

M0130 – 5. PRAKTIKUM : M0130pr05 (*Identifikace periodických komponent*)

A. REGRESNÍ MODELY CYKlickÝCH TRENDŮ.

1. NELINEÁRNÍ MODEL:

Uvažujme následující model s nekorelovanými a homoskedastickými rezidui:

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

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

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

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

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

nebo ekvivalentně

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

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

$$(a) \quad \alpha_1, \dots, \alpha_p \quad \beta_1, \dots, \beta_p \quad \mu \quad \lambda_1, \dots, \lambda_p$$

$$(b) \quad \gamma_1, \dots, \gamma_p \quad \mu \quad \lambda_1, \dots, \lambda_p \quad \omega_1, \dots, \omega_p$$

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

$$(a) \quad S(\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p, \lambda_1, \dots, \lambda_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$$

$$(b) \quad S(\mu, \gamma_1, \dots, \gamma_p, \omega_1, \dots, \omega_p, \lambda_1, \dots, \lambda_p) = \sum_{i=1}^n (Y_i - f(t_i))^2$$

Numericky lze systém nelineárních rovnic řešit např. pomocí Gaussovy-Newtonovy iterační metody.

V prostředí R lze v základním statistickém balíčku **stats** najít funkci **nls()** (*Nonlinear Least Squares*). Prostudujte ji např. pomocí příkazu **?nls**.

K určení period (resp. frekvencí), které se na tvorbě periodického trendu výrazně uplatňují je výhodné užít metod *spektrální analýzy časových řad*, např. tzv. *metodu skrytých period*, která vychází z hodnot periodogramu.

2. LINEÁRNÍ MODEL PRO ZNÁMÉ FREKVENCE

Situace se zjednoduší, pokud frekvence $\boxed{\lambda_1, \dots, \lambda_p}$ jsou známé. Pak model (a) je lineární, vektor neznámých parametrů lze psát takto $\beta = (\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p)'$ a matice plánu je tvaru

$$\mathbf{X}_{n \times (2p+1)} = \begin{pmatrix} 1 & \cos \lambda_1 t_1 & \sin \lambda_1 t_1 & \cdots & \cos \lambda_p t_1 & \sin \lambda_p t_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \cos \lambda_1 t_n & \sin \lambda_1 t_n & \cdots & \cos \lambda_p t_n & \sin \lambda_p t_n \end{pmatrix}.$$

Pak

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & \frac{n}{2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \cdots & \frac{n}{2} \end{pmatrix} \Rightarrow \hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \text{ tj. } \begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n Y_i \\ \hat{\alpha}_j = \frac{2}{n} \sum_{i=1}^n Y_i \cos \lambda_j t_i \\ \hat{\beta}_j = \frac{2}{n} \sum_{i=1}^n Y_i \sin \lambda_j t_i \end{cases}$$

pro $j = 1, \dots, p$.

Neznámé parametry modelu (b) získáme ze vztahů

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

3. METODA SKRYTÝCH PERIOD.

Uvažujme nyní takové časové řady, které můžeme rozložit na **součet harmonických frekvencí**, jejichž **délky period** lze 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
 $\omega_k = 2\pi f_k = 2\pi \frac{k}{n}$ k -tá úhlová frekvence

Maximální délka periody, kterou jsme schopni určit, 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 $T_{\min} = 2$. Této délce odpovídá frekvence $\omega_{\max} = \pi$, tzv. **Nyquistova frekvence**.

Z předchozích úvah vyplývá, že 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.

Pokud n je **sudé číslo**, může být mezi vybranými periodami i perioda délky $T_{\min} = 2$, což odpovídá frekvenci $\omega_{\max} = \pi$.

Vzhledem k tomu, že platí

$$\sin n\pi = 0 \text{ a } \cos n\pi = (-1)^n,$$

proto zapisujeme koeficient odpovídající této frekvenci zvlášť.

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 $\beta_0 = 0$.

V dalším pro jednoduchost předpokládejme, že n je **liché číslo**, jednoduchými úpravami dostaneme

$$\begin{aligned} s_t &= \frac{\alpha_0}{2} + \sum_{k=1}^p (\alpha_k \cos t\omega_k + \beta_k \sin t\omega_k) \\ &= \frac{\alpha_0}{2} + \sum_{k=1}^p \left[\frac{1}{2} (e^{it\omega_k} + e^{-it\omega_k}) \alpha_k + \frac{1}{2i} (e^{it\omega_k} - e^{-it\omega_k}) \beta_k \right] \\ &= \underbrace{\frac{\alpha_0}{2}}_{c_0} + \sum_{k=1}^p \left[\underbrace{\frac{1}{2}(\alpha_k - i\beta_k)}_{c_k} e^{it\omega_k} + \frac{1}{2}(\alpha_k + i\beta_k) e^{-it\omega_k} \right] = \sum_{k=-p}^p c_k e^{it\omega_k}. \end{aligned}$$

Všimněme si periodogramu náhodné posloupnosti $\{Y_t = s_t + \varepsilon_t, t \in \{t_1, \dots, t_n\}\}$

$$\begin{aligned} \boxed{I_n(\omega)} &= \frac{1}{2\pi n} \left| \sum_{t=1}^n \left(\sum_{j=-p}^p c_j e^{it\omega_j} + \varepsilon_t \right) e^{-it\omega} \right|^2 \\ &= \frac{1}{2\pi} \left| \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{j=-p}^p c_j e^{it(\omega_j - \omega)}}_{I_n^{(1)}(\omega)} + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t e^{-it\omega}}_{I_n^{(2)}(\omega)} \right|^2 \end{aligned}$$

První část

$$\boxed{I_n^{(1)}(\omega)} = \frac{1}{\sqrt{n}} \sum_{j=-p}^p c_j \underbrace{\sum_{t=1}^n e^{it(\omega_j - \omega)}}_{\text{součet geom.řady}} = \frac{1}{\sqrt{n}} \sum_{j=-p}^p c_j e^{i(\omega_j - \omega)} \underbrace{\frac{e^{in(\omega_j - \omega)} - 1}{e^{i(\omega_j - \omega)} - 1}}_{g_n(\omega_j - \omega)}.$$

Je-li ω různé od všech $\omega_1, \dots, \omega_p$, pak platí zřejmě $\lim_{n \rightarrow \infty} I_n^{(1)}(\omega) = 0$.

Existuje-li takové k , že platí $\omega = \omega_k$, pak

$$I_n^{(1)}(\omega_k) = n \frac{c_k}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{j=-p, j \neq k}^p g_j(\omega_j - \omega) = \underbrace{\sqrt{n}c_k}_{\rightarrow \infty} + \underbrace{\frac{1}{\sqrt{n}} \sum_{j=-p, j \neq k}^p g_n(\omega_j - \omega)}_{\rightarrow 0}$$

a pro $n \rightarrow \infty$ vzrůstá jeho absolutní hodnota nade všechny meze.

Druhý člen $I_n^{(2)}(\omega)$ je náhodná veličina s nulovou střední hodnotou a s rozptylem

$$\begin{aligned} DI_n^{(2)}(\omega) &= E \left| I_n^{(2)}(\omega) \right|^2 = E \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t e^{-it\omega} \right|^2 = \frac{1}{n} E \left(\sum_{t=1}^n \varepsilon_t e^{-it\omega} \right) \left(\sum_{s=1}^n \varepsilon_s e^{is\omega} \right) \\ &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n e^{-i(t-s)\omega} \underbrace{E\varepsilon_t \varepsilon_s}_{\text{nekorel.}} = \frac{1}{n} \sum_{t=1}^n E\varepsilon_t^2 = \frac{1}{n} n \sigma^2 = \sigma^2 \end{aligned}$$

Pokud tedy bude platit, že náhodná posloupnost splňuje model

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

bude mít **periodogram** pro velká n v bodech $\omega = \lambda_1, \dots, \lambda_p$ výrazně velké hodnoty, jinde jeho hodnoty budou relativně malé, budou kolísat kolem hodnoty $\frac{\sigma^2}{2\pi}$. **Periodogram je tedy dobrým ukazatelem periodicit.**

Z výše uvedených vlastností vyplývá, že významná lokální maxima v průběhu periodogramu by měla identifikovat periodickou strukturu uvažovaného modelu tak, že vyznačí neznámé frekvence $\omega_1, \dots, \omega_p$.

Nějaký vhodný statistický test by pak měl rozhodnout, jaké hodnoty periodogramu můžeme opravdu považovat za významně velké ve srovnání s hodnotami ostatními.

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 .

Budeme **testovat hypotézu**:

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

Předpokládejme, že $n = 2m + 1$ je liché číslo (je-li n sudé, obvykle se vynechá první člen jakožto časově nejvzdálenější od současnosti) 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 . Položme

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

(tzv. **Fisherova statistika**), která nabývá hodnot mezi nulou a jedničkou.

Budou-li všechny veličiny **téměř stejné**, bude hodnota W blízká číslu $\frac{1}{m}$.

Bude-li naopak veličina V_1 nabývat **velmi vysokých hodnot ve srovnání s ostatními veličinami** V_2, \dots, V_m , bude hodnota W blízká číslu 1 .

Je tedy vidět, že velké hodnoty (které jsou blízké 1) budou tvořit kritický obor naší 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 (viz Anděl, J.: *Statistická analýza časových řad*, Praha 1976, str. 79-86) 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

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

Pak **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*.

V případě, že pomocí Fisherova testu **zjistíme signifikantní periodicitu určité frekvence**, je na místě otázka, jak statisticky testovat případné další periodicity o jiných frekvencích.

Whittle doporučil, aby se v případě významnosti největší hodnoty periodogramu V_1 tato hodnota vynechala. Dále pak 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.

SIEGELŮV TEST

V praxi se ukázalo, že při platnosti alternativní hypotézy s $p > 1$ (tzn., že se uplatňuje více než jedna periodičita) nezamítá Fisherův test nulovou hypotézu s příliš velkou pravděpodobností.

Jinými slovy lze říci, že Fisherův test nemá při alternativní hypotéze pro $p > 1$ takovou sílu jako v případě, kdy $p = 1$. (Tehdy je jeho síla přijatelná a dokonce v jistém smyslu optimální).

Proto byly hledány modifikace, které tento nedostatek odstraňují. Uvedeme tzv. **Siegelův test**. Zde se místo statistiky W 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 $(x)_+ = \max(x, 0)$, $\boxed{g_F}$ 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 pak zamítáme, pokud $T_\lambda > t_\lambda$, kde $\boxed{t_\lambda}$ 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 .

Je třeba však mít na paměti, že 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**.

4. ITERATIVNÍ METODA (DAMSLETH A SPJØTVOLL, 1982).

Popíšeme nyní iterativní metodu, kterou v roce 1982 navrhli Norové Damsleth a Spjøtvoll.

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

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

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**

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

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

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

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

a s řadou Z_i jdeme na bod (1).

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

PŘÍKLAD 1

Opět se vrátíme k témuž datovému souboru jako v předešlém praktiku. Jde o průměrné roční průtoky vody v řece Nigeru v Coulicouro (Mali). Údaje se vztahují k období 1907 až 1957 a jsou uvedeny v $\text{cfs} \cdot 10^{-3}$. Data jsou uložena v souboru `DataNiger.dat`. První řádek obsahuje popis časové řady, pak následuje volný řádek, teprve potom dva sloupce dat: rok a průměrný roční průtok.

Nejprve načteme nadpis.

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

```
[1] "Prumerne rocni prutoky vody v rece Nigeru v Coulicoure (Mali) v letech 1907 az 1957"
```

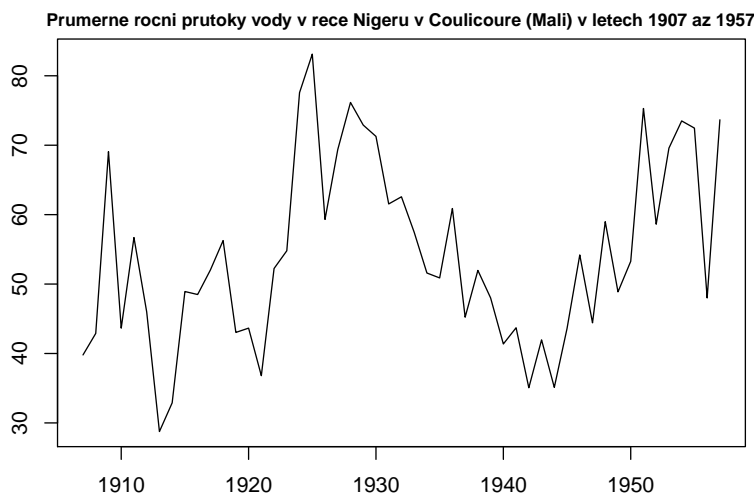
Pak načteme vlastní data do datového rámce, vypíšeme jeho strukturu.

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

'data.frame':      51 obs. of  2 variables:
 $ V1: int   1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 ...
 $ V2: num   39.8 42.9 69.1 43.7 56.7 ...
```

Z načtených dat vytvoříme časovou řadu a vykreslíme ji.

```
> NigerTS <- ts(dataNiger[, 2], start = 1907, frequency = 1)
> par(mar = c(2, 2, 1, 0) + 0.5)
> plot(NigerTS, main = TXT, cex.main = 0.85)
```



Obrázek 1: Průměrné roční průtoky vody v řece Nigeru v Coulicouro (Mali).

Hledání významných period je založeno na periodogramu. Nejprve je však třeba odstranit lineární trend. Do grafu kromě lineárního trendu vykreslíme také intervaly spolehlivosti kolem trendu a predikční interval.

```

> y <- as.vector(NigerTS)
> x <- as.vector(time(NigerTS))
> 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
-21.667  -9.092  -2.412   7.895  30.325

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) -351.6125    239.0573  -1.471  0.1477
x              0.2101     0.1237   1.698  0.0959 .
---

```

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

Residual standard error: 13.01 on 49 degrees of freedom

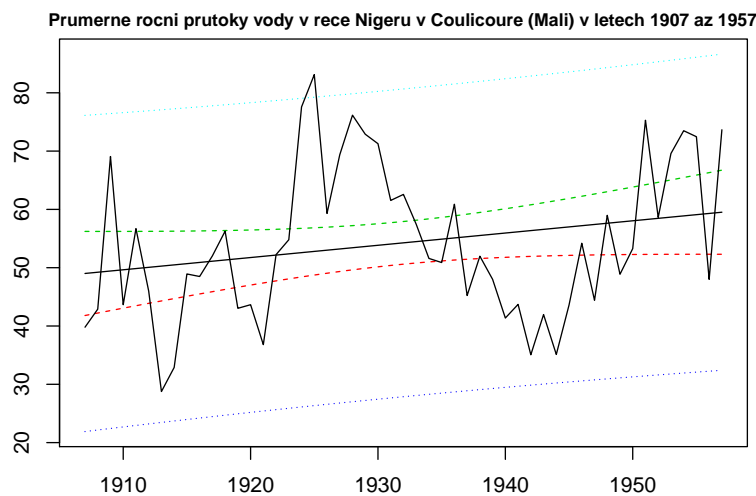
Multiple R-squared: 0.05556, Adjusted R-squared: 0.03629

F-statistic: 2.883 on 1 and 49 DF, p-value: 0.09587

```

> 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)
> matplot(new$x, cbind(pred.w.clim, pred.w.plim[, -1]), lty = c(1,
  2, 2, 3, 3), type = "l", ylab = "predicted y")
> lines(x, y)
> title(main = TXT, cex.main = 0.85)

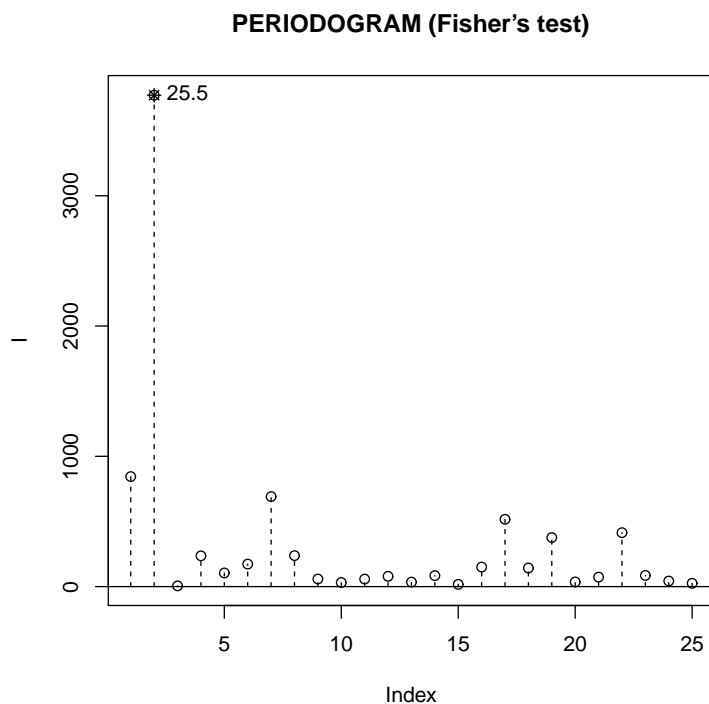
```



Obrázek 2: Lineární trend pro průměrné roční průtoky vody v řece Nigeru v Coulicoure (Mali).

I když není lineární trend statisticky významný, přesto jej odstraníme a vstupem do periodogramu (funkce `\tperiod()`) budou rezidua. Tato funkce na požádání periodogram s významnými periodami vykreslí.

```
> res <- resid(LinTrend)
> pdgV <- tperiod(res, TRUE)
```



Obrázek 3: Periodogram pro průměrné roční průtoky vody v řece Nigeru v Coulicouro (Mali).

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

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

```
3
25.5
```

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

```
3
25.5
```

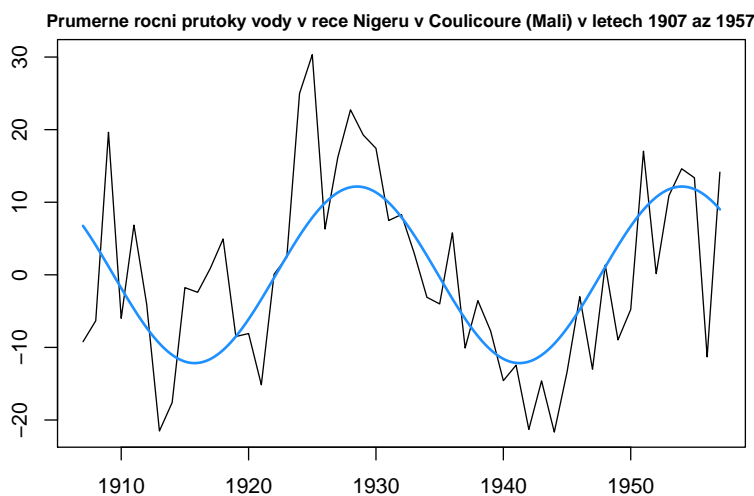
Chceme-li explicitně zjistit a do grafu vykreslit trigonometrický trend, použijeme k tomu dvě funkce, a to `\trigofit()` (jde o metodu skrytých period) a `\trigoval()`

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

List of 10

```
$ periodogram: num [1:25] 67.193 300.132 0.456 18.79 8.333 ...
$ freq       : num [1:25] 0.123 0.246 0.37 0.493 0.616 ...
$ signA      : num 9.01
$ signB      : num -8.17
$ signFreq   : num 0.246
$ amplituda  : num [1:25] 1.1395 2.4084 0.0939 0.6026 0.4013 ...
$ faze       : num [1:25] 6.779 -1.277 -0.381 0.909 0.729 ...
$ signL      : logi [1:25] FALSE TRUE FALSE FALSE FALSE ...
$ A          : num [1:25] -2.9 9.01 -0.13 -2.45 1.71 ...
$ B          : num [1:25] 4.973 -8.168 0.456 -1.807 1.08 ...
```

```
> 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 rezidua pro průměrné roční průtoky vody v řece Nigeru v Coulicouro (Mali).

Chceme-li použít Iterativní metodu (Damsleth a Spjøtvoll, 1982) použijeme k tomu funkci `\damsleth()` a graficky znázorníme klasickou metodu i iterativní do jednoho grafu.

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

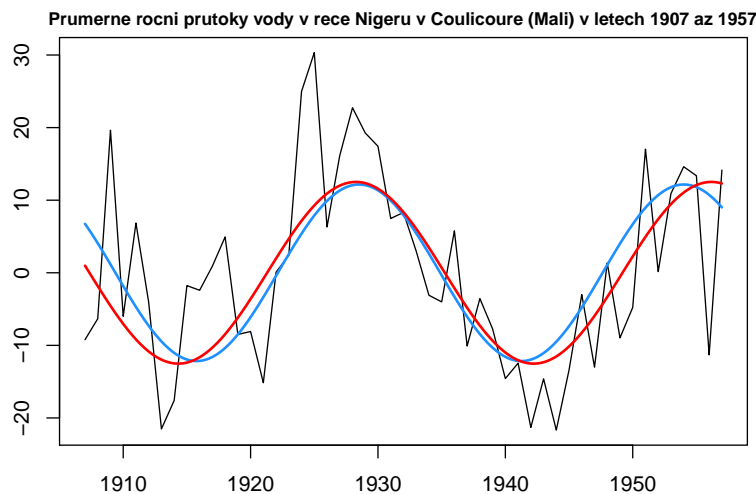
List of 3

```
$ A : Named num 3.74
..- attr(*, "names")= chr "A"
$ B : Named num -11.9
..- attr(*, "names")= chr "B"
$ freq: Named num 0.225
..- attr(*, "names")= chr "freq"
```

```

> 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")

```



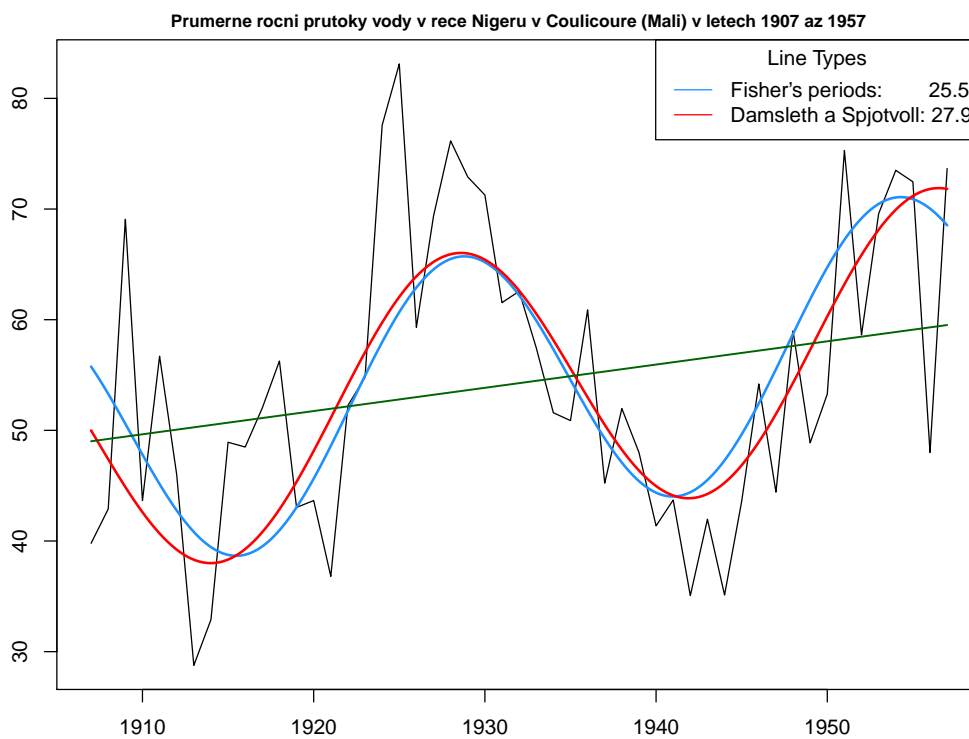
Obrázek 5: Trigonometrický trend (metoda skrytých period i iterativní metoda) na reziduua pro průměrné roční průtoky vody v řece Nigeru v Coulicouro (Mali).

Na závěr přidáme i lineární trend a vše vykreslíme

```

> 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")
> lines(new$x, pred.lintrend, lwd = 1.5, col = "darkgreen")
> FisherPeriods <- 2 * pi/TrigoTrend$signFreq
> DamslethPeriods <- 2 * pi/betaD$freq
> TFP <- paste("Fisher's periods:      ", paste(round(FisherPeriods,
  1), collapse = ", "))
> TDS <- paste("Damsleth a Spjotvoll:", paste(round(DamslethPeriods,
  1), collapse = ", "))
> legend("topright", legend = c(TFP, TDS), lty = 1, xjust = 1,
  yjust = 1, col = c("dodgerblue", "red"), title = "Line Types")

```



Obrázek 6: Lineární i trigonometrický trend (metoda skrytých period i iterativní metoda) pro průměrné roční průtoky vody v řece Nigeru v Coulicoure (Mali).