

# *Časové rady - úvod*

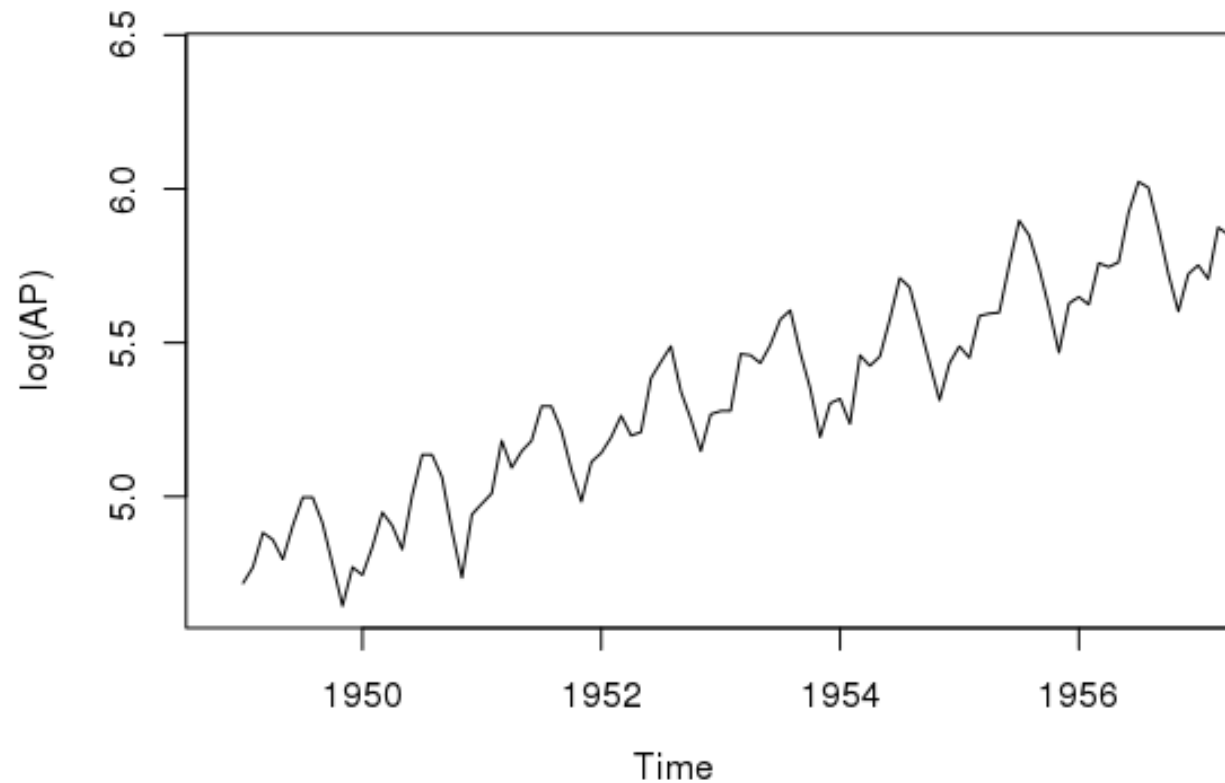
Beáta Stehlíková

Časové rady, FMFI UK

# Analýza časových radov

---

- Máme mesačné dáta - počty cestujúcich aerolinkami:



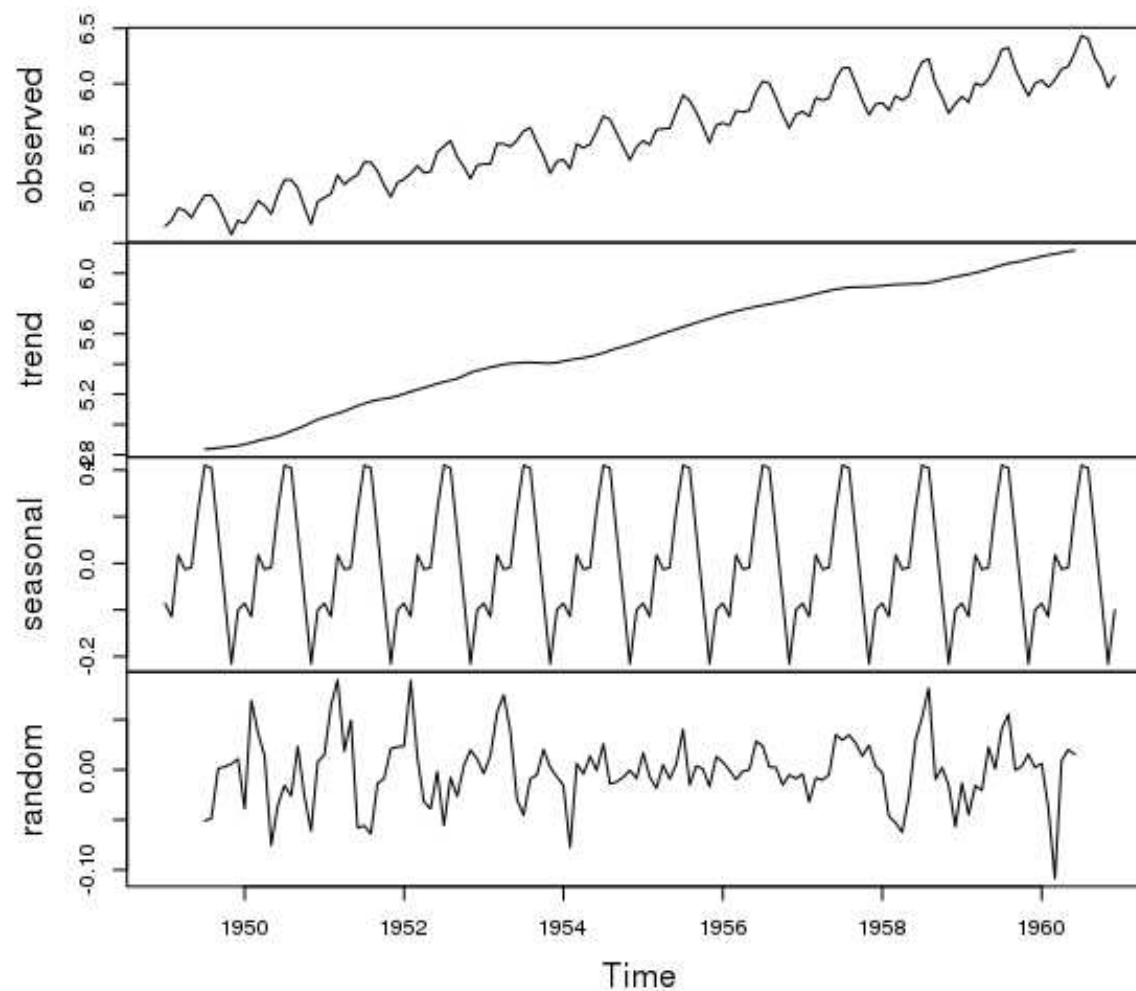
G. E. P. Box, G. M. Jenkins: **Time Series Analysis: Forecasting and Control.**

- Otázka: aký bude ďalší vývoj?

# Analýza časových radov

---

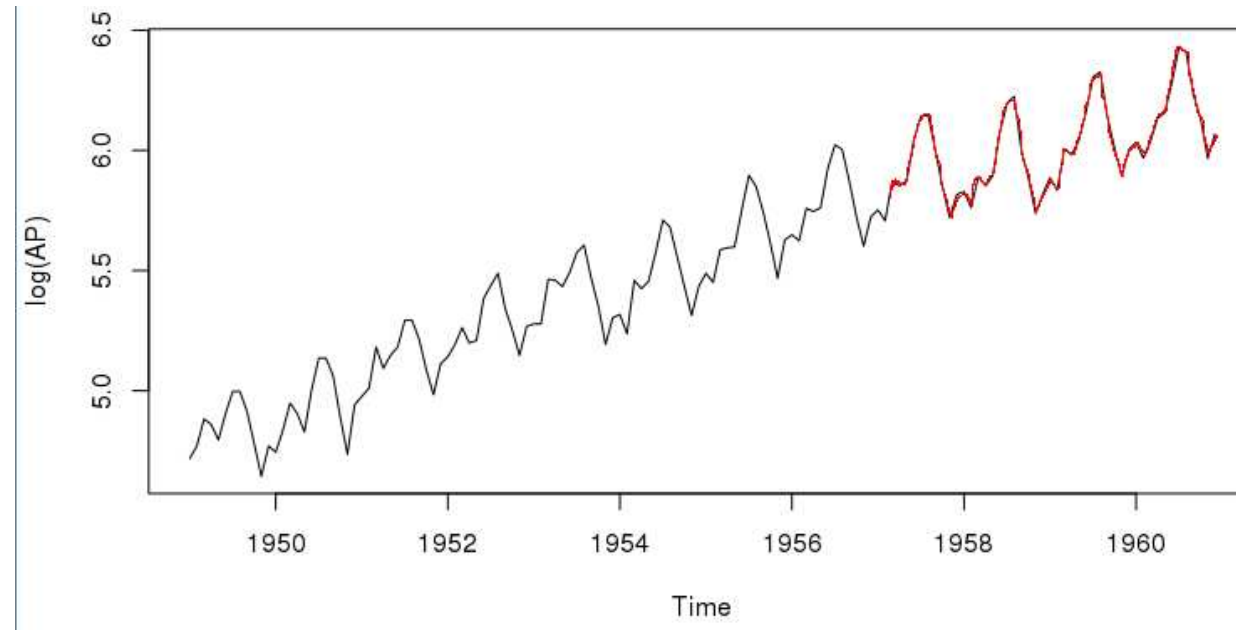
- Vidíme **rastúci trend** a **sezónnosť** (mesiace).
- Rozklad v R-ku: `plot(decompose(log(AP)))`



# Analýza časových radov

---

- Intuitívne: ak nenastane nejaký šok, zachová sa rastúci trend a sezónnosť, teda nejak takto:



- Otázky: Ako to vyjadriť kvantitatívne? Ako určiť presnosť odhadov, ako zostrojiť interval spoľahlivosti?

# *Box a Jenkins*

---

Budeme sa zaobert' najmä prístupom od **Boxa a Jenkinsa**

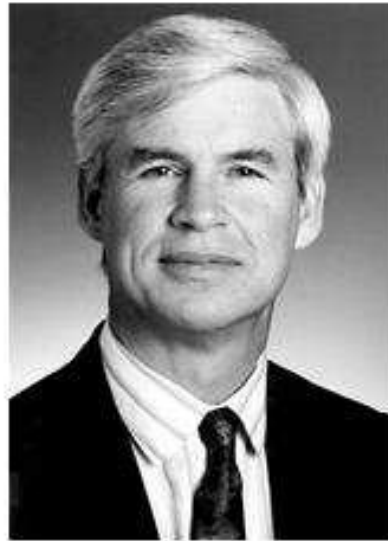


Rozhovor s G. E. P. Boxom  
po oslave jeho 80. narodenín  
(1999):

*"The first paper you wrote  
with Jenkins has been  
considered as a  
**breakthrough in statistics.**  
How do you become  
interested in time series?"*

# Modelovanie volatility

---



Robert F. Engle III



Clive W.J. Granger

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2003 was divided equally between Robert F. Engle III "for methods of analyzing economic time series with time-varying volatility (ARCH)" and Clive W.J. Granger "for methods of analyzing economic time series with common trends (cointegration)".

<http://www.nobelprize.org>

- ARCH model a jeho zovšeobecnenia - tiež súčasťou tohto kurzu

# Zaujímavosť

---

- Manželka R. Engla sa narodila v Prešove, v r. 2005 spolu navštívili Prešov



Rob, Marianne, Jordan and Lindsey in Williamstown 2002.

[http://www.nobelprize.org/nobel\\_prizes/economics/laureates/2003/engle-autobio.html](http://www.nobelprize.org/nobel_prizes/economics/laureates/2003/engle-autobio.html)

# *Základné pojmy - obsah prednášky*

---

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



# Časový rad, momenty

---

- Stochastický proces  $x_1, \dots, x_T$  - úplne je charakterizovaný  $T$ -rozmernou distribučnou funkciou
- Obvykle sa zameriavame len na prvé dva momenty:
  - ◇ stredná hodnota  $E[x_t]$
  - ◇ variancia  $Var[x_t]$
  - ◇ kovariancie  $Cov[x_t, x_s]$ , tzv. autokovariancie

# Stacionarita a ergodicita

---

- Väčšinou máme len jeden časový rad - jednu realizáciu náhodného procesu → 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  $E[x_t] = \mu$ ,  $Var[x_t] = \sigma^2, \dots$  pre  $\forall t$

# Stacionarita a ergodicita

---

- **Silná stacionarita:** združená distribučná funkcia sa nemení pri posune v čase
- Obvykle sa pracuje so slabším predpokladom → **slabá stacionarita:**

$$(1) \quad E[x_t] = \mu \quad \forall t$$

$$(2) \quad Cov[x_t, x_s] = \gamma(|t - s|) \quad \forall t, s$$

z (2) vyplýva:  $Var[x_t] = const.$  pre  $\forall t$

- Ďalej budeme pod stacionaritou rozumieť slabú stacionaritu

# *Stacionarita - dáta*

---

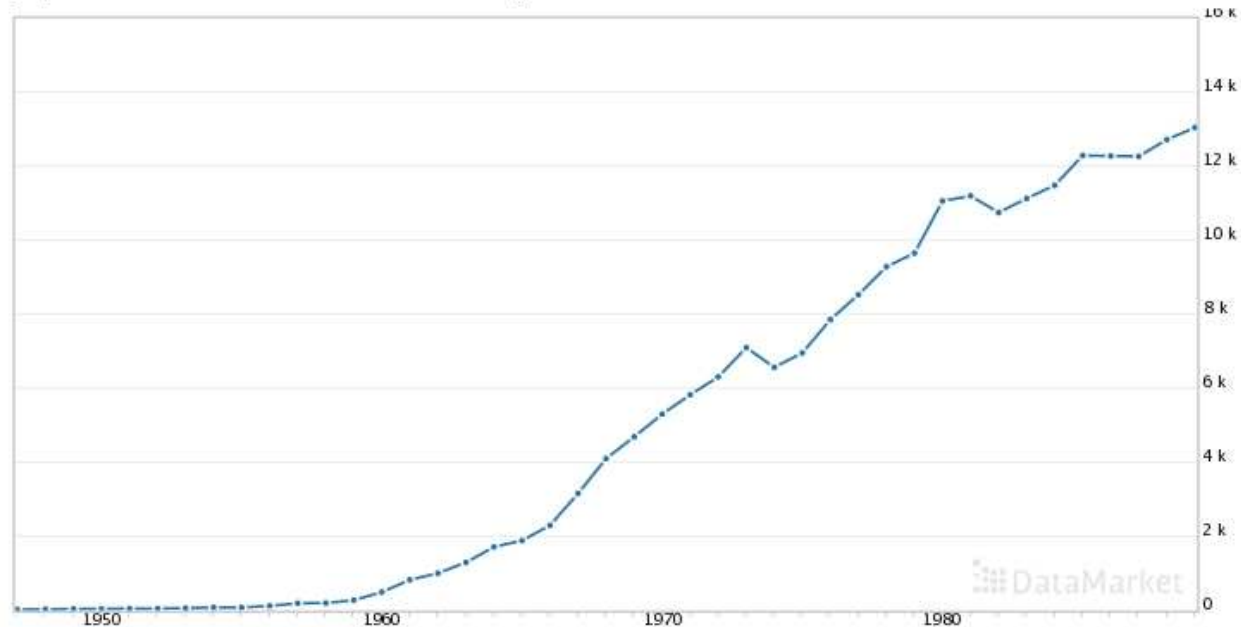
- Stacionárny časový rad: dáta sú prit'ahované k určitej rovnovážnej hodnote, okolo ktorej oscilujú
- Nestacionárny časový rad: napríklad trend

# Stacionarita - dáta

---

- PRÍKLAD 1

Japanese annual motor vehicle production for 1947-1989



<http://data.is/RVcjbL>

Vidíme:

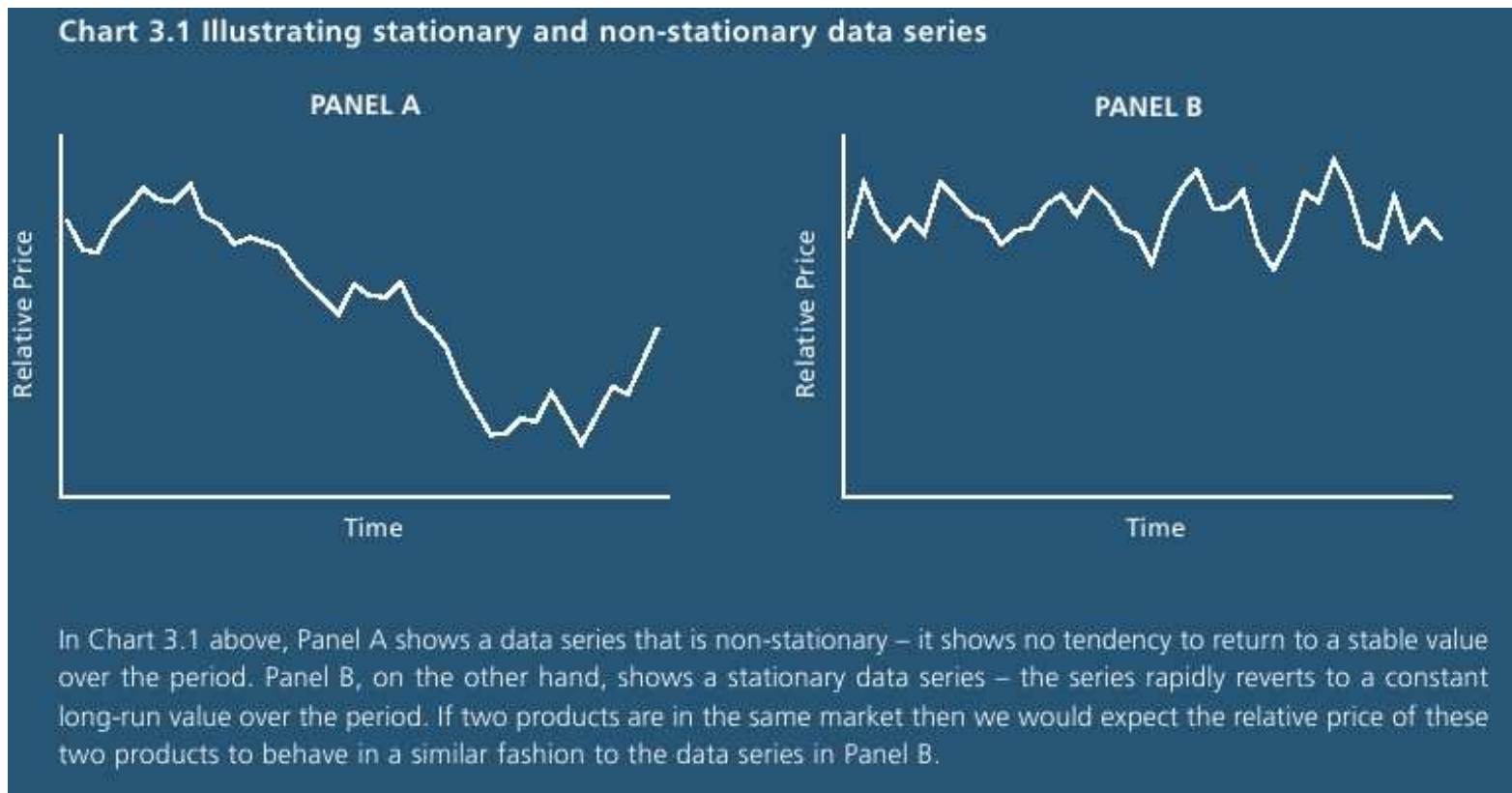
rastúci trend  $\Rightarrow$  stredná hodnota nie je konštantná  $\Rightarrow$   
časový rad nie je stacionárny

# Stacionarita - dáta

---

- PRÍKLAD 2

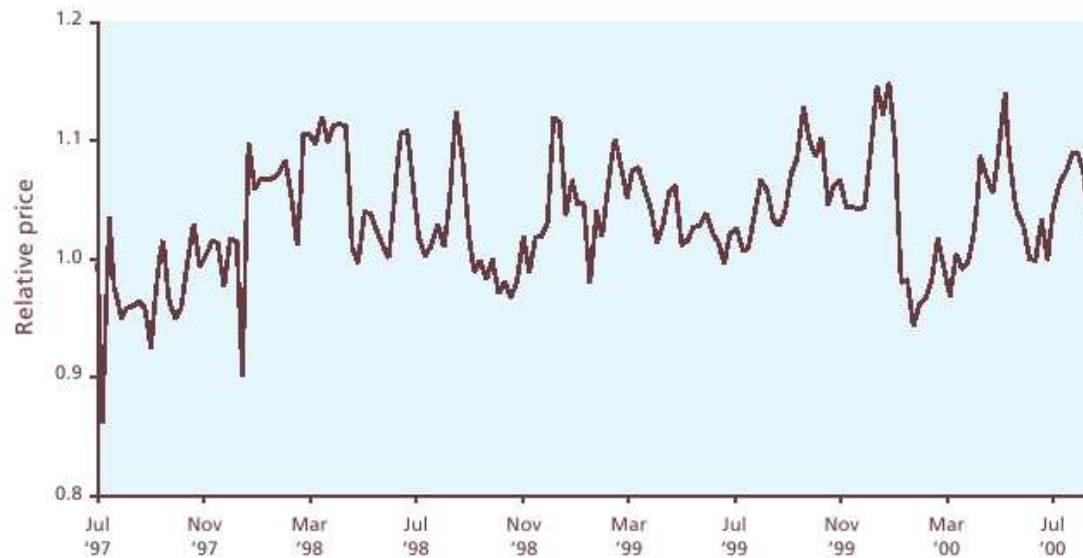
**Relatívna cena:** ak sa dva tovary predávajú na tom istom trhu, ich relatívna cena by sa mala pohybovať okolo určitej rovnovážnej hodnoty



# Stacionarita - dáta

---

Chart 3.2: Price of Scottish Salmon relative to the price of Norwegian salmon in the UK



The above chart shows that the price of Scottish salmon relative to the price of Norwegian salmon in the UK appears to vary randomly around a constant long-run value, which suggests that the relative price is stationary. The econometric test for stationarity confirms that the relative price of Scottish salmon is stationary, which is what we would expect to observe if Scottish and Norwegian gutted salmon compete in the same product market in the UK.

**An Introduction to Quantitative Techniques in Competition Analysis, Lexecon, 2005**

# Biely šum

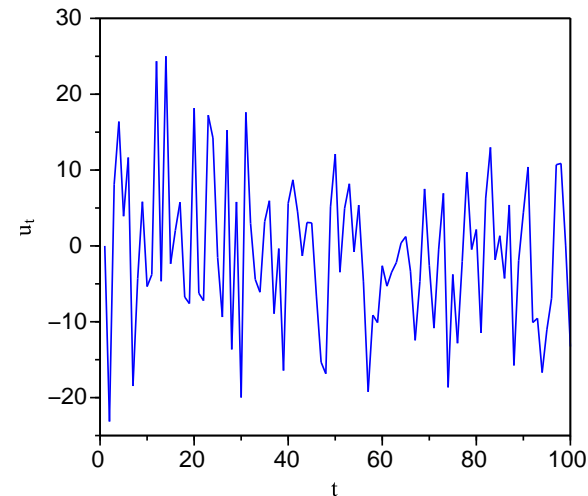
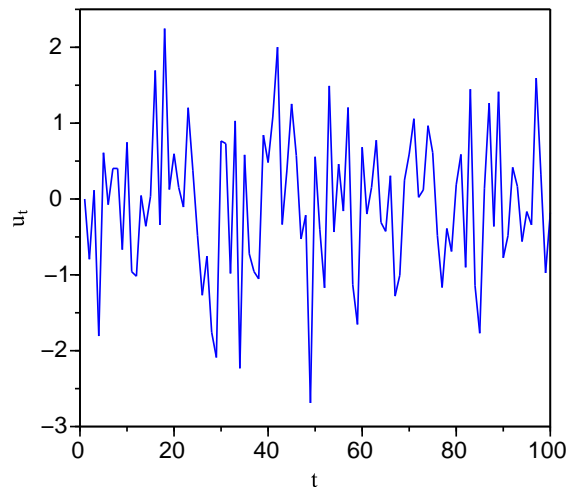
---

- Biely šum  $\{u_t\}$  - dôležitý príklad stacionárneho procesu

$$E[u_t] = 0 \quad \forall t$$

$$Var[u_t] = \sigma^2 \quad \forall t$$

$$Cov[u_t, u_s] = 0 \quad \forall t \neq s$$



- Základný náhodný proces, pomocou ktorého budeme definovať ďalšie, aj modely pre dáta



# Stacionarita - príklad 1

---

- Nech  $u_t$  je biely šum, definujme

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

- Vypočítame:

$$E[x_t] = 0, \quad Var[x_t] = 2\sigma^2$$

$$Cov[x_t, x_s] = \begin{cases} \sigma^2 & \text{pre } |t - s| = 1 \\ 0 & \text{pre } |t - s| = 2, 3, \dots \end{cases}$$

→ proces je stacionárny

# Stacionarita - príklad 2

---

- Nech  $u_t$  je biely šum; definujme

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

- $y_t$  sa dá zapísať ako  $y_t = \sum_{i=1}^t u_i$
- Vypočítame:

$$E[y_t] = 0, \quad Var[y_t] = t \sigma^2$$

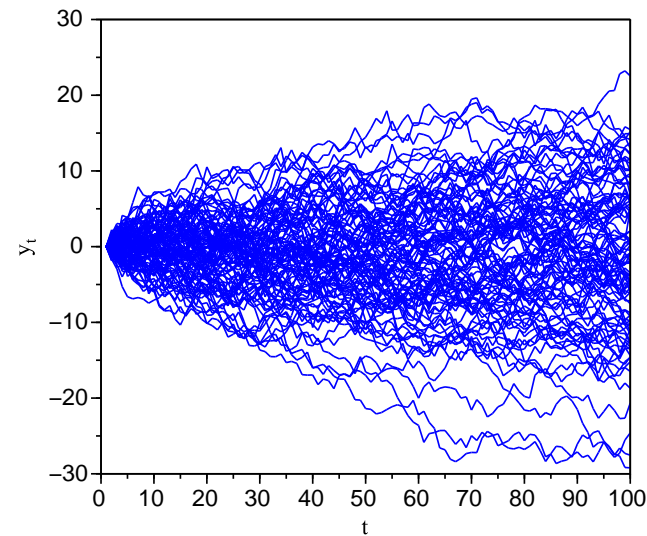
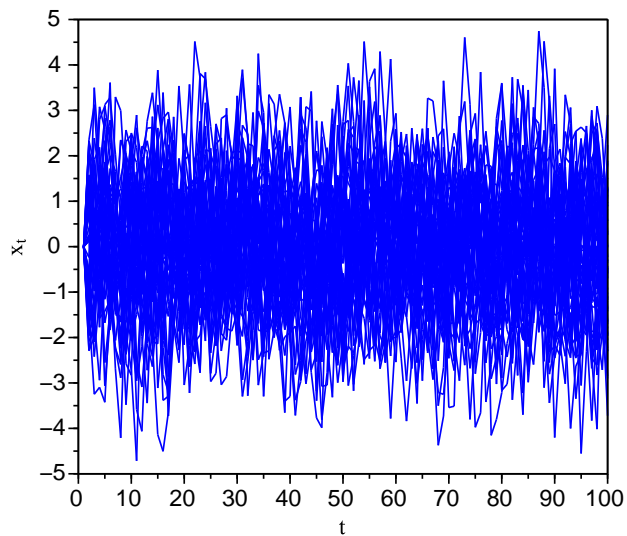
$$Cov[y_t, y_s] = \sigma^2 \min(t, s)$$

→ proces nie je stacionárny

# Stacionarita - príklady 1, 2

---

- Porovnanie trajektórií procesov z predchádzajúcich dvoch príkladov:
  - ◇ vľavo - stacionárny proces (pr. 1)
  - ◇ vpravo - nestacionárny proces (pr. 2), vidíme rastúcu disperziu



# Stacionarita - príklad 3

---

- Nech  $\{u_t\}_{t=-\infty}^{\infty}$  je biely šum, definujme

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

pričom koeficienty  $\psi_j$  splňajú  $\sum_{j=0}^{\infty} \psi_j^2 < \infty, \psi_0 = 1$

- Vypočítame:

$$E[x_t] = \mu, \quad Var[x_t] = \sum_{j=0}^{\infty} \psi_j^2$$

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

→ proces je stacionárny

# Woldova reprezentácia

---

- Predchádzajúci príklad: proces  $x_t$ , ktorý má tvar (3), je stacionárny
- Dá sa dokázať:  
Každý stacionárny proces  $x_t$  sa dá zapísať v tvare (3), t. j.

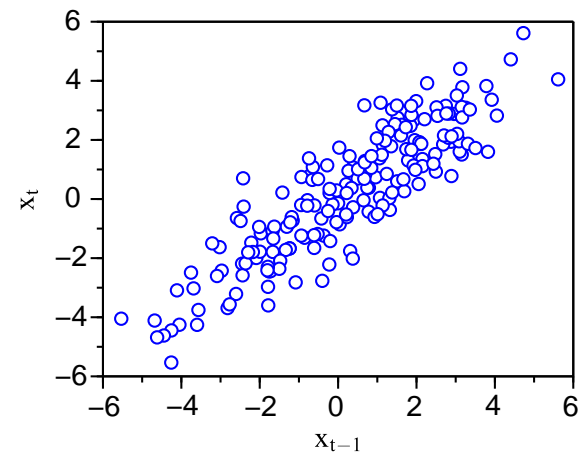
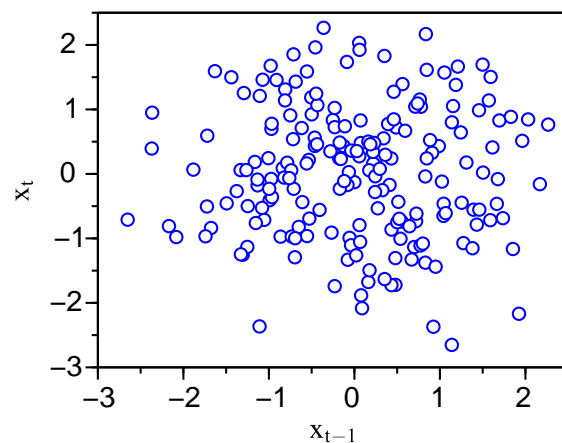
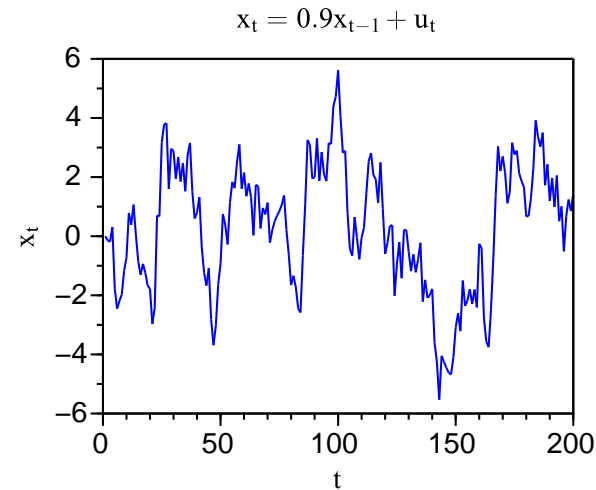
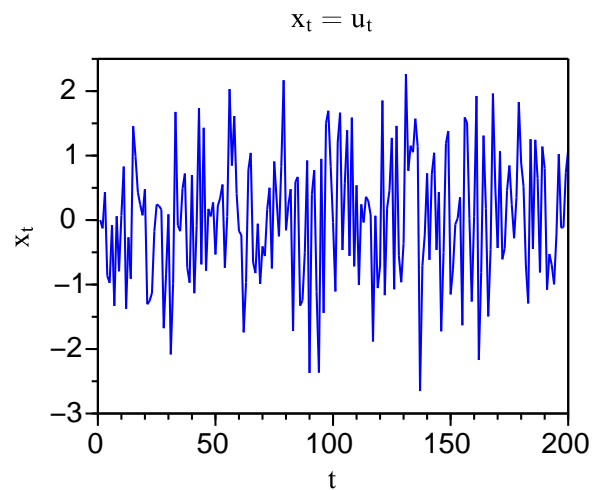
$$x_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j},$$

pričom koeficienty  $\psi_j$  spĺňajú  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ ,  $\psi_0 = 1$   
a  $u_t$  je biely šum

- Toto vyjadrenie stacionárneho procesu sa nazýva Woldova reprezentácia (Wold, 1938)

# Autokorelačná funkcia - motivácia

- Vľavo:  $y_t = u_t$ , vpravo:  $y_t = 0.9y_{t-1} + u_t$
- Realizácia procesu a ako závisí  $y_t$  od  $y_{t-1}$



# Autokorelačná funkcia

---

- Autokorelačná funkcia (ACF) stacionárneho procesu:

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

t. j.

$$\rho(\tau) = \text{Cor}(x_t, x_{t+\tau}) = \frac{\text{Cov}(x_t, x_{t+\tau})}{\sqrt{\text{Var}(x_t) \text{Var}(x_{t+\tau})}}$$

- Platí:

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

→ stačí nám počítať  $\rho(\tau)$  pre  $\tau = 1, 2, \dots$

# Autokorelačná funkcia

---

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

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

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

$\rightarrow$  konzistentný odhad autokorelačnej funkcie:

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

- je asymptoticky nevychýlený



# Autokorelačná funkcia - testy

---

- Odhad ACF v prípade bieleho šumu:
  - ◇ asymptoticky nevychýlený
  - ◇ variancia  $\approx 1/T$
  - ◇  $\Rightarrow$  približný 95 % interval spoľahlivosti:  $\pm 2/\sqrt{T}$ , často sa zobrazuje spolu s odhadnutými autokoreláciami
- V prípade stochastického procesu, pre ktorý  $\rho(\tau) = 0$  pre  $\tau > k$ , pre tieto  $\tau$  platí

$$\text{Var}[\hat{\rho}(\tau)] \approx \frac{1}{T} \left( 1 + 2 \sum_{j=1}^k \hat{\rho}(j)^2 \right)$$

# Autokorelačná funkcia - testy

---

- Testovanie, či daný časový rad je biely šum:
  1. interval spoľahlivosti  $\pm 2/\sqrt{T}$  pre každú autokoreláciu samostatne
  2. testovanie nulovosti  $\rho(1), \dots, \rho(m)$  súčasne:
    - ◇ **Box & Pierce, 1970**: ak platí  $H_0$ , asymptoticky

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

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

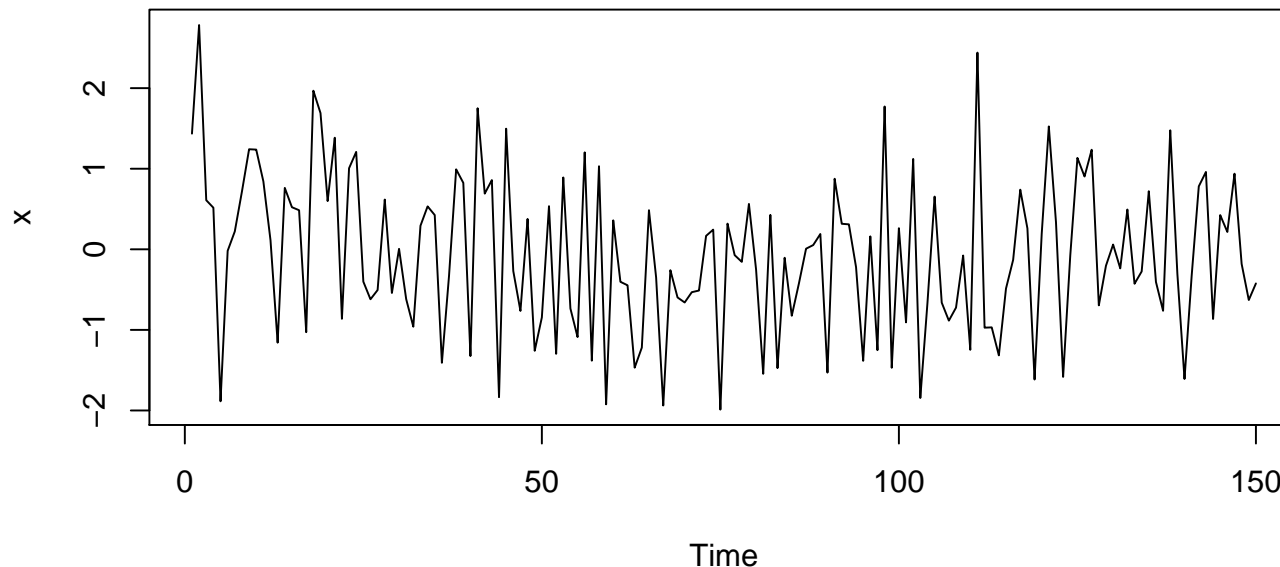
$$Q = T(T + 2) \sum_{j=1}^m \frac{\rho(j)^2}{T - j} \sim \chi_m^2$$

- ◇ *počet stupňov voľnosti sa zmení, ak ide o rezíduá z modelu*

# Autokorelácie časového radu - príklad 1

---

- Vygenerujeme 150 nezávislých realizácií  $N(0, 1)$  rozdelení (teda biely šum):

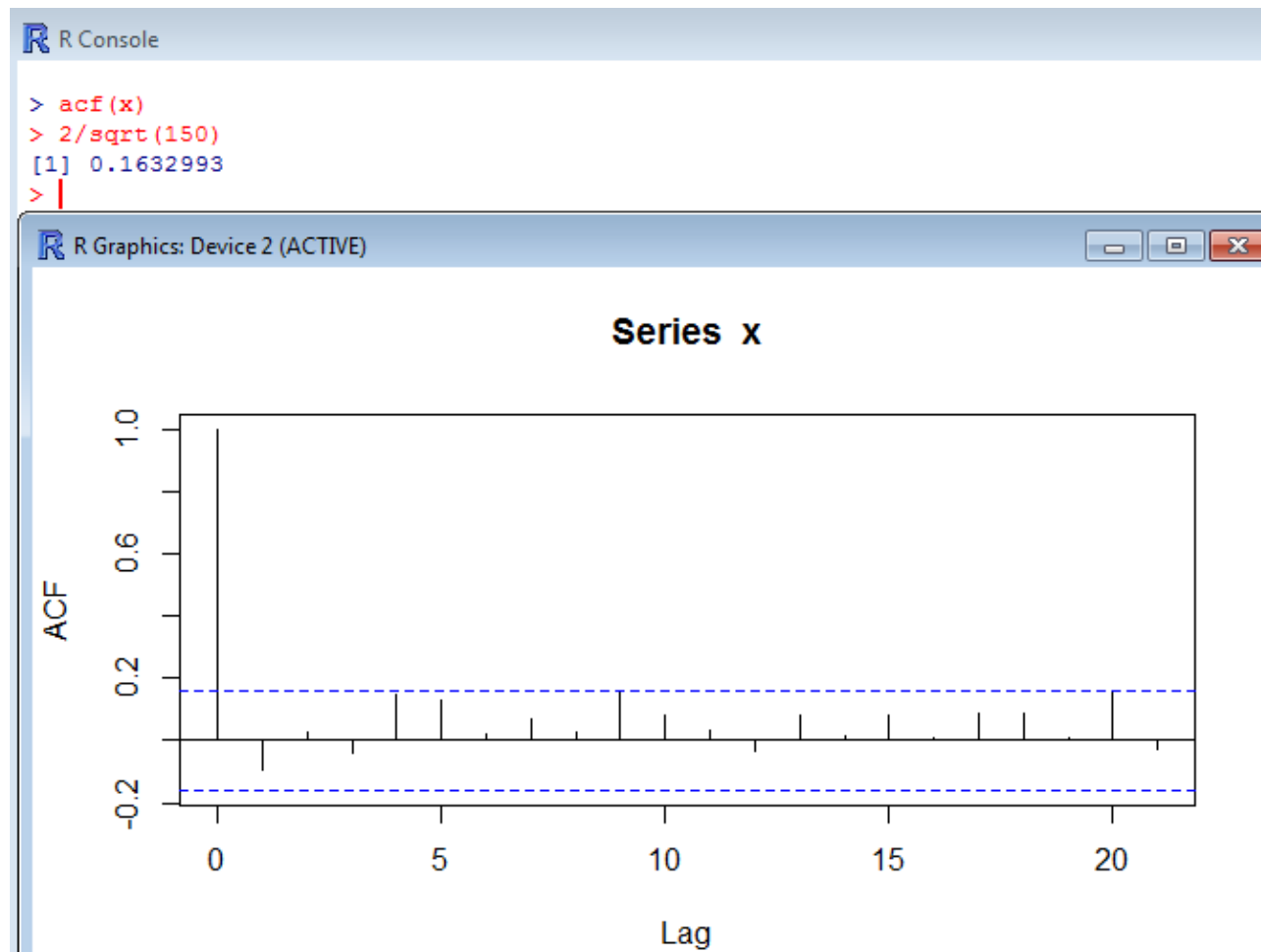


- Autokorelačná funkcia v softvéri R:
  - ◇ funkcia `acf`, napr. `acf(x)`
  - ◇ autokorelácia pre lag 0 je vždy 1, môžeme ju vynechať

# Autokorelácie časového radu - príklad 1

---

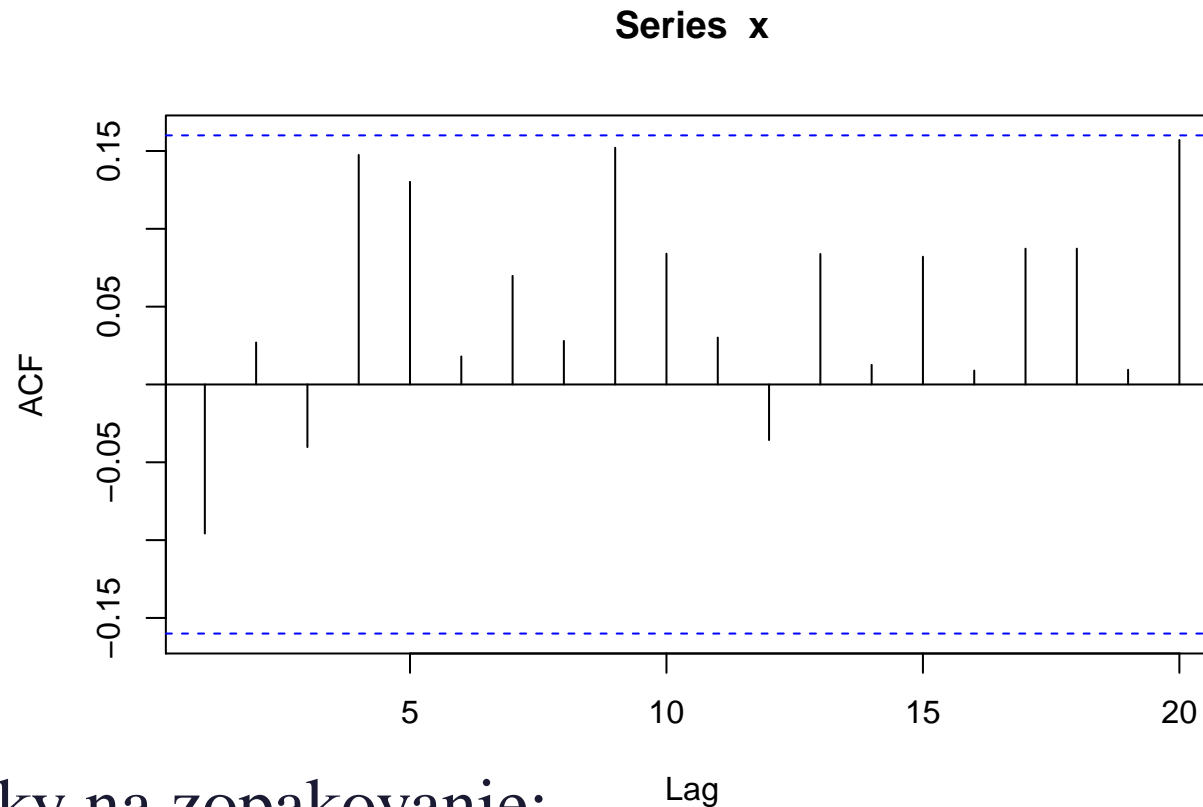
- Použitie funkcie `acf` + hranice intervalov spoľahlivosti :



# Autokorelácie časového radu - príklad 1

---

- Bez nultej autokorelácie:



- Otázky na zopakovanie:
  - ◇ Akú hypotézu vieme testovať pre každú autokoreláciu?
  - ◇ Kedy sa táto nulová hypotéza zamietá?
  - ◇ Čo dostávame v našom prípade?

# Autokorelačná funkcia - príklad

---

- Ljung-Boxova štatistika:
  - ◇ v R-ku funkcia `Box.test`
  - ◇ testujme napríklad hypotézu, že prvé tri autokorelácie sú nulové:

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

- Otázky na zopakovanie:
  - ◇ Aka sa vypočíta testovacia štatistika?
  - ◇ Aké je pravdepodobnostné rozdelenie štatistiky za platnosti nulovej hypotézy?
  - ◇ Pre aké hodnoty štatistiky sa hypotéza zamietá?
  - ◇ Ako sa vypočíta p-hodnota?

# Autokorelačná funkcia - príklad

---

- Dostaneme: 

```
> Box.test(x,lag=3,type="Ljung")
```

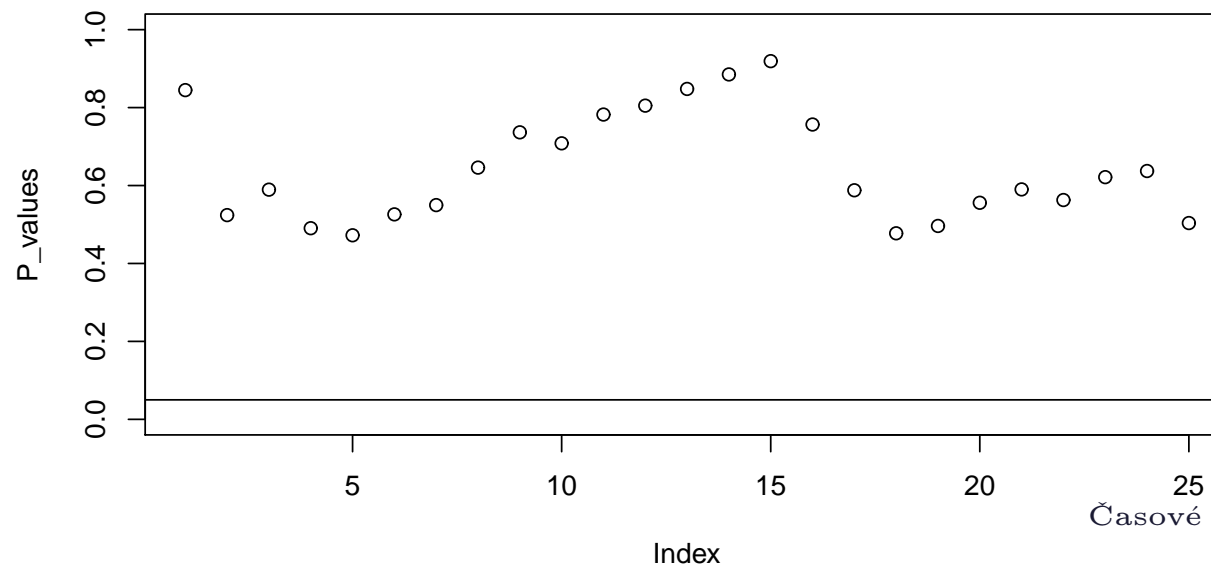
Box-Ljung test

```
data: x  
X-squared = 1.7655, df = 3, p-value = 0.6225
```

- Prístup k p-hodnotám:

`Box.test(x,lag=3,type="Ljung")$p.value`

→ dáme do cyklu a vykreslíme p-hodnoty pre jednotlivé lags a porovnáme s 0.05 → čo z tohto vyplýva?



# Operátor posunu

---

- Operátor posunu (lag operator)  $L$  - užitočný pri práci s časovými radmi
- Vrátí hodnotu procesu posunutú o jedno obdobie dozadu:

$$Lx_t = x_{t-1}$$

- Vlastnosti:
  - ◇ mocniny:  $L^2x_t = L(Lx_t) = x_{t-2}$
  - ◇  $L^0 = 1$  je identita:  $(1 - L)x_t = x_t - x_{t-1}$
  - ◇ počítanie s mocninami:  $L^2(L^3) = L^5$
  - ◇ násobenie:  $(1 - 0.5L)(1 - 0.3L) = 1 - 0.8L + 0.15L^2$
  - ◇ ak  $c$  je konštanta, tak napr.  
 $(1 - 0.1L + 2L^2)c = (1 - 0.1 + 2)c = 2.9c$