Autoregresné (AR) procesy

Beáta Stehlíková
Časové rady, FMFI UK

Príklad 1

- AR(2) proces z prednášky:

$$x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$$

- V R-ku:
  - korene charakteristického polynómu - overíme stacionaritu procesu
  - hodnoty ACF a PACF
  - vygenerujeme realizáciu procesu

Príklad 1(a): Stacionarita AR procesu

- Z informácií o skúške na webe:

  - súčasťou písomky budú dve otázky z kostry predmetu (zisťovanie stacionarity a invertovateľnosti ARMA procesu; nájdenie vyhovujúceho ARIMA modelu pre zadané "pekné" dáta a vysvetlenie použitých testov), ich zodpovedanie aspoň na 60 percent je nutnou podmienkou spravenia skúšky. Ukážky takýchto zadaní budú zverejnené po odprednášaní príslušných tém.

- Stacionarita AR procesu:
  - ◇ špeciálny prípad uvedeného zadania
  - ◇ možný príklad z kostry predmetu na skúške
  - ◇ nutné na vyriešenie všeobecnejšieho zadania

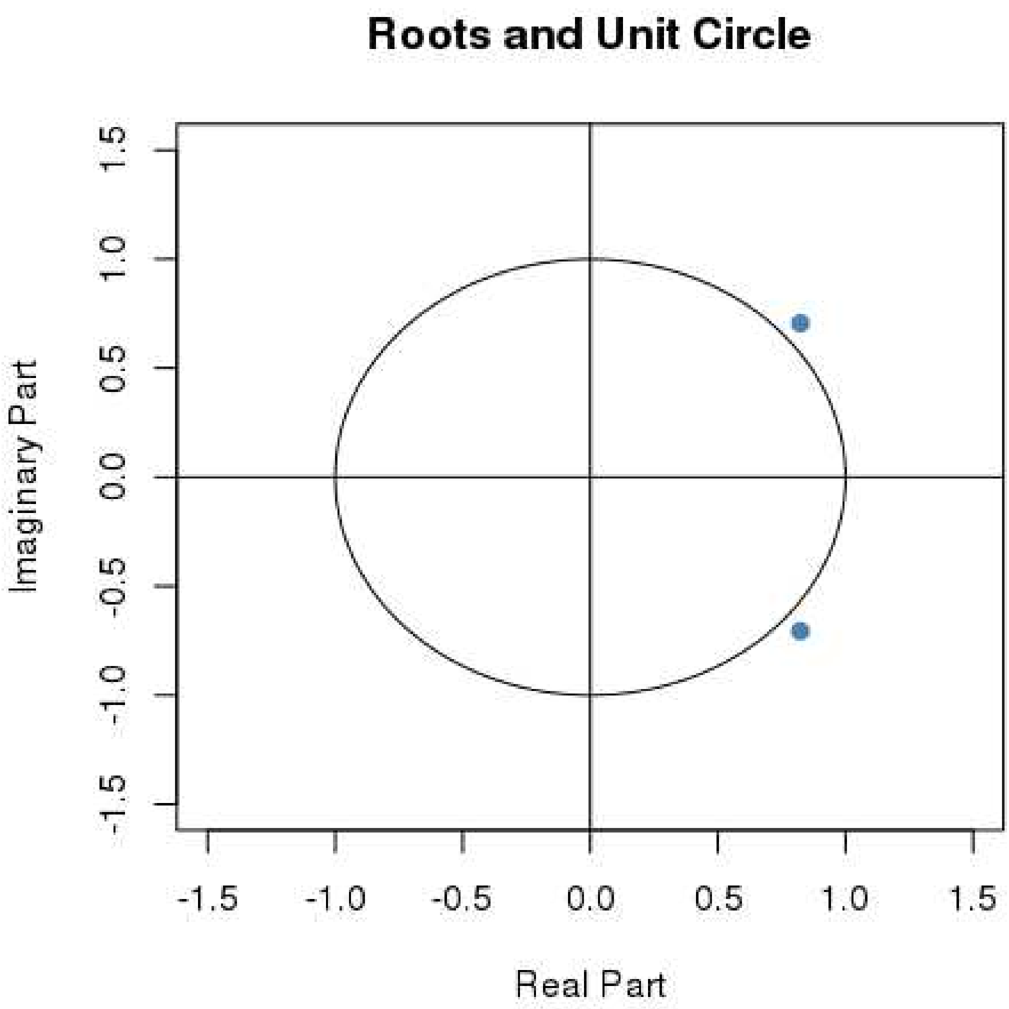
Príklad 1(a): Stacionarita AR procesu

- Máme proces: $x_t = 1.4x_{t-1} - 0.85x_{t-2} + u_t$
- Prepíšeme: $(1 - 1.4L + 0.85L^2)x_t = u_t$
- Korene polynómu $1 - 1.4L + 0.85L^2 = 0$ - majú byť mimo jednotkového kruhu
- V R-ku:
  - knižnica library(fArma) (ak treba, najprv nainštalovať)
  - potom (pozor na znamienka): armaRoots(c(1.4,-0.85))
- Výstup: tabuľka s koreňmi, aj ich absolútnymi hod.

```
> armaRoots(c(1.4,-0.85))
      re      im   dist
1 0.8235  0.7059 1.0847
2 0.8235 -0.7059 1.0847
```

Príklad 1(a): Stacionarita AR procesu

- Druhý výstup: grafické znázornenie koreňov



Cvičenia 1(a): Stacionarita AR procesu

Zistite, či sú nasledovné AR procesy stacionárne. Pre každý z procesov napíšte:

- polynóm, ktorého korene majú byť mimo jednotkového kruhu
- stupeň polynómu, a teda počet jeho koreňov
- absolútnu hodnotu každého koreňa

1. $x_t = 2 + 0.2x_{t-1} + 0.5x_{t-2} + u_t$
2. $(1 - 0.25L + 0.4L^2)x_t = 1 + u_t$
3. $x_t = 0.35x_{t-2} + u_t$
4. $(1 - 0.7L - 0.4L^2)(1 + 0.5L)x_t = 10 + u_t$
5. $(1 - 0.2L)^2(1 + 0.25L)x_t = 1 + u_t$
6. $(1 - 0.2L^2)(1 + 0.25L)x_t = 1 + u_t$

Príklad 1(b): Komplexné korene

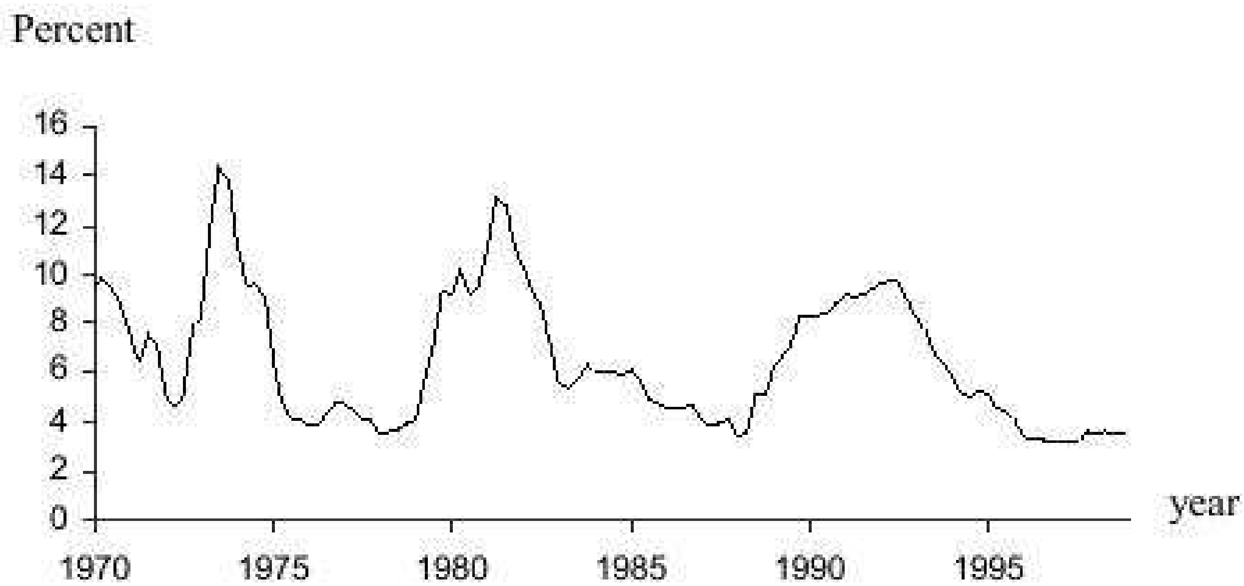
- Komplexné korene $\rightarrow$ periodický charakter procesu $\rightarrow$ chceme vypočítať periódu
- Využívame pritom zápis komplexného čísla v tvare $a + ib = \rho(\cos\varphi + i \sin\varphi) = \rho\, e^{i\,\varphi}$ a z neho hodnotu φ
- Získame ju napr. takto:

```
> x=complex(real=0.8235, imaginary=0.7059)
> Arg(x)
[1] 0.7086563
```

- *Opakovanie z prednášky:* Ako z tohto vypočítame periódu?

Cvičenia 1(b): Komplexné korene

- Model z prednášky, úroková miera v Nemecku:



- Odhadnutý model:

$$\begin{aligned} GSR_t &= \underset{(2.82)}{0.577} + \underset{(17.49)}{1.407}\, GSR_{t-1} - \underset{(-6.16)}{0.498}\, GSR_{t-2} + \hat{u}_t, \end{aligned}$$

$$\overline{R}^2 = 0.910, \quad SE = 0.812, \quad Q(6) = 6.431 \quad (p = 0.377)$$

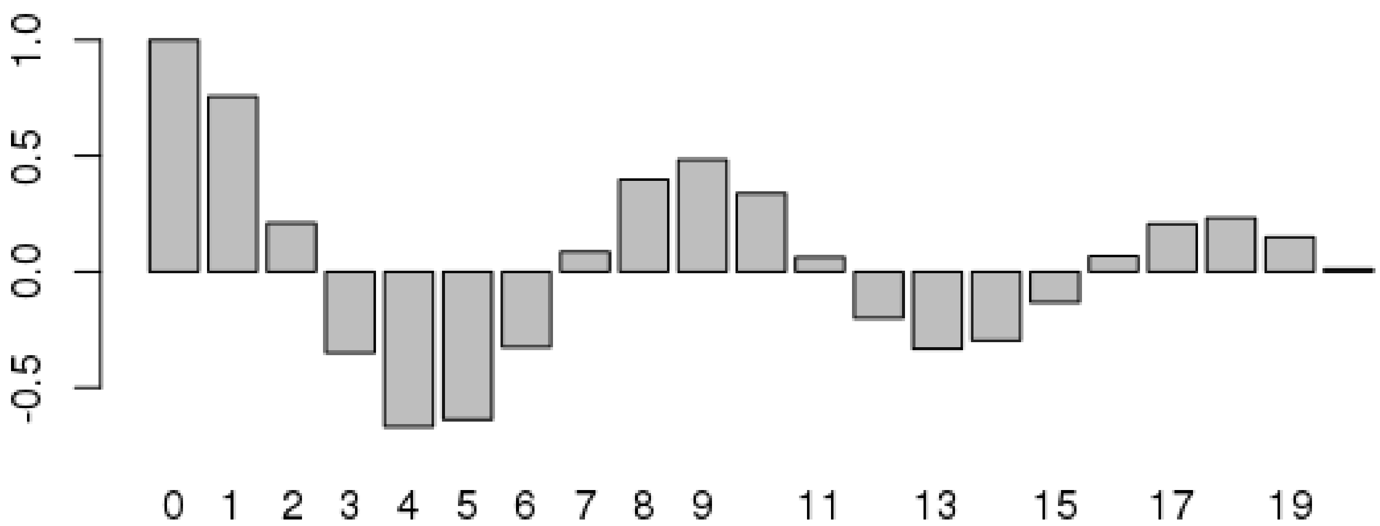
Cvičenia 1(b): Komplexné korene

- Overte stacionaritu procesu.
- Vysvetlite nasledovné tvrdenie z knihy (str.49) a vypočítajte uvedené hodnoty: *"The two roots of the process are 0.70 +/- 0.06i, i.e. they indicate cycles ... the frequency* $f = 0.079$ *corresponds to a period of 79.7 quarters and therefore of nearly 20 years."*

Príklad 1(c): ACF a PACF

- Vypísanie hodnôt:
  - ◇ ARMAacf(ar=c(1.4,-0.85),lag.max=10) (vratane nulového lagu)
  - ◇ ARMAacf(ar=c(1.4,-0.85),lag.max=10, partial="true")
- Graficky napríklad takto:

barplot(ARMAacf(ar=c(1.4,-0.85),lag.max=20))



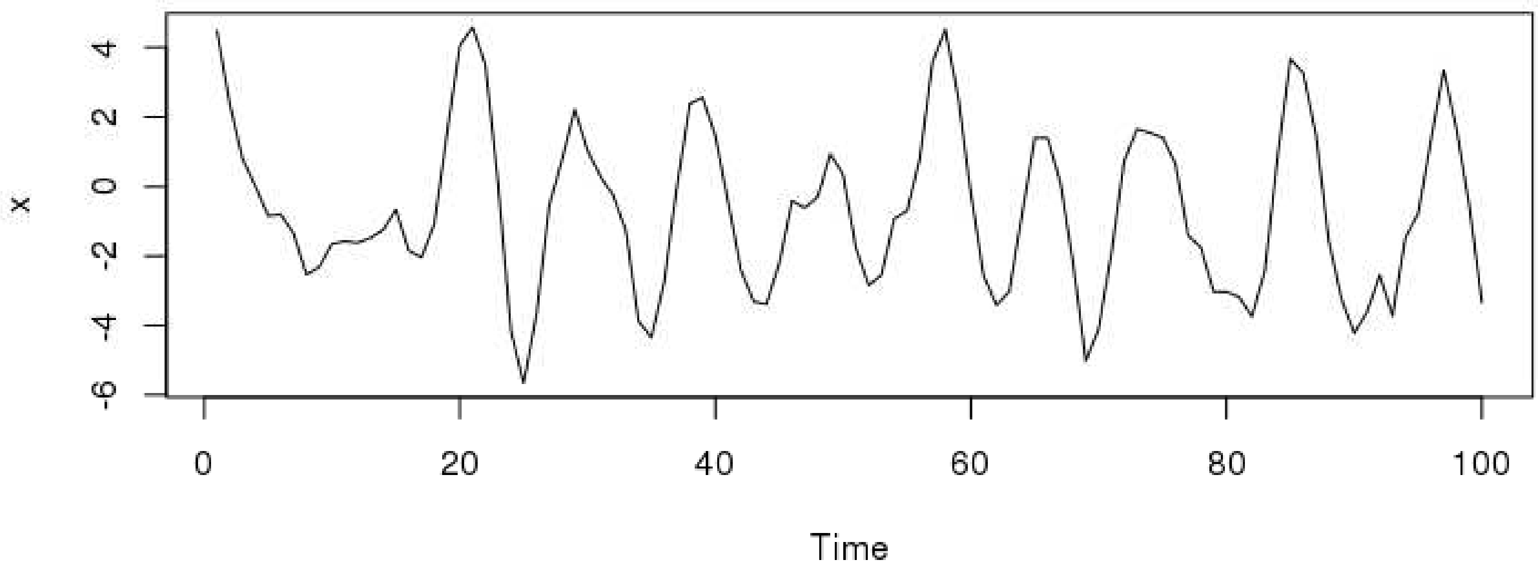
Cvičenia 1(c): ACF a PACF

- Zobrazte ACF a PACF stacionárnych procesov zo str. 6.
- Čo vieme dopredu povedat' o PACF pre každý z týchto procesov?

Príklad 1(d): Realizácia procesu

- Realizácia procesu:

  x=arima.sim(n=100, list(ar = c(1.4,-0.85)),sd=0.1)



Príklad 2: Odhadovanie parametrov

- Budeme používať knižnicu astsa
- Postup:
  - výberová ACF a PACF $\rightarrow$ identifikácia modelu
  - odhadnutie modelu
  - kontrola rezíudí - ak nie sú bielym šumom, treba skúsiť iný model
  - keď už máme dobrý model, môžeme spraviť predikcie

Príklad 2(a): Výberová (P)ACF

- Dve možnosti:
  - acf(x), acf(x, type="partial")
  - acf2(x) z knižnice astsa - spolu ACF (bez nulového lagu) a PACF v jednom obrázku
- Vyskúšame pre dáta v premennej x - vygenerované z AR(2) procesu

Príklad 2(b): Odhad parametrov

- Budeme používať funkciu sarima z knižnice astsa
- AR(p) model pre dáta v premennej x:

  sarima(x,p,0,0)

- Výstup pre naše dáta:

```
Coefficients:
         ar1      ar2   xmean
      1.3571  -0.7737  0.0049
s.e.  0.0877   0.0883  0.0282
```

To znamená, že odhadnutý model je:

$$x_t = \delta + 1.3571 x_{t-1} - 0.7737 x_{t-2} + u_t$$

Dopočítajte parameter δ tak, aby platilo $\mathbb{E}[x_t] = 0.0049$ (ako vidíme vo výstupe). Sú rezíduá bielym šumom?

Príklad 2(c): Konštrukcia predikcií

- Budeme používať funkciu sarima.for z knižnice astsa
- AR(p) model pre dáta v premennej x, predikcie na N období do budúcnosti:

  sarima.for(x,N,p,0,0)

- Výstup obsahuje graf, na ktorom sú
  - dáta
  - predikcie - odhadnutá stredná hodnota v budúcnosti
  - intevaly spoľahlivosti pre predikcie

Cvičenie 2

- Proces 1 zo str. 6: $x_t = 2 + 0.2x_{t-1} + 0.5x_{t-2} + u_t$
- Vygenerujte realizáciu procesu s dĺžkou 250, pričom $\mathbb{D}[u_t] = 0.0625$
  NÁVOD: (set.seed - aby sme mali rovnaké výsledky:

  set.seed(12345)
  x=arima.sim(n=..., list(ar = c(...)), sd=...) + c

  pričom sd je štandardná odchýlka u_t a c je vhodná konštanta
- Vypočítajte výberovú ACF a PACF, porovnajte ich s presnými.

Príklad 3: Reálne dáta

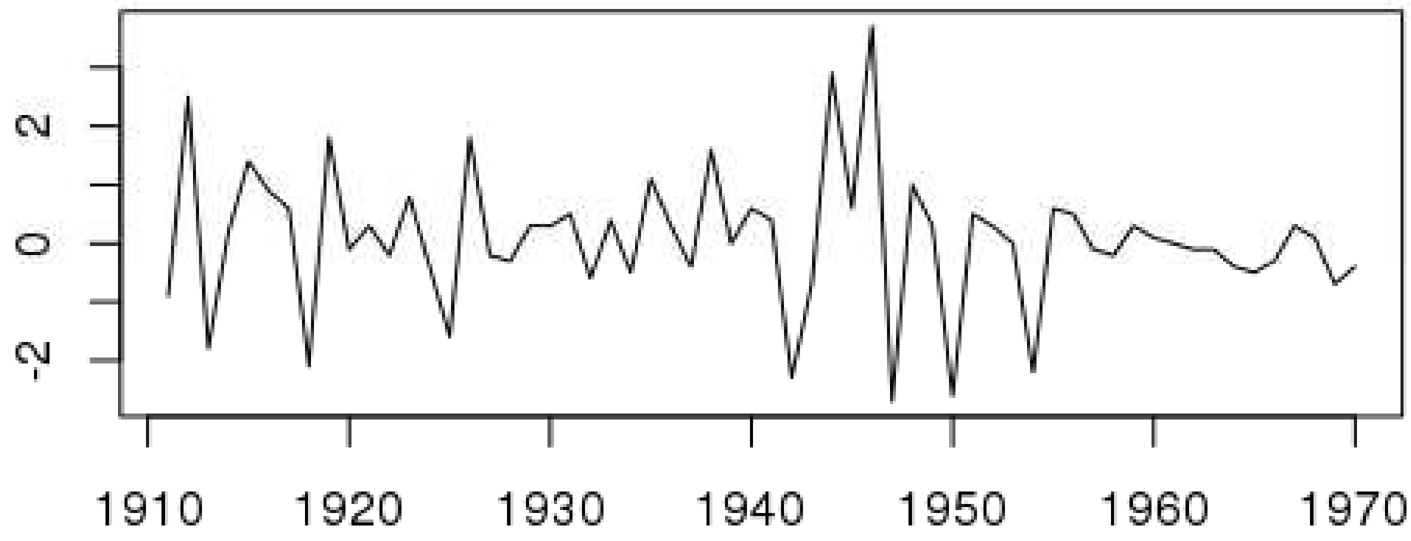
- Spotreba kávy v USA v rokoch 1910 - 1970
- Zdroj: http://robjhyndman.com/tsdldata/data/coffee.dat
- Na stránke cvičení: kava.txt
- Načítanie dát:
  ```
  x <-
  read.table("http://www.iam.fmph.uniba.sk/institute/
  stehlikova/cr14/data/kava.txt")
  ```
- Z týchto dát spravíme časový rad pomocou funkcie ts:

  ```
  x <- ts(x, frequency=..., start=c(...))
  ```

  namiesto start sa dá použiť end (podrobnejšie na stránke)
- Vykreslíme → zdá sa, že v dátach je trend → budeme pracovať s diferenciami (potrebujeme stacionárne dáta)

Príklad 3: Reálne dáta

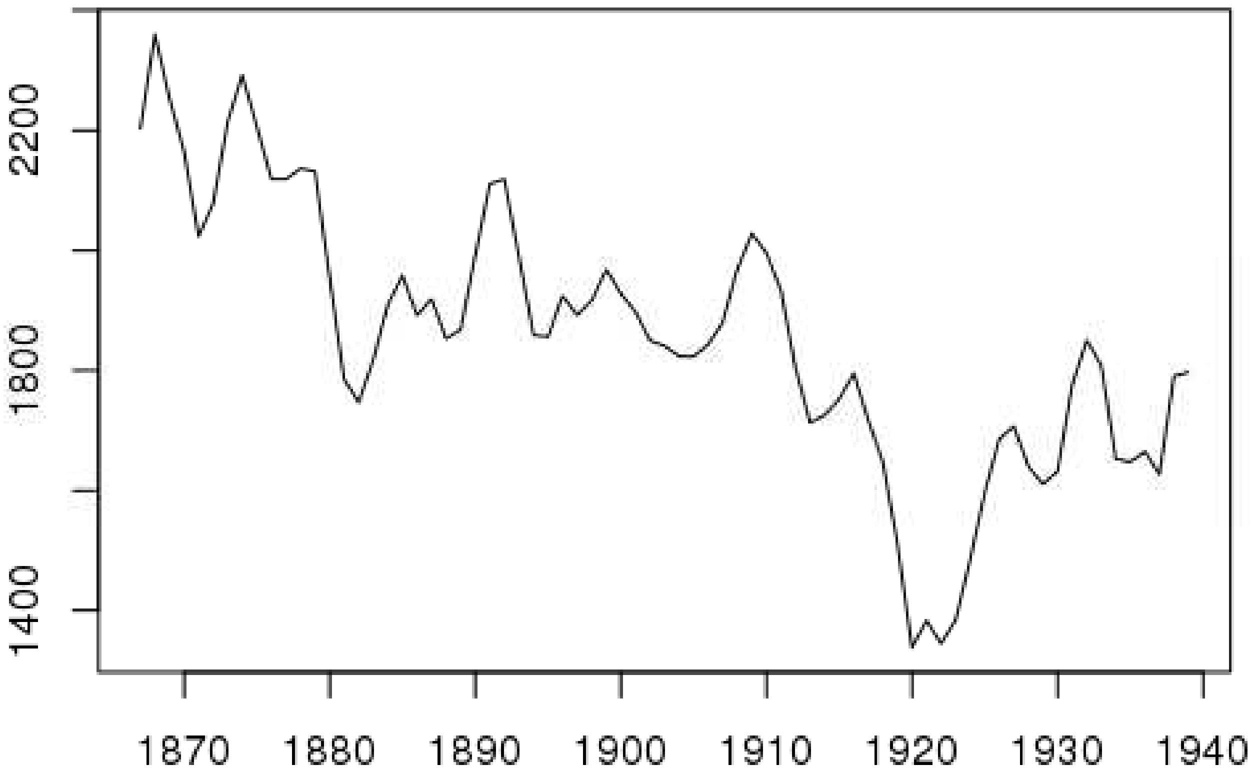
- Priebeh diferencií - plot(diff(x)):



- Ďalej je postup analogický ako pri simulovaných dátach v predchádzajúcom príklade
- Ak odhadujeme **AR(p)** model pre **k**-te diferencie, príkazy sú: sarima(x,p,k,0), resp. sarima.for(x,N,p,k,0)

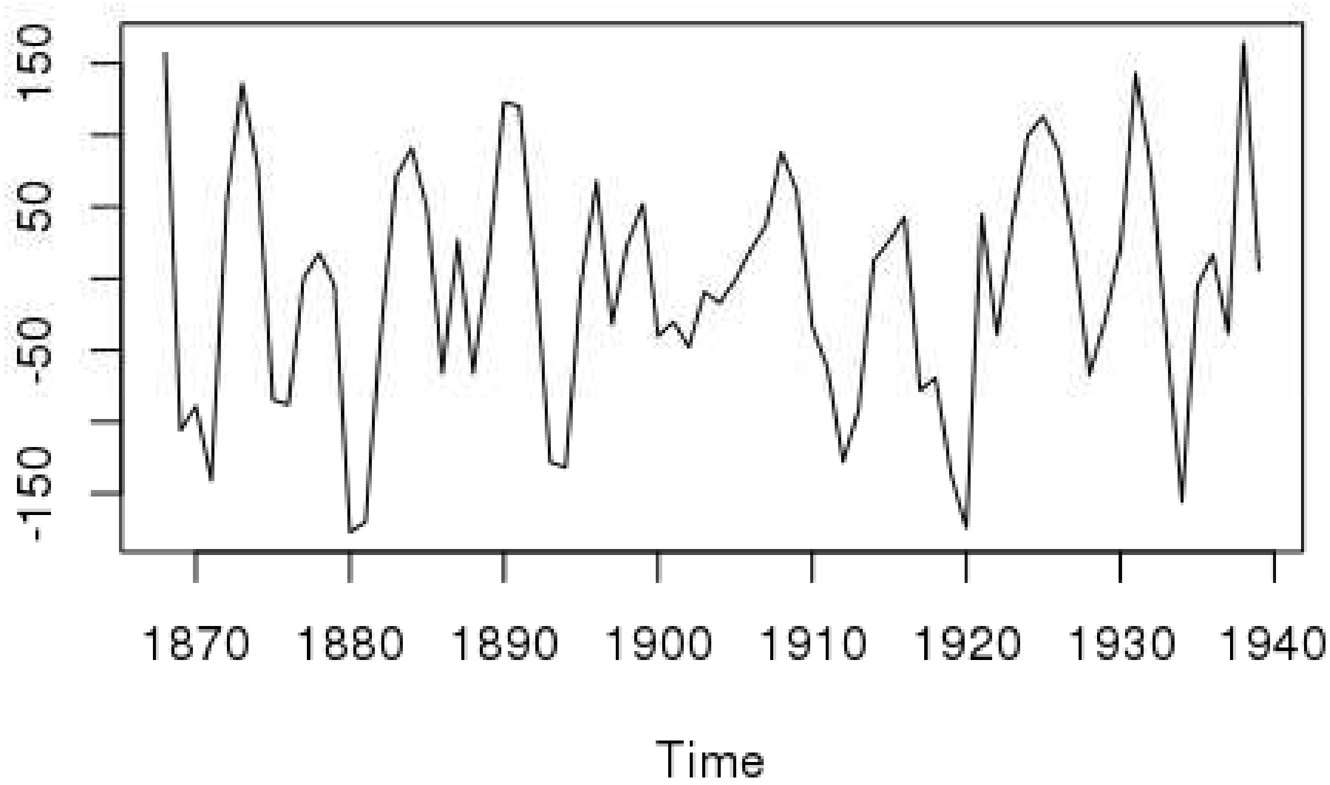
Cvičenie 3: Reálne dáta

- Počet oviec v tisícoch v Anglicku a Walese, ročné dáta, 1867-1939
- Zdroj: http://datamarket.com/data/set/22px/
- Na stránke: ovce.txt



Cvičenie 3: Reálne dáta

- Zdá sa, že dáta nie sú stacionárne, ale majú klesajúci trend $\rightarrow$ budeme pracovať s diferenciami:



Cvičenie 3: Reálne dáta

súčasťou písomky budú dve otázky z kostry predmetu (zisťovanie stacionarity a invertovateľnosti ARMA procesu; nájdenie vyhovujúceho ARIMA modelu pre zadané "pekné" dáta a vysvetlenie použitých testov), ich zodpovedanie aspoň na 60 percent je nutnou podmienkou spravenia skúšky. Ukážky takýchto zadaní budú zverejnené po odprednášaní príslušných tém.

Zatiaľ poznáme AR modely, na tieto dáta stačia. Úlohy:

- Zobrazte výberovú ACF a PACF. Použite ich na nájdenie vhodného modelu.
- Overte, že získaný model je stacionárny. Aké má korene?
- Skomentujte rezíduá; vysvetlite použité testy a prečo sa model dá považovať za vyhovujúci.
- Spravte predikcie.