t$  a je zvolen měsíční časový interval začínající hodnotou 1.09, což znamená první měsíc v roce 2009. Do kolonky Seasonality, neboli v jakých intervalech se opakují sezónní vlivy, je zadána hodnota 12. Počet předpovědí, které chceme zobrazit, je zadán hodnotou 24 (tedy 2 roky) a 12 pozorování z konce časové řady bude použito pro validační účely.

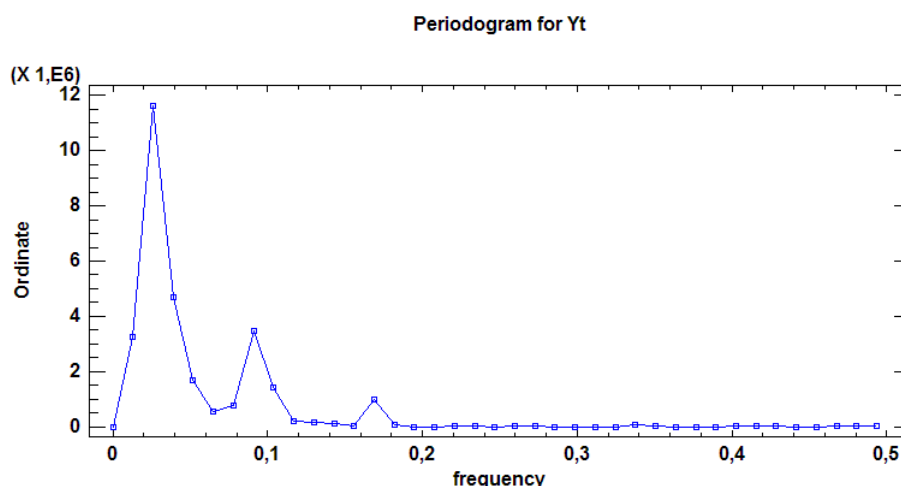
Po vyplnění vstupního panelu se zobrazí nabídka Automatic Forecasting options, kde mohou být zvoleny modely, které budou zahrnuty do analýzy. Dále je zde i možnost volby metody kritéria při výběru vhodného modelu předpovědi. Procedura automatické předpovědi je založena na indikování a výběru modelu, který má nejnížší výslednou hodnotu výběrového kritéria.

#### Identifikace sezónnosti

Identifikaci sezónní složky provádíme podle Arlta, Arltové, Rublíkové (2002) pomocí periodogramu ve Statgraphicsu v proceduře Descriptive Methods.

Periodogram časové řady očištěné od lineárního trendu je na následujícím grafu, kde vidíme 3 výrazné vrcholy. Je proto možné očekávat 3 cykly s různou délkou.

Obrázek 18 - Periodogram řady reziduí (řada očištěná od lineárního trendu)



(zdroj: výstup programu Statgraphics)

Z grafu se však frekvence jednotlivých vrcholů nedají přesně zjistit, proto použijeme hodnoty periodogramu z tabulky. Zde vidíme 3 významné vrcholy, celá tabulka je k nahlédnutí v Příloze 8.

Tabulka 10 - Hodnoty periodogramu pro  $Y_t$

$i$	Frequency	Period	Ordinate	Cumulative Sum	Integrated Periodogram
3	0,038961	25,6667	4,69986E6	1,95734E7	0,662944
8	0,103896	9,625	1,44273E6	2,75069E7	0,931648
14	0,181818	5,5	57178,9	2,91058E7	0,985803

(zdroj: upravený výstup z programu Statgraphics)

Maximální hodnota je ve frekvenci 0,038961, což signalizuje přítomnost periody délky asi 26 let (viz sloupec Period). V řadě se dále vyskytuje druhý vrchol s frekvencí 0,103896, což odpovídá délce cyklu přibližně 10 let, frekvence třetího vrcholu je 0,181818 odpovídající délce cyklu asi 5 až 6 let.

„Jelikož je v časové řadě identifikována sezónnost, je vhodné ji modelovat. Sezónní dekompozice se využívá především k získání odhadů sezónních výkyvů a sezónně očištěné časové řady. Sezónní indexy jsou interpretovány v procentech.“[2]. Sezónní dekompozici provedeme v programu Statgraphics pomocí procedury Seasonal Decomposition. Tato procedura provede každý krok sezónní dekompozice multiplikační metodou. Trendovou a cyklickou složku získá pomocí centrovaných klouzavých průměrů délky 12 aplikovaných na hodnoty  $Y_t$ .

Sezónnost je získána podílem hodnot pozorování klouzavými průměry a následně vynásobením 100x. Sezónní indexy jsou následně sečteny za každý měsíc a zprůměrovány skrze všechna pozorování v tomto měsíci. Nakonec jsou data vydělena hodnotami trendové, cyklické a sezónní složky, pro získání nesystematického komponentu (reziduální složka je také vynásobena 100x). Všechny zmíněné výsledky uvádí Příloha 9.

### **Vygenerování prognóz**

V případě, že máme všechny požadavky zadány, můžeme Statgraphicsu potvrdit naši volbu. Program následně zkonstruuje všechny zvolené modely, otestuje je podle výběrového kritéria a navrhne nejvhodnější model pro danou časovou řadu. „Při hledání modelu procedura vyzkouší všechny modely, které byly zaškrtnuty v dialogovém okně Analysis Options. Vyjma Wintersova exponenciálního vyrovnání a ARIMA modelů jsou sezónní data nejprve očištěna od sezónních vlivů, poté je provedena předpověď. Nejdříve je vygenerována první předpověď, poté je sezónní faktor vrácen zpět pro tvorbu finální předpovědi.“ (STATPOINT TECHNOLOGIES, INC. Automatic Forecasting; 2010).

Pro analyzovaná data počtu nezaměstnaných procedura zvolila ARIMA(0,1,0)x(0,1,2)<sub>12</sub> model. Jak je vidět z grafu testujících hodnot a prognóz, výsledky jsou velmi uspokojivé. V grafu můžeme pozorovat pokračující mírný klesající trend se silnou sezónní oscilací. I když se předpovědní limity mohou zdát poměrně široké, musíme brát v potaz různé vybočující situace, které mohou nastat, navíc z důvodu, že byly pozorovány již minulosti. Tím mám na mysli například velkou ekonomickou krizi v letech 2007 a 2008, která měla za následek prudký nárůst nezaměstnanosti v České republice.

Detailně, graf zobrazuje pozorované údaje  $Y_t$  včetně případných hodnot nahrazujících chybějící data, zobrazené jako bodové symboly. Dále zobrazuje takzvané předpovědi jeden krok vpřed neboli  $\hat{Y}_t(1)$ , které jsou znázorněny, jako plná čára skrze body. Tyto předpovědi jsou vytvořeny s použitím zvoleného modelu a předpovídají hodnotu  $t+1$  pro každou časovou periodu pouze pomocí informací dostupných v čase  $t$ .  $\hat{Y}_t(1)$  chyby prognóz neboli  $e_t$  jsou pozorovatelné jako svislá vzdálenost mezi pozorováním a plnou čarou.

Předpovědi pro budoucí hodnoty vytvořené v čase  $t$ , kdy jsou pozorování dostupná, jsou zobrazeny prodloužením plné čáry předpovědi za posledním pozorováním.

Graf také znázorňuje pravděpodobnostní limity pro předpovědi u  $100(1 - \alpha)\%$  intervalu spolehlivosti (v našem případě 95% interval spolehlivosti), kde jsou meze dány vzorcem:

$$F_{n+m}(k) \pm z_{\alpha/2} \sqrt{\hat{V}(k)},$$

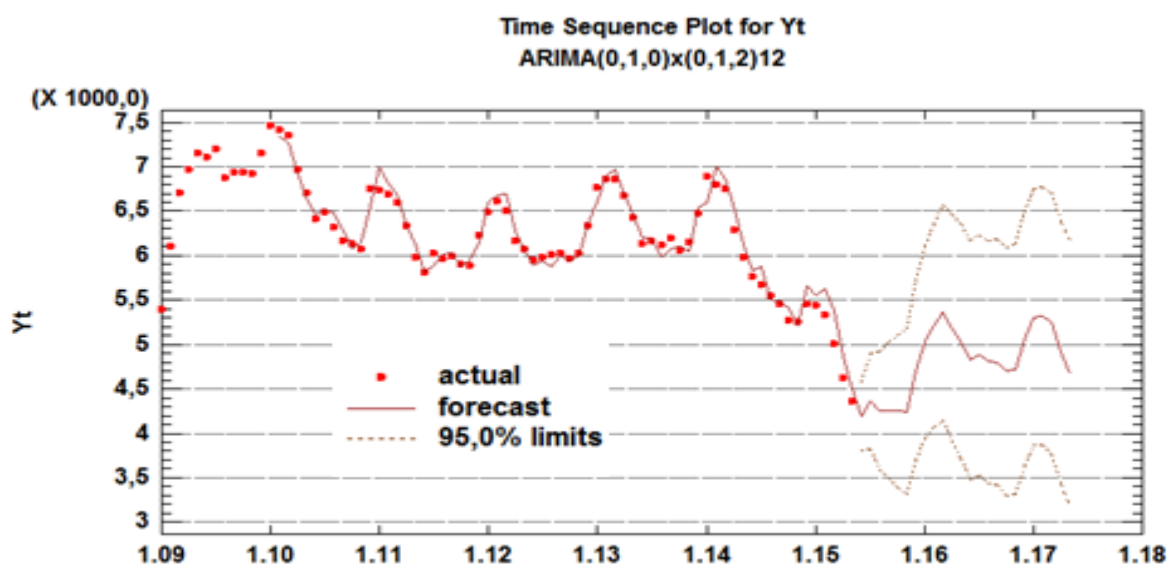
kde  $\hat{V}(k)$  je rovno odhadovanému rozptylu předpovědi k period. Je nutné poznamenat, že limity jsou platné jen v případě dodržení několika předpokladů, včetně:

- Byl vybrán odpovídající model
- Vybraný model byl validní pro všechna historická data
- Vybraný model je nadále validním i v budoucnosti
- Model byl odhadnut z tak dlouhé časové řady, že chyba odhadu modelu je zanedbatelná ve srovnání s rozptylem chyb (s výjimkou pro modely odhadované pomocí lineární regrese, které již zahrnují model chyby odhadu)

[35]

V praxi jsou limity považovány za přiblížení toho, jak daleko mohou časové řady vybočovat od předpovídaných hodnot v budoucnosti.

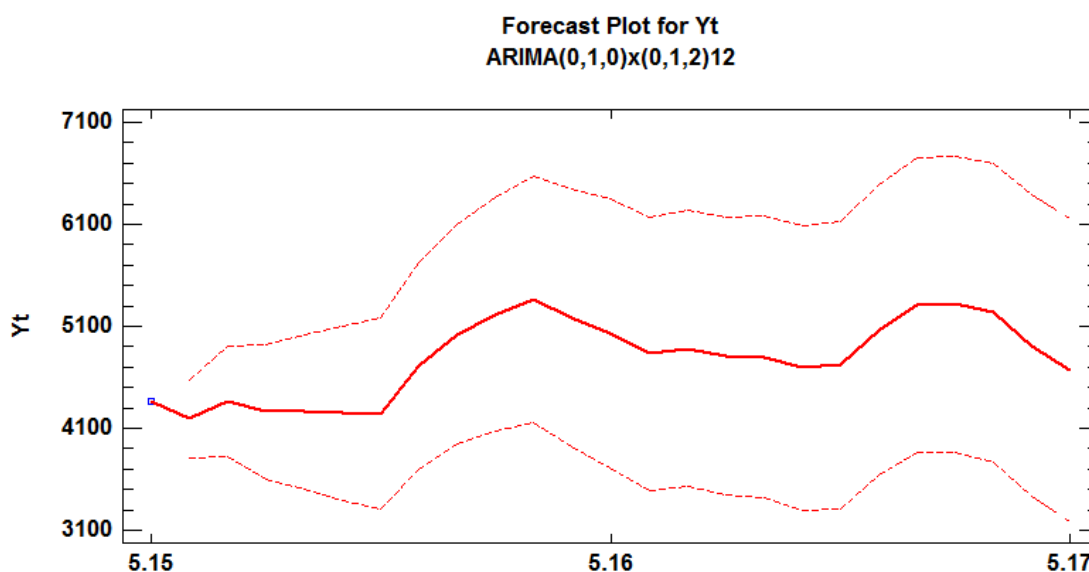
Obrázek 19 - Původní časová řada s předpověďmi a předpovědními intervaly



(zdroj: výstup z programu Statgraphics)

Procedura vygenerovala předpovědi ovlivněné sezónním faktorem, což je lépe vidět na grafu, který zobrazuje předpovědi a předpovědní limity. Největší výkyvy tvoří sezónní faktor vždy v zimních měsících (leden, únor, březen).

Obrázek 20 - Graf předpovědí pro období 6/2015 - 6/2017



(zdroj: výstup z programu Statgraphics)

Hodnoty předpovědí z grafu, který zobrazuje Obrázek 20, jsou uvedeny v příloze 10.

### Výběrové kritérium a testy kvality modelu

Procedura Automatické předpovědi nám vygenerovala tabulku porovnávající modely podle výběrových kritérií. Jelikož jsme jako srovnávací kritérium zvolili Akaikeho informační kritérium (AIC), byl zvolen model s nejnižší hodnotou tohoto kritéria. Hodnota AIC pro model  $ARIMA(0,1,0) \times (0,1,2)_{12}$  je 9,52023, a proto byl tento model (ve výstupu označený písmenem N) zvolen jako výchozí pro generování prognózy.

Procedura též shrnuje výsledky pěti testů založených na residuu, a tím testuje, který model je pro data adekvátní. "OK" znamená, že model testem prošel, "\*" znamená, že test selhal na 95% intervalu spolehlivosti, "\*\*\*" znamenají, že model selhal na 99% intervalu spolehlivosti, "\*\*\*\*" znamenají, že model selhal na 99,9% intervalu spolehlivosti. Celkově v těchto testech dopadly nejlépe, jak znázorňuje Tabulka 11, modely ARIMA a Brownovo exponenciální vyrovnání. Model  $ARIMA(0,1,0) \times (0,1,2)_{12}$ , aktuálně vybraný model, v tomto případě prošel všemi pěti testy.



Tabulka 11 - Srovnání modelů podle výběrových kritérií

<i>Model</i>	<i>RMSE</i>	<i>RUNS</i>	<i>RUNM</i>	<i>AUTO</i>	<i>MEAN</i>	<i>VAR</i>
(A)Random walk	195,568	OK	OK	**	OK	***
(B)Random walk with drift	196,387	OK	OK	***	OK	***
(C)Constant mean	462,255	***	***	***	***	***
(D)Linear trend	434,936	***	***	***	OK	***
(E)Quadratic trend	434,219	***	***	***	OK	***
(F)Exponential trend	434,196	***	***	***	OK	***
(G)S-curve trend	433,679	***	***	***	OK	***
(H)Simple moving avg of 2 terms	243,606	OK	**	***	OK	***
(I)Simple exponential smoothing	195,578	OK	OK	***	OK	***
(J) Brown's linear exp. smoothing	214,779	OK	OK	OK	OK	***
(K)Holt's linear exp. smoothing	184,44	OK	*	**	OK	***
(L)Brown's quadratic exp. smoothing	250,839	OK	OK	OK	OK	***
(M)Winters' exp. smoothing	233,205	OK	**	***	OK	OK
(N)ARIMA(0,1,0)x(0,1,2) <sub>12</sub>	113,222	OK	OK	OK	OK	OK
(O)ARIMA(1,2,0)x(0,1,2) <sub>12</sub>	114,764	OK	OK	OK	OK	OK
(P)ARIMA(0,1,2)x(0,1,2) <sub>12</sub>	113,165	OK	OK	OK	OK	OK
(Q)ARIMA(2,1,1)x(0,1,2) <sub>12</sub>	112,703	OK	OK	OK	OK	OK
(R)ARIMA(1,2,0)x(2,1,2) <sub>12</sub>	112,951	OK	OK	OK	OK	OK

(zdroj: výstup z programu *Statgraphics*)

### 3.4.5. ARIMA model v programu IBM SPSS Statistics

#### Popisná statistika

Tabulka 12 - Popisná statistika

		<b>Descriptives</b>		
		Statistic	Std. Error	
Počet uchazečů o zaměstnání	Mean	6271,38	71,030	
	95% Confidence Interval for Mean	Lower	6129,91	
		Upper	6412,85	
	5% Trimmed Mean	6292,76		
	Median	6193,00		
	Variance	388486,159		
	Std. Deviation	623,287		
	Minimum	4358		
	Maximum	7460		
	Range	3102		
	Interquartile Range	781		
	Skewness	-,512	,274	
	Kurtosis	,479	,541	

(zdroj: výstup programu *IBM SPSS Statistics*)

Pomocí cesty z hlavní nabídky Analyze/Descriptive Statistics/Explore program vygeneruje tabulku se základními výpočty popisné statistiky. Tabulka 12 zobrazuje hodnotu průměru pro danou časovou řadu 6271,38 uchazečů; 95% interval spolehlivosti pro průměr je v intervalu (6129,91;6412,85) a průměrná odchylka od průměru je 71,030. Za zmínku stojí ještě směrodatná odchylka (Std. Deviation) časové řady, která značí relativně vysoké výkyvy od průměru.

Tabulka 12 znázorňuje i rozsah minimální a maximální hodnoty časové řady, který má za období leden/2009 – květen/2015 hodnotu 3102 uchazečů. Velký rozsah je zřejmě dán klesajícím trendem v daném období, jak ukazuje Tabulka 13 extrémních hodnot, která zobrazuje pět nejnižších a pět nejvyšších hodnot v období a přiřadí tuto hodnotu k období, kdy extrém nastal.

Tabulka 13 - Extrémní hodnoty časové řady

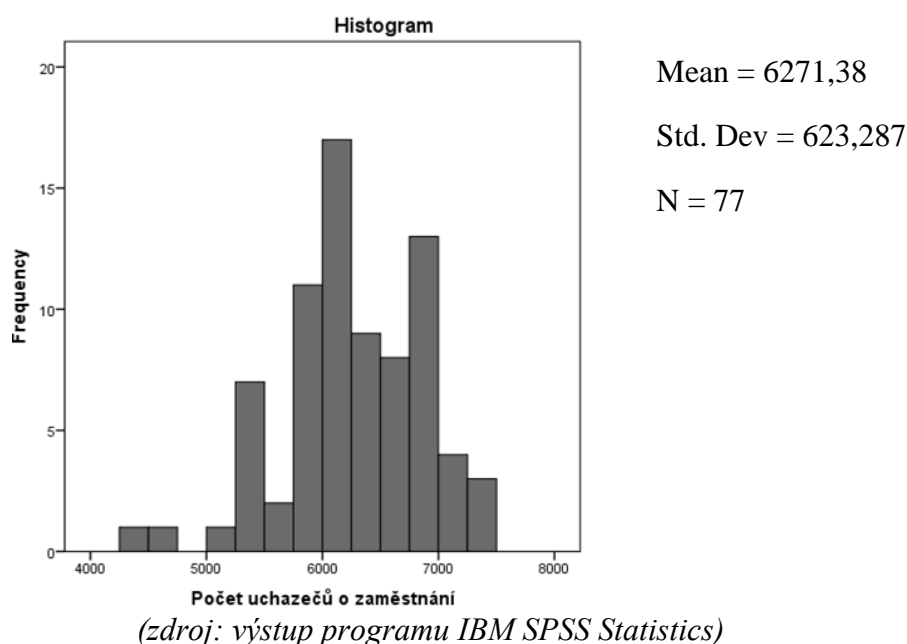
<b>Extreme Values</b>				
		Case Number	Date. Format: "MMM YYYY"	Value
Highest	1	13	JAN 2010	7460
	2	14	FEB 2010	7415
	3	15	MAR 2010	7353
	4	7	JUL 2009	7197
	5	12	DEC 2009	7151
Lowest	1	77	MAY 2015	4358
	2	76	APR 2015	4628
	3	75	MAR 2015	5009
	4	71	NOV 2014	5256
	5	70	OCT 2014	5268

*(zdroj: výstup programu IBM SPSS Statistics)*

### Grafická analýza

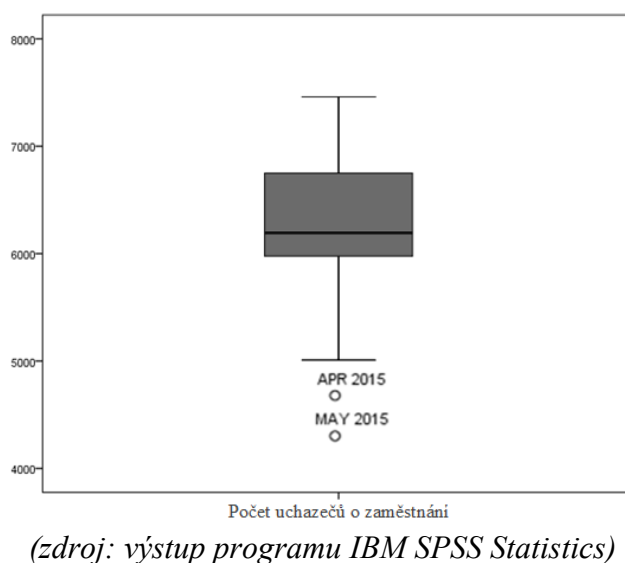
Ve volbě Explore program vygeneruje dva způsoby grafického zobrazení rozložení jednotlivých pozorování. Prvním z nich je histogram, který ukazuje frekvenci výskytu hodnot podle nastavených intervalů. V grafu je vidět největší výskyt hodnot v intervalu (6000; 6250). Již v tomto grafu jsou vidět relativně odlehlé hodnoty v rozmezí (4250; 4750), které lépe znázorňuje krabicový graf.

Obrázek 21 - Histogram počtu uchazečů o zaměstnání



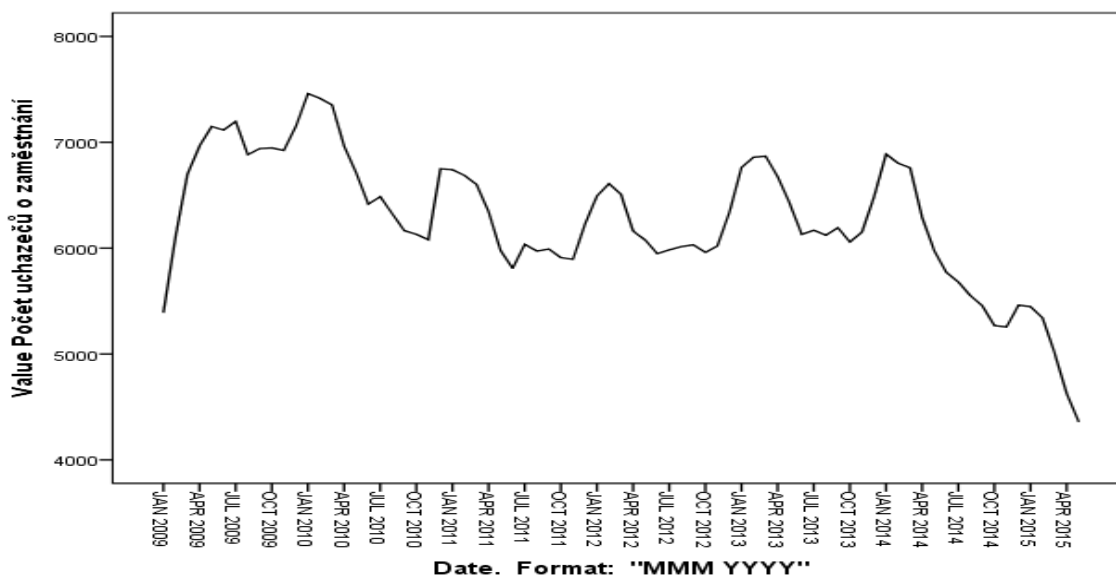
Druhou možností zobrazení je krabicový graf, který nalezneme pod názvem Boxplot. Krabicový graf zobrazuje rozpětí a rozložení dat kolem číselné osy. Modrá oblast je ohraničena kvartily  $x_{0,25}$  a  $x_{0,75}$  a obsahuje medián časové řady (vodorovná čára uvnitř „krabice“). Vodorovné čáry mimo modrou oblast značí nejnižší údaj 1,5 IQR spodního kvartilu a nejvyšší údaj 1,5 IQR horního kvartilu a vyjadřují variabilitu dat pod prvním a nad třetím kvartilem. Odlehlé hodnoty jsou pak vykresleny jako samostatné body s popiskem data, kdy hodnota nastala. V grafu jsou vidět dvě hodnoty, které jsou nižší než 1,5 IQR spodního kvartilu. Tyto hodnoty jsou z konce sledované časové řady, přesněji v dubnu a květnu roku 2015, a jsou způsobeny klesajícím trendem celé časové řady.

Obrázek 22 - Krabicový graf počtu uchazečů o zaměstnání



Možnost zobrazení spojitého grafu časové řady nalezneme v nabídce Graphs/Legacy Dialogs/Line, která zakreslí jednotlivé hodnoty do souřadných os a zobrazí pozorování ve formě spojitě linie. Na vodorovné ose je vynesena časová proměnná  $t$  a na svislé ose jsou počty uchazečů o zaměstnání. Při znázornění lineárního grafu časové řady počtu uchazečů o zaměstnání je přehledně vidět sezónní vliv v zimním období a klesající trend ke konci období.

Obrázek 23 - Spojitý graf počtu uchazečů o zaměstnání



(zdroj: výstup programu IBM SPSS Statistics)

### Volba předpovědního modelu

Pro využití funkce Expert Modeler pro předpověď dat zvolíme v hlavním menu cestu Analyze/Forecasting/Create Models. V zobrazeném dialogovém okně v záložce Variables zvolíme za závislou proměnnou (Dependent Variables) časovou řadu Počet uchazečů o zaměstnání. V rolovacím menu vybereme metodu Expert Modeler, která automaticky vybere vyhovující model pro předpověď. V záložce Options u volby Forecast Period zaškrtneme možnost First case after end of estimation period through a specified date a do políčka Date vepíšeme rok 2017 a měsíc 5. S tímto nastavením bude vytvořena prognóza pro období od června 2015 do května 2017.

V záložce Statistics je možné nastavit statistiku pro porovnání modelů. Ve standardním výběru je možnost Goodness of fit, která vytvoří tabulku s výběrovými statistikami, jako jsou R-kvadrát, střední absolutní procentuální chyba a normalizované BIC (Bayesovo informační kritérium). Tabulka poskytne shrnutí, jak dobře se zvolené modely hodí k datům.

Pro vybraný model ARIMA a naše data bylo vypočteno  $BIC = 10,168$ . Pro toto kritérium platí, že čím menší je, tím lepší byl zvolen model.

Tabulka 14 - Model Description znázorňuje model  $ARIMA(0,2,1)(0,1,1)$ , který byl podle zmíněných kritérií zvolen jako nejvhodnější pro data počtu uchazečů o zaměstnání.

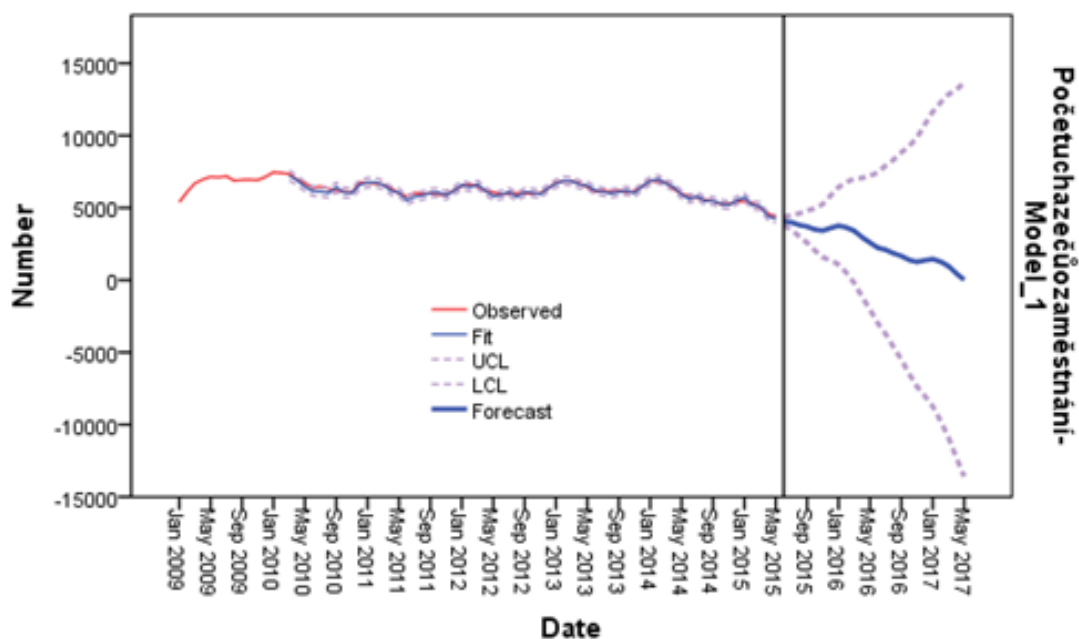
Tabulka 14 - Volba modelu pro prognózu počtu uchazečů o zaměstnání

Model Description			Model Type
Model ID	Počet uchazečů o zaměstnání	Model_1	$ARIMA(0,2,1)(0,1,1)$

(zdroj: výstup programu IBM SPSS Statistics)

Příloha 10 zobrazuje vygenerovanou tabulku prognóz vývoje počtu uchazečů o zaměstnání v období červen/2015 – květen/2017, kde jsou vidět přesné hodnoty grafického zobrazení. Model je z pohledu budoucích hodnot nekvalitní, jelikož hodnoty klesají velmi rychle až na nepřírozených 22 uchazečů o zaměstnání v květnu roku 2017. Model je tedy použitelný pouze pro krátkodobé predikce.

Obrázek 24 - Původní časová řada s předpověďmi a předpovědními intervaly



(zdroj: výstup programu IBM SPSS Statistics)

## 3.5. Statistické doplňky pro MS Excel

Cílem této části práce je představit možnost využití volně dostupných statistických balíčků, které mohou být vestavěny přímo do programu Microsoft Excel.

Proč Excel? Uživatelé volí program MS Excel z důvodu, že je široce rozšířený, a tak vědí jak s ním pracovat. Není proto potřeba učit se nové metody manipulací s daty a tvorby grafů. Další výhodou oproti speciálním statistickým programům je, že MS Excel uživatelům poskytuje více možností kontroly a flexibility. Je totiž mnohem lépe vidět do funkcí probíhajících pomocí MS Excel než u jiných statistických programů, kde analýza probíhá skrytě vůči uživateli.

MS Excel však nebyl původně vyvinut pro statistické účely a je v základním balíčku limitován, co se týká obsahu statistických funkcí a nástrojů pro analýzu dat. Naštěstí jsou pro základní obsah MS Excel vytvořena i mnohá vylepšení a rozšiřující doplňky.

[32][39]

### 3.5.1. Real Statistics

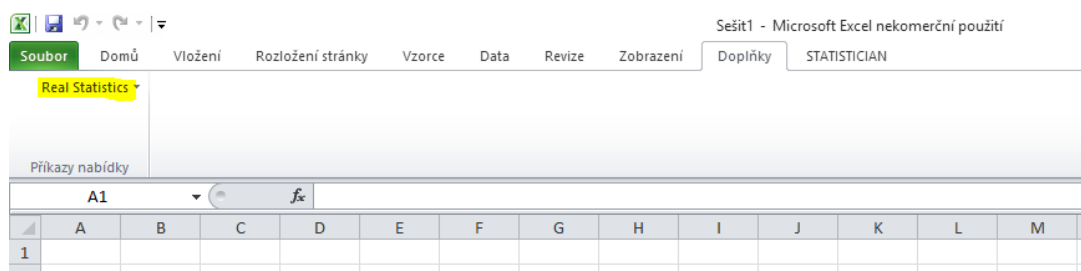
Real Statistics je volně dostupný statistický software, který rozšiřuje vestavěné statistické funkce MS Excel a tím umožňuje snadněji provádět celou řadu statistických analýz v MS Excel.[39]

#### Instalace doplňku

Balíček Real Statistics Resource Pack je volně stažitelný z Real Statistics Website (dostupné z: <http://www.real-statistics.com>), kde jsou k dispozici návody statistických analýz s použitím vestavěných funkcí Excelu a využitím rozšířených možností Real Statistics.

Instalace doplňku Real Statistics kompatibilního s MS Excel 2010 je jednoduchá, přímo ze stránek produktu, a proběhla během jedné minuty. Při otevření staženého souboru s názvem RealStats.xlam se zobrazí nový dokument v programu Excel, kde se v liště základní nabídky ve volbě Doplňky objeví nový doplněk Real Statistics, jak znázorňuje Obrázek 25.

Obrázek 25 - Doplňek Real Statistics v programu MS Excel

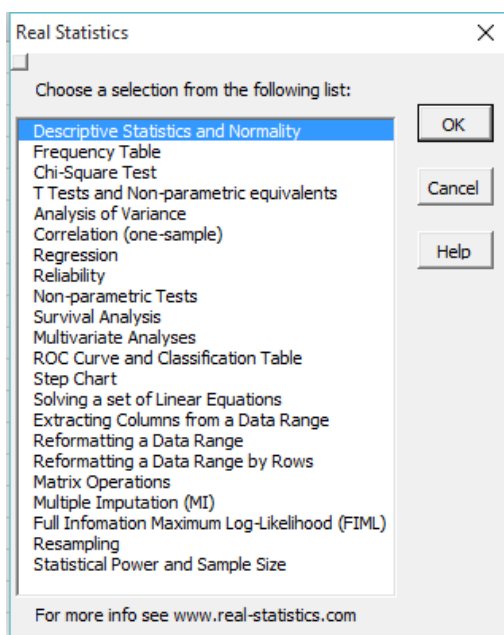


(zdroj: výstřížek z programu MS Excel)

## Nabídka funkcí

Pokud instalace doplňku proběhla v pořádku, je možno použít všechny funkce Real Statistics stejným způsobem jako se používají vestavěné funkce MS Excel. Kompletní seznam funkcí dodávaných doplňkem MS Excel lze nalézt na stránkách produktu (dostupné z: <http://www.real-statistics.com>) v částech Real Statistics Functions a Real Statistics Multivariate Functions.

Obrázek 26 - Nabídka funkcí doplňku Real Statistics v MS Excel



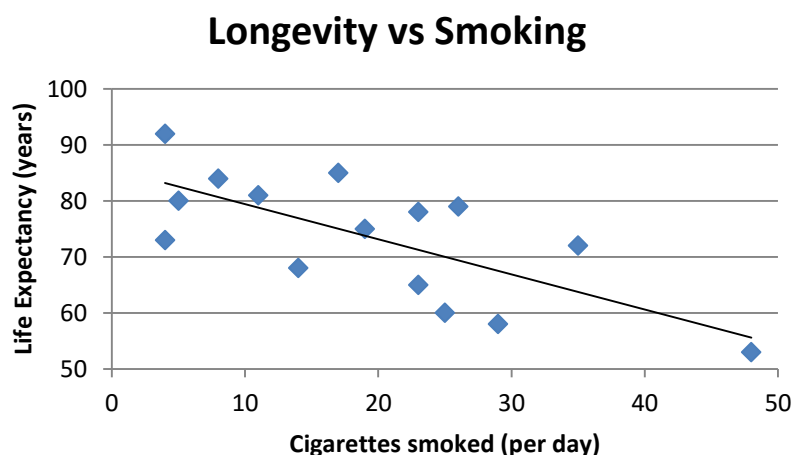
(zdroj: MS Excel)

## Práce s programem

Doplňek Real Statistics bude v následujícím odstavci představen na konkrétním příkladu. Obecně bude představena funkce z oblasti lineární regrese, regresní analýza – speciálně půjde o využití regresní přímky k předpovědi budoucích hodnot.

Příklad řeší počet vykouřených cigaret za den (Smoking) ve vztahu k očekávané délce života (Longevity). Obrázek 27 ukazuje rozložení hodnot a jimi proloženou regresní přímku, která znázorňuje nepřímou úměru mezi veličinami.

Obrázek 27 - Graf vztahu počtu vykouřených cigaret denně a očekávanou délkou života



(zdroj: Real Statistics Examples Part 2 - <http://www.real-statistics.com>)

Spojnice trendu v grafu je vyjádřena rovnicí  $y = -0,6282x + 85,72$ , jejíž parametry jsou vypočteny za pomoci funkcí covariance a variability z původních hodnot. Původní hodnoty z příkladu jsou uvedeny v tabulce v Příloze 11.

cov(x,y)	-90,9467	=COVAR(X1:X15,Y1:Y15)
varp(x)	144,7733	=VARP(X1:X15)
slope b	-0,6282	= cov(x,y)/ varp(x)
intercept a	85,72042	=PRŮMĚR(Y1:Y15)-slope b*PRŮMĚR(X1:X15)

Kontrolní predikce ( $\hat{y}$ ) jsou obsaženy v tabulce v Příloze 12, která obsahuje i výpočet chyby předpovědi (Error – e). Predikce je v příkladu popsána třemi různými způsoby výpočtu, které umožňuje Real Statistics. Všechny možnosti generují stejné hodnoty, a proto jsou plně nahraditelné.

Tabulka 15 - Hodnoty předpovědí očekávaného věku dožití

x	$\hat{y}$
4	83,20762
24	70,64361
44	58,0796

(zdroj: Real Statistics Examples Part 2 - <http://www.real-statistics.com>)

a) Predikce pomocí rovnice trendu

Metoda využívá rovnici trendu ve formě  $y = a + b \cdot x$ ; kde jsou parametry určeny pomocí metody popsané výše.



### b) Predikce pomocí funkce FORECAST

Funkce forecast se používá ve formě: =FORECAST(x; pole\_y; pole\_x). Pro tento příklad bude zadána pro x zkoumaná hodnota počtu vykouřených cigaret, pro pole\_y bude matice původních hodnot očekávaného věku dožití a pro pole\_x bude matice původních hodnot počtu vykouřených cigaret za den.

### c) Predikce pomocí funkce TREND

Funkce trend se používá ve formě: =LINTREND(pole\_y;[pole\_x];[nová\_x];[b]), kde pole\_y je množina známých hodnot y. Jedná se o volitelnou množinu hodnot x, které mohou být známy ve vztahu  $y = mx + b$ ; tedy matice původních hodnot počtu vykouřených cigaret za den. Nová\_x je povinný argument. Jedná se o nové hodnoty x, pro něž se mají určit odpovídající hodnoty y, tj. výsledek funkce LINTREND. Funkce TREND vrátí množinu hodnot y, které na přímce odpovídají hodnotám nová\_x, po proložení přímky množinou bodů, zadaných souřadnicemi pole\_y a pole\_x (metoda nejmenších čtverců).

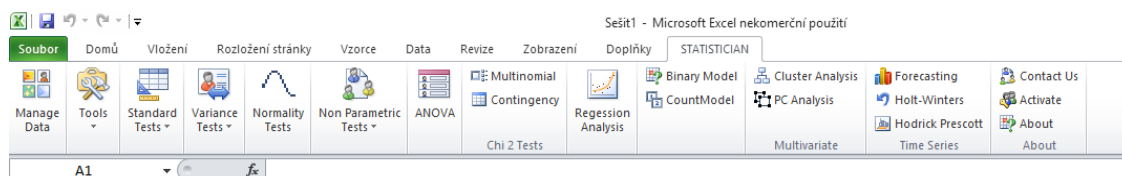
## 3.5.2. Statistician

Statistician je komplexní a snadno použitelný doplněk pro MS Excel 2007 a novější verze. Byl vyvinut v prostředí Microsoft Visual Studio. Tento software provádí vysoce kvalitní statistickou analýzu, která se opírá o funkce použitelné prostřednictvím MS Excel. [32]. Program je k dispozici ve dvou verzích. První je Statistician Lite, který je dostupný zdarma ke stažení i s komplexní dokumentací a ukázkovými daty. Lite verze bude v následujících řádcích blíže popsána. Druhá verze se nazývá Statistician Standard, který už je zpoplatněn a obsahuje mnoho dalších statistických modelů a funkcí.

### Instalace doplňku

Po instalaci Statistician Lite z oficiálních stránek produktu, se rozšíří nabídka MS Excel o samostatnou záložku s názvem STATISTICIAN (viz Obrázek 28).

Obrázek 28 - Záložka doplňku STATISTICIAN v programu MS Excel



(zdroj: Microsoft Excel)

### Nabídka funkcí

Statistický software Statistician obsahuje funkce pro analýzu dat, například vícerozměrnou regresi, logistickou regresi, Poissonovu regresi, základní t-test. [32]

Statistician rozšiřuje funkce Excelu o statistické analýzy, které jsou v záložce rozděleny do několika skupin.

První nabídka Tools obsahuje deskriptivní statistiku souboru, jako je průměr, medián, kvartily, rozptyl, šikmost. Dále nabízí výpočet kovariance, korelace a autokorelace. Volba Statistical Tables umožňuje výstup podstatných statistických tabulek (normální rozdělení, studentovo t rozdělení, poissonovo rozdělení a dalších) přímo do listu MS Excel. Možnost tool obsahuje ještě funkce pro práci s daty, a to ve formě uspořádání, hodnocení, vzorkování a optimalizaci.

Součástí Statistician záložky jsou statistické testy, mezi které patří Standard Tests, Variance Tests, Normality Tests a Non Parametric Tests. Tyto testy jsou založeny na definici hypotéz a jedná se například o test o shodě průměrů, test o shodě rozptylů, Kolmogorov-Smirnov test, Chi kvadrát test dobré shody.

Dalšími funkcemi, které jsou součástí doplňku Statistician, jsou jednofaktorová a dvoufaktorová analýza rozptylu neboli ANOVA, regresní analýza, Count modely a binární modely.

Time series sekce uživateli nabízí množství formulářů pro analýzu chování všech druhů časových řad. V nastavení je možnost vybrat z devíti metod očištění časové řady od náhodných vlivů (jednoduchá metoda průměrování, čtyři metody klouzavých průměrů a čtyři trendové metody). Statistician modeluje dva základní modely časových řad, které jsou založeny na klasické dekompozici, jsou jimi aditivní a multiplikatívni model. Další možností je výpočet sezónních indexů pomocí již dříve vyhlazených časových řad. Volba Time series osahuje tři podsekce: Forecasting, Holt-Winters a Hodrick-Prescott filter.[33]

## **Práce s programem**

Statistician je od ostatních statistických doplňků jedinečným v tom, že umožňuje uložit soubor dat a provádět s ním více analýz bez opětovného nahrávání souboru. Práce s programem při vytváření analýz je velmi intuitivní a skládá se v souhrnu ze tří základních kroků.

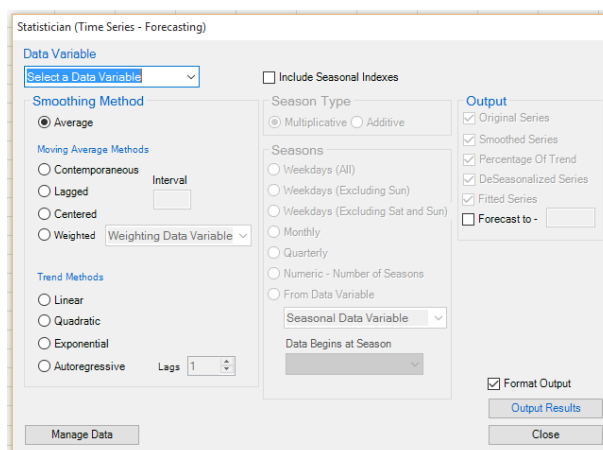
### a) Uspořádání dat

Veškeré statistické analýzy jsou v programu Statistician prováděny na základě zadaných proměnných. Řady proměnných jsou získány přímo z původně uložených hodnot v MS Excel a lze je definovat, odstraňovat, exportovat a přejmenovávat pomocí formuláře Manage Data v záložce Statistician. Při vyplňování formuláře je nutno dát pozor na zaškrťávací políčko s názvem „Data Variable Names in First Row“, které automaticky zvolí první vybrané pole za název. Pro dokončení stačí potvrdit tlačítkem Close a při uložení dokumentu zůstanou proměnné v nabídce Manage Data.

b) Vstupní formulář

Chce-li uživatel použít libovolnou funkci na kteroukoliv předem uloženou proměnnou, stačí vybrat vhodný příkaz ze záložky Statistician. Pokud chce uživatel vytvořit prognózu budoucích hodnot, zvolí metodu Forecasting a zobrazí se formulář (Obrázek 29).

Obrázek 29 - Vstupní formulář pro tvorbu předpovědi v programu Statistician



(zdroj: statistický software Statistician)

Uživatel může ve vstupním formuláři vybrat z již navolených proměnných a zvolit další funkce použité metody včetně prezentace výstupu.

c) Výstup

Výstupní hodnoty jsou poslány do zvoleného místa v listu MS Excel, ovšem za předpokladu, že je vybrána volba „Output results“. Výstup je pak uveden v čitelné a uživatelsky přívětivé podobě. Pro příklad jsou použita data počtu uchazečů o zaměstnání v okrese Česká Lípa a výstupní hodnoty metody Forecasting uvádí Tabulka 16. Ve vstupním formuláři byla pro vyhlazení časové řady zvolena metoda centrovaných klouzavých průměrů řádu 12, pro zjištění trendu byla z nabídky vybrána autoregresivní metoda a pro výstup bylo zvoleno 24 hodnot předpovědi. Pro názornost Tabulka 16 uvádí výsledné hodnoty předpovědi  $Y_t$  pomocí metody Forecasting.

Tabulka 16 - Hodnoty předpovědí s použitím funkce Forecasting programu Statistician

t	Rok/Měsíc	$Y_t$
<b>78</b>	2015/6	4367
<b>79</b>	2015/7	4593
<b>80</b>	2015/8	4510
<b>81</b>	2015/9	4807
<b>82</b>	2015/10	4883
<b>83</b>	2015/11	5154
<b>84</b>	2015/12	5555

(zdroj: statistický software Statistician)

Zbylé výstupní hodnoty uvádí Příloha 13, kde je vidět vstupní volba parametrů, koeficienty autoregresivní metody, sezónní indexy jak pro vyhlazenou, tak nevyhlazenou časovou řadu, a v neposlední řadě tabulka obsahující originální hodnoty v porovnání s hodnotami vyhlazené časové řady, procentuálním vyjádřením trendové složky, hodnotami časové řady očištěné od sezónních vlivů a hodnotami předpovědi.[32]

### 3.5.3. Unistat

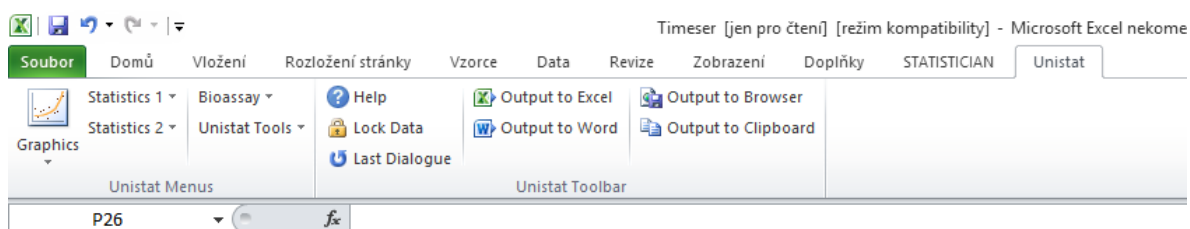
Doplňek Unistat verze 6.5 je kompatibilní s MS Excel 1997-2013 a při stažení programu je k dispozici jak samostatná verze, tak i verze doplňku pro Excel. Práce s doplňkem je velmi intuitivní, stačí pouze vybrat data a poté funkci. Výstup je standardně vytvořen v novém listu MS Excel, Unistat však nabízí možnosti exportovat výstup do MS Word nebo HTML formátu pro web. [38]

Unistat je však zdarma dostupný pouze v demo verzi, kde lze použít vestavěná příkladová data. Pro použití statistických analýz vlastních dat je potřeba licenční verze doplňku Unistat.

#### Instalace doplňku

Po instalaci doplňku jsou vytvořeny dvě ikony Unistat 6.5 for Excel a Unistat 6.5.[38] Spuštění doplňku se aktivuje dvojklikem na ikonu, následně je doplněk spuštěn přímo v prostředí MS Excel, jak zobrazuje Obrázek 30. Při spuštění MS Excel vyžaduje povolení maker.

Obrázek 30 - Záložka doplňku Unistat 6.5 for Excel v editoru MS Excel



(zdroj: Microsoft Excel)

#### Nabídka funkcí

Verze doplňku Unistat Standard Edition obsahuje funkci grafického zobrazení dat ve formě 2D grafů, 3D grafů, histogramů, grafů deskriptivní statistiky, diagramů a různých funkčních grafů. Další funkce v programu Unistat jsou z oblasti deskriptivní statistiky a distribuce. Unistat z této oblasti nabízí především komplexní statistickou analýzu dat, intervaly spolehlivosti a kvartily.

Ze statistických testů jsou součástí programu Unistat testy korelace, parametrické i neparametrické testy, testy dobré shody. V doplňku nechybí statistické funkce regrese a ANOVA. Pro analýzu časových řad je dostupná metoda Box-Jenkins ARIMA, metody

předpovědí a vyhlazení časových řad (klouzavé průměry, Brownovo exponenciální vyrovnání, Winterovo aditivní sezónní vyrovnání) a kontrola kvality předpovědí.[38]

### Práce s programem

Pro příklad práce s programem Unistat budou použita data z měsíčních průměrů hotelových pokojů pro roky 1973-1986, která jsou součástí příkladů k programu Unistat. Data nalezneme ve složce Unistat tools/Example files pod názvem TIMESER a budou představena v použití s ARIMA modelem.

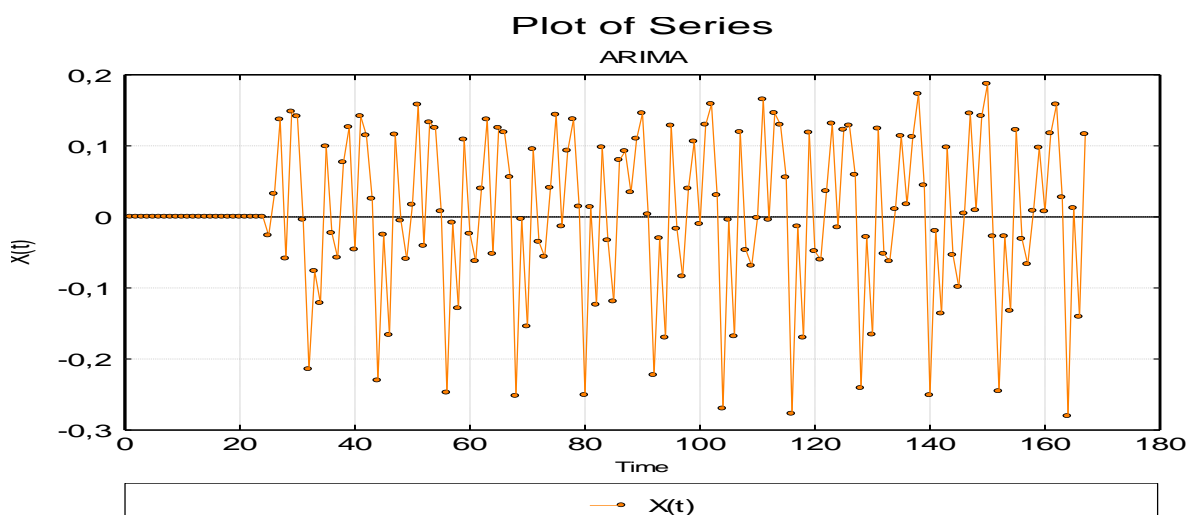
V první fázi je potřeba označit analyzovaná data, z nabídky Unistat/Statistics 2 zvolit funkci ARIMA a ve vstupním formuláři vybrat Room Averages jako vstupní proměnnou. V dialogovém okně Differencing Input Options se zvolí následující hodnoty:

- 0 Nonseasonal Differencing
- 1 Seasonal Differencing
- 12 Seasonal Period
- 0 Lambda (0 Log(X), else  $X^\Lambda$ )
- 0 Offset (minimum 480)
- 25 Maximum Lag

Po volbě výstupu se zvolené tabulky a grafy zobrazí v samostatném listu. Graf hodnot časové řady vypadá následovně (viz Obrázek 31).

[37]

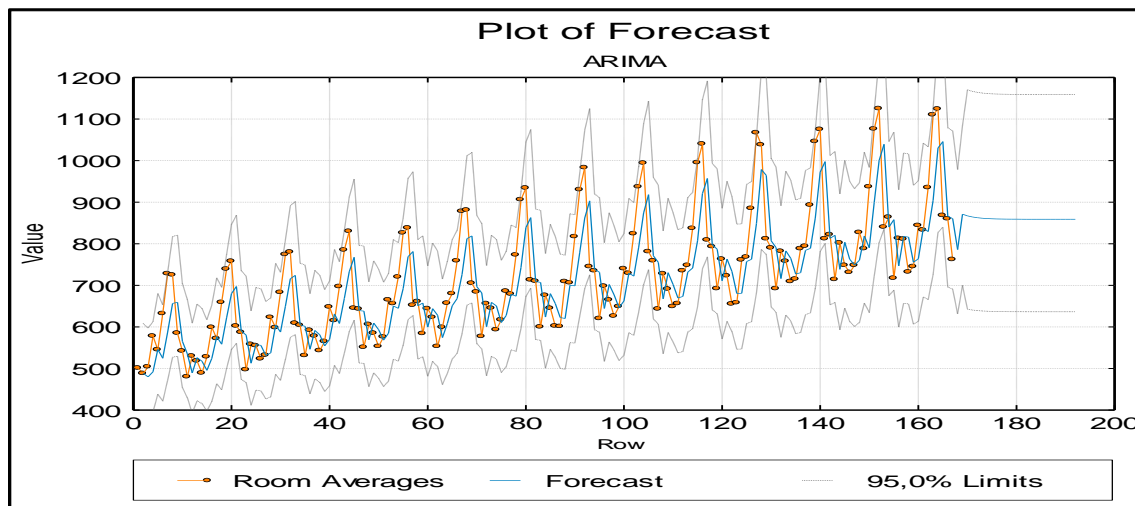
Obrázek 31 - Grafické znázornění časové řady z příkladů doplňku Unistat



(zdroj: Example Files z programu Unistat 6.5 for Excel)

Pro vytvoření předpovědí se po volbě parametrů ve formuláři Differencing Input Option zvolí možnost Fit Model a Forecasting. Výstup této funkce obsahuje, kromě tabulky předpovědí a předpovědních intervalů, graf předpovědí s 95% předpovědními intervaly.

Obrázek 32 - Graf předpovědí pro časovou řadu z příkladů doplňku Unistat



(zdroj: Example Files z programu Unistat 6.5 for Excel)

## 4. Shrnutí výsledků

V následujících kapitolách jsou shrnuty dosažené výsledky jednotlivých metod předpovědi analyzované časové řady. Tyto výsledky jsou následně porovnány dvojím způsobem. V první části jsou srovnávány s již dostupnými skutečnými hodnotami počtu nezaměstnaných. V části druhé jsou srovnány s odhady předpokládaného vývoje nezaměstnanosti, které provádí úřad práce v České Lípě.

### Dosažené výsledky

Pro porovnání výsledků jsou vybrány získané předpovědi pro období od června roku 2015 do června roku 2016, tedy roční předpověď počtu uchazečů o zaměstnání. V tomto období jsou dostupné a hlavně srovnatelné hodnoty všech metod předpovědi včetně Brownova dvojitěho exponenciálního vyrovnání, které umožňuje pouze krátkodobé předpovědi. Zkoumané metody předpovědi a legendu k následujícímu grafu srovnání popisuje Obrázek 33.

Obrázek 33 - Legenda grafu porovnání předpovědí

- A - Klasický model dekompozice
- B - Brownovo dvojitě exponenciální vyrovnání
- C - ARIMA model v programu Statgraphics Centurion XVI
- D - ARIMA model v programu IBM SPSS Statistics

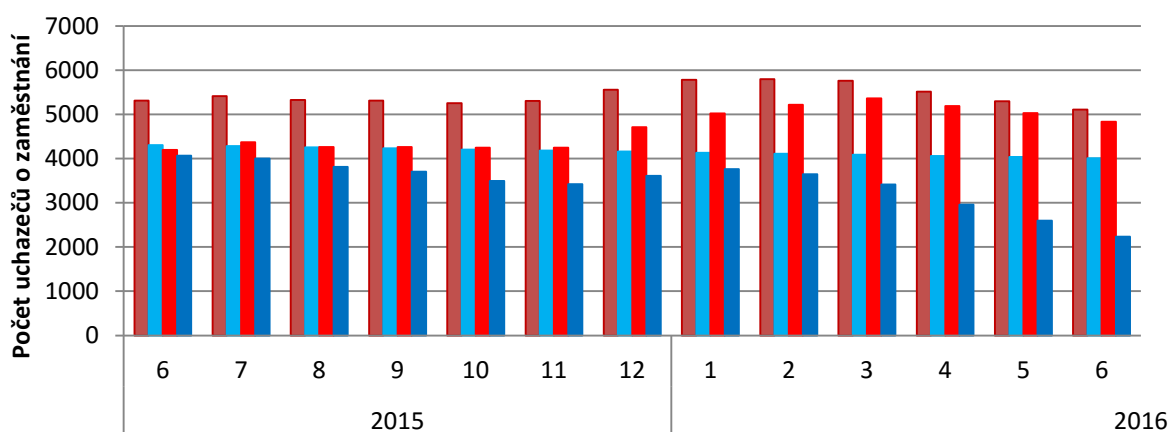
*(zdroj: vlastní tvorba v programu MS Excel)*

Obrázek 34 znázorňuje grafické srovnání použitých metod pro predikci budoucích hodnot časové řady počtu uchazečů o zaměstnání. Predikované hodnoty se z větší části nevychylují z rozmezí hodnot 4000 až 6000 uchazečů o zaměstnání, což jsou reálné hodnoty sledovaného ukazatele.

Z grafu je možné pozorovat, že Brownovo dvojitě exponenciální vyrovnání nemá v předpovědi zakomponovanou sezónní složku, jelikož je založen na průběhu a následné extrapolaci trendové složky. Zbylé tři metody obsahují viditelný sezónní vliv zvyšující se v zimních měsících, což koresponduje s vlastnostmi původní časové řady.

Dále je v grafu možné pozorovat prudký pokles předpovědních hodnot modelu ARIMA provedeném v programu IBM SPSS Statistics. Jak již bylo řečeno v odstavci 3.4.5 ARIMA model v programu IBM SPSS Statistics je vhodné použití tohoto modelu pouze pro krátkodobější předpovědi (ideálně 6 období). Přesné hodnoty výsledných předpovědí jsou uvedeny v Příloze 14.

Obrázek 34 - Grafické porovnání předpovědních metod



(zdroj: vlastní tvorba v programu MS Excel)

### Srovnání výsledků s již dostupnými skutečnými hodnotami

V této části práce jsou již z portálu mpsv.cz dostupné skutečné hodnoty počtu uchazečů o zaměstnání. Tyto hodnoty nebyly při počátečním sběru dat ještě známy, a byly tudíž získávány pomocí vybraných předpovědních metod. Nyní lze předpovědi  $\hat{Y}_t$ , které jsou uvedeny v Příloze 14, porovnat se skutečnými statistickými hodnotami  $Y_t$ . Hodnoty skutečného počtu uchazečů o zaměstnání uvádí Tabulka 17.

Tabulka 17 – Dostupné hodnoty počtu nezaměstnaných

t	Rok	Měsíc	$Y_t$
78	2015	6	4 115
79		7	4 002
80		8	3 910
81		9	3 800

(zdroj: měsíční zprávy MPSV)

Srovnání skutečných a vypočtených hodnot nejlépe vyjadřují odchylky, které uvádí Tabulka 18. Tabulka obsahuje odchylky hodnot předpovědních metod od reálných hodnot pro období červen až září roku 2015. Například při porovnání červnových hodnot vychází nejlépe použití ARIMA modelu v programu IBM SPSS Statistics. Tyto hodnoty se nejméně liší od skutečných statistických hodnot. Velmi dobře vychází i ARIMA model v programu Statgraphics a Brownovo dvojitě vyrovnání.



Tabulka 18 – Hodnoty odchylek předpovědí od již dostupných skutečných hodnot

odchylky $\hat{Y}_t$ od $Y_t$				
	Klasický model dekompozice	Brownovo dvojité exponenciální vyrovnání	ARIMA (Statgraphics)	ARIMA (IBM SPSS Statistics)
6/2015	1 195	190	78	-50
7/2015	1 407	278	364	-4
8/2015	1 412	346	350	-98
9/2015	1 507	431	459	-99

(zdroj: vlastní výpočty v programu MS Excel)

### Srovnání výsledků s dosavadní formou předpovědi

Dosavadní forma předpovědi, kterou vykonává analytický pracovník kontaktního pracoviště úřadu práce v České Lípě, uvádí optimističtější hodnotu 5 300 uchazečů o zaměstnání. Odchylka této hodnoty je 1 185, což se dá považovat za velice nepřesnou předpověď. Navržené formy předpovědi B, C a D jsou v tomto ohledu kvalitnější.

Tabulka 19 – Srovnání očekávaného stavu se skutečnou hodnotou

Počet uchazečů o zaměstnání		
Očekávaný stav		Stav k 30.6.2015
Optimističtější varianta	Pesimističtější varianta	
5 300	6 000	4 115

(zdroj: kontaktní pracoviště Česká Lípa)

Tabulka 20 srovnává výsledné hodnoty podle velikosti odchylky od skutečného počtu uchazečů o zaměstnání. Z tabulky vyplývá, že použité modely předpovědi jsou ve třech případech přesnější než forma předpovědi používaná na Úřadu práce v České Lípě. Nejpřesnější předpovědi jsou získané pomocí ARIMA modelů vytvořených ve specializovaných softwarech.

Tabulka 20 – Finální porovnání výsledků

Porovnání výsledků	
ARIMA (IBM SPSS Statistics)	-50
ARIMA (Statgraphics)	78
Brownovo dvojité exp. vyrovnání	190
Úřad práce	1185
Klasický model dekompozice	1195

(zdroj: výsledky šetření)

## 5. Závěr

Obsahem diplomové práce je především hledání alternativních forem předpovědi budoucích hodnot časové řady. Tvorbě předpovědí však předchází důkladná analýza dat a průběhu časové řady. V této oblasti zkoumání práce bylo nalezeno několik problémů, které bylo potřeba vyřešit. Před samotnou analýzou časové řady došlo k několika nesrovnalostem při sběru dat. V první řadě bylo nutné jasně definovat data, se kterými se dále pracuje. Podmínkou analytické práce s časovými řadami je totiž srovnatelnost hodnot.

Původní volbou pro analýzu byly hodnoty podílu nezaměstnanosti, se kterými pracuje i Úřad práce v České Lípě. Z důvodu změny metodiky během zkoumaného období bylo nakonec přistoupeno k práci s počty uchazečů o zaměstnání. Po definování této proměnné byla seskupena data z měsíčních statistik MPSV do jedné časové řady. Vypočtené předpovědi v hlavní části diplomové práce dokazují velké nepřesnosti v hodnotách předpovědí konstruovaných na Úřadě práce v České Lípě. Tento fakt poukazuje na to, že je dosavadní forma předpovědi velice nekvalitní a neudává tak hodnotné informace o budoucích počtech uchazečů o zaměstnání. Ze čtyř předpovědních metod dopadly tři lépe než metoda používaná Úřadem práce, z nichž nejlepší výsledků dosáhla metoda s použitím ARIMA modelu zkonstruovaná v programu IBM SPSS Statistics.

Metoda byla doporučena Úřadu práce v České Lípě jako doplněk k dosavadní metodě. Mělo by být však zohledněno použití této metody na kratší časové období, nežli tomu bylo doposud. Návrhem tedy bude tvorba prediktivních modelů, pomocí metody ARIMA v programu IBM SPSS Statistics, v horizontu šesti měsíců. Při tvorbě predikcí se pak pracovníci úřadu mohou opřít o krátkodobé výsledky metody. Kombinací kvalitativní a kvantitativní metody predikce mohou analýzy trhu práce dosáhnout přesnějších výsledků.

Statistické doplňky neboli add-ins jsou vhodnou volbou pro uživatele, kteří vyžadují sofistikovanější metody statistické analýzy než je MS Excel, ale zároveň nechtějí nebo nemohou vynaložit tolik financí a času na učení se práci s novým softwarem. K práci se statistickými doplňky je obvykle vyžadována pouze základní znalost MS Excel a anglického jazyka.

Hlavním přínosem diplomové práce je bližší seznámení se s podmínkami analýzy časové řady a hlubší poznání metod předpovědi budoucích hodnot. Obsah práce vysvětluje zacházení s časovou řadou, která má být podkladem pro predikci nadcházejících hodnot.

Dalším přínosem je prozkoumání dostupných statistických softwarů na trhu a představení funkcí sloužících k analýze časových řad a tvorbě předpovědi. Informace obsažené v práci mohou být využity pro změny v tvorbě predikcí nejen pro úřady práce, ale mohou být inspirací pro statistické analýzy soukromých firem, neziskových organizací podporujících nezaměstnanost, ministerstva práce a sociálních věcí, státní správy, podnikatelů a jiných úřadů.

## 6. Seznam použité literatury

### Knižní zdroje:

- [1] ARLT, Josef; ARLTOVÁ, Markéta. *Finanční časové řady (Vlastnosti, metody modelování, příklady a aplikace)*. 1. vyd. Praha: Grada. 2003. 220 s. ISBN 80-247-0330-0.
- [2] ARLT, Josef; ARLTOVÁ, Markéta; RUBLÍKOVÁ, Eva. *Analýza ekonomických časových řad s příklady*. Praha: Oeconomica. 2004.
- [3] ARLT, Josef; ŠKUTHANOVÁ, Markéta. *Úvod do problematiky sezónního očišťování ekonomických časových řad*. Acta oeconomica pragensia 3: (1), str. 15-23, 1995. ISSN 0572-3043.
- [4] ARLTOVÁ, Markéta; MATUŠŮ, Martin. *STATGRAPHICS Plus for Windows (Zadávání úloh)*. 2. vyd. Praha: Oeconomica. 2004. 91 s. ISBN 80-245-0731-5.
- [5] ARLT, Josef; ARLTOVÁ, Markéta. *Ekonomické časové řady: vlastnosti, metody modelování, příklady a aplikace*. 1. vyd. Praha: Grada Publishing. 2007. 288 s. ISBN 978-80-247-1319-9.
- [6] BORŮVKOVÁ, Jana; HORÁČKOVÁ, Petra; HANÁČEK, Miroslav. *Statistika v SPSS*. Vysoká škola polytechnická v Jihlavě: Katedra matematiky. 2014. 108 s.
- [7] BUDÍKOVÁ, Marie; KRÁLOVÁ, Maria; MAROŠ, Bohumil. *Průvodce základními statistickými metodami*. 1. vyd. Praha: Grada Publishing a.s. 2010. 272 s. ISBN 978-80-247-3243-5.
- [8] CIPRA, Tomáš. *Analýza časových řad s aplikacemi v ekonomii*. 1. vyd. Praha: SNTL. 1986. 248 s. ISBN 99-00-00157-X.
- [9] CIPRA, Tomáš. *Finanční ekonometrie*. 1. vyd. Praha: Ekopress. 2008. 538 s. ISBN 978-80-86929-43-9.
- [10] DANEL, Roman. *Predikce časové řady pomocí autoregresního modelu*. Brno: Marek Konečný. 2005. 1 elektronický optický disk.
- [11] DUFEK, Jaroslav. *Ekonometrie*. 1. vyd. Brno: MZLU. 2003. 136 s. ISBN 80-7157-654-9.
- [12] GURINOVÁ, Kateřina; VALENTOVÁ, Vladimíra. *Základy práce s programem Statgraphics Centurion XVI*. Liberec: TUL. 2012. 118 s. ISBN 978-80-7372-810-6.
- [13] HANČLOVÁ, Jana; TVRDÝ, Lubor. *Úvod do analýzy časových řad*. Ostrava: Fakulta ekonomická, katedra matematických metod v ekonomice a katedra regionální ekonomiky, VŠB-TU Ostrava, 2003.
- [14] CHATFIELD, Christopher. *The Analysis of Time Series: an Introduction*. 6. vyd. CRC Press. 2013. 352 s. ISBN 978-02-034-9168-3.
- [15] KOZÁK, Josef; HINDLS, Richard; ARLT, Josef. *Úvod do analýzy ekonomických časových řad*. 1. vyd. Praha: VŠE. 1994. 208 s. ISBN 80-707-9760-6.

- [16] KROPÁČ, Jiří. *Statistika B: jednorozměrné a dvourozměrné datové soubory, regresní analýza, časové řady*. 2. dopl. vyd. Brno: Vysoké učení technické v Brně, Fakulta podnikatelská. 2009. 145 s. ISBN 978-80-214-3984-9.
- [17] KŘIVÝ, Ivan. *Analýza časových řad*. Ostrava: OU. 2006. 77 s.
- [18] MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven C.; HYNDMAN, Rob J. *Forecasting: Methods and Applications*. 3. vyd. New York: John Wiley & Sons. 1998. 656 s. ISBN 978-0-471-53233-9.
- [19] PACÁKOVÁ, Viera a kolektiv. *Štatistika pre ekonómov*. 1. vyd. Bratislava: Iura Edition. 2003. 358 s. ISBN 80-89047-74.
- [20] RAJTROVÁ, Barbora. *Analýza vývoje na trhu práce v okrese Česká Lípa za rok 2014 – Roční zpráva*. Úřad práce: Kontaktní pracoviště Česká Lípa. 2015. 79 s.
- [21] ROJÍČEK, Marek a kolektiv. *Příčiny a odstraňování nesrovnatelnosti časových řad*. Statistika: Ekonomicko-statistický časopis. Č. 2. 2009. 95-111.
- [22] VESELÝ, Vítězslav a kolektiv. *Úvod do časových řad*. 2004.
- [23] YAFFEE, Robert A.; MCGEE, Monnie. *An Introduction to Time Series Analysis and Forecasting: With Applications of SAS® and SPSS®*. Academic Press. 2000, 528 s. ISBN 978-0-080-47870-8.

#### Internetové zdroje:

- [24] ČESKÝ STATISTICKÝ ÚŘAD. Metodika ukazatele vyhledána dne 15. 6. 2015. In: Metodika ukazatelů (nejdůležitější údaje) [online]. Datum poslední aktualizace: 10. 11. 2014. Dostupné z: < [https://www.czso.cz/csu/xl/metodika-ukazatelu-nejdulezitejsi-udaje#podil\\_nezamestnanychosob\\_na\\_obyvatelstvu\\_ve\\_veku\\_15\\_64\\_let](https://www.czso.cz/csu/xl/metodika-ukazatelu-nejdulezitejsi-udaje#podil_nezamestnanychosob_na_obyvatelstvu_ve_veku_15_64_let)>
- [25] ČESKÝ STATISTICKÝ ÚŘAD. Hrubý domácí produkt (HDP) - Metodika vyhledána dne 13. 9. 2015. In: HDP, národní účty [online]. Datum poslední aktualizace: 19. 2. 2015. Dostupné z: <[https://www.czso.cz/csu/czso/hruby\\_domaci\\_produk\\_t\\_-hdp-](https://www.czso.cz/csu/czso/hruby_domaci_produk_t_-hdp-)>
- [26] IBM Corporation. IBM SPSS Forecasting [online-pdf]. In: 2015-01 [cit. 2015-08-23]. Dostupné z: <<http://www-03.ibm.com/software/products/cs/>>
- [27] IBM Corporation. IBM SPSS Statistics Base [online-pdf]. In: 2015-01 [cit. 2015-08-23]. Dostupné z: <<http://www-03.ibm.com/software/products/cs/spss-stats-base>>
- [28] MINISTERSTVO PRÁCE A SOCIÁLNÍCH VĚCÍ. Statistická data vyhledána dne 11. 6. 2015. In: Vývoj nezaměstnanosti od července 2004 [online]. ©2015. Dostupné z: <[http://portal.mpsv.cz/sz/stat/nz/vyvoj\\_od\\_072004](http://portal.mpsv.cz/sz/stat/nz/vyvoj_od_072004)>
- [29] MINISTERSTVO PRÁCE A SOCIÁLNÍCH VĚCÍ. Vyhledáno dne 15. 7. 2015. In: Upozornění na změnu metodiky [online]. ©2015. Dostupné z: <[http://portal.mpsv.cz/sz/stat/nz/zmena\\_metodiky](http://portal.mpsv.cz/sz/stat/nz/zmena_metodiky)>

- [30] POLHEMUS, N., W. Statistical software development [online]. Vyhledáno dne 2. 7. 2015. Dostupné z: <<http://www.neilpolhemus.com/statistical-software-development/statgraphics-centurion-for-windows/>>
- [31] STATGRAPHICS – oficiální stránky produktu. Time series analysis and forecasting [online]. Vyhledáno dne 7.7.2015. Dostupné z: <<http://www.statgraphics.com/time-series-analysis-and-forecasting>>
- [32] STATISTICIAN – oficiální stránky produktu. Statistician add-in [online]. In: 2009-2014 [cit. 2015-10-15]. Dostupné z: <http://www.statisticianaddin.com>
- [33] STATISTICIAN MANUAL. User manual for Statistician (Lite), Statistician (Standard). In: 2012 [cit. 2015-10-20]. Dostupné z: <<http://www.statisticianaddin.com/downloads/Statistician-Manual.pdf>>
- [34] STATPOINT TECHNOLOGIES, INC. Automatic Forecasting [online]. In: 2010-06-12 [cit. 2015-08-16]. Dostupné z: <<http://www.statgraphics.com/time-series-analysis-and-forecasting>>
- [35] STATPOINT TECHNOLOGIES, INC. Forecasting [online]. In: 2010-06-12 [cit. 2015-08-17]. Dostupné z: <<http://www.statgraphics.com/time-series-analysis-and-forecasting>>
- [36] ŠETEK, David. Ekonometrie1 - Analýza časových řad. In: Youtube [online]. Zveřejněno 3. 9. 2012 [vid. 2015-06-07]. Kanál uživatele ekospace. Dostupné z: <<https://www.youtube.com/watch?v=ptCFVltNE7E>>
- [37] UNISTAT STATISTICS SOFTWARE. Users Guide. In: [www.unistat.com](http://www.unistat.com) [online]. 2015 [cit. 2015-10-21]. Dostupné z: <<https://www.unistat.com/guide/time-series-analysis-box-jenkins-arima/>>
- [38] UNISTAT LTD. Unistat statistics software [online]. 2015 [cit. 2015-10-21]. Dostupné z: <<https://www.unistat.com/>>
- [39] ZAIONTZ, Charles. Real Statistics using Excel [online]. In: 2013-2015 [cit. 2015-10-14]. Dostupné z: <<http://www.real-statistics.com>>

## 7. Seznamy

### Seznam obrázků

Obrázek 1 - Úvodní dialog funkce Explore .....	9
Obrázek 2 - Metody výběrových kritérií používaných pro výběr vhodného modelu .....	13
Obrázek 3 - Metody výběrových kritérií používaných pro výběr vhodného modelu .....	13
Obrázek 4 - Rozhraní při volbě metody vyrovnání časové řady .....	26
Obrázek 5 - Nabídka Tables and Graphs pro volbu výstupů .....	26
Obrázek 6 - Vstupní panel při volbě Automatic Forecasting .....	27
Obrázek 7 - Vstupní panel k metodě Explore (popisná statistika) .....	29
Obrázek 8 - Poloha města Česká Lípa v rámci okresu .....	32
Obrázek 9 - Spojnicový graf časové řady počtu uchazečů o zaměstnání .....	35
Obrázek 10 - Graf sezónních hodnot měsíční časové řady počtu uchazečů o zaměstnání .....	36
Obrázek 11 - Graf sezónních indexů pro $Y_t$ získaných pomocí multiplikační metody dekompozice .....	37
Obrázek 12 - Grafické srovnání metod sezónní dekompozice .....	38
Obrázek 13 - Graf vyrovnání časové řady hodnot $Y_t$ metodou klouzavých průměrů (CMA) .....	42
Obrázek 14 - Hodnoty sezónního faktoru pro $Y_t$ získaných pomocí aditivní metody .....	43
Obrázek 15 - Předpovědi klasického modelu dekompozice a původní hodnoty .....	44
Obrázek 16 - Graf kontrolních predikcí Brownova dvojitého exponenciálního vyrovnání .....	45
Obrázek 17 - Predikce pomocí metody Brownova dvojitého exponenciálního vyrovnání .....	45
Obrázek 18 - Periodogram řady reziduí (řada očištěná od lineárního trendu) .....	47
Obrázek 19 - Původní časová řada s předpověďmi a předpovědními intervaly .....	49
Obrázek 20 - Graf předpovědi pro období 6/2015 - 6/2017 .....	50
Obrázek 21 - Histogram počtu uchazečů o zaměstnání .....	53
Obrázek 22 - Krabicový graf počtu uchazečů o zaměstnání .....	53
Obrázek 23 - Spojitý graf počtu uchazečů o zaměstnání .....	54
Obrázek 24 - Původní časová řada s předpověďmi a předpovědními intervaly .....	55
Obrázek 25 - Doplněk Real Statistics v programu MS Excel .....	56
Obrázek 26 - Nabídka funkcí doplňku Real Statistics v MS Excel .....	57
Obrázek 27 - Graf vztahu počtu vykouřených cigaret denně a očekávanou délkou života .....	58
Obrázek 28 - Záložka doplňku STATISTICIAN v programu MS Excel .....	59
Obrázek 29 - Vstupní formulář pro tvorbu předpovědi v programu Statistician .....	61
Obrázek 30 - Záložka doplňku Unistat 6.5 for Excel v editoru MS Excel .....	62
Obrázek 31 - Grafické znázornění časové řady z příkladů doplňku Unistat .....	63
Obrázek 32 - Graf předpovědi pro časovou řadu z příkladů doplňku Unistat .....	64
Obrázek 33 - Legenda grafu porovnání předpovědí .....	65
Obrázek 34 - Grafické porovnání předpovědních metod .....	66

## Seznam tabulek

<i>Tabulka 1 - Význam termínů ARIMA modelu</i>	21
<i>Tabulka 2 - Zjišťování nenáhodných vzorů v časové řadě pomocí metody Run Charts</i>	25
<i>Tabulka 3 - Výpočet popisné statistiky pro Počet uchazečů o zaměstnání</i>	29
<i>Tabulka 4 – Srovnání sezónních indexů pro <math>Y_t</math></i>	37
<i>Tabulka 5 - Roční průměrné počty uchazečů o zaměstnání v okrese Česká Lípa</i>	39
<i>Tabulka 6 - Meziroční koeficienty růstu s měsíčním krokem pro měsíc květen</i>	40
<i>Tabulka 7 - Předpokládaný vývoj nezaměstnanosti v dalším období</i>	41
<i>Tabulka 8 - Koeficienty získané pomocí jednoduché lineární regrese</i>	43
<i>Tabulka 9 - Bodové odhady počtu uchazečů o zaměstnání pro období 6/2015 - 6/2016</i>	46
<i>Tabulka 10 - Hodnoty periodogramu pro <math>Y_t</math></i>	47
<i>Tabulka 11 - Srovnání modelů podle výběrových kritérií</i>	51
<i>Tabulka 12 - Popisná statistika</i>	51
<i>Tabulka 13 - Extrémní hodnoty časové řady</i>	52
<i>Tabulka 14 - Volba modelu pro prognózu počtu uchazečů o zaměstnání</i>	55
<i>Tabulka 15 - Hodnoty předpovědí očekávaného věku dožití</i>	58
<i>Tabulka 16 - Hodnoty předpovědí s použitím funkce Forecasting programu Statistician</i>	61
<i>Tabulka 17 – Dostupné hodnoty počtu nezaměstnaných</i>	66
<i>Tabulka 18 – Hodnoty odchylek předpovědí od již dostupných skutečných hodnot</i>	67
<i>Tabulka 19 – Srovnání očekávaného stavu se skutečnou hodnotou</i>	67
<i>Tabulka 20 – Finální porovnání výsledků</i>	67

## 8. Přílohy

**Příloha 1:** Časová řada počtu uchazečů o zaměstnání (Yt), kde t je počet pozorování.

t	Rok	Měsíc	Yt
1	2009	1	5 389
2		2	6 104
3		3	6 701
4		4	6 965
5		5	7 149
6		6	7 116
7		7	7 197
8		8	6 884
9		9	6 941
10		10	6 947
11		11	6 923
12		12	7 151
13	2010	1	7 460
14		2	7 415
15		3	7 353
16		4	6 964
17		5	6 711
18		6	6 416
19		7	6 486
20		8	6 325
21		9	6 165
22		10	6 131
23		11	6 079
24		12	6 750
25	2011	1	6 741
26		2	6 687
27		3	6 601
28		4	6 345
29		5	5 976
30		6	5 810
31		7	6 036
32		8	5 972
33		9	5 992
34		10	5 910
35		11	5 895
36		12	6 225
37	2012	1	6 493
38		2	6 609
39		3	6 507
40		4	6 161
41		5	6 077
42		6	5 948
43		7	5 984
44		8	6 014
45		9	6 032
46		10	5 960
47		11	6 022
48		12	6 340
49	2013	1	6 762
50		2	6 859
51		3	6 869
52		4	6 675
53		5	6 423
54		6	6 130
55		7	6 169
56		8	6 122
57		9	6 193
58		10	6 058
59		11	6 151
60		12	6 484
61	2014	1	6 888
62		2	6 802
63		3	6 759
64		4	6 287
65		5	5 977
66		6	5 773
67		7	5 680
68		8	5 552
69		9	5 455
70		10	5 268
71		11	5 256
72		12	5 461
73	2015	1	5 448
74		2	5 341
75		3	5 009
76		4	4 628
77		5	4 358

(zdroj: portal.mpsv.cz, vlastní úprava)



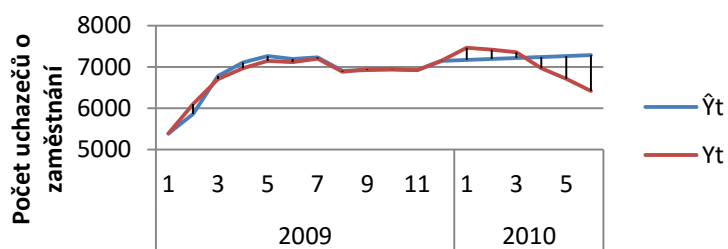
**Příloha 2:** Sezónní dekompozice – srovnání aditivní ( $Y_t - S_t$ ) a multiplikativní ( $Y_t / S_t$ ) metody očištění od sezónní složky.

<b>t</b>	<b>Rok</b>	<b>Měsíc</b>	<b>Yt/St</b>	<b>Yt-St</b>
1	2009	1	5055	4941
2		2	5693	5627
3		3	6270	6257
4		4	6785	6826
5		5	7230	7236
6		6	7436	7392
7		7	7359	7301
8		8	7132	7097
9		9	7187	7152
10		10	7250	7214
11		11	7132	7146
12		12	7008	7006
13	2010	1	6998	7012
14		2	6916	6938
15		3	6880	6909
16		4	6784	6825
17		5	6787	6798
18		6	6704	6692
19		7	6632	6590
20		8	6552	6538
21		9	6384	6376
22		10	6398	6398
23		11	6263	6302
24		12	6615	6605
25	2011	1	6323	6293
26		2	6237	6210
27		3	6176	6157
28		4	6181	6206
29		5	6043	6063
30		6	6071	6086
31		7	6172	6140
32		8	6187	6185
33		9	6205	6203
34		10	6168	6177
35		11	6073	6118
36		12	6100	6080
37	2012	1	6090	6045
38		2	6164	6132
39		3	6088	6063
40		4	6002	6022
41		5	6146	6164
42		6	6215	6224
43		7	6119	6088
44		8	6230	6227
45		9	6246	6243
46		10	6220	6227
47		11	6204	6245
48		12	6213	6195
49	2013	1	6343	6314
50		2	6397	6382
51		3	6427	6425
52		4	6503	6536
53		5	6495	6510
54		6	6405	6406
55		7	6308	6273
56		8	6342	6335
57		9	6413	6404
58		10	6322	6325
59		11	6337	6374
60		12	6354	6339
61	2014	1	6461	6440
62		2	6344	6325
63		3	6324	6315
64		4	6125	6148
65		5	6044	6064
66		6	6032	6049
67		7	5808	5784
68		8	5752	5765
69		9	5649	5666
70		10	5498	5535
71		11	5415	5479
72		12	5352	5316
73	2015	1	5110	5000
74		2	4981	4864
75		3	4687	4565
76		4	4509	4489
77		5	4407	4445

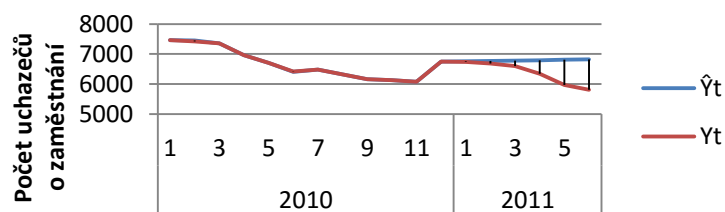
(zdroj: vlastní úprava, výpočty v Excelu)

Příloha 3: Predikce pomocí Brownova dvojitého exponenciálního vyrovnání

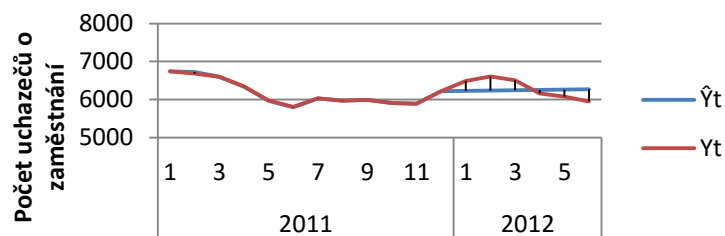
Rok	Měsíc	$Y_t$	$\hat{Y}_t$	Error
2010	1	7460	<b>7169</b>	291
	2	7415	<b>7192</b>	223
	3	7353	<b>7214</b>	139
	4	6964	<b>7236</b>	-272
	5	6711	<b>7259</b>	-548
	6	6416	<b>7281</b>	-865



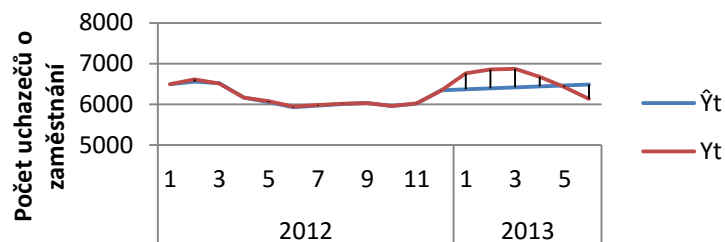
2011	1	6741	<b>6761</b>	-20
	2	6687	<b>6772</b>	-85
	3	6601	<b>6784</b>	-183
	4	6345	<b>6796</b>	-451
	5	5976	<b>6808</b>	-832
	6	5810	<b>6820</b>	-1010



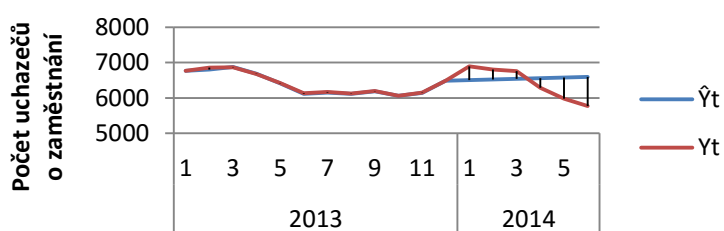
2012	1	6493	<b>6232</b>	261
	2	6609	<b>6240</b>	369
	3	6507	<b>6248</b>	259
	4	6161	<b>6256</b>	-95
	5	6077	<b>6264</b>	-187
	6	5948	<b>6273</b>	-325



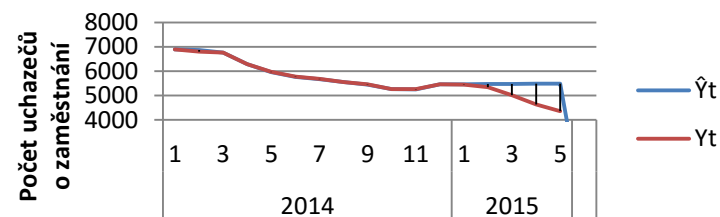
2013	1	6762	<b>6366</b>	396
	2	6859	<b>6390</b>	469
	3	6869	<b>6414</b>	455
	4	6675	<b>6437</b>	238
	5	6423	<b>6461</b>	-38
	6	6130	<b>6485</b>	-355



2014	1	6888	<b>6504</b>	384
	2	6802	<b>6523</b>	279
	3	6759	<b>6541</b>	218
	4	6287	<b>6559</b>	-272
	5	5977	<b>6578</b>	-601
	6	5773	<b>6596</b>	-823



2015	1	5448	<b>5464</b>	-16
	2	5341	<b>5469</b>	-128
	3	5009	<b>5475</b>	-466
	4	4628	<b>5480</b>	-852
	5	4358	<b>5485</b>	-1127



(zdroj: vlastní tvorba)

**Příloha 4:** Tabulka předpovědí pro  $Y_t$  pomocí modelu ARIMA (0,1,0)x(0,1,2)<sub>12</sub>

<i>Period</i>	<i>Forecast</i>	<i>Lower 95,0%</i>	<i>Upper 95,0%</i>
		<i>Limit</i>	<i>Limit</i>
6.15	4192,52	3810,57	4574,48
7.15	4365,73	3825,56	4905,9
8.15	4259,78	3598,21	4921,35
9.15	4259,47	3495,55	5023,38
10.15	4247,41	3393,33	5101,5
11.15	4242,32	3306,71	5177,92
12.15	4708,53	3697,97	5719,1
1.16	5020,69	3940,35	6101,03
2.16	5213,26	4067,39	6359,13
3.16	5360,45	4152,59	6568,3
4.16	5186,96	3920,15	6453,77
5.16	5030,88	3707,74	6354,02
6.16	4829,32	3491,12	6167,53
7.16	4883,87	3530,77	6236,97
8.16	4807,37	3439,53	6175,21
9.16	4796,68	3414,26	6179,1
10.16	4696,28	3299,44	6093,13
11.16	4718,5	3307,38	6129,63
12.16	5059,21	3633,94	6484,47
1.17	5308,8	3869,54	6748,07
2.17	5319,97	3866,85	6773,1
3.17	5235,61	3768,76	6702,47
4.17	4911,21	3430,75	6391,67
5.17	4668,92	3174,97	6162,86

(zdroj: statistický program Statgraphics)

**Příloha 5:** Výpočet elementárních charakteristik pomocí Statgraphics

Rok	Mesic	Y <sub>t</sub>	Absolutni_prirustky	Koeficienty_rustu	$\bar{k}_t$	$\overline{\Delta Y}_t$
2009	1	5389			0,998	-13,566
	2	6104	715	1,133		
	3	6701	597	1,098		
	4	6965	264	1,039		
	5	7149	184	1,026		
	6	7116	-33	0,995		
	7	7197	81	1,011		
	8	6884	-313	0,957		
	9	6941	57	1,008		
	10	6947	6	1,001		
	11	6923	-24	0,997		
	12	7151	228	1,033		
2010	1	7460	309	1,043		
	2	7415	-45	0,994		
	3	7353	-62	0,992		
	4	6964	-389	0,947		
	5	6711	-253	0,964		
	6	6416	-295	0,956		
	7	6486	70	1,011		
	8	6325	-161	0,975		
	9	6165	-160	0,975		
	10	6131	-34	0,994		
	11	6079	-52	0,992		
	12	6750	671	1,110		
2011	1	6741	-9	0,999		
	2	6687	-54	0,992		
	3	6601	-86	0,987		
	4	6345	-256	0,961		
	5	5976	-369	0,942		
	6	5810	-166	0,972		
	7	6036	226	1,039		
	8	5972	-64	0,989		
	9	5992	20	1,003		
	10	5910	-82	0,986		
	11	5895	-15	0,997		
	12	6225	330	1,056		
2012	1	6493	268	1,043		
	2	6609	116	1,018		
	3	6507	-102	0,985		
	4	6161	-346	0,947		
	5	6077	-84	0,986		
	6	5948	-129	0,979		

	7	5984	36	1,006
	8	6014	30	1,005
	9	6032	18	1,003
	10	5960	-72	0,988
	11	6022	62	1,010
	12	6340	318	1,053
2013	1	6762	422	1,067
	2	6859	97	1,014
	3	6869	10	1,001
	4	6675	-194	0,972
	5	6423	-252	0,962
	6	6130	-293	0,954
	7	6169	39	1,006
	8	6122	-47	0,992
	9	6193	71	1,012
	10	6058	-135	0,978
	11	6151	93	1,015
	12	6484	333	1,054
2014	1	6888	404	1,062
	2	6802	-86	0,988
	3	6759	-43	0,994
	4	6287	-472	0,930
	5	5977	-310	0,951
	6	5773	-204	0,966
	7	5680	-93	0,984
	8	5552	-128	0,977
	9	5455	-97	0,983
	10	5268	-187	0,966
	11	5256	-12	0,998
	12	5461	205	1,039
2015	1	5448	-13	0,998
	2	5341	-107	0,980
	3	5009	-332	0,938
	4	4628	-381	0,924
	5	4358	-270	0,942

(zdroj: statistický program Statgraphics)

**Příloha 6:** Předpovědi pro 24 období pomocí modelu klasické dekompozice

<b>t</b>	<b>Rok</b>	<b>Měsíc</b>	<b>Ÿt</b>
78	2015	6	<b>5 310</b>
79		7	<b>5 409</b>
80		8	<b>5 322</b>
81		9	<b>5 307</b>
82		10	<b>5 249</b>
83		11	<b>5 300</b>
84		12	<b>5 554</b>
85	2016	1	<b>5 783</b>
86		2	<b>5 797</b>
87		3	<b>5 760</b>
88		4	<b>5 514</b>
89		5	<b>5 294</b>
90		6	<b>5 107</b>
91		7	<b>5 201</b>
92		8	<b>5 117</b>
93		9	<b>5 102</b>
94		10	<b>5 045</b>
95		11	<b>5 094</b>
96		12	<b>5 337</b>
97	2017	1	<b>5 557</b>
98		2	<b>5 570</b>
99		3	<b>5 533</b>
100		4	<b>5 296</b>
101		5	<b>5 084</b>

(zdroj: vlastní tvorba)

**Příloha 7:** Tabulka porovnávající modely podle výběrových kritérií

<i>Model</i>	<i>RMSE</i>	<i>MAE</i>	<i>MAPE</i>	<i>ME</i>	<i>MPE</i>	<i>AIC</i>	<i>HQC</i>	<i>SBIC</i>
(A)	195,568	113,428	1,73297	17,2776	0,25084	10,8903	11,0355	11,2583
(B)	196,387	115,505	1,76387	1,19179	0,00162559	10,9294	11,0878	11,3308
(C)	462,255	311,756	4,79641	-4,06119	-0,41497	12,6415	12,7999	13,0429
(D)	434,936	283,566	4,41368	-4,13566	-0,35952	12,5504	12,722	12,9853
(E)	434,219	249,363	3,88876	-4,69314	-0,349435	12,5779	12,7627	13,0462
(F)	434,196	283,117	4,39397	7,39847	-0,184781	12,547	12,7186	12,9819
(G)	433,679	281,567	4,37069	7,33822	-0,18443	12,5446	12,7162	12,9795
(H)	243,606	130,232	1,96312	15,666	0,196893	11,3603	11,5187	11,7618
(I)	195,578	111,688	1,70638	17,0125	0,246985	10,9211	11,0795	11,3226
(J)	214,779	129,467	2,03007	11,2073	0,227677	11,1084	11,2668	11,5099
(K)	184,44	121,272	1,83829	-29,0841	-0,412471	10,8346	11,0062	11,2695
(L)	250,839	160,973	2,50711	3,85371	0,123672	11,4189	11,5772	11,8203
(M)	233,205	179,236	2,8036	-78,2862	-1,27915	10,9961	11,0357	11,0965
(N)	113,222	89,7033	1,39709	-14,3921	-0,2288	9,52023	9,54663	9,58714
(O)	114,764	89,9762	1,41695	10,2386	0,166363	9,57807	9,61766	9,67842
(P)	113,165	90,2707	1,42375	-13,961	-0,213189	9,58078	9,63357	9,71458
(Q)	112,703	87,7197	1,3804	-22,7184	-0,358913	9,60336	9,66935	9,77062
(R)	112,951	87,6031	1,38985	23,7908	0,382036	9,60775	9,67375	9,77501

Vysvětlivky modelů:

- (A) Random walk
- (B) Random walk with drift = 16,0509
- (C) Constant mean = 6440,13
- (D) Linear trend = 13021,4 + -8,88155 t
- (E) Quadratic trend = 147020, + -370,784 t + 0,244199 t<sup>2</sup>
- (F) Exponential trend = exp(9,70302 + -0,00126152 t)
- (G) S-curve trend = exp(7,82569 + 697,979 /t)
- (H) Simple moving average of 2 terms
- (I) Simple exponential smoothing with alpha = 0,9999
- (J) Brown's linear exp. smoothing with alpha = 0,9334
- (K) Holt's linear exp. smoothing with alpha = 0,9999 and beta = 0,0995
- (L) Brown's quadratic exp. smoothing with alpha = 0,7162
- (M) Winters' exp. smoothing with alpha = 0,9481, beta = 0,0157, gamma = 0,9999
- (N) ARIMA(0,1,0)x(0,1,2)<sub>12</sub>
- (O) ARIMA(1,2,0)x(0,1,2)<sub>12</sub>
- (P) ARIMA(0,1,2)x(0,1,2)<sub>12</sub>
- (Q) ARIMA(2,1,1)x(0,1,2)<sub>12</sub>
- (R) ARIMA(1,2,0)x(2,1,2)<sub>12</sub>

(zdroj: výstup ze statistického programu Statgraphics)

Příloha 8: Periodogram pro Yt

				<i>Cumulative</i>	<i>Integrated</i>
<i>i</i>	<i>Frequency</i>	<i>Period</i>	<i>Ordinate</i>	<i>Sum</i>	<i>Periodogram</i>
0	0,0		1,4524E-23	1,4524E-23	4,91923E-31
1	0,012987	77,0	3,25344E6	3,25344E6	0,110193
2	0,025974	38,5	1,16201E7	1,48735E7	0,503761
3	0,038961	25,6667	4,69986E6	1,95734E7	0,662944
4	0,0519481	19,25	1,69915E6	2,12725E7	0,720494
5	0,0649351	15,4	559331,	2,18319E7	0,739438
6	0,0779221	12,8333	765399,	2,25973E7	0,765362
7	0,0909091	11,0	3,46685E6	2,60641E7	0,882783
8	0,103896	9,625	1,44273E6	2,75069E7	0,931648
9	0,116883	8,55556	207365,	2,77142E7	0,938671
10	0,12987	7,7	150160,	2,78644E7	0,943757
11	0,142857	7,0	138727,	2,80031E7	0,948456
12	0,155844	6,41667	40698,6	2,80438E7	0,949834
13	0,168831	5,92308	1,0048E6	2,90486E7	0,983866
14	0,181818	5,5	57178,9	2,91058E7	0,985803
15	0,194805	5,13333	10678,3	2,91165E7	0,986165
16	0,207792	4,8125	424,804	2,91169E7	0,986179
17	0,220779	4,52941	44024,1	2,91609E7	0,98767
18	0,233766	4,27778	46767,4	2,92077E7	0,989254
19	0,246753	4,05263	2363,79	2,921E7	0,989334
20	0,25974	3,85	19930,2	2,923E7	0,990009
21	0,272727	3,66667	50742,2	2,92807E7	0,991728
22	0,285714	3,5	6217,4	2,92869E7	0,991938
23	0,298701	3,34783	7072,99	2,9294E7	0,992178
24	0,311688	3,20833	6020,85	2,93E7	0,992382
25	0,324675	3,08	176,131	2,93002E7	0,992388
26	0,337662	2,96154	60010,6	2,93602E7	0,99442
27	0,350649	2,85185	23443,6	2,93837E7	0,995214
28	0,363636	2,75	180,926	2,93838E7	0,995221
29	0,376623	2,65517	6165,93	2,939E7	0,995429
30	0,38961	2,56667	4729,57	2,93947E7	0,99559
31	0,402597	2,48387	14582,3	2,94093E7	0,996084
32	0,415584	2,40625	15812,3	2,94251E7	0,996619
33	0,428571	2,33333	16994,8	2,94421E7	0,997195
34	0,441558	2,26471	11518,9	2,94536E7	0,997585
35	0,454545	2,2	2770,76	2,94564E7	0,997679
36	0,467532	2,13889	12360,3	2,94688E7	0,998097
37	0,480519	2,08108	41697,7	2,95105E7	0,99951
38	0,493506	2,02632	14478,4	2,95249E7	1,0

(zdroj: výstup ze statistického programu Statgraphics)



**Příloha 9:** Tabulka dat sezónní dekompozice (multiplikativní metoda) pro  $Y_t$ 

<i>Period</i>	<i>Data</i>	<i>Trend-Cycle</i>	<i>Seasonality</i>	<i>Irregular</i>	<i>Seasonally Adjusted</i>
1.9	5 389				5 053,66
2.9	6 104				5 697,72
3.9	6 701				6 284,49
4.9	6 965				6 838,44
5.9	7 149				7 273,30
6.9	7 116				7 466,12
7.9	7 197	6 875,21	104,68	106,89	7 348,56
8.9	6 884	7 016,13	98,12	101,87	7 147,28
9.9	6 941	7 097,92	97,79	101,49	7 203,41
10.9	6 947	7 125,04	97,50	102,16	7 279,14
11.9	6 923	7 106,75	97,41	101,30	7 198,75
12.9	7 151	7 059,33	101,30	99,37	7 015,14
1.10	7 460	7 000,54	106,56	99,93	6 995,80
2.10	7 415	6 947,63	106,73	99,62	6 921,46
3.10	7 353	6 892,00	106,69	100,06	6 895,97
4.10	6 964	6 825,67	102,03	100,17	6 837,46
5.10	6 711	6 756,50	99,33	101,05	6 827,69
6.10	6 416	6 704,63	95,70	100,40	6 731,68
7.10	6 486	6 657,96	97,42	99,47	6 622,59
8.10	6 325	6 597,67	95,87	99,53	6 566,90
9.10	6 165	6 536,00	94,32	97,89	6 398,08
10.10	6 131	6 478,88	94,63	99,16	6 424,13
11.10	6 079	6 422,46	94,65	98,42	6 321,13
12.10	6 750	6 366,58	106,02	104,01	6 621,76
1.11	6 741	6 322,58	106,62	99,98	6 321,54
2.11	6 687	6 289,13	106,33	99,25	6 241,91
3.11	6 601	6 267,21	105,33	98,78	6 190,71
4.11	6 345	6 250,79	101,51	99,66	6 229,71
5.11	5 976	6 233,92	95,86	97,53	6 079,91
6.11	5 810	6 204,38	93,64	98,25	6 095,87
7.11	6 036	6 172,17	97,79	99,85	6 163,11
8.11	5 972	6 158,58	96,97	100,68	6 200,40
9.11	5 992	6 151,42	97,41	101,09	6 218,54
10.11	5 910	6 139,83	96,26	100,86	6 192,56
11.11	5 895	6 136,38	96,07	99,89	6 129,80
12.11	6 225	6 146,33	101,28	99,36	6 106,74
1.12	6 493	6 149,92	105,58	99,01	6 088,97
2.12	6 609	6 149,50	107,47	100,32	6 169,10
3.12	6 507	6 152,92	105,76	99,18	6 102,55

4.12	6 161	6 156,67	100,07	98,25	6 049,05
5.12	6 077	6 164,04	98,59	100,30	6 182,66
6.12	5 948	6 174,13	96,34	101,08	6 240,66
7.12	5 984	6 190,13	96,67	98,71	6 110,01
8.12	6 014	6 211,75	96,82	100,52	6 244,00
9.12	6 032	6 237,25	96,71	100,37	6 260,05
10.12	5 960	6 273,75	95,00	99,54	6 244,95
11.12	6 022	6 309,58	95,44	99,24	6 261,86
12.12	6 340	6 331,58	100,13	98,23	6 219,55
1.13	6 762	6 346,88	106,54	99,91	6 341,23
2.13	6 859	6 359,08	107,86	100,68	6 402,46
3.13	6 869	6 370,29	107,83	101,13	6 442,05
4.13	6 675	6 381,08	104,61	102,71	6 553,71
5.13	6 423	6 390,54	100,51	102,26	6 534,68
6.13	6 130	6 401,92	95,75	100,46	6 431,61
7.13	6 169	6 413,17	96,19	98,22	6 298,91
8.13	6 122	6 416,04	95,42	99,07	6 356,13
9.13	6 193	6 409,08	96,63	100,28	6 427,13
10.13	6 058	6 388,33	94,83	99,36	6 347,64
11.13	6 151	6 353,58	96,81	100,67	6 396,00
12.13	6 484	6 320,13	102,59	100,64	6 360,82
1.14	6 888	6 284,88	109,60	102,78	6 459,39
2.14	6 802	6 240,75	108,99	101,74	6 349,26
3.14	6 759	6 186,25	109,26	102,47	6 338,89
4.14	6 287	6 122,58	102,69	100,82	6 172,76
5.14	5 977	6 052,38	98,75	100,47	6 080,92
6.14	5 773	5 972,46	96,66	101,42	6 057,05
7.14	5 680	5 869,83	96,77	98,80	5 799,61
8.14	5 552	5 748,96	96,57	100,27	5 764,33
9.14	5 455	5 615,17	97,15	100,82	5 661,23
10.14	5 268	5 473,13	96,25	100,85	5 519,87
11.14	5 256	5 336,54	98,49	102,41	5 465,35
12.14	5 461				5 357,25
1.15	5 448				5 108,99
2.15	5 341				4 985,50
3.15	5 009				4 697,66
4.15	4 628				4 543,91
5.15	4 358				4 433,77

(zdroj: výstup ze statistického programu Statgraphics)

**Příloha 10:** Prognóza vývoje počtu uchazečů o zaměstnání v období červen/2015 – květen/2017

**Forecast**

Model		Jun 2015	Jul 2015	Aug 2015	Sep 2015	Oct 2015
Počet uchazečů o zaměstnání-Model_1	Forecast	4065	3998	3812	3701	3496
	UCL	4335	4515	4611	4817	4960
	LCL	3795	3482	3014	2586	2032

Model		Nov 2015	Dec 2015	Jan 2016	Feb 2016	Mar 2016
Počet uchazečů o zaměstnání-Model_1	Forecast	3423	3612	3764	3642	3411
	UCL	5265	5859	6442	6776	7023
	LCL	1581	1364	1085	508	-201

Model		Apr 2016	May 2016	Jun 2016	Jul 2016	Aug 2016
Počet uchazečů o zaměstnání-Model_1	Forecast	2959	2592	2232	2099	1846
	UCL	7071	7226	7446	7931	8331
	LCL	-1154	-2041	-2982	-3734	-4640

Model		Sep 2016	Oct 2016	Nov 2016	Dec 2016	Jan 2017
Počet uchazečů o zaměstnání-Model_1	Forecast	1667	1395	1255	1377	1462
	UCL	8838	9282	9887	10780	11661
	LCL	-5504	-6492	-7376	-8026	-8738

Model		Feb 2017	Mar 2017	Apr 2017	May 2017
Počet uchazečů o zaměstnání-Model_1	Forecast	1273	975	456	22
	UCL	12294	12856	13189	13641
	LCL	-9749	-10907	-12277	-13596

(zdroj: výstup programu IBM SPSS Statgraphics)

**Příloha 11:** Hodnoty očekávaného věku dožití v závislosti na počtu vykouřených cigaret denně

Cig (x)	Life Exp (y)
5	80
23	78
25	60
48	53
17	85
8	84
4	73
26	79
11	81
19	75
14	68
35	72
29	58
4	92
23	65

(zdroj: <http://www.real-statistics.com>)

**Příloha 12:** Vytvoření hodnot předpovědí pro očekávaný věk dožití a hodnoty chyby předpovědi

Data		Prediction	
Cig (x)	Life Exp (y)	Life Exp ( $\hat{y}$ )	Error (e)
5	80	82,58	-2,58
23	78	71,27	6,73
25	60	70,02	-10,02
48	53	55,57	-2,57
17	85	75,04	9,96
8	84	80,69	3,31
4	73	83,21	-10,21
26	79	69,39	9,61
11	81	78,81	2,19
19	75	73,78	1,22
14	68	76,93	-8,93
35	72	63,73	8,27
29	58	67,50	-9,50
4	92	83,21	8,79
23	65	71,27	-6,27

(zdroj: <http://www.real-statistics.com>)

Příloha 13: Výstup funkce Forecasting doplňku Statistician Lite

<b>Time Series:</b> Forecasting				
<b>Data Variable:</b> Yt				
<b>Smoothing Method:</b> Autoregressive Model (Lags = 2)				
<b>Seasonal Index Type:</b> Multiplicative				
<b>Sample Range:</b> 1 to 77				
<b>Number of Observations:</b> 77				
<b>Dependent Variable:</b> Yt				
	<b>Coefficient</b>	<b>Std Error</b>	<b>t-Ratio</b>	<b>p-Value</b>
<b>Constant</b>	404,3302784	256,7021	1,575095	0,119618612
<b>AR(1)</b>	1,463994049	0,09435	15,5167	1,1556E-24
<b>AR(2)</b>	-0,530921675	0,098341	-5,3988	8,20589E-07
<b>R squared:</b> 0,9067				
<b>Seasonal Indexes:</b>	<b>Unadj'd</b>	<b>Adj'd</b>		
<b>January</b>	1,0100	1,0088		
<b>February</b>	0,9856	0,9844		
<b>March</b>	0,9957	0,9945		
<b>April</b>	0,9689	0,9677		
<b>May</b>	0,9962	0,9950		
<b>June</b>	0,9894	0,9882		
<b>July</b>	1,0278	1,0265		
<b>August</b>	0,9798	0,9786		
<b>September</b>	1,0078	1,0066		
<b>October</b>	0,9879	0,9867		
<b>November</b>	1,0090	1,0078		
<b>December</b>	1,0565	1,0552		
<b>MAD:</b> 109,4358217				
<b>RMSE:</b> 142,0548419				
<b>MAPE:</b> 1,74 %				

	<b>Original</b>	<b>Smoothed Series</b>	<b>% of Trend</b>	<b>Season</b>	<b>DeSeas'd Series</b>	<b>Fitted Series</b>
<b>1</b>	5389			January	5342	5436
<b>2</b>	6104			February	6201	6009
<b>3</b>	6701	6479	103,42%	March	6738	6664
<b>4</b>	6965	6974	99,87%	April	7197	6740
<b>5</b>	7149	7043	101,50%	May	7185	7114
<b>6</b>	7116	7173	99,21%	June	7201	7032
<b>7</b>	7197	7027	102,43%	July	7011	7388
<b>8</b>	6884	7163	96,11%	August	7035	6737
<b>9</b>	6941	6661	104,20%	September	6896	6987
<b>10</b>	6947	6911	100,52%	October	7041	6854

11	6923	6890	100,49%	November	6869	6977
12	7151	6851	104,38%	December	6777	7546
13	7460	7198	103,64%	January	7395	7525
14	7415	7529	98,48%	February	7532	7299
15	7353	7299	100,74%	March	7394	7313
16	6964	7232	96,29%	April	7196	6739
17	6711	6696	100,23%	May	6744	6678
18	6416	6532	98,23%	June	6493	6340
19	6486	6234	104,04%	July	6319	6658
20	6325	6493	97,41%	August	6463	6190
21	6165	6221	99,11%	September	6125	6206
22	6131	6072	100,98%	October	6214	6049
23	6079	6107	99,54%	November	6032	6126
24	6750	6049	111,59%	December	6397	7123
25	6741	7059	95,50%	January	6682	6800
26	6687	6689	99,96%	February	6793	6583
27	6601	6615	99,79%	March	6637	6565
28	6345	6518	97,35%	April	6557	6140
29	5976	6189	96,56%	May	6006	5946
30	5810	5784	100,44%	June	5879	5741
31	6036	5737	105,21%	July	5880	6196
32	5972	6156	97,01%	August	6103	5844
33	5992	5943	100,83%	September	5953	6031
34	5910	6006	98,40%	October	5990	5831
35	5895	5875	100,34%	November	5849	5941
36	6225	5897	105,57%	December	5899	6569
37	6493	6388	101,65%	January	6436	6550
38	6609	6605	100,06%	February	6714	6506
39	6507	6633	98,11%	March	6543	6471
40	6161	6422	95,94%	April	6366	5962
41	6077	5969	101,80%	May	6107	6047
42	5948	6030	98,64%	June	6019	5878
43	5984	5886	101,67%	July	5830	6143
44	6014	6007	100,12%	August	6146	5885
45	6032	6032	100,00%	September	5993	6072
46	5960	6042	98,64%	October	6041	5880
47	6022	5927	101,60%	November	5975	6069
48	6340	6056	104,69%	December	6008	6690
49	6762	6489	104,21%	January	6703	6821
50	6859	6938	98,86%	February	6968	6752
51	6869	6856	100,19%	March	6907	6831
52	6675	6819	97,89%	April	6898	6460
53	6423	6530	98,37%	May	6455	6391
54	6130	6264	97,87%	June	6203	6058
55	6169	5969	103,36%	July	6010	6332
56	6122	6181	99,04%	August	6256	5991

57	6193	6092	101,66%	September	6153	6234
58	6058	6221	97,39%	October	6140	5977
59	6151	5985	102,77%	November	6103	6199
60	6484	6193	104,70%	December	6145	6842
61	6888	6631	103,87%	January	6828	6948
62	6802	7046	96,54%	February	6910	6696
63	6759	6705	100,80%	March	6796	6722
64	6287	6688	94,00%	April	6497	6084
65	5977	6020	99,29%	May	6007	5947
66	5773	5817	99,25%	June	5842	5705
67	5680	5683	99,95%	July	5533	5831
68	5552	5655	98,18%	August	5674	5433
69	5455	5517	98,88%	September	5419	5491
70	5268	5443	96,79%	October	5339	5198
71	5256	5220	100,68%	November	5215	5297
72	5461	5302	103,00%	December	5175	5763
73	5448	5609	97,14%	January	5401	5496
74	5341	5481	97,45%	February	5426	5258
75	5009	5331	93,96%	March	5037	4981
76	4628	4902	94,41%	April	4782	4479
77	4358	4520	96,41%	May	4380	4336
78		4420		June		4367
79		4475		July		4593
80		4609		August		4510
81		4776		September		4807
82		4949		October		4883
83		5114		November		5154
84		5264		December		5555
85		5396		January		5443
86		5509		February		5423
87		5604		March		5573
88		5684		April		5501
89		5751		May		5722
90		5805		June		5737
91		5850		July		6005
92		5887		August		5761
93		5917		September		5956
94		5941		October		5861
95		5960		November		6007
96		5976		December		6306
97		5989		January		6041
98		5999		February		5906
99		6007		March		5974
100		6014		April		5820
101		6019		May		5990

(zdroj: výstup doplňku Statistician Lite)



Příloha 14: Porovnání výsledných předpovědí

t	Rok	Měsíc	$\hat{Y}_t$			
			Klasický model dekompozice	Brown	ARIMA (Statgraphics)	ARIMA (IBM SPSS Statistics)
78	2015	6	5 310	4 305	4 193	4 065
79		7	5 409	4 280	4 366	3 998
80		8	5 322	4 256	4 260	3 812
81		9	5 307	4 231	4 259	3 701
82		10	5 249	4 206	4 247	3 496
83		11	5 300	4 182	4 242	3 423
84		12	5 554	4 157	4 709	3 612
85	2016	1	5 783	4 132	5 021	3 764
86		2	5 797	4 107	5 213	3 642
87		3	5 760	4 083	5 360	3 411
88		4	5 514	4 058	5 187	2 959
89		5	5 294	4 033	5 031	2 592
90		6	5 107	4 009	4 829	2 232

(zdroj: vlastní tvorba)

Podklad pro zadání DIPLOMOVÉ práce studenta

PŘEDKLÁDÁ:	ADRESA	OSOBNÍ ČÍSLO
Zajícová Tereza	Klášteří 2192, Česká Lípa	1100934

**TÉMA ČESKY:**

Návrh metod prognózy nezaměstnanosti v okrese Česká Lípa

**TÉMA ANGLICKY:**

Suggestion of forecasting methods for unemployment in Česká Lípa region.

**VEDOUcí PRÁCE:**

prof. RNDr. Hana Skalská, CSc. - KIKM

**ZÁSADY PRO VYPRACOVÁNÍ:**

použití Box-Jenkinsovy metodologie - ARIMA modely

**SEZNAM DOPORUČENÉ LITERATURY:**

MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven C.; HYNDMAN, Rob J. Forecasting: Methods and Applications. 3. vyd. New York: John Wiley & Sons. 1998. 656 s. ISBN 978-0-471-53233-9.

Podpis studenta:

*Tereza Zajícová*

Datum:

*4.3.2014*

Podpis vedoucího práce:

*Hana Skalská*

Datum:

*4.3.2014*