Sezónní dekompozice časových řad

Zadání příkladu:

Pro časovou řadu počtu zakoupených zájezdů v nejmenované cestovní kanceláři v letech 2001-2004 určete, zdali:

1. má sezónní výkyvy,
2. pokud ano, určete jejich velikost,
3. odhadněte parametry lineárního trendu,
4. odhadněte hodnoty řady v roce 2005.

.

Tab. 1 - Měsíční nákupy letních dovolených v cestovní kanceláři v letech 2001-2004

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
2001	335	312	352	354	398	400	404	405	373	365	324	329
2002	333	315	340	339	383	396	404	399	377	370	334	345
2003	354	327	367	368	408	405	412	402	378	371	338	343
2004	352	327	366	367	408	406	414	392	354	351	323	321

Zdroj dat: vlastní zpracování

Vypracování příkladu

Sezónní složka časové řady jsou periodické odchylky jejích hodnot od trendu. Délka periody je přitom maximálně 1 rok. Oscilace vznikají působením různých příčin. Ať už přímých nebo nepřímých (střídání ročních období, společenské zvyky atd.). Cílem je porovnat vývoj ČŘ v rámci roku v jednotlivých sezónách (letech) a odkrýt tak základní dynamiku jejího vývoje. Postup je následující:
1. Kvantifikovat sezónní výkyvy.
2. Očistit ČŘ od těchto výkyvů.

Připomeneme, že pro časové řady používáme dva základní modely:

- **Aditivní** $y_t = T_t + S_t + C_t + \epsilon_t$
- **Multiplikativní** $y_t = T_t.S_t.C_t.\epsilon_t$

V dalším výkladu bude použit model aditivní.

Abychom vyjádřili rok a dílčí období, budeme používat pro jednotlivé hodnoty ČŘ dvou indexů: i, j, kde i značí pořadí roku a j pořadí dílčího období v rámci roku. Hodnoty indexů jsou $i = 1, \ldots, m$; $j = 1, \ldots, r$.

Aditivní model (model konstantní sezónnosti)

Příslušný model po zavedení dvou indexů má tuto podobu:

$$y_{ij} = T_{ij} + S_{ij} + +\epsilon_{ij} \tag{1}$$

Pro aditivní sezónní faktory Sij se předpokládá, že $S_{ij} = Sj$, charakter sezónních faktorů se v jednotlivých letech $i = 1, \dots, m$ nemění. Aby se vliv sezónních faktorů vykompenzoval, musí platit

$$\sum_{j=1}^{r} S_j = 0$$

Program STATGRAPHICS Centurion XVII

U časových řad je přímo nutné grafické zpracování. Proto nejprve vytvoříme graf. V tomto případě v menu *Plot/Time Sequence Plots/Run Charts/Individuals*. Do povinného pole **Observations** je třeba zadat název proměnné, kde jsou hodnoty sledované veličiny. Do pole **(Date/Time/Labels:)** je možné zadat odpovídající hodnoty času. Z grafu je patrné, že řada vykazuje roční periodicitu.

V programu STATGRAPHICS je pro sezónní dekompozici ČŘ vytvořena funkcionalita v menu *Describe/Time Series/Seasonal Decomposition*. Do povinného pole **Data** je třeba zadat název proměnné, kde jsou hodnoty sledované veličiny y_t. Do nepovinného pole **(Time Indices:)** je možné zadat odpovídající hodnoty času. Zde vybrat volbu Month(s) a do pole Starting At: zadat 1.01 - řada začíná v lednu 2001. Ještě je nutné vyplnit pole Seasonality, kam zadáme 12 (roční periodicita). Po kliknutí na tlačítko OK se objeví další formulář *Adjustment Options*, kde pouze klikneme na tlačítko OK.
V dalším formuláři *Tables and Graphs* zaškrtneme všechny tři tabulky a grafy *Trend-Cycle*, *Seasonal Indices*, *Seasonally Adjusted Data*.

Výsledky:
Hodnoty sezónní složky pro jednotlivé měsíce jsou v tabulce 2 (uvádíme pouze počátek a konec roku). Ve STATGRAPHICS je najdeme v tabulce *Seasonal Indices*.

s
Tab. 2 - Sezónní složka S_j

Měsíc	S_j
I	-22,11
II	-45,40
III	-10,29
...	...
XI	-36
XII	-29,22

Zdroj dat: STATGRAPHICS, vlastní zpracování

Rozklad členů časové řady na jednotlivé složky podle modelu (1) je v tabulce *Data Table*.

Tab. 3 - Dekompozice časové řady

Období	y_{ij}	T_{ij}	S_{ij}	ϵ_{ij}	$y_{ij} - S_j$
1.01	335				357,11
2.01	312				357,4
3.01	352				362,29
4.01	354				363,5
5.01	398				365,63
6.01	400				364,83
7.01	404	362,5	41,5	1,29	363,79
8.01	405	362,54	42,46	7,36	369,9
9.01	373	362,17	10,83	2,14	364,31
10.01	365	361,04	3,96	2,97	364,01
11.01	324	359,79	-35,79	0,21	360
...	...	...	...	...	...

Zdroj dat: STATGRAPHICS, vlastní zpracování

Odhad lineárního trendu

Pokud odhadneme parametry lineárního trendu $T_t = \alpha_0 + \alpha_1 t$ (viz téma 9), dostaneme model

$$\hat{T}_t = 357,287 + 0,447t \tag{2}$$

kde jsou oba parametry statisticky významné (P-value obou se rovná 0).

Předpovědi pro rok 2005

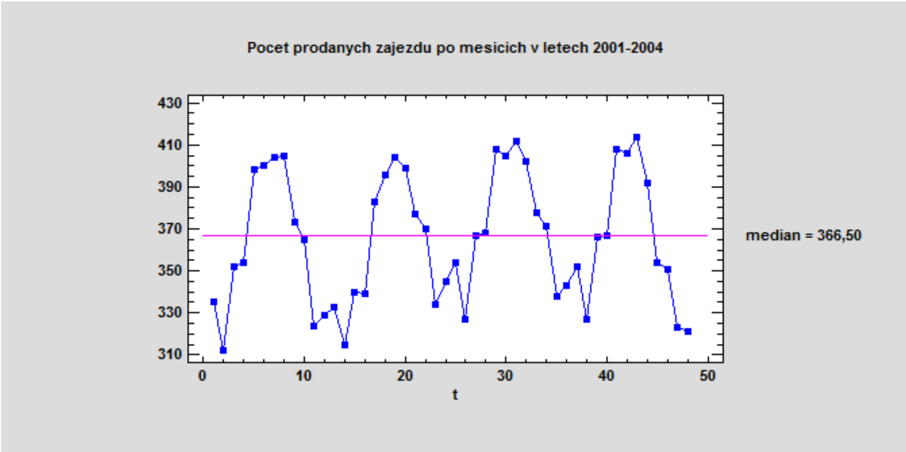
Předpovědi pro rok 2005 získáme tak, že do modelu 2 dosadíme za t čísla 49,...,60 a k takto získaným odhadům trendové složky $\hat{T}_t$ přičteme sezónní složku S_j z tabulky 2 pro příslušný měsíc. Výsledky jsou v tabulce 4.

Tab. 4 - Předpovědi $\hat{Y}_t$ pro rok 2005

Měsíc	S_j	$\hat{T}_t$	$\hat{Y}_t$
1.05	-22,11	368,72	346,56
2.05	-45,40	368,81	323,4
3.05	-10,29	368,94	358,65
...	...		
11.05	-36	370	334
12.05	-29,22	370,13	340,91

Zdroj dat: STATGRAPHICS, vlastní zpracování

Obrázek 1: Počet prodaných zájezdů v letech 2001-2004 po měsících



Zdroj: vlastní zpracování

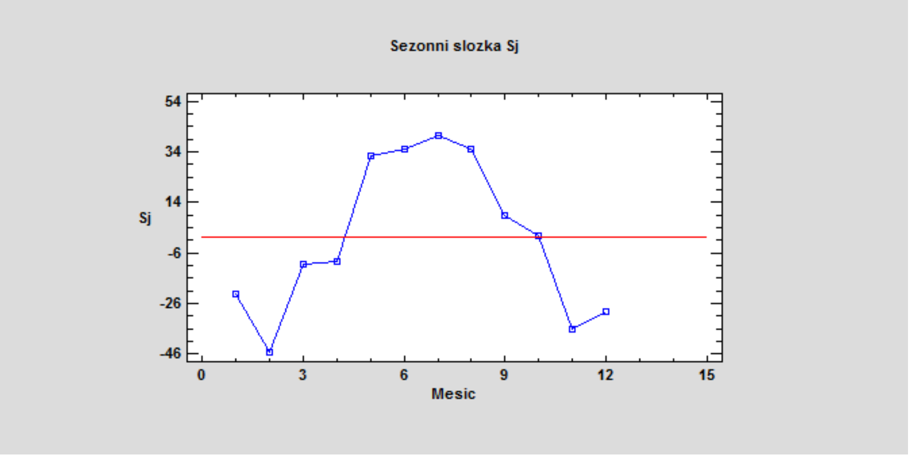
Interpretace výsledků

Analyzovaná ČŘ pokrývá 4 sezóny, kde každá má 12 období. Z obr. 1 je zřejmé, že řada vykazuje sezónnost, konkrétně roční. Z tabulky *Seasonal Indices* a obr. 3 je patrné, že minimum, respektive maximum sezónní složky S_j je v únoru, respektive v červenci. V tabulce *Data Table*, část z ní je v tabulce 3, jsou uvedeny jednotlivé složky, přičemž pro hodnoty v sloupci y_{ij} platí

$$y_{ij} = T_{ij} + S_j + \epsilon_{ij}$$

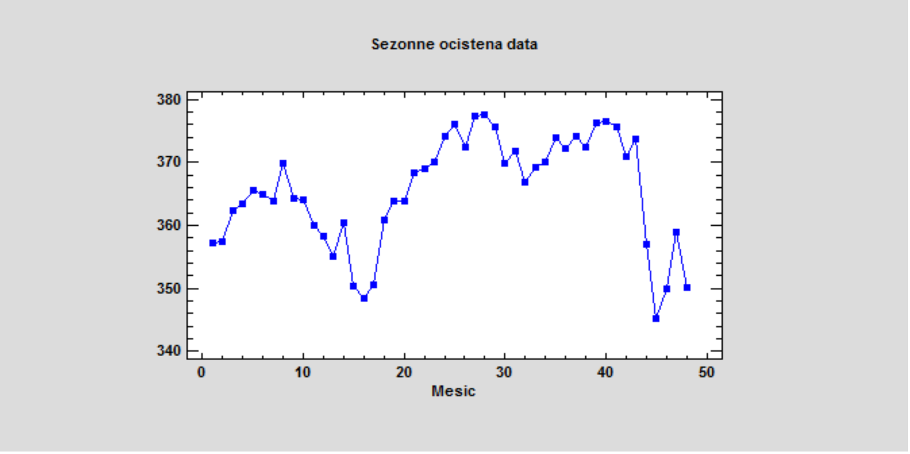
Sezónně očištěné hodnoty ČŘ jsou v posledním sloupci tabulky 3 a platí, že se rovnají rozdílu $y_{ij} - S_j$. Grafické znázornění je na obr. 3. Na těchto údajích byla provedena trendová analýza (lineárné model). Oba jeho parametry jsou statisticky významné, přičemž směrnice je kladná. Trend je tudíž rostoucí. Na obr. 4 je potom původní ČŘ doplněná o předpovědi na rok 2005. Je vidět, že odhady pro tento rok jsou podobné jako hodnoty z předchozích let.

Obrázek 2: Sezónní složka - Počet prodaných zájezdů



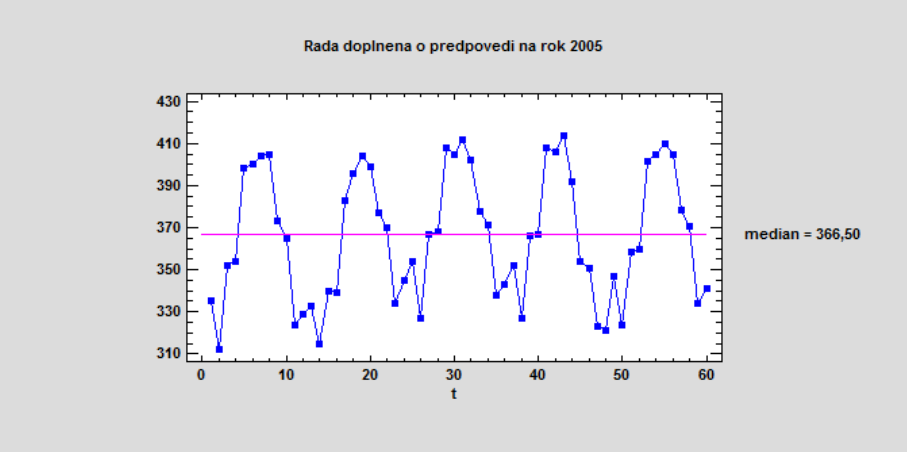
Zdroj: vlastní zpracování

Obrázek 3: Sezónně očištěná data - Počet prodaných zájezdů



Zdroj: vlastní zpracování

Obrázek 4: Počet prodaných zájezdů - Řada doplněná o předpovědi na rok 2005



Zdroj: vlastní zpracování