Úvod, základné pojmy, testovanie bieleho šumu

Beáta Stehlíková

2-PMS-10 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Analýza časových radov: úvod

Klasický vzorový príklad

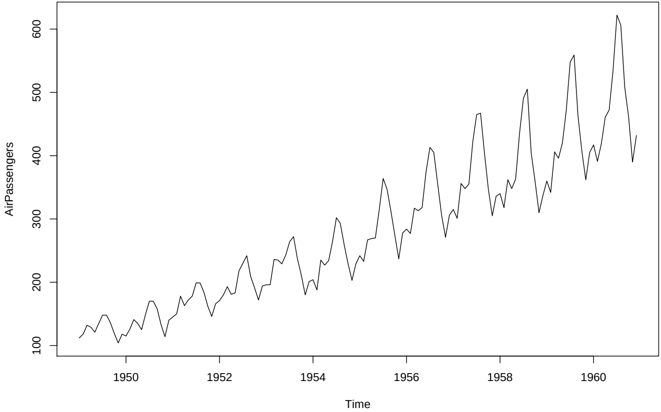
- Pozrieme sa na dáta - počty cestujúcich aerolinkami
- Dáta `AirPassengers` z balíka `datasets`
- Popis dát v dokumentácii (pomocou `?AirPassengers`):
  - The classic Box & Jenkins airline data. Monthly totals of international airline passengers, 1949 to 1960.
  - A monthly time series, in thousands.

```r
library(datasets)
AirPassengers
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
```

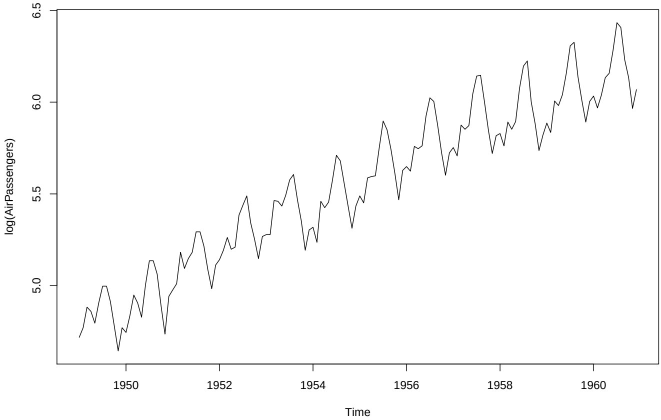
Klasický vzorový príklad - priebeh dát

Stačí `plot(AirPassengers)` a vďaka časovej štruktúre dát máme:



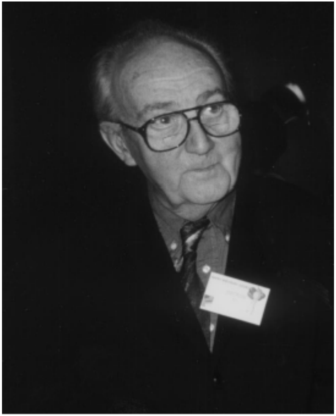
Klasický vzorový príklad - priebeh dát

Po zlogaritmovaní sa stabilizuje volatilita:



Box a Jenkins

Budeme sa zaoberať **prístupom od Boxa a Jenkinsa**.



Rozhovor s G. E. P. Boxom po oslave jeho 80. narodenín (1999), ako sa začal zaujímať o štatistiku a ďalšie otázky.

> *The first paper you wrote with Jenkins has been considered as a breakthrough in statistics.*

*P*eña, D. (2001). George Box: An interview with the International Journal of Forecasting. International Journal of Forecasting, 17(1), 1-9.

Odkaz na článok s rozhovorom:
https://www.sciencedirect.com/science/article/pii/S0169207000000613

Informačný list predmetu

Výsledky vzdelávania:

Študenti sa oboznámia s Box-Jenkinsovou metodológiou modelovania časových radov pomocou ARIMA procesov. Budú poznať teoretické vlastnosti týchto modelov a budú ich vedieť použiť na analýzu reálnych dát.

Základné pojmy

Obsah

- Časový rad, momenty
- Stacionarita, ergodicita
- Biely šum
- Autokorelačná funkcia
- Woldova reprezentácia
- Testy o autokorelačnej funkcii

Momenty časového radu

- Náhodný proces $x_1, x_2, \ldots, x_T$ je úplne charakterizovaný svojou T-rozmernou distribučnou funkciou
- Obvykle sa zameriavame na **prvé dva momenty**:
  - stredná hodnota $E(x_t)$
  - variancia $D(x_t)$
  - kovariancie $Cov(x_t, x_s)$, tzv. autokovariancie

Stacionarita a ergodicita

- Väčšinou máme len jeden časový rad - jednu realizáciu náhodného procesu $\rightarrow$ aby sa dala robiť štatistická inferencia, potrebujeme dodatočné predpoklady
- Napríklad: na to, aby sme odhadli strednú hodnotu, ... potrebujeme viac ako jednu realizáciu tejto náhodnej premennej
- Ergodický proces - výberové momenty počítané z časového radu s T pozorovaniami konvergujú pre $T \rightarrow \infty$ k zodpovedajúcim momentom
- Tento koncept má zmysel iba ak predpokladáme, že $E(x_t) = \mu$, $D(x_t) = \sigma^2, \ldots$ pre každé t

Stacionarita a ergodicita

- Silná stacionarita: združená distribučná funkcia sa nemení pri posune v čase
- Obvykle sa pracuje so slabším predpokladom $\rightarrow$ slabá stacionarita:

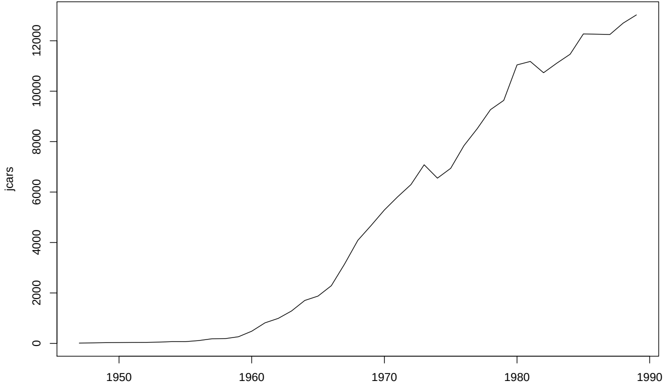
$$\begin{aligned} E(x_t) &= \mu \ \forall t & (1) \\ Cov(x_t, x_s) &= \gamma(|t-s|) \ \forall t, s & (2) \end{aligned}$$

- Z (2) vyplýva, že $D(x_t) = const.$ pre všetky t.
- Ďalej budeme pod stacionaritou rozumieť slabú stacionaritu.

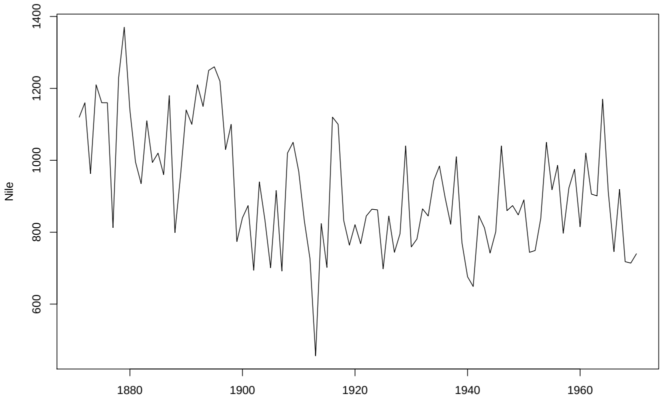
Stacionarita - dáta

- *Stacionárny časový rad*:
  - dáta sú priťahované k určitej rovnovážnej hodnote, okolo ktorej oscilujú
- *Nestacionárny časový rad*:
  - napríklad trend: rastúci trend $\rightarrow$ stredná hodnota nie je konštantná $\rightarrow$ proces nie je stacionárny
  - neskôr budeme vidieť aj iné druhy nestacionarity (napr. zatiaľ nejasne znejúci pojem *jednotkový koreň* v sylabe predmetu)

```r
# PRIKLAD 1
# Japanese motor vehicle production in thousand (1947-1989)
library(fma)
plot(jcars)
```



```
# PRIKLAD 2
# Rocny prietok Nilu v Aswane v 10^8 m^3 (stavba priehrad)
library(datasets)
plot(Nile)
```

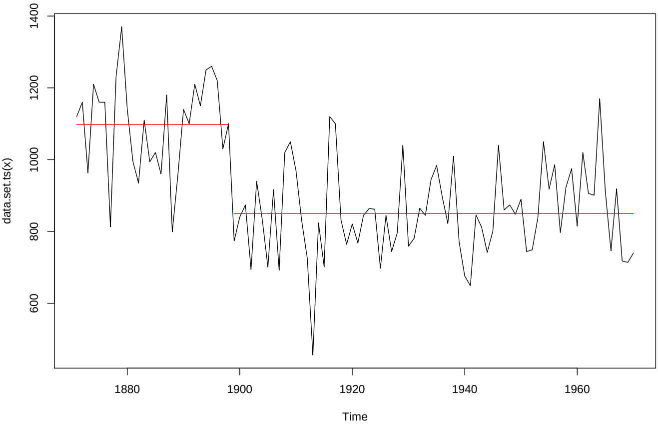


Nepovinné poznámky: Hľadanie bodov zmeny

- Tutoriál (useR! International R User 2017 Conference): https://www.youtube.com/watch?v=I7jUBro78RM
- Diplomovka Lucie Macháčkovej, zmeny v prietokoch slovenských riek (mEMM 2020, školiteľ: doc. Pekár)
- Článok s prehľadom rôznych metód: *Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. Knowledge and information systems, 51(2), 339-367.* https://link.springer.com/article/10.1007/s10115-016-0987-z

```r
# pozrime sa na prietoky Nilu:
library(changepoint)
zmena <- cpt.mean(Nile)
plot(zmena)
```

Výstup z predchádzajúceho slajdu:



Biely šum

Definícia bieleho šumu

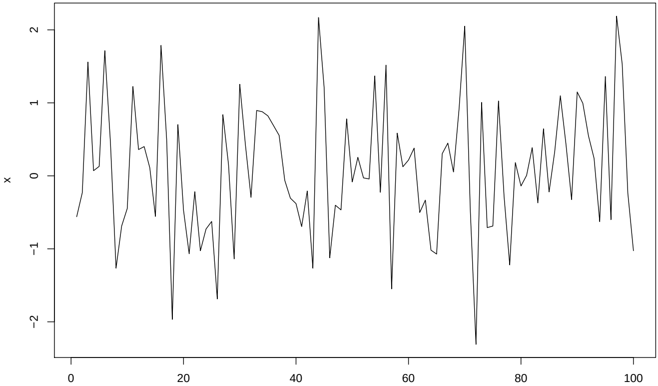
- Dôležitý príklad stacionárneho procesu, pomocou ktorého budeme definovať aj rôzne modely pre dáta
- Biely šum u_t je náhodný proces s nasledujúcimi vlastnosťami

$$
\begin{aligned}
\mathbb{E}(u_t) &= 0 \ \forall t \\
\mathbb{D}(u_t) &= \sigma^2 \ \forall t \\
\mathrm{Cov}(u_t, u_s) &= 0 \ \forall t \neq s
\end{aligned}
$$

- Napríklad postupnosť nezávislých náhodných premenných s rovnakým rozdelením (a konečnou strednou hodnotou a disperziou), ale nie je to jediná možnosť

Príklad

```
x <- rnorm(100)
plot(x, type = "l")
```



Príklady: zisťovanie stacionarity procesu

Príklad 1

- Nech u_t je biely šum, definujme

$$x_t = u_t + u_{t-1}$$

- Platí:

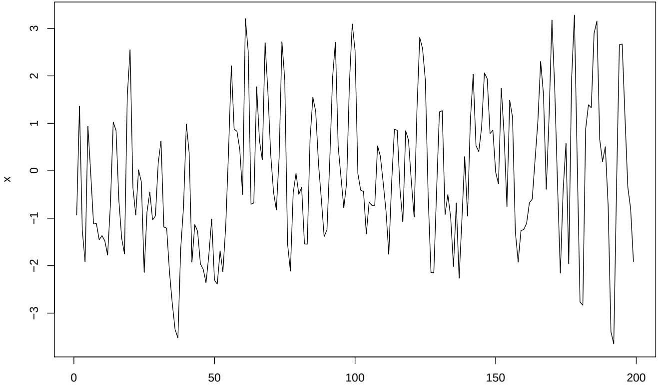
$$\mathbb{E}(x_t) = 0, \mathbb{D}(x_t) = 2\sigma^2$$

$$\mathrm{Cov}(x_t, x_{t+k}) = \begin{cases} \sigma^2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \dots \end{cases}$$

- Proces teda **je stacionárny**

Príklad 1: simulácia priebehu procesu, N = 200 pozorovaní

```
u <- rnorm(N + 1)
x <- u[2:N] + u[1:(N - 1)]
```



Príklad 2

- Nech u_t je biely šum, definujme

$$x_t = \begin{cases} u_1 & \text{pre } t = 1, \\ x_{t-1} + u_t & \text{pre } t = 2, 3, \dots \end{cases}$$

- x_t sa dá zapísať v tvare $x_t = \sum_{i=1}^{t} u_i$
- Platí:

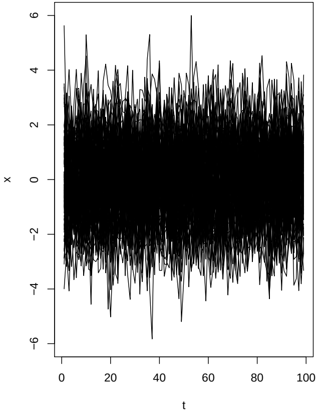
$$\mathbb{E}(x_t) = 0, \mathbb{D}(x_t) = t\sigma^2$$

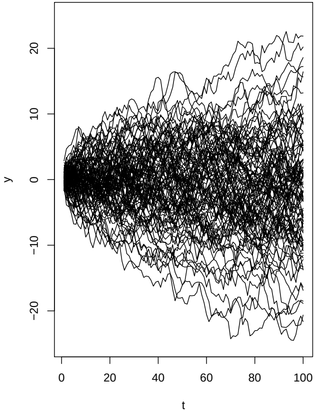
$$\mathrm{Cov}(x_t, x_{t+k}) = t\sigma^2 \quad \text{pre } k > 0$$

- Proces teda **nie je stacionárny** (to vieme povedať už po výpočte disperzie)

Príklady 1, 2: porovnanie simulácií procesov

- Vľavo: stacionárny proces z príkladu 1, vpravo: proces s rastúcou disperziou z príkladu 2





Príklad 3

- Nech u_t je biely šum, definujme

$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}, \tag{3}$$

kde koeficienty ψ_j spĺňajú $\psi_0 = 1$, $\sum \psi_j^2 < \infty$

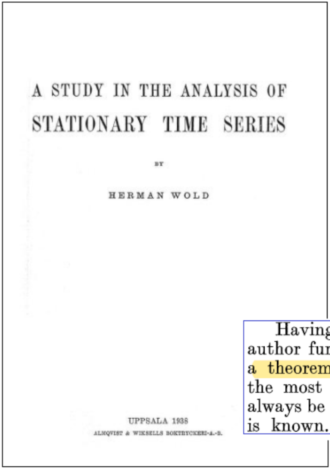
- Platí:

$$\mathbb{E}(x_t) = \mu, \mathbb{D}(x_t) = \sigma^2 \sum_{j=1}^{\infty} \psi_j^2$$

$$\mathrm{Cov}(x_t, x_{t+k}) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{k+j}$$

- Proces **je stacionárny**.

Woldova reprezentácia


A STUDY IN THE ANALYSIS OF

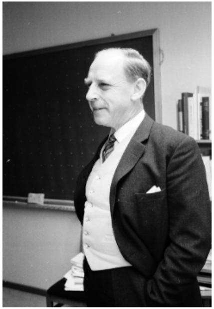
STATIONARY TIME SERIES

BY

HERMAN WOLD

UPPSALA 1938

ALMQVIST & WIKSELLS BOKTRYCKERI-A.-B.



Having discussed a number of types of random processes, the author furnishes proofs of their various properties, and then gives a theorem of considerable interest, concerning the structure of the most general discrete stationary process. This, in fact, can always be presented as a sum of two components the nature of which is known.

J. N. (1939). Journal of the Royal Statistical Society, 102(2), 295-298.

https://archive.org/details/in.ernet.dli.2015.262214
https://www.jstor.org/stable/298000

https://digitaltmuseum.se/021016543711/professor-herman-wold-uppsala-1969

Woldova reprezentácia stacionárneho procesu

- V príklade 3: Proces tvaru (3) je stacionárny
- Dá sa dokázať: Každý stacionárny proces x_t sa dá zapísať v tvare

$$x_t = \mu_t + \sum_{j=0}^{\infty} \psi_j u_{t-j}$$

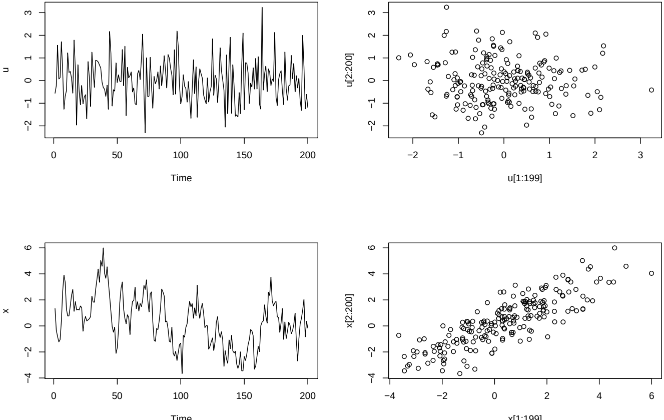
kde u je biely šum, $\psi_0 = 1$, $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ a μ_t sa dá presne predikovať z predchádzajúcich hodnôt procesu x (v našich aplikáciách to bude konštanta)

- Toto vyjadrenie sa nazýva Woldova reprezentácia

Autokorelačná funkcia (ACF)

Motivácia

- Hore: u_t, dolu: $x_t = 0.9x_{t-1} + u_t$ - simulácia a závislosť x_t od x_{t-1}



Definícia základné vlastnosti

- Autokorelačná funkcia (ACF) stacionárneho procesu je definovaná ako

$$\rho(\tau) = \mathrm{cor}(x_t, x_{t+\tau}) = \frac{\mathrm{cov}(x_t, x_{t+\tau})}{\sqrt{\mathbb{D}(x_t)\mathbb{D}(x_{t+\tau})}}$$

- ACF sa teda dá vyjadriť pomocou autokovariančnej funkcie γ:

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)},$$

- Platí:

$$\rho(0) = 1, \rho(-\tau) = \rho(\tau),$$

stačí nám teda počítať $\rho(\tau)$ pre $\tau = 1, 2, \ldots$

Príklad

- Nech u_t je biely šum, definujme

$$x_t = u_t + u_{t-1}$$

- Pre tento proces sme odvodili stacionaritu a vlastnosti:

$$\mathbb{D}(x_t) = 2\sigma^2, \mathrm{Cov}(x_t, x_{t+k}) = \begin{cases} \sigma^2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \dots \end{cases}$$

- ACF teda je

$$\rho(k) = \mathrm{Cor}(x_t, x_{t+k}) = \begin{cases} 1/2 & \text{pre } k = 1, \\ 0 & \text{pre } k = 2, 3, \dots \end{cases}$$

Odhadovanie ACF z dát

- Ergodický proces $\rightarrow$ stredná hodnota, disperzia a autokovariancie sa dajú konzistentne odhadnúť z dát $x_1, \ldots, x_T$:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t, \quad \hat{\gamma}(0) = \frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{\mu})^2$$

$$\hat{\gamma}(\tau) = \sum_{t=1}^{T-\tau}(x_t - \hat{\mu})(x_{t+\tau} - \hat{\mu})$$

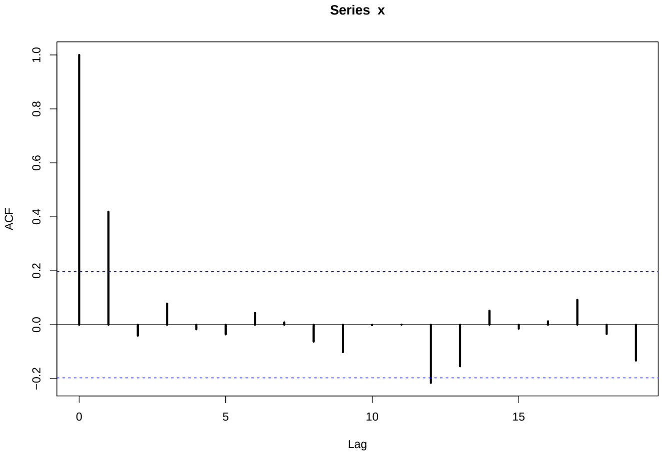
- Z toho - konzistentný odhad autokorelačnej funkcie:

$$\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$$

- je asymptoticky nevychýlený

Odhadovanie ACF z dát v R-ku: funkcia acf

```
acf(x) # pre data x <- u[2:N] + u[1:(N - 1)]
```



Využitie na kontrolu výpočtov

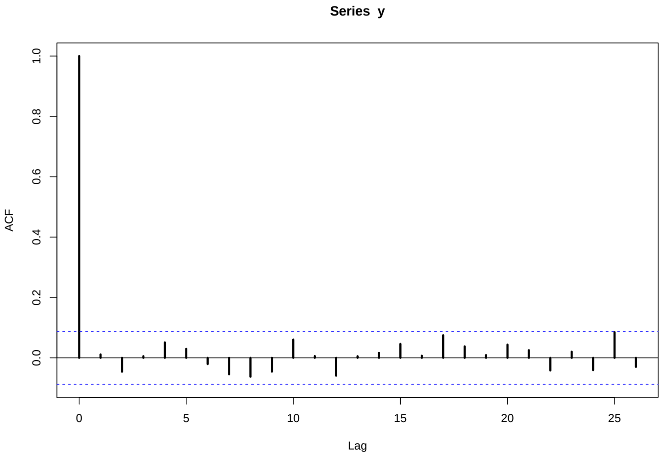
- *Zadanie:* Nech z_t je proces, ktorého hodnoty sú nezávislé náhodné premenné s rozdelením $N(0,1)$. Ukážte, že nasledujúci proces je stacionárny a vypočítajte jeho ACF:

$$y_t = \begin{cases} z_t & \text{pre } t \text{ nepárne} \\ \frac{1}{\sqrt{2}}(z_{t-1}^2 - 1) & \text{pre } t \text{ párne} \end{cases}$$

- Vygenerujeme si daný proces a zobrazíme odhadnutú ACF (náš výpočet by mal dať podobný výsledok):

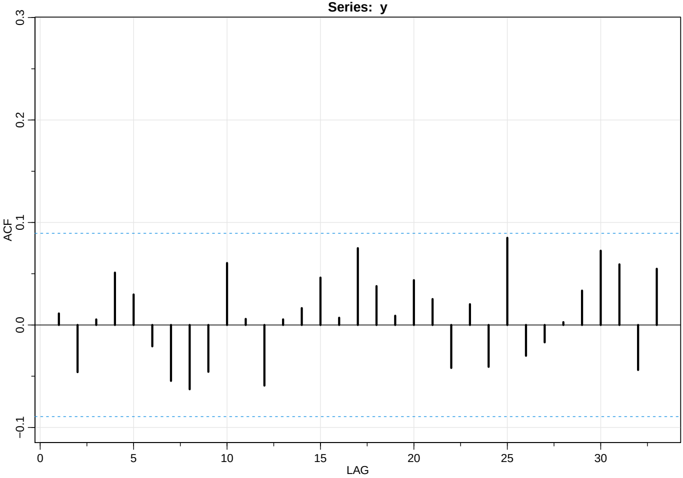
```
set.seed(1234)
N <- 500
z <- rnorm(N, mean = 0, sd = 1)
y <- z  # pre neparne indexy zostane, parne upravime
ind.parne <- seq(from = 2, to = N, by = 2)
y[ind.parne] <- (1/sqrt(2)) * (z[ind.parne - 1]^2 - 1)
```

```
acf(y)
```



Alternatíva: funkcia acf1 z balíka astsa

```
library(astsa); acf1(y)
```



Testovanie nulovosti autokorelácií

Testovanie nulovosti autokorelácií - každej samostatne

- Biely šum má nulovú ACF $\Rightarrow$ pri testovaní, či sú dáta bielym šumom, budeme testovať, či majú nulové autokorelácie
- Odhad ACF v prípade bieleho šumu
  - asymptoticky nevychýlený
  - disperzia $\approx 1/T$
  - $\Rightarrow$ približný 95 % interval spoľahlivosti: $\pm 1.96/\sqrt{T}$, resp. $\pm 2/\sqrt{T}$ - často sa zobrazuje spolu s odhadnutými autokoreláciami (aj v prípade funkcií `acf` a `acf1`)
- Pre každú autokoreláciu samostatne:
  - Testujeme, či sa rovná nule.
  - Nulovú hypotézu zamietame, ak je jej odhad mimo intervalu spoľahlivosti

Ak testujeme nulovosť autokorelácie, nie pre biely šum:

- V prípade procesu, pre ktorý platí $\rho(\tau) = 0$ pre $\tau > k$, pre tieto τ platí
$$\mathbb{D}(\hat{\rho}(\tau)) \approx \frac{1}{T}\left(1 + 2\sum_{j=1}^{k}\hat{\rho}(j)^2\right)$$

Testovanie nulovosti autokorelácií - Ljung-Boxov test

- Netestujeme nulovosť každej autokorelácie samostatne, ale testujeme hypotézu $\rho(1) = \rho(2) = \cdots = \rho(m) = 0$
- Box & Pierce, 1970: ak platí H_0, asymptoticky

$$Q = T \sum_{j=1}^{m} \hat{\rho}(j)^2 \sim \chi^2_m$$

- Ljung & Box, 1978: modifikácia s lepšími vlastnosťami pri menšom počte dát

$$Q = T(T+2) \sum_{j=1}^{m} \frac{\hat{\rho}(j)^2}{T-j} \sim \chi^2_m$$

- *Poznámka, ktorú využijeme neskôr: Počet stupňov voľnosti sa zmení, ak ide o rezíduá z modelu*

Ljung-Boxov test v R-ku: funkcia Box.test

- Testujme pre dáta x, že prvé tri autokorelácie sú nulové

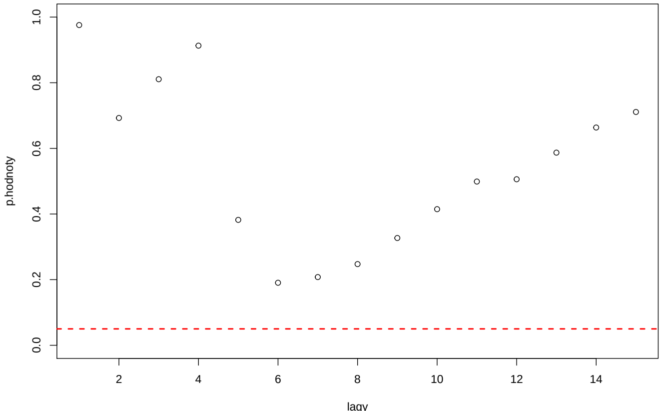
```
set.seed(12345)
x <- rnorm(N)

Box.test(x, lag = 3, type = "Ljung-Box") # staci `Ljung`
```

```
##
##  Box-Ljung test
##
## data:  x
## X-squared = 0.9611, df = 3, p-value = 0.8107
```

Užitočný výstup pre LB test: rôzny počet lagov

- Zobrazíme hranicu 0.05 a P hodnoty (prístup napr. `Box.test(x, lag = 3, type = "Ljung")$p.value`):



Cvičenia

Cvičenie 1: Ljung-Boxov test - opakovanie

```
##
##  Box-Ljung test
##
## data:  x
## X-squared = 4.4469, df = 3, p-value = 0.2171
```

- Aká hypotéza sa tu testuje?
- Vysvetlite, čo presne znamenajú hodnoty uvedené s popisom `X-squared`, `df` a `p-value`.
- Zamietame nulovú hypotézu alebo nie?
- Čo nám tento test hovorí o dátach? Môže ísť o biely šum alebo je to nepravdepodobné? Aké ďalšie testy by ste spravili, aby ste vedeli lepšie odpovedať na túto otázku?

- Ako sa počíta testovacia štatistika?
- Vypočítajte jej hodnotu, ak viete, že dáta obsahujú 50 pozorovaní a prvé hodnoty autokorelačnej funkcie sú nasledovné:

```
acf(x, lag.max = 5, plot = FALSE)
```

```
##
## Autocorrelations of series 'x', by lag
##
##      0      1      2      3      4      5
##  1.000  0.026 -0.171  0.226 -0.106  0.048
```

Až na zaokrúhľovacie chyby nám má výjsť výsledok z výstupu testu (ten sa počíta z presných, nie zaokrúhlených hodnôt ACF).

Cvičenie 2: Ljung-Boxov test pre prietoky Nílu

- Uvažujme dáta o prietoku Nílu z prednášky po postavení priehrad, napr. od roku 1910. Sú v premennej `Nile`, ktorá je typu `ts` (*time series*):

```
class(Nile)
```

```
## [1] "ts"
```

- Kratší časový rad z nej vytvoríme pomocou funkcie `window` s parametrami `start` a/alebo `end`:

```
x <- window(Nile, start = 1910)
```

- Zobrazte ACF a LB testom testujte hypotézu, že hladiny Nílu sú nekorelované, výsledok zobrazte graficky (ako na str.44).

Cvičenie 3: Ljung-Boxov test pre úrokové miery

- Budeme analyzovať dáta dlhodobých úrokových mierach z Európskej centrálnej banky, pričom použijeme mesačné dáta z obdobia 20 rokov (2001-2021). Využijeme ich aj neskôr, pri ďalších témach.
  https://sdw.ecb.europa.eu/browse.do?node=bbn4864
- Na stránke predmetu je csv-súbor stiahnutý z webu ECB v auguste (pri aktuálnom stiahnutí bude obsahovať aj novšie dáta).

	A	B	C	
1	Data Source in SDW: https://sdw.ecb.europa.eu/browse.do?node=bbn4864			
2		IRS.M.AT.L.L40.CI.0000.EUR.N.Z	IRS.M.BE.L.L40.CI.0000.EUR.N.Z	IRS.M.BG.L.L40.(
3		Austria, Euro	Belgium, Euro	Bulgaria, Bulgaria
4	Collection:	Average of observations through period (A)	Average of observations through period (A)	Average of observ
5	Period\Unit:	[Percent]	[Percent]	[Percent]
6	2022Jul	1.70	1.80	1.85
7	2022Jun	2.07	2.13	1.77
8	2022May	1.54	1.58	1.62
9	2022Apr	1.29	1.30	1.62
10	2022Mar	0.72	0.79	1.09
11	2022Feb	0.54	0.59	0.61
12	2022Jan	0.18	0.26	0.57

- Načítame dáta:

```
# DOPLNTE
data <- read.csv(   ,       # cesta k suboru
                 skip =   , # vynechanie riadkov,
                            # chceme iba data
                 header = FALSE, # vynechali sme riadky
                 row.names = 1)  # prvy sltpec nie su data
                                 # tieto datumy pouzijeme
                                 # ako nazvy riadkov
```

```
head(data)
```

```
##             V2   V3   V4   V5   V6    V7    V8   V9  V1
## 2021Dec -0.05  0.03 0.44 0.64 2.62 -0.38 -0.08 0.11 0.4
## 2021Nov -0.01  0.07 0.25 0.55 2.62 -0.31  0.01 0.20 0.4
## 2021Oct  0.08  0.16 0.25 0.39 2.34 -0.21  0.10 0.18 0.4
```

- Získajte zo súboru s dátami názvy (štáty a meny):

```
# DOPLNTE
nazvy <- read.csv(...)
```

```
head(nazvy)
```

```
##       [,1]
## [1,] "Austria, Euro"
## [2,] "Belgium, Euro"
## [3,] "Bulgaria, Bulgarian lev"
## [4,] "Cyprus, Euro"
## [5,] "Czech Republic, Czech koruna"
## [6,] "Germany, Euro"
```

- Teraz môžeme priradiť stĺpcom ich názvy a získať dáta, s ktorými budeme ďalej pracovať:

```
colnames(data) <- nazvy
head(data)
```

```
##          Austria, Euro Belgium, Euro Bulgaria, Bulgarian
## 2021Dec          -0.05          0.03                   (
## 2021Nov          -0.01          0.07                   (
## 2021Oct           0.08          0.16                   (
## 2021Sep          -0.08          0.01                   (
## 2021Aug          -0.23         -0.14                   (
## 2021Jul          -0.13         -0.03                   (
##          Czech Republic, Czech koruna Germany, Euro Denma
## 2021Dec                          2.62         -0.38
## 2021Nov                          2.62         -0.31
## 2021Oct                          2.34         -0.21
```

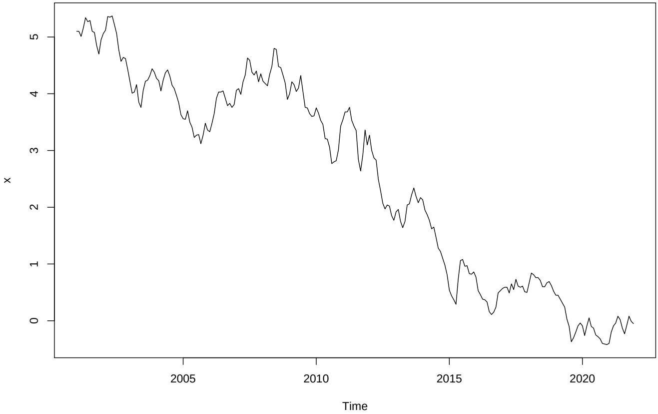
- V tomto cvičení zoberieme dáta pre Rakúsko:

```
x <- data[, "Austria, Euro"]
x <- x[length(x):1]  # usporadame od najstarsich
```

- Z vektora spravíme časový rad funkciou `ts` s parametrami:
  - použité dáta
  - `frequency` - frekvencia dát: 1 pre ročné dáta, 4 pre kvartálne, 12 pre mesačné
  - `start` alebo `end` - v závislosti od frekvencie, napr. `start = 2000` pri ročných dátach, `start = c(2000, 1)` pri kvartálnych znamená prvý kvartál roku 2000
- My máme mesačné dáta so začiatkom v januári 2001 (a koncom v decembri 2021)

```
# DOPLNTE
x <- ts(x, ....)
```

```
plot(x)
```



- Vidíme klesajúci trend, dáta preto nie sú stacionárne (a nemôžeme napríklad odhadovať ACF). **Budeme preto pracovať s diferenciami** (teda medzimesačnými zmenami).
- Vytvoríme diferencie pomocou funkcie `diff` - napríklad:

```
# numericky vektor
y <- c(1, 4, 6, 2)
diff(y)
```

```
## [1]  3  2 -4
```

```
# casove rady
z <- ts(y, frequency = 12, start = c(2022, 1))
diff(z)
```

```
##      Feb Mar Apr
## 2022   3   2  -4
```

- Vytvorte vektor diferencií úrokových mier a zobrazte ich.

```
# DOPLNTE
xDif <- ...
plot(xDif)
```

```
head(xDif)  # pre kontrolu
```

```
##          Feb   Mar   Apr   May   Jun   Jul
## 2001   0.00 -0.09  0.15  0.18 -0.07  0.02
```

- **Zaujíma nás, či sú tieto diferencie korelované alebo či ich môžeme považovať za biely šum.**
  - Zobrazte ACF. Ktoré autokorelácie sú signifikantné?
  - Zobrazte výsledky LB testu pre rôzne lagy a skomentujte.
  - Zhodnoťte výsledky. Aký je váš záver ohľadom charakteru daného časového radu?

Cvičenie 4: Testovacia štatistika Ljung-Boxovho testu

- Vygenerujme si dáta, ktoré **sú** nekorelované. Napríklad:

```
x <- rnorm(100, mean = 10, sd = 3)  # iid N(10, 3^2)
```

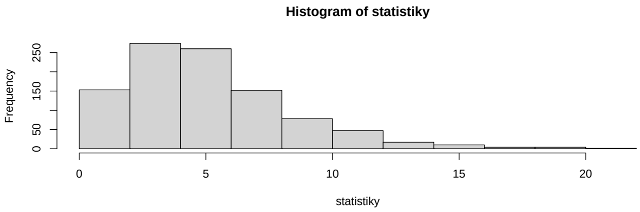
- Budeme testovať hypotézu, že prvých 5 autokorelácií je nulových a zaznamenáme hodnotu testovacej štatistiky:

```
# DOPLNTE
simulacia <- function(){
  x <- rnorm(100, mean = 10, sd = 3)
  LB <- Box.test(x, ...)
  return(...)
}
```

- Zopakujeme 1000 krát - pri opakovaní simulácií je užitočná funkcia `replicate` - a zobrazíme histogram:

```
statistiky <- replicate(1000, simulacia())
hist(statistiky)
```

- Aká je teoretická hustota rozdelenia štatistiky v Ljung-Boxovom teste? Porovnajte so získaným histogramom.
- Ukážka výstupu (kvôli náhodnosti môžete dostať iný):



- Zopakujte simulácie, histogram a jeho porovnanie s hustotou, kvantil a jeho porovnanie s kritickou hodnotou testu:
  - pre iný počet dát
  - pre iné rozdelenie generovaných iid náhodných veličín
  - pre dáta, ktoré nie sú iid, ale sú bielym šumom
  - pre iný počet lagov

Cvičenie 5: Sila Ljung-Boxovho testu

- Vypočítajte ACF procesu $x_t = u_t - \beta u_{t-1}$, kde u_t je biely šum (na prednáške sme mali prípad $\beta = -1$).
- Pre $\beta \neq 0$ dostaneme nenulovú prvú hodnotu ACF. LB testom budeme testovať, či sú prvé dve hodnoty ACF nulové. Skutočnosť je samozreme taká, že nie sú.

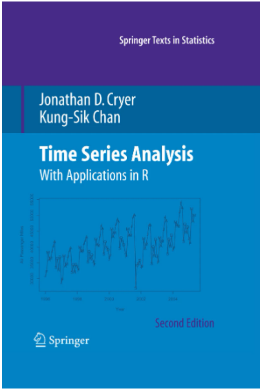
```
set.seed(123) # kvoli reprodukovatelnosti
N <- 101
u <- rnorm(N, mean = 0, sd = 1)
x <- u[1:(N - 1)] + 3 * u[2:N] # 100 hodnot procesu
                               # beta = -3
LB <- Box.test(x, lag = 2, type = "Ljung-Box")
LB$p.value
```

```
## [1] 0.03218388
```

- Dal Ljung-Boxov test v predchádzajúcej simulácii správny výsledok?
- Napíšte funkciu, ktorá pre zadaný parameter procesu βa dĺžku časového radu vráti TRUE/FALSE podľa toho, či sa nulová hypotéza zamietla. Pomocou `simulate` odhadnite, aká je pravdepodobnosť toho, že sa testovaná hypotéza zamietne.
- Graficky zobrazte odhadnutú silu testu ako funkciu parametra procesu β.
- Zobrazte aj funkciu, ktorá vyjadruje závislosť ACF(1) od parametra β a vysvetlite, ako súvisia priebehy týchto dvoch funkcií .

Cvičenie 6: Teoretické príklady

- Výpočet strednej hodnoty, disperzie, autokovariancií a autokorelácií.
- Overovanie stacionarity.



- Jonathan D. Cryer, Kung-Sik Chan: *Time Series Analysis With Applications in R. Second Edition.* Springer, New York, 2008
- Cvičenia ku kapitole 2: **2.4–2.16**
- Dostupné zo školskej siete: https://link.springer.com/book/10.1007/978-0-387-75959-3