

# *ARIMA modelovanie*

Beáta Stehlíková

Časové rady, FMFI UK, 2014/2015

# Príklad: dáta

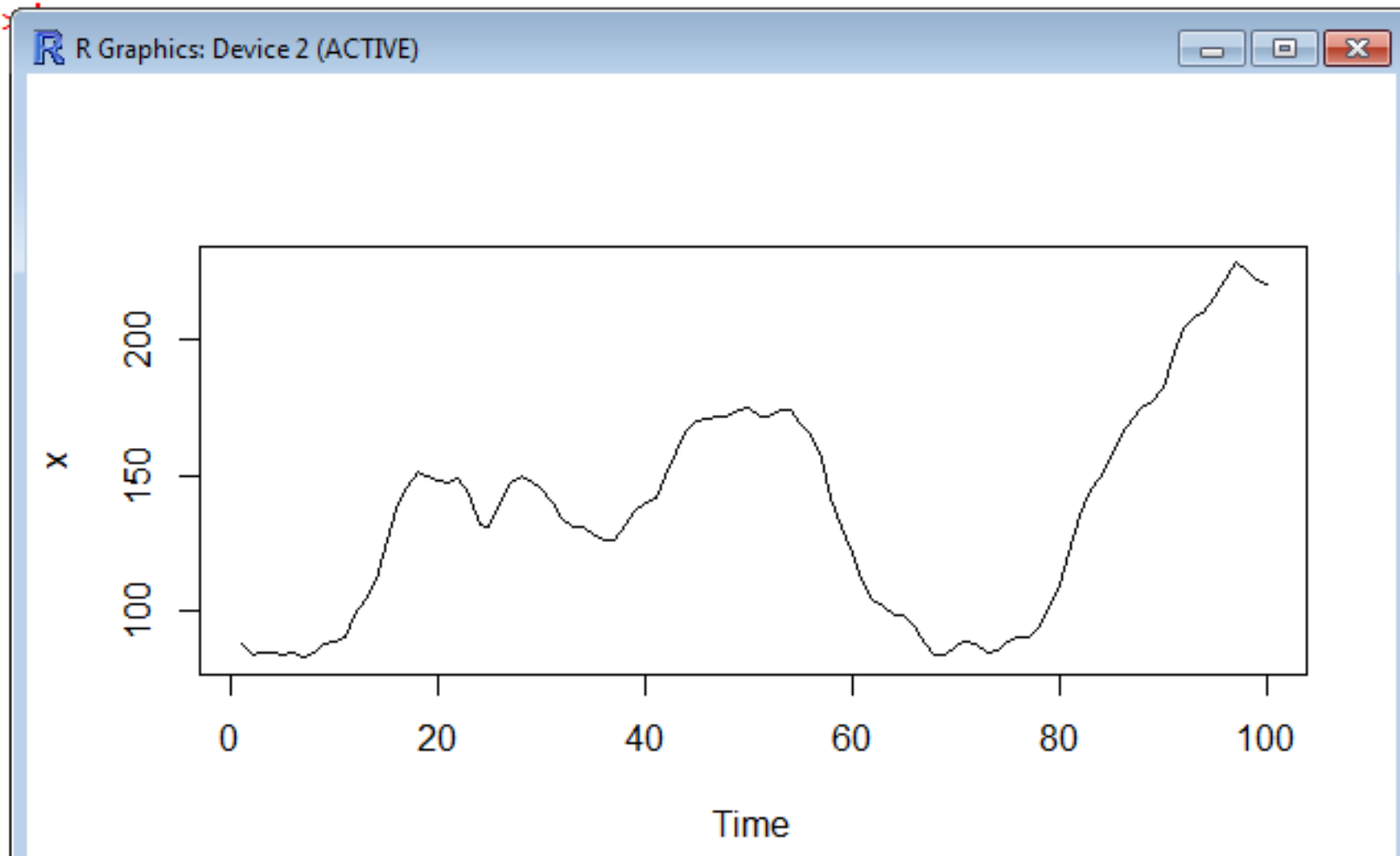
---

- Použité dáta:
  - ◇ Počet používateľov prihlásených na server, dáta po minútach, spolu 100 minút
  - ◇ Zdroj:  
<http://robjhyndman.com/tsdldata/data/computer.dat>
  - ◇ Na stránke predmetu: [comp.txt](#)
- Načítanie dát v R-ku:
  - ◇ Najskôr klasicky: uložíme, nastavíme pracovný adresár, potom `x=read.table("comp.txt");` alebo `x=read.table("http://www.iam.fmph.uniba.sk/institute/stehlikova/cr14/data/comp.txt")`
  - ◇ Potom: ešte musíme spraviť z vektora časový rad: `x=ts(s)` (funkcia `ts` - *time series*)
- Vykreslime ich príbeh.

# Príklad: dáta

---

```
> x=read.table("comp.txt")  
> x=ts(x)  
> plot(x,ylab="x")
```



# Príklad: testovanie jednotkového koreňa

---

- Potrebujeme mať stacionárne dáta → prvý krok pri modelovaní: potrebujeme zistiť, či môžeme pracovať s pôvodnými dátami alebo či ich treba zdiferencovať
- Kedy treba dáta diferencovať:
  - ◇ čo tomu nasvedčuje: ACF klesá veľmi pomaly a PACF má výraznú prvú hodnotu
  - ◇ máme štatistický test: ak ADF test nezamietne jednotkový koreň
- V dátach nemôže zostať trend; ani vtedy, keď ADF zamietne jednotkový koreň:
  - ◇ Aj tak diferencie → nie kvôli unit root, ale kvôli trendu
  - ◇ Alebo (nebudeme v tomto kurze robiť): modelovať samostatne trend a zostávajúcu stacionárnu časť

# Príklad: testovanie jednotkového koreňa

---

ADF (*augmented Dickey-Fuller*) test v R-ku:

- Knižnica `urca` (ur - *unit root*, ca - *cointegration*)
- Funkcia `ur.df` (ur - *unit root*, df - *Dickey-Fuller*), ktorá má parametre:
  - ◇ `type`: možnosti sú `drift` (konštanta bez lineárneho trendu), `trend` (konštanta aj lineárny trend), `none` (nič)
  - ◇ `lags`: maximálny počet lagov
  - ◇ `selectlags`: kritérium, podľa ktorého sa vyberá počet lagov (informačné kritériá: AIC, BIC)
- Hypotézu o jednotkovom koreni zamietame, ak je štatistika menšia ako kritická hodnota.

# Príklad: testovanie jednotkového koreňa

---

- V našom prípade:

- ◇ Príkaz v R-ku:

```
summary(ur.df(x,type="drift",lags=10,selectlags="BIC"))
```

(hodnota parametra `type` na základe priebehu dát, `lags` zvolíme "dost' veľké")

- ◇ Dostaneme:

```
Value of test-statistic is: -2.4442 3.1656
```

```
Critical values for test statistics:
```

	1pct	5pct	10pct
tau2	-3.51	-2.89	-2.58
phi1	6.70	4.71	3.86

- ◇ Záver: nulovú hypotézu o jednotkovom koreni nezamietame → dáta treba diferencovať

## Príklad: testovanie jednotkového koreňa

---

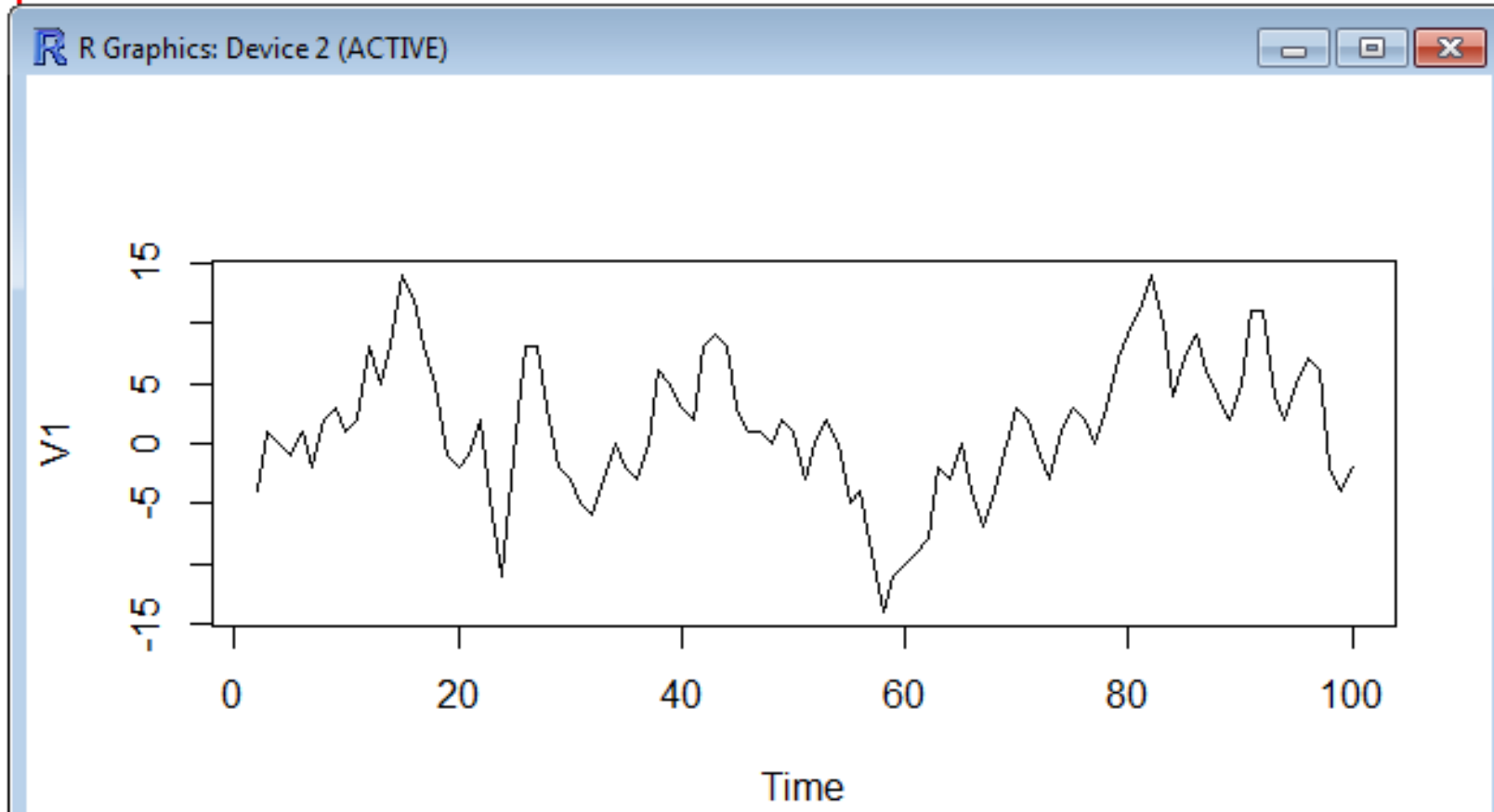
- Pokračujeme:
  - ◇ Stačí diferencovať dáta raz alebo majú aj diferencie jednotkový koreň?
  - ◇ Testujeme teda hypotézu o jednotkovom koreni pre diferencie
  - ◇ Pokračujeme v diferencovaní, kým sa jednotkový koreň nezamietne
- Priebeh diferencií: `plot(diff(x))`
- CVIČENIE: Testujte prítomnosť jednotkového koreňa v časovom rade diferencií.

# Príklad: testovanie jednotkového koreňa

---

```
> plot(diff(x))
```

```
>
```





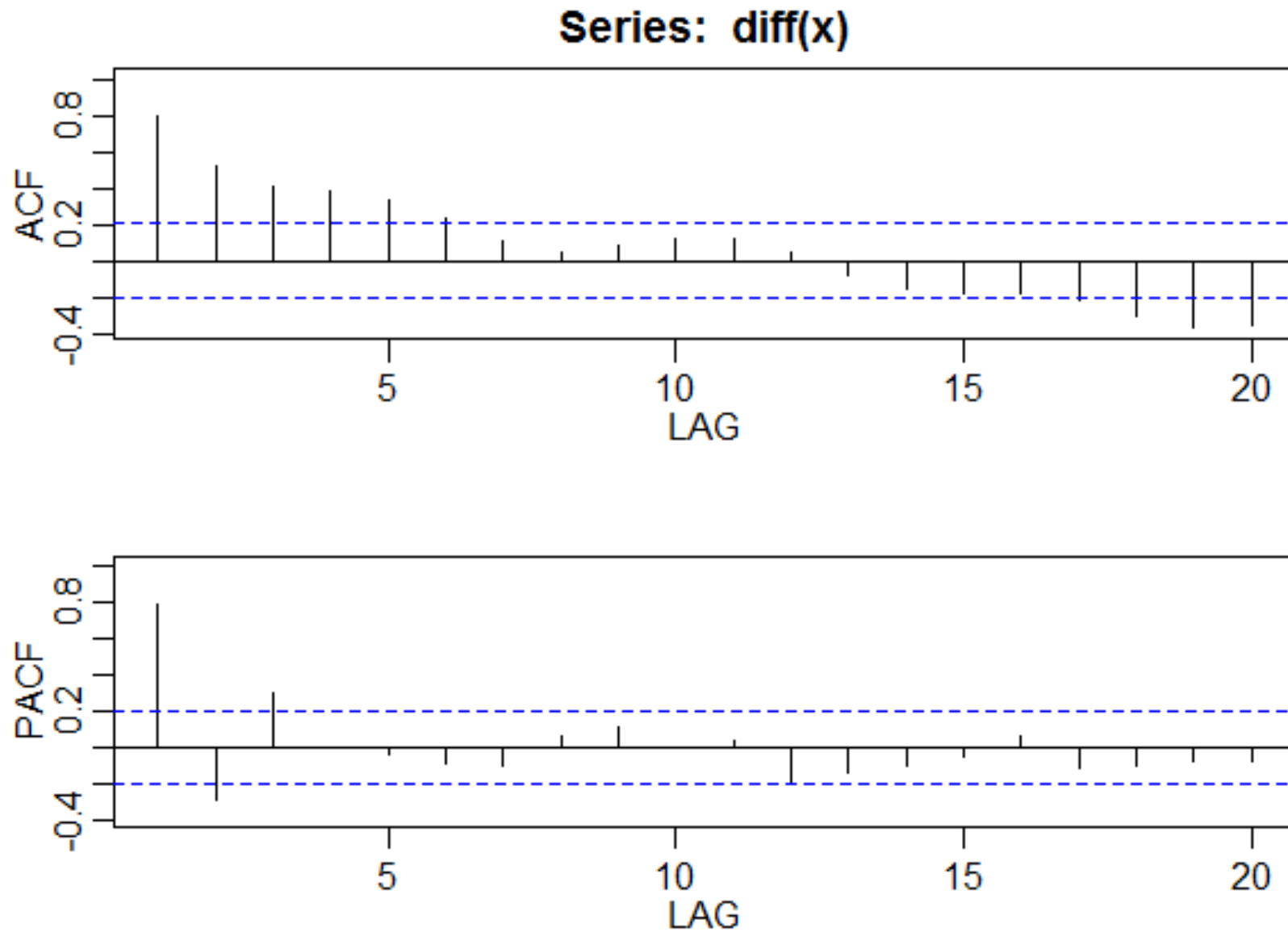
# Príklad: identifikácia ARMA modelu

---

- Ak už máme stacionárne dáta:
  - ◇ Načítame balík `astsa` (*applied statistical time series analysis*)
  - ◇ Zobrazíme ACF (bez nulového lagu) a PACF príkazom `acf2(data)`, v našom prípade `acf2(diff(x))`
  - ◇ Vyberieme vhodný model (ak sa dá), resp. rozhodneme sa skúšať viaceré modely

# Príklad: identifikácia ARMA modelu

---



# Príklad: odhadovanie a testovanie

---

- Odhadujeme **ARMA(p,q)** model pre **k-te** diferencie (teda **ARIMA(p,k,q)** pre pôvodné dáta): príkaz **sarima(x,p,k,q)**
- Výstup poznáme prednášky a minulého cvičenia:
  - ◇ odhadnuté koeficienty
  - ◇ ACF rezíduí
  - ◇ p-hodnoty Ljung-Boxovej štatistiky
- Užitočná funkcia pre **overovanie stacionarity a invertovateľnosti**: funkcia **armaRoots** z balíka **fArma**, napr. **armaRoots(c(1.5,-0.9))**, p. predchádzajúce cvičenie - vypočíta korene AR/MA polynómu a ich absolútne hodnoty - *pozor na znamienka!*
- **CVIČENIE**: Nájdite vhodný model pre naše dáta.

## Príklad: predikcie

---

- Máme  $ARIMA(p,k,q)$  model a chceme spraviť predikcie na  $N$  periód dopredu:
  - ◇ príkaz `sarima.for(x,N,p,k,q)`
  - ◇ dostaneme predikcie a intervaly spoľahlivosti
- CVIČENIE: Spravte vo vašom modeli predikcie na nasledujúcich 10 minút

# *Cvičenia*

---

- Zadania na nájdenie modelu pre dáta zo starých skúšok  
→ rozdane na cvičení