INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Periodicita v časové řadě, její popis a identifikace

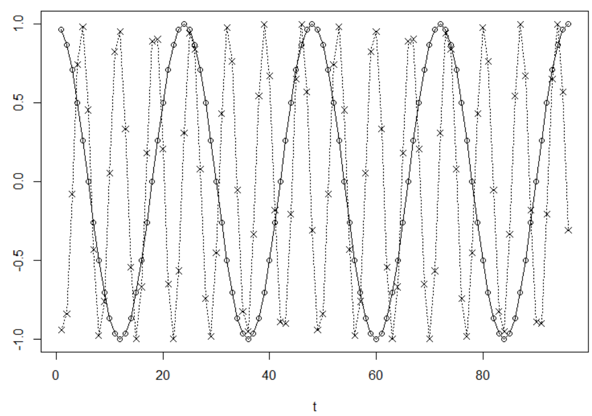
1 Periodicita

Některé časové řady obsahují periodickou složku. Pomocí vybraných nástrojů **spektrální analýzy** budeme tuto složku identifikovat. Mějme funkci periodickou funkci

$$R\cos(2\pi ft + \phi),$$

R je **amplituda**, f je **frekvence** a ϕ označuje **fázový posuv**. Tato funkce se opakuje každou časovou jednotku $T = \frac{1}{f}$ – **perioda**.

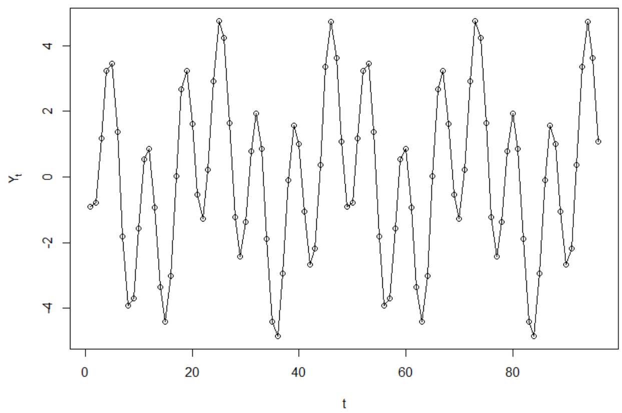
Graf zobrazuje dvě tyto funkce v diskrétním čase $t = 1, \ldots, 96$ s frekvencemi 4/96 a 14/96. Funkce s nižší frekvencí má nulový fázový posuv, ta s vyšší frekvencí má fázový posuv 0,6π.



Vytvořme lineární kombinaci těchto funkcí

$$Y_t = 2\cos\left(2\pi t\frac{4}{96}\right) + 3\cos\left[2\pi\left(t\frac{14}{96} + 0{,}3\right)\right]. \quad (1)$$

Periodicita je nyní „skrytá“.



Operační program Vzdělávání pro konkurenceschopnost
Název projektu: Inovace magisterského studijního programu Fakulty ekonomiky a managementu
Registrační číslo projektu: CZ.1.07/2.2.00/28.0326

PROJEKT JE SPOLUFINANCOVÁN EVROPSKÝM SOCIÁLNÍM FONDEM A STÁTNÍM ROZPOČTEM ČESKÉ REPUBLIKY.

Platí

$$R\cos(2\pi ft+\phi) = A\cos(2\pi ft) + B\sin(2\pi ft),$$

kde

$$R=\sqrt{A^2+B^2},\quad \phi = \arctan\left(-\frac{B}{A}\right)$$

a obráceně

$$A = R\cos\phi,\quad B=-R\sin\phi.$$

Pro pevně danou hodnotu frekvence f lze použít $\cos(2\pi ft)$ a $\sin(2\pi ft)$ jako prediktory a odhadnout A a B pomocí metody nejmenších čtverců. Obecnou kombinaci m kosinových funkcí s libovolnými amplitudami a frekvencemi lze zapsat ve tvaru[1]

$$Y_t = A_0 + \sum_{j=1}^{m}\left[A_j\cos(2\pi f_j t) + B_j(2\sin f_j t)\right].$$

Metodu nejmenších čtverců lze použít pro odhady A_j a B_j, pokud mají frekvence speciální tvar, regrese jsou jednoduché. Předpokládejme že n je liché, $n=2k+1$. Frekvence $1/n, 2/n, \ldots, k/n$ se nazývají **fourierovské frekvence**. Prediktory tvořené funkcemi sinus a kosinus v těchto frekvencích jsou ortogonální, dostáváme odhady

$$\widehat{A}_0 = \overline{Y_t}$$

$$\widehat{A}_j = \frac{2}{n}\sum_{t=1}^{n} Y_t\cos\left(\frac{2\pi tj}{n}\right)\quad \text{a}\quad \widehat{B}_j = \frac{2}{n}\sum_{t=1}^{n} Y_t\sin\left(\frac{2\pi tj}{n}\right)$$

Je-li n sudé, $n=2k$, předchozí rovnice platí pro $j=1,2\ldots,k-1$, ale pro $j=k$ dostáváme

$$\widehat{A}_k = \frac{1}{n}\sum_{t=1}^{n}(-1)^t Y_t,\quad \text{a}\quad \widehat{B}_k = 0. \tag{2}$$

1.1 Periodogram

Pro liché $n=2k+1$ je **periodogram** I pro frekvenci $f=j/n, j=1,2,\ldots,k$ definován

$$I\left(\frac{j}{n}\right) = \frac{n}{2}\left(\widehat{A}_j^2 + \widehat{B}_j^2\right).$$

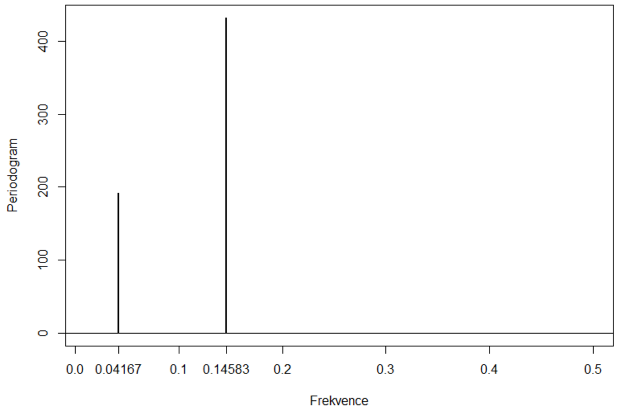
Pro sudé $n=2k$ získáme hodnoty periodogramu pro $j=1,2,\ldots,k-1$ podle předchozího vztahu, pro $j=k$, tedy frekvenci $f=k/n=1/2$ je

$$I\left(\frac{1}{2}\right) = n\widehat{A}_k^2,$$

viz výraz (2).

Pozn. Pro dlouhé časové řady se pro výpočet používá FFT (fast Fourier transform). Graf zobrazuje periodogram lineární kombinace funkcí (1)

[1] A_0 je koeficient kosinové funkce pro nulovou frekvenci, B_0 je koeficient kosinové funkce pro nulovou frekvenci, je tedy nulový a ve vzorci se neobjevuje.



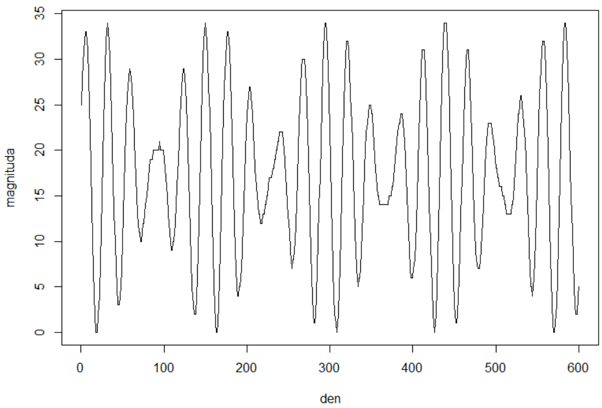
Rozšíříme nyní definici periodogramu na všechny frekvence z intervalu $0 \leq f \leq 1/2$

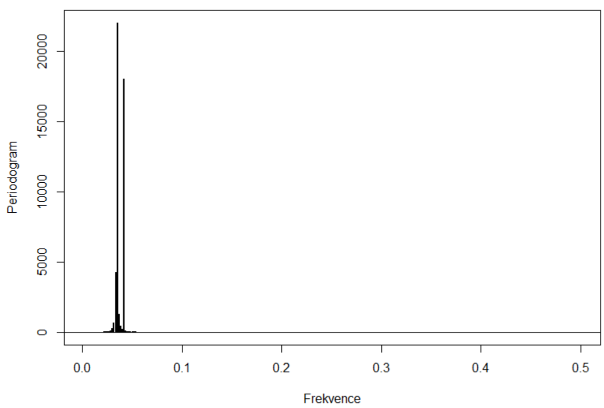
$$I(f) = \frac{n}{2}\left(\widehat{A}_f^2 + \widehat{B}_f^2\right),$$

kde

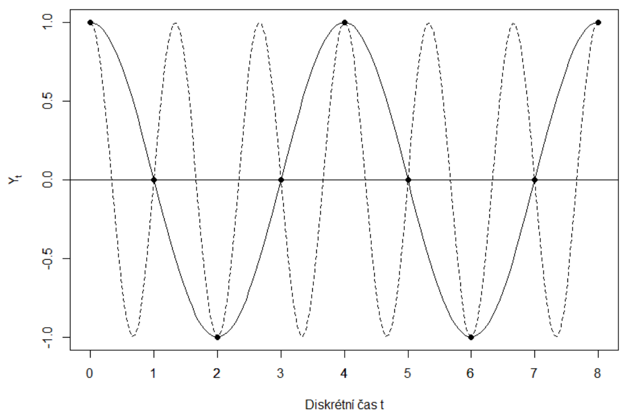
$$\widehat{A}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \cos(2\pi t f) \quad \text{a} \quad \widehat{B}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \sin(2\pi t f).$$

Příklad. Půlnoční magnitudy (jas) jisté hvězdy v 600 po sobě jdoucích dnech jsou zobrazeny na levém grafu, periodogram je vpravo.

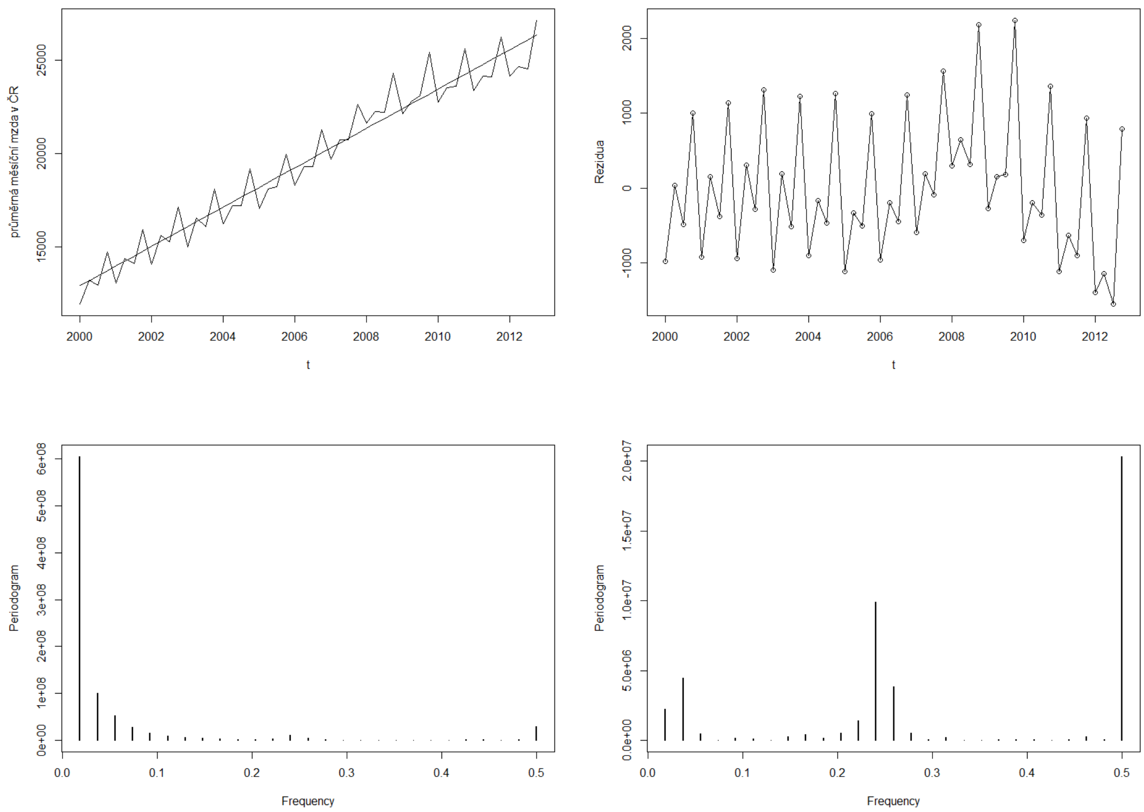




Proč frekvence omezovat na interval $0 \leq f \leq 1/2$? V grafu jsou zobrazeny dvě kosinové funkce, jedna s frekvencí $f = 1/4$ a druhá s frekvencí $f = 3/4$ (čárkovaná čára). Měříme-li hodnoty těchto funkcí pouze v časech $t = 0, 1, 2, \ldots$, dostáváme identické hodnoty. V diskrétním čase nemůže od sebe tyto dvě funkce rozlišit – **aliasing** frekvencí. **Nyquistova frekvence** – $f = 1/2$.

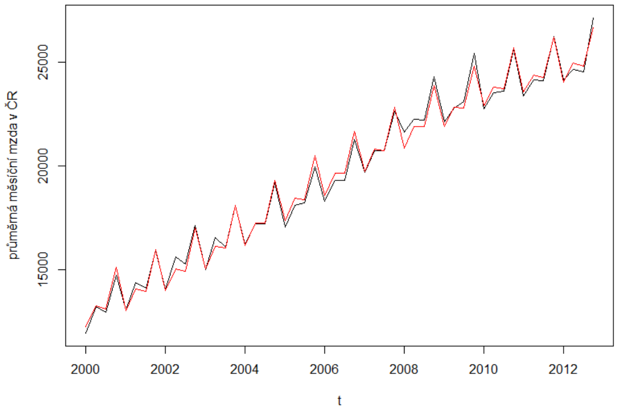


Příklad. Máme k dispozici údaje o mzdách v ČR za období 2000–2012. Odhadnutý lineární trend pomocí lineární regrese je znázorněn na následující grafu (nahoře vlevo). Po odečtení toho trendu získáme rezidua (graf vpravo nahoře). Pod těmito grafy jsou zobrazeny odpovídající periodogramy.



S využitím vypočítaných hodnot periodogramu (z reziduí), popíšeme vývoj mezd pomocí modelu

$$Y = \beta_1 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 \sin 4\pi t + \beta_5 \cos 4\pi t + \beta_6 \sin 2\pi t + \beta_7 \cos 2\pi t.$$



2 Exponenciální vyrovnávání

Chceme-li predikovat (předpovídat) hodnotu časové řady v čase $t = \tau$, je přirozené vzít v úvahu předcházející hodnoty a onu predikci určit jako vážený součet předchozích pozorování.

$$\hat{y}_{t=\tau} = \lambda_0 y_\tau + \lambda_1 y_{\tau-1} + \lambda_2 y_{\tau-2} + \cdots .$$

zdá se být rozumné dát nedávným pozorováním větší váhu než pozorováním v čase hodně vzdáleným. Jednou z možností je použití následujících vah

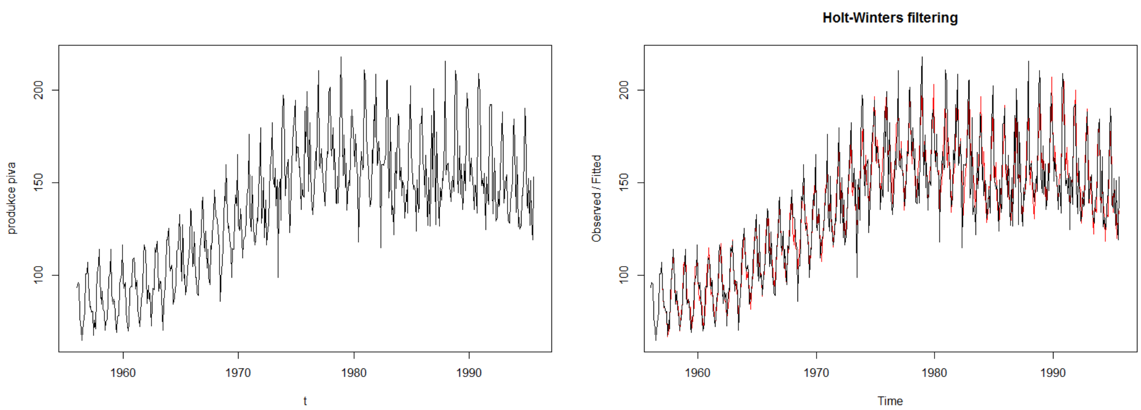
$$\lambda_i = \alpha(1-\alpha)^i, \quad 0 < \alpha < 1,$$

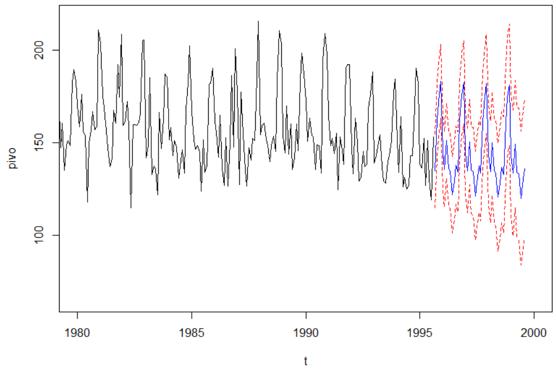
potom

$$\hat{y}_{t=\tau} = \alpha y_\tau + (1-\alpha) y_{\tau-1} + (1-\alpha)^2 y_{\tau-2} + \cdots .$$

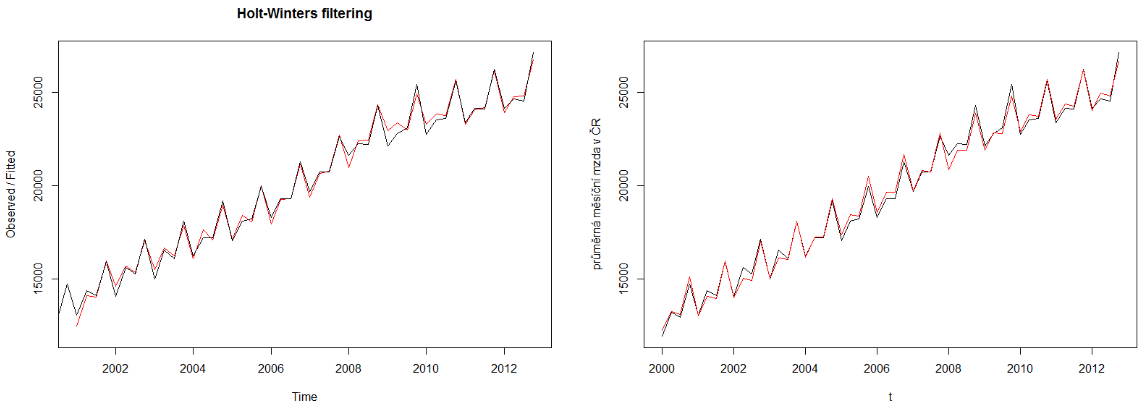
Exponenciální vyrovnávání (název pochází z faktu, že váhy klesají exponenciálně) v tomto základním tvaru může být použito pouze pro časové řady bez trendu a sezónní složky. Zobecněním uvedené procedury je tzv. Holt-Wintersovo vyrovnávání, které již uvažuje i trend a sezónní složku. Obsahuje tři parametry: α pro úroveň, β pro trend a γ pro sezónní složku (funkce `HoltWinters`).

Příklad. Měsíční produkce piva v Austrálii.





Příklad. Mzda v České republice.



Obrázek vlevo ukazuje výsledek exponenciálního vyrovnávání, obrázek vpravo potom fit pomocí periodických funkcí – viz periodogram.

Příklady k procvičení

---

1. Najděte významné periody časových řad uvedených v souboru perioda.txt. Pomocí lineárního regresního modelu odhadněte trend pomocí funkcí sinus a kosinus pro významné periody.
   [Datový soubor: perioda.txt]

2. Datový soubor airpass z balíčku „TSA“ obsahuje měsíční údaje o počtu pasažérů na mezinárodních linkách letech 1960 až 1971. Pomocí periodogramu identifikujte významné periodické složky v časové řadě logaritmů počtu pasažéru, popište trend pomocí funkcí sinus a cosinus s významnými periodami. Výsledek porovnejte s odhadnutým proložením pomocí dummy proměnných.
   [Příkaz pro R: data(airpass,package=”TSA”)]

3. Datový soubor airpass z balíčku „TSA“ obsahuje měsíční údaje o počtu pasažérů na mezinárodních linkách letech 1960 až 1971. Pomocí klouzavých průměrů vhodné délky proveďte vyhlazení této časové

řady. Proveďte dekompozici této časové řady (příkazy decompose, stl). Využijte funkci HoltWinters pro výpočet exponenciálního vyrovnání. Určete předpovědi pomocí vhodného typu exponenciálního vyrovnání. [Příkaz pro R: data(airpass,package=”TSA”)]

4. Datový soubor co2 z balíčku „TSA“ obsahuje atmosférické koncentrace CO_2. Jedná se o měsíční data od roku 1994 do roku 2004 (Alert, Northwest Territories, Kanada). Pomocí klouzavých průměrů vhodné délky proveďte vyhlazení této časové řady. Proveďte dekompozici této časové řady (příkazy decompose, stl). Využijte funkci HoltWinters pro výpočet exponenciálního vyrovnání. Určete předpovědi pomocí zvoleného typu exponenciálního vyrovnání. [Příkaz pro R: data(co2, package=”TSA”)]