

M5201 – 4. CVIČENÍ:
Identifikace periodických komponent

1 Regresní modely cyklických trendů

1.1 Nelineární model

Uvažujme následující model s nekorelovanými a homoskedastickými rezidui (tj. majícími konstantní rozptyl):

$$\boxed{Y_i = f(t_i) + \varepsilon_i} \quad E\varepsilon_i = 0; \quad D\varepsilon_i = \sigma^2; \quad C(\varepsilon_i, \varepsilon_j) = 0; \quad i \neq j; \quad i, j = 1, \dots, n$$

Je-li $f(t)$ periodická funkce s periodou T , pak **frekvencí** rozumíme veličinu

$$\boxed{\lambda = \frac{2\pi}{T}}.$$

Dále uvažujme takovou $f(t_i)$, kterou lze zapsat ve tvaru

$$(a) \quad f(t_i) = \mu + \sum_{j=1}^p (\alpha_j \cos \omega_j t_i + \beta_j \sin \omega_j t_i)$$

nebo ekvivalentně

$$(b) \quad f(t_i) = \mu + \sum_{j=1}^p \rho_j \cos(\omega_j t_i + \theta_j) \quad \rho_j = \sqrt{\alpha_j^2 + \beta_j^2}; \quad \theta_j = \arctan \frac{\beta_j}{\alpha_j}.$$

Jde o **nelineární regresní model** vzhledem k $(3p + 1)$ neznámých parametrů:

$$(a) \quad \alpha_1, \dots, \alpha_p \quad \beta_1, \dots, \beta_p \quad \mu \quad \omega_1, \dots, \omega_p$$

$$(b) \quad \rho_1, \dots, \rho_p \quad \mu \quad \omega_1, \dots, \omega_p \quad \theta_1, \dots, \theta_p$$

Odhad vektoru **neznámých parametrů** pomocí *metody nejmenších čtverců* minimalizuje výraz

$$(a) \quad S(\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p, \omega_1, \dots, \omega_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$$

$$(b) \quad S(\mu, \rho_1, \dots, \rho_p, \theta_1, \dots, \theta_p, \omega_1, \dots, \omega_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$$

Numericky lze systém nelineárních rovnic řešit např. pomocí Gaussovy-Newtonovy iterační metody. V prostředí R lze v základním statistickém balíčku `stats` najít funkci `nls()`, kterou je možno použít k odhadu parametrů nelineárních regresních modelů jako je tento.

1.2 Lineární model pro známé frekvence

Pokud jsou *frekvence* $\omega_1, \dots, \omega_p$ *známé*, situace je podstatně jednodušší. Z modelu (a) se v tom případě stane *lineární regresní model*, jehož koeficienty můžeme odhadnout metodou nejmenších čtverců.

Vektor neznámých parametrů lze psát takto $\beta = (\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p)'$ a matice plánu je tvaru

$$\mathbf{X}_{n \times (2p+1)} = \begin{pmatrix} 1 & \cos \omega_1 t_1 & \sin \omega_1 t_1 & \cdots & \cos \omega_p t_1 & \sin \omega_p t_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos \omega_1 t_n & \sin \omega_1 t_n & \cdots & \cos \omega_p t_n & \sin \omega_p t_n \end{pmatrix}.$$

Platí

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & \frac{n}{2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \frac{n}{2} \end{pmatrix} \Rightarrow \text{ze vztahu } \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \text{ dostaneme}$$

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n Y_i \\ \hat{\alpha}_j &= \frac{2}{n} \sum_{i=1}^n Y_i \cos \omega_j t_i \quad \text{pro } j = 1, \dots, p \\ \hat{\beta}_j &= \frac{2}{n} \sum_{i=1}^n Y_i \sin \omega_j t_i \end{aligned}$$

Neznámé parametry modelu (b) získáme na základě odhadů pro model (a) ze vztahů

$$\begin{aligned} \hat{\gamma}_j &= \sqrt{\hat{\alpha}_j^2 + \hat{\beta}_j^2} \\ \hat{\omega}_j &= \arctan \frac{\hat{\beta}_j}{\hat{\alpha}_j} \end{aligned} \quad \text{pro } j = 1, \dots, p$$

1.3 Metoda skrytých period

Metoda skrytých period slouží k nalezení významných period $\omega_1, \dots, \omega_p$, pokud nejsou předem známé.

Uvažujme takové časové řady, které lze rozložit na **součet harmonických frekvencí**, jejichž **délky period** můžeme vyjádřit jako podíl $T_k = \frac{n}{k}$, kde n je počet naměřených hodnot a $0 < k \leq n$.

Označme dále $f_k = \frac{1}{T_k} = \frac{k}{n}$ k -tá frekvence

$$\omega_k = 2\pi f_k = 2\pi \frac{k}{n} \quad k\text{-tá úhlová frekvence}$$

Maximální délka periody je rovna počtu pozorování, tj. $T_{\max} = n$, tedy $k = 1$ a **minimální frekvence** má velikost $\omega_{\min} = \frac{2\pi}{n}$.

Nejkratší zjistitelná perioda je $T_{\min} = 2$, jí odpovídající frekvence je $\omega_{\max} = \pi$ (tzv. *Nyquistova frekvence*).

k může nabývat hodnot:

$$k = \begin{cases} 1, 2, \dots, \frac{n}{2} & \text{pro } n \text{ sudé} \\ 1, 2, \dots, \frac{n-1}{2} & \text{pro } n \text{ liché.} \end{cases}$$

Model časové řady pak můžeme zapsat ve tvaru $Y_t = s_t + \varepsilon_t$, kde

- t označuje ekvidistatní **časové okamžiky** měření, pro jednoduchost předpokládáme, že intervaly mají jednotkovou velikost;

- n je **počet naměřených hodnot**;
- ε_t je **bílý šum**: $\varepsilon_t \sim WN(0, \sigma^2)$, tj. $E\varepsilon_t = 0$, $D\varepsilon_t = \sigma^2$, $C(\varepsilon_t, \varepsilon_s) = E\varepsilon_t \varepsilon_s = 0$; $t \neq s$,
- s_t je **periodická funkce** tvaru

$$s_t = \begin{cases} \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) & \text{pro } n \text{ liché} \\ \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) + \alpha_{\frac{n}{2}} (-1)^t & \text{pro } n \text{ sudé} \end{cases}$$

kde $\alpha_j, \beta_j \in \mathbb{R}$, $p \in \mathbb{N}$ jsou daná čísla a nazýváme je **parametry** modelu.

Ekvivalentně můžeme s_t napsat ve tvaru:

$$s_t = \begin{cases} \sum_{j=1}^p \rho_j \cos(t\omega_j - \theta_j) & \text{pro } n \text{ liché} \\ \sum_{j=1}^p \rho_j \cos(t\omega_j - \theta_j) + \alpha_{\frac{n}{2}} (-1)^t & \text{pro } n \text{ sudé} \end{cases},$$

kde

$$\rho_k = \sqrt{\alpha_k^2 + \beta_k^2}$$

je **amplituda** k -té harmonické složky,

$$\theta_k = \begin{cases} \arctan \frac{\beta_k}{\alpha_k} & \alpha_k > 0 \\ \arctan \frac{\beta_k}{\alpha_k} + \pi & \alpha_k < 0 \end{cases}$$

je **fázový posun** k -té harmonické složky.

Pokud zahrneme do počtu frekvencí i frekvenci nulovou (kdy $k = 0$), pak přibude člen, který označíme $\frac{\alpha_0}{2}$, přitom $\beta_0 = 0$.

1.4 Periodogram

Ukazatelem významných periodicit náhodné posloupnosti $\{Y_t, t \in \{1, \dots, n\}\}$ je **periodogram**:

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n Y_t e^{-it\omega} \right|^2$$

Za platnosti uvažovaného modelu

$$Y_t = s_t + \varepsilon_t = \frac{\alpha_0}{2} + \sum_{j=1}^p (\alpha_j \cos t\omega_j + \beta_j \sin t\omega_j) + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2),$$

bude mít **periodogram** pro velká n v bodech $\omega = \omega_1, \dots, \omega_p$ **výrazně velké hodnoty**, jinde jeho hodnoty budou relativně malé, budou kolísat kolem hodnoty $\frac{\sigma^2}{2\pi}$.

Odtud plyne, že významná lokální maxima v průběhu periodogramu by měla identifikovat periodickou strukturu uvažovaného modelu tak, že vyznačí neznámé frekvence $\omega_1, \dots, \omega_p$.

Vhodným statistickým testem pak rozhodneme, které hodnoty periodogramu můžeme opravdu považovat za významně velké ve srovnání s hodnotami ostatními.

1.5 Test R. A. Fishera

R. A. Fisher odvodil test, kterým se dá zjistit významnost nejvyšších hodnot periodogramu. Uvažujme posloupnost nezávislých náhodných veličin Y_1, \dots, Y_n .

Nulová hypotéza:

$$H_0 : Y_t = \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2).$$

Předpokládejme, že $n = 2m + 1$ je liché číslo (je-li n sudé, obvykle se vynechá první člen) a uvažujme hodnoty periodogramu $I_n(\omega_k)$ v bodech $\omega_k = \frac{2\pi}{n}k$ ($k = 1, \dots, m$). Srovnajme je podle velikosti a označme V_1, \dots, V_m . Test se provádí na základě tzv. **Fisherovy statistiky**

$$W = \frac{V_1}{V_1 + \dots + V_m}$$

která nabývá hodnot mezi nulou a jedničkou.

Velké hodnoty (blízké 1) budou tvořit kritický obor nulové hypotézy proti alternativě

$$H_1 : Y_t = \sum_{j=1}^p \rho_j \cos(t\omega_j - \theta_j) + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2)$$

R. A. Fisher odvodil distribuční funkci statistiky W za platnosti hypotézy H_0 : (za předpokladu, že uvažujeme gaussovský bílý šum)

$$1 - F_{W|H_0}(x) = P(W > x|H_0) = m(1-x)^{m-1} - \binom{m}{2}(1-2x)^{m-1} + \dots,$$

kde $0 < x < 1$ a na pravé straně sčítáme tak dlouho, dokud jsou členy $(1 - kx)$ kladné, což lze také zapsat takto

$$P(W > x|H_0) = \sum_{k=1,2,\dots} \binom{m}{k} [\max(0, 1 - kx)]^{m-1} = \sum_{k=1,2,\dots} \binom{m}{k} [(1 - kx)_+]^{m-1}.$$

Hypotézu $H_0 : Y_t = \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma^2)$ zamítáme na hladině významnosti α , pokud

$$1 - F_{W|H_0}(w) = P(W > w|H_0) = \alpha_{W|H_0} \leq \alpha,$$

kde w je skutečná hodnota Fisherovy statistiky při daných hodnotách konkrétní časové řady a $\alpha_{W|H_0}$ je tzv *p-value*.

1.5.1 Modifikace Fisherova testu pro testování dalších periodicit

V případě, že pomocí Fisherova testu **zjistíme signifikantní periodicitu určité frekvence**, postupujeme při testování dalších periodicit následujícím způsobem.

Vynecháme V_1 , na základě veličin V_2, \dots, V_m položíme

$$W^{(2)} = \frac{V_2}{V_2 + \dots + V_m}$$

a stanovíme $P(W^{(2)} > w^{(2)})$ podle stejného vzorce, kde místo m dosadíme $m - 1$. Vyjde-li i tato druhá největší hodnota významná, opět se vynechá a m se zmenší o další jedničku.

Když takto stanovíme frekvence $\omega_1, \dots, \omega_p$, získáme model se známými frekvencemi a zbylé neznámé koeficienty odhadneme metodou nejmenších čtverců (viz odstavec Lineární model pro známé frekvence).

1.6 Siegelův test

V případě více významných period není Fisherův test dostatečně silný. Vhodnější by měl být **Siegelův test**.

Zde se používá statistika T_λ tvaru

$$T_\lambda = \sum_{j=1}^m \left(\frac{I_n(\omega_j)}{\sum_{i=1}^m I_n(\omega_i)} - \lambda g_F \right)_+$$

kde $(x)_+ = \max(x, 0)$, g_F je $100(1 - \alpha)\%$ kritická hodnota Fisherova testu (tj. $P(W > g_F | H_0) = \alpha$) a $0 < \lambda < 1$ je předem zvolená konstanta (doporučuje se volit $\lambda = 0.6$).

Nulovou hypotézu zamítáme, pokud $T_\lambda > t_\lambda$, kde t_λ je kritická hodnota tohoto testu. Kritické hodnoty bývají tabelovány pro různá λ a m . Jako významné jsou uznány pouze ty periodicity, jejichž odpovídající sčítance přispěly do celkové hodnoty testovací statistiky T_λ .

Pokud **počet dat odpovídá násobku délky periody, najde** Fisherův (či Siegelův) test **skutečnou periodu přesně**. V případě, že však **počet dat není násobkem délky periody** nebo snad **je dat méně než je délka periody**, uvedené testy **skutečnou periodu neobjeví**.

1.7 Iterativní metoda (Damsleth a Spjøtvoll, 1982)

Následující iterativní metoda představuje alternativní způsob určení významných frekvencí a odhadu regresních koeficientů v modelech s trigonometrickým trendem. Postupujeme takto:

- (1) Pomocí Fisherova (nebo jiného) testu zjistíme významnou frekvenci ω_0 . Pokud taková neexistuje, končíme.
- (2) Minimalizujeme nelineární regresní model

$$Y_i = \mu + \alpha \cos t_i \omega + \beta \sin t_i \omega + \varepsilon_i$$

vzhledem k neznámým parametrům $(\mu, \alpha, \beta, \omega)$, přičemž jako **počáteční hodnoty** bereme

- ω_0 frekvenci nabídnutou v kroku (1) Fisherovým či jiným testem,
- μ_0, α_0, β_0 získáme jako řešení **lineárního regresního modelu**

$$Y_i = \mu_0 + \alpha_0 \cos t_i \omega_0 + \beta_0 \sin t_i \omega_0 + \zeta_i$$

Získané hodnoty označme $\hat{\mu}, \hat{\alpha}, \hat{\beta}, \hat{\omega}$.

- (3) Z řady odstraníme vliv $\hat{\omega}$, tj.

$$Z_i = Y_i - \hat{\mu} - \hat{\alpha} \cos t_i \hat{\omega} - \hat{\beta} \sin t_i \hat{\omega} + \eta_i$$

a s řadou Z_i jdeme na bod (1).

Algoritmus při odhadu frekvencí vychází z významných frekvencí, které určil Fisherův či jiný test. Současně s odhadem dalších parametrů odhadnuté frekvence ještě zpřesňuje. Jeho nevýhodou je, že odhaluje frekvence postupně, tj. v každém kroku hledá pouze jedinou významnou frekvenci.

2 Příklad

Datový soubor, kterým se budeme v příkladu zabývat, se týká vzdáleností letokruhů jednoho stromu, který rostl v Kašmíru v letech 1753-1980. Vzdálenosti byly transformovány tak, že jsou vyjádřeny v bezrozměrných jednotkách. Data jsou uložena v souboru `treerings.txt`.

První řádek obsahuje nadpis, druhý řádek je volný a od třetího řádku dál jsou v jednom sloupci uvedeny vzdálenosti letokruhů.

Nejprve načteme nadpis.

```
> fileDat <- paste(data.library,"treerings.txt",sep="")
> (TXT <- paste(scan(fileDat,what="", nlines=1),collapse=" "))
```

```
[1] "Tree rings of Gulmarg, Kashmir (Himalayan Fir), Years: 1753-1980"
```

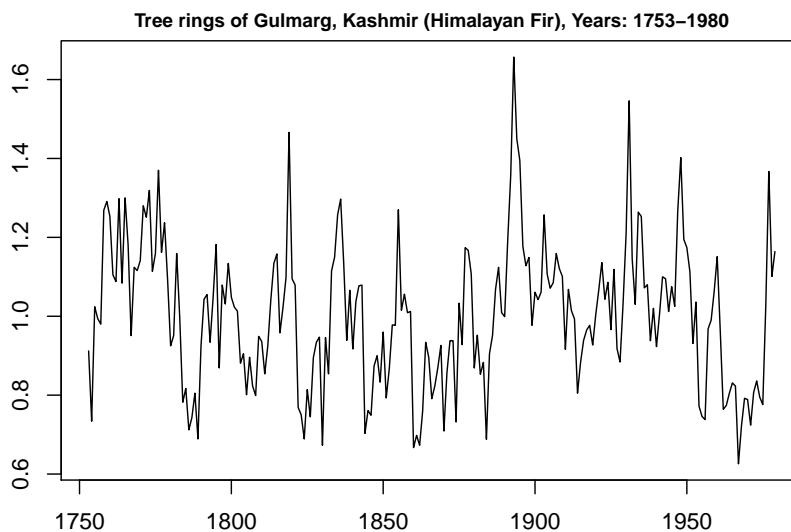
Pak načteme vlastní data do datového rámce a vypíšeme jeho strukturu.

```
> rings<-read.table(fileDat,header=F,skip=2)
> str(rings)
```

```
'data.frame':      227 obs. of  1 variable:
 $ V1: num  0.912 0.734 1.024 0.994 0.98 ...
```

Z načtených dat vytvoříme časovou řadu a vykreslíme ji.

```
> ringsTS<-ts(rings,start=1753,frequency=1)
> par(mar=c(2,2,1,0)+0.5)
> plot(ringsTS,main=TXT,cex.main=0.85)
```



Obrázek 1: Šířky letokruhů

Abychom zjistili významné periody, budeme chtít vypočítat periodogram. Než tak budeme moci učinit, bude nutné z časové řady odstranit lineární trend. Proto nejprve odhadneme model s lineárním trendem a rezidua, která na základě tohoto modelu získáme, použijeme k výpočtu periodogramu.

```
> y <- as.vector(ringsTS)
> x <- as.vector(time(ringsTS))
> n<-length(y)
> nn<-300
> data<-data.frame(x,y)
> LinTrend <- lm( y ~ x,data=data)
> print(summary(LinTrend))
```

Call:

```
lm(formula = y ~ x, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.35750	-0.13113	0.00535	0.10955	0.65982

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.3470184	0.3420518	3.938	0.00011 ***
x	-0.0001848	0.0001832	-1.009	0.31415

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

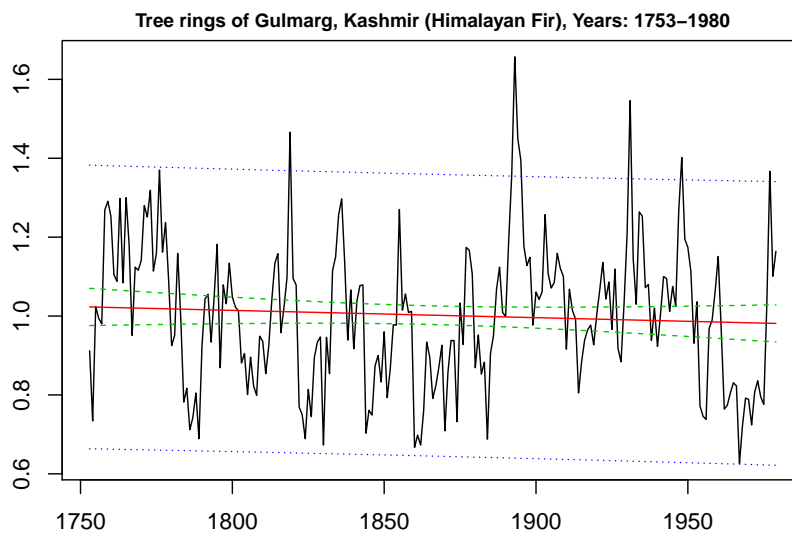
Residual standard error: 0.1809 on 225 degrees of freedom

Multiple R-squared: 0.004503, Adjusted R-squared: 7.826e-05

F-statistic: 1.018 on 1 and 225 DF, p-value: 0.3142

Odhadnutý lineární trend zakreslíme do grafu a spolu s ním vyneseme také intervaly spolehlivosti kolem trendu a predikční interval.

```
> new <- data.frame(x = seq(x[1],x[n],length.out=nn))
> pred.w.plim <- predict(LinTrend, new, interval="prediction")
> pred.w.clim <- predict(LinTrend, new, interval="confidence")
> par(mar=c(2,2,1,0)+0.5)
> plot(x,y,type="l")
> matlines(new$x,cbind(pred.w.clim, pred.w.plim[,-1]),col=c(2,3,3,4,4),
  lty=c(1,2,2,3,3), type="l", ylab="predicted y")
> title(main=TXT,cex.main=0.85)
```



Obrázek 2: Lineární trend pro šířky letokruhů.

Přestože není lineární trend statisticky významný (p – value = 0.31415 > 0.05), odstraníme jej z časové řady a vstupem do funkce, která vypočítává periodogram, budou rezidua. Ta získáme pomocí funkce `resid`.

```
> res<-resid(LinTrend)
```

K výpočtu periodogramu slouží funkce `tperiod()`. Tato funkce není součástí R, ale je obsažena v souboru `FunkceM5201.R`. Abychom ji mohli použít, musíme zmíněný soubor nejprve načíst. Načtení provedeme příkazem `source`, kde jako argument zadáme cestu k souboru.

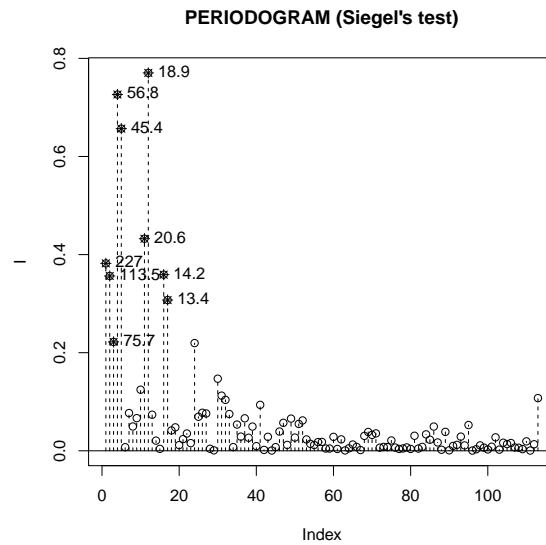
```
> fileFun<-paste(data.library,"FunkceM5201.R",sep="")
> source(fileFun)
```

Nyní již můžeme funkci `tperiod()` použít. Tato funkce na požádání (pokud zadáme argument `fig=TRUE`) periodogram také vykreslí do grafu.

```
> pdgV<-tperiod(res,fig=TRUE)
> str(pdgV)
```

List of 3

```
$ I : Named num [1:113] 0.382 0.356 0.222 0.726 0.657 ...
.. attr(*, "names")= chr [1:113] "2" "3" "4" "5" ...
$ kF: Named num [1:8] 12 4 5 11 1 16 2 17
.. attr(*, "names")= chr [1:8] "13" "5" "6" "12" ...
$ kS: Named num [1:9] 12 4 5 11 1 16 2 17 3
.. attr(*, "names")= chr [1:9] "13" "5" "6" "12" ...
```

Obrázek 3: Periodogram pro šířky letokruhů

Na obrázku jsou vyneseny hodnoty periodogramu pro periody $\frac{n}{k}$, na ose x jsou uspořádány podle k . Ty periody, které jsou významné na základě Fisherova testu, jsou v grafu označeny hvězdičkou a popsány (číslo udává periodu).

Výstupem funkce `tperiod` je (kromě grafu) objekt typu *seznam* obsahující tři položky. První položka `I` obsahuje hodnoty periodogramu pro všechny periody $\frac{n}{k}$. Ve druhé položce `kF` jsou uložena ta k , jimž odpovídající periody označil Fisherův test jako významné. A třetí položka `kS` obsahuje k , jimž přísluší periody, které jsou významné podle Siegelova testu.

Vypišme hodnoty významných period, jak nám je nabízí Fisherův a Siegelův test.

```
> (TkF<-n/pdgV$kF)
```

```
      13      5      6      12      2      17      3      18
18.91667 56.75000 45.40000 20.63636 227.00000 14.18750 113.50000 13.35294
```

```
> (TkS<-n/pdgV$kS)
```

```
      13      5      6      12      2      17      3      18
18.91667 56.75000 45.40000 20.63636 227.00000 14.18750 113.50000 13.35294
      4
75.66667
```

Chceme-li explicitně zjistit a do grafu vykreslit trigonometrický trend, použijeme k tomu dvě funkce, a to `trigofit()` a `trigoval()`. Opět se jedná o funkce, které nejsou součástí R. Protože jsou ale uloženy ve stejném souboru jako `tperiod()`, máme je již načtené, a tudíž je můžeme rovnou používat.

První z těchto dvou funkcí – `trigofit()` – opět hledá pro zadanou časovou řadu významné periody, ale navíc odhadne koeficienty v lineárním regresním modelu s trigonometrickým trendem, v němž byly použity nalezené významné periody.

Její výstupem je seznam s položkami:

- `periodogram` (hodnoty periodogramu),
- `freq` (odpovídající úhlové frekvence),
- `signA` (odhady koeficientů α_j pro významné periody v modelu dle zápisu (a)),
- `signB` (odhady koeficientů β_j pro významné periody v modelu dle zápisu (a)),
- `signFreq` (významné frekvence),
- `amplituda` (odhady amplitud ρ_j pro v modelu ve tvaru (b) se všemi periodami),
- `faze` (odhady fází θ_j v modelu (b) se všemi periodami),
- `signL` (logický vektor s hodnotami TRUE na pozicích k odpovídajících významným frekvencím),
- `A` (odhady koeficientů α_j v modelu (a) se všemi periodami)
- `B` (odhady koeficientů β_j v modelu (a) se všemi periodami)

```
> TrigoTrend<-trigofit(res)
> str(TrigoTrend)
```

```
List of 10
 $ periodogram: num [1:113] 0.0304 0.0284 0.0177 0.0578 0.0523 ...
 $ freq       : num [1:113] 0.0277 0.0554 0.083 0.1107 0.1384 ...
 $ signA      : num [1:8] -0.0704 -0.0497 0.0241 0.0566 0.0261 ...
 $ signB      : num [1:8] 0.0428 0.0627 0.0722 0.0245 -0.0518 ...
 $ signFreq   : num [1:8] 0.3322 0.1107 0.1384 0.3045 0.0277 ...
 $ amplituda  : num [1:113] 0.00545 0.00526 0.00415 0.00751 0.00714 ...
 $ faze       : num [1:113] 2.26 -5.6 6.58 -3.13 -0.15 ...
 $ signL      : logi [1:113] TRUE TRUE FALSE TRUE TRUE FALSE ...
 $ A          : num [1:113] 0.0261 -0.0327 0.0255 -0.0497 0.0241 ...
 $ B          : num [1:113] -0.0518 0.0455 0.0362 0.0627 0.0722 ...
```

```
> (FisherPeriods<-2*pi/TrigoTrend$signFreq)
```

```
[1] 18.91667 56.75000 45.40000 20.63636 227.00000 14.18750 113.50000
[8] 13.35294
```

Významné frekvence, které vypočítá funkce `trigofit()`, jsou úhlové frekvence, zatímco funkce `tperiod()` dává jako výsledek indexy k odpovídající významným periodám. Abychom mohli výsledky navzájem srovnat, musíme je přepočítat.

Indexy k odpovídající významným periodám:

```
> pdgV$kF # v prvním radku jsou jména složek vektoru, výsledky jsou až ve druhém radku
```

```
13 5 6 12 2 17 3 18
12 4 5 11 1 16 2 17
```

```
> TrigoTrend$signFreq*n/(2*pi)
```

```
[1] 12  4  5 11  1 16  2 17
```

Významné frekvence ve tvaru $\frac{k}{n}$:

```
> pdgV$kF/n
```

```
      13      5      6      12      2      17
0.052863436 0.017621145 0.022026432 0.048458150 0.004405286 0.070484581
      3      18
0.008810573 0.074889868
```

```
> TrigoTrend$signFreq/(2*pi)
```

```
[1] 0.052863436 0.017621145 0.022026432 0.048458150 0.004405286 0.070484581
[7] 0.008810573 0.074889868
```

Významné úhlové frekvence (tvar $2\pi\frac{k}{n}$):

```
> 2*pi*pdgV$kF/n
```

```
      13      5      6      12      2      17      3
0.33215077 0.11071692 0.13839615 0.30447153 0.02767923 0.44286769 0.05535846
      18
0.47054692
```

```
> TrigoTrend$signFreq
```

```
[1] 0.33215077 0.11071692 0.13839615 0.30447153 0.02767923 0.44286769 0.05535846
[8] 0.47054692
```

Významné periody ve tvaru $\frac{n}{k}$:

```
> n/pdgV$kF
```

```
      13      5      6      12      2      17      3      18
18.91667  56.75000  45.40000  20.63636 227.00000  14.18750 113.50000  13.35294
```

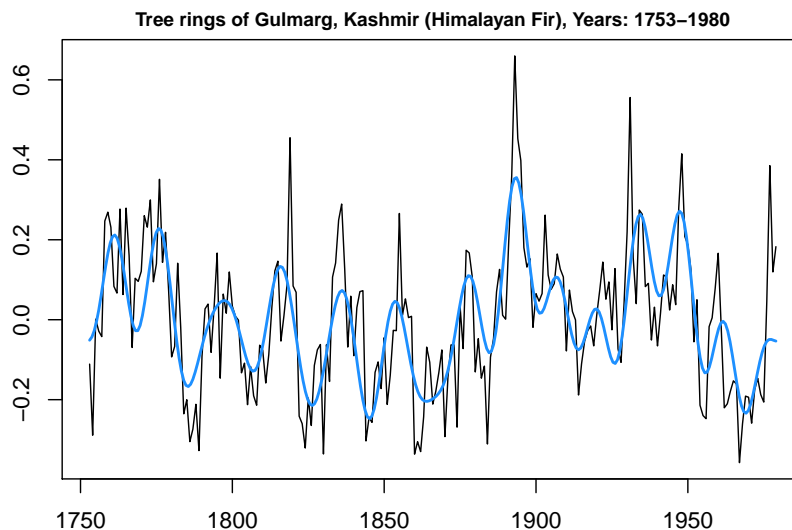
```
> 2*pi/TrigoTrend$signFreq
```

```
[1] 18.91667  56.75000  45.40000  20.63636 227.00000  14.18750 113.50000
[8] 13.35294
```

Nyní s využitím funkce `trigoval()` vykreslíme pro rezidua do grafu trigonometrický trend se zjištěnými významnými periodami. Za účelem vytvoření hladké křivky nejprve vytvoříme vektor `TTIME` se 300 ekvidistantními časovými okamžiky. Pak na něj aplikujeme již zmíněnou funkci `trigoval`, která vyžaduje ještě další tři argumenty: `A` – vektor odhadů α_j , `B` – odhady koeficientů β_j , `freq` – vektor s významnými úhlovými frekvencemi. Dosadíme do nich výsledky funkce `trigoval()`. Výsledkem funkce `trigoval()` pak bude vektor s hodnotami odhadnutého modelu s trigonometrickým trendem pro zadané časové okamžiky.

Výsledky znázorníme graficky obvyklým způsobem.

```
> TTIME<-seq(1,length(res),length=nn)
> predTrigoTrend<-trigoval(TTIME,A=TrigoTrend$signA,B=TrigoTrend$signB,freq=TrigoTrend$signFreq)
> par(mar=c(2,2,1,0)+0.5)
> plot(res~x, type="l",main=TXT,cex.main=0.85)
> lines(new$x,predTrigoTrend,col="dodgerblue",lwd=2)
```



Obrázek 4: Trigonometrický trend na reziduích pro šířky letokruhů

Na tatáž data zkusíme aplikovat iterativní metodu Damsletha a Spjøtvolla a porovnat výsledky obou metod. K tomu použijeme funkci `damsleth()` (další funkce ze souboru `FunkceM5201.R`). Jako vstup musíme této funkci zadat data, pro která má být model odhadován (v našem případě rezidua z modelu s lineárním trendem), a hladinu významnosti pro Fisherův test významných period (argument `pstlevel`).

```
> betaD<-damsleth(res,pstlevel=0.05)
> str(betaD)
```

List of 3

```
$ A : Named num [1:7] -0.0818 -0.0478 -0.0182 0.0454 0.0715 ...
.. attr(*, "names")= chr [1:7] "A" "A" "A" "A" ...
$ B : Named num [1:7] 0.01663 0.06242 0.07955 0.05614 -0.00469 ...
.. attr(*, "names")= chr [1:7] "B" "B" "B" "B" ...
$ freq: Named num [1:7] 0.3367 0.1107 0.1432 0.3094 0.0373 ...
.. attr(*, "names")= chr [1:7] "freq" "freq" "freq" "freq" ...
```

Funkce vrací jako výsledek seznam se třemi vektory: **freq** – významné úhlové frekvence, A a B – odhady koeficientů α_j a β_j pro trigonometrický trend s těmito periodami.

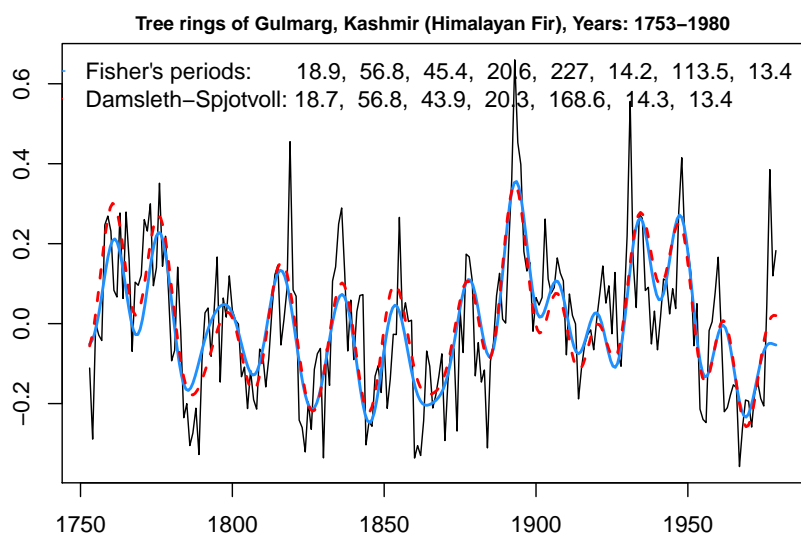
Podívejme se, k jakým významným periodám dospěl iterační algoritmus.

```
> (DamslethPeriods<-2*pi/betaD$freq)
```

```
      freq      freq      freq      freq      freq      freq      freq
18.66019  56.77548  43.87882  20.30496 168.56434  14.27805  13.41903
```

Nyní pro rezidua graficky znázorníme výsledky klasické i iterativní metody do jednoho grafu.

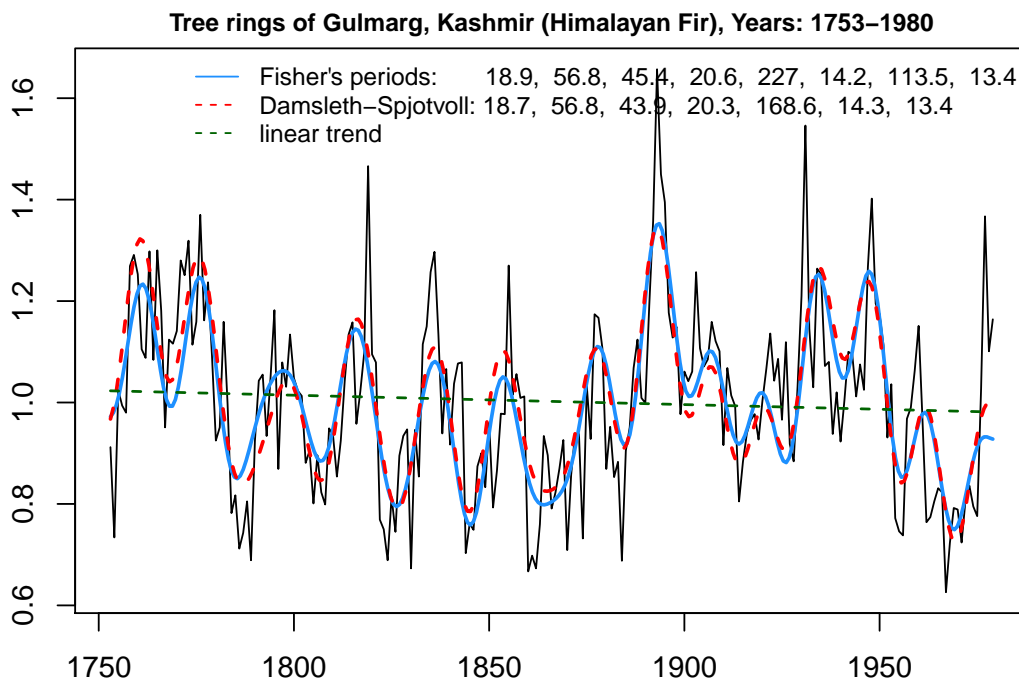
```
> predTrigD<-trigoval(TTIME,A=betaD$A,B=betaD$B,freq=betaD$freq)
> par(mar=c(2,2,1,0)+0.5)
> plot(res~x, type="l",main=TXT,cex.main=0.85)
> lines(new$x,predTrigoTrend,col="dodgerblue",lwd=2)
> lines(new$x,predTrigD,lwd=2,col="red",lty=2)
> TFP<-paste("Fisher's periods:      ",
  paste(round(FisherPeriods,1),collapse=" ", ))
> TDS<-paste("Damsleth-Spjotvoll:",
  paste(round(DamslethPeriods,1),collapse=" ", ))
> legend("topright", legend = c(TFP,TDS),bty="n",
  lty = c(1,2), xjust = 1, yjust = 1,col=c("dodgerblue","red"))
```



Obrázek 5: Trigonometrický trend (metoda skrytých period i iterativní metoda) na reziduích pro šířky letokruhů.

Na závěr přidáme i lineární trend a vše vykreslíme pro původní hodnoty časové řady.

```
> pred.lintrend<-predict(LinTrend, new)
> par(mar=c(2,2,1,0)+0.5)
> plot(y~x, type="l",main=TXT,cex.main=0.85)
> lines(new$x,predTrigoTrend+pred.lintrend,col="dodgerblue",lwd=2)
> lines(new$x,predTrigD+pred.lintrend,lwd=2,col="red",lty=2)
> lines(new$x,pred.lintrend,lwd=1.5,col="darkgreen",lty=2)
> legend("topright", legend = c(TFP,TDS,"linear trend"),bty="n",cex=0.8,
  lty = c(1,2,2), xjust = 1, yjust = 1,col=c("dodgerblue","red","darkgreen"))
```



Obrázek 6: Lineární i trigonometrický trend (metoda skrytých period i iterativní metoda) pro šířky letokruhů odpovídajících rokům 1753–1980

3 Úkol

Datový soubor `andrews2.dat` obsahuje údaje o ročním počtu získaných rysích kožešin týkající se kanadské společnosti Hudson's Bay company.

- Datový soubor načtěte.
- Časovou řadu vykreslete do grafu.
- Vypočítejte periodogram. (Předtím nezapomeňte z řady odstranit lineární trend.)
- Pokud naleznete nějaké významné periody:
 - Vykreslete do grafu časovou řadu (po odečtení lineárního trendu) spolu s odhadnutým trigonometrickým trendem.
 - Vyzkoušejte také odhadnout parametry v nelineární model s využitím algoritmu, který navrhli Damsleth a Spjøtvoll a výsledky znázorněte graficky.
 - Regresní křivky získané oběma metodami vynesete do jednoho grafu, pracujte při tom jak s časovou řadou po odstranění lineárního trendu, tak s časovou řadou bez odstraněného lineárního trendu.