

Cvičenie v R-ku I.: ARIMA modely

Beáta Stehlíková

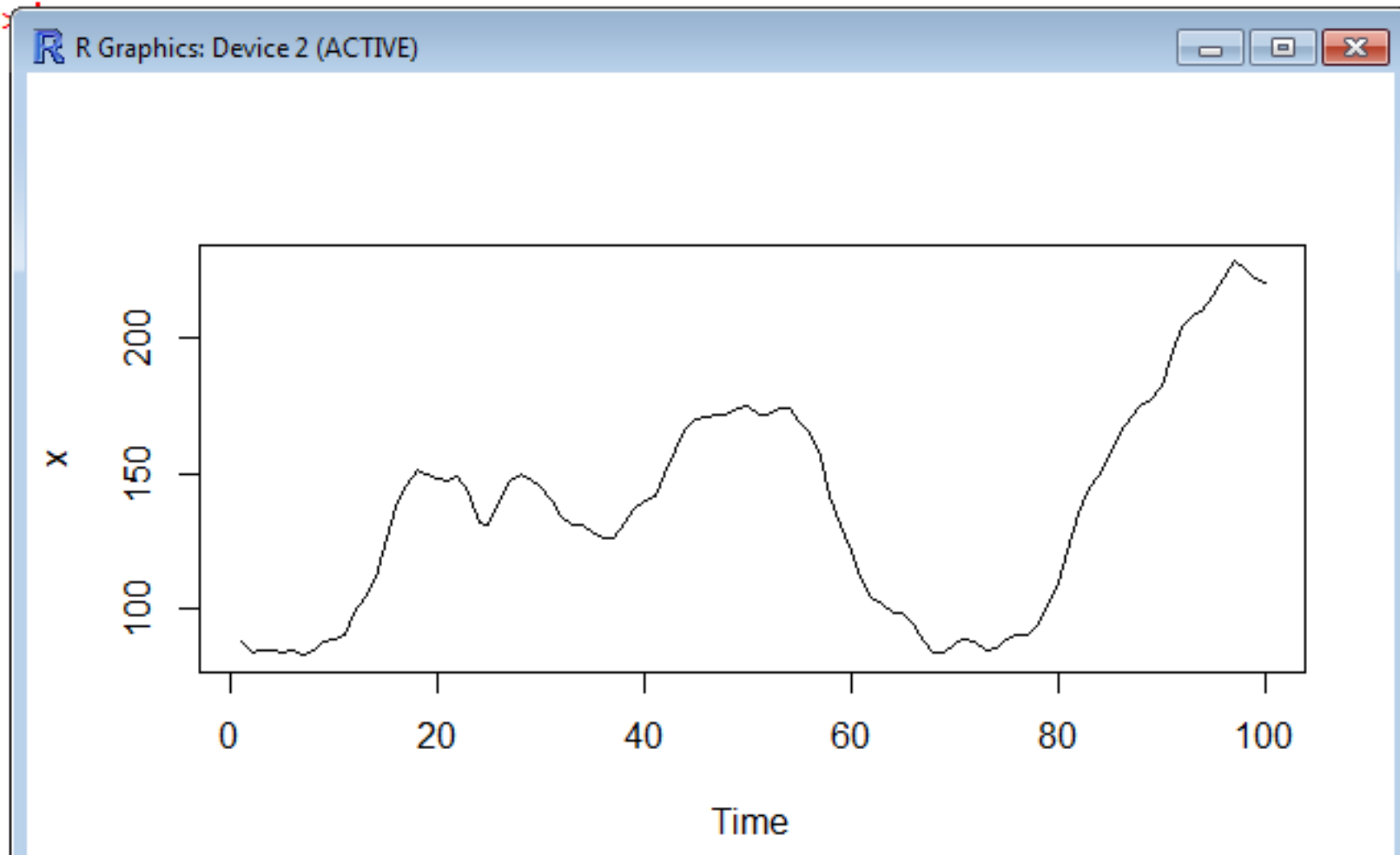
Časové rady, FMFI UK, 2013/2014

Príklad 1: dáta

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

Príklad 1: dáta

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



Príklad 1: 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:
 - ◇ ak ACF klesá veľmi pomaly a PACF má výraznú prvú hodnotu
 - ◇ ak ADF test nezamietne jednotkovú koreň
- V dátach nemôže zostať trend; ani vtedy, keď ADF zamietne jednotkový koreň:
 - ◇ Môžeme skúsiť aj tak diferencie, ak tomu nasvečuje ACF a PACF
 - ◇ Alebo (nebudeme v tomto kurze robiť): modelovať samostatne trend a zostávajúcu stacionárnu časť

Príklad 1: 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 1: 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

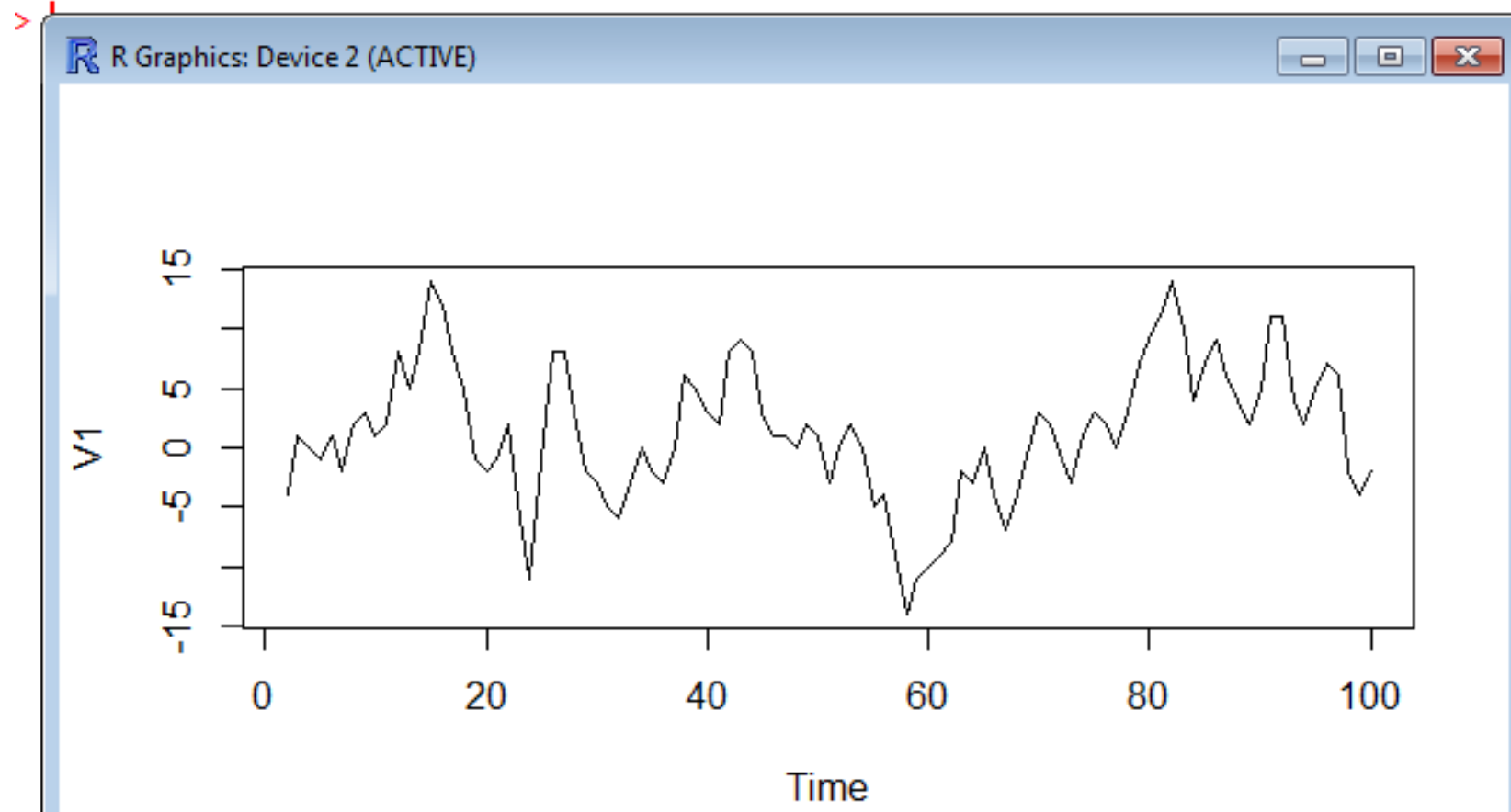
- ◇ Záver: nulovú hypotézu o jednotkovom koreni nezamietame → dáta treba diferencovať
- ◇ Zvyšok výstupu budeme analyzovať neskôr

Príklad 1: 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 1: testovanie jednotkového koreňa

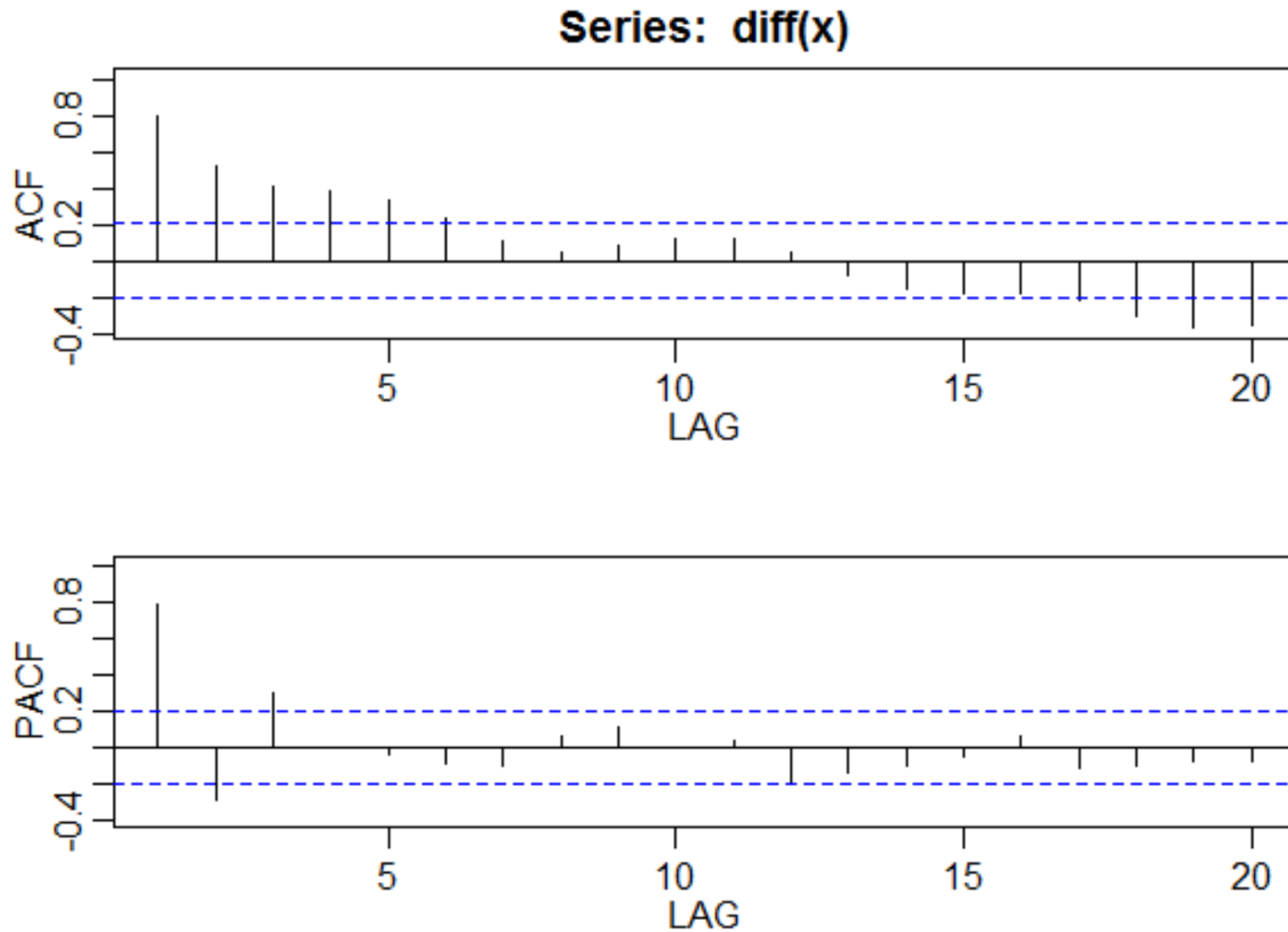
```
> plot(diff(x))
```



Príklad 1: 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 1: identifikácia ARMA modelu



Príklad 1: 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:
 - ◇ 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))** - vypočíta korene AR/MA polynómu a ich absolútne hodnoty
- **CVIČENIE**: Nájdite vhodný model pre naše dáta.

Príklad 1: predikcie

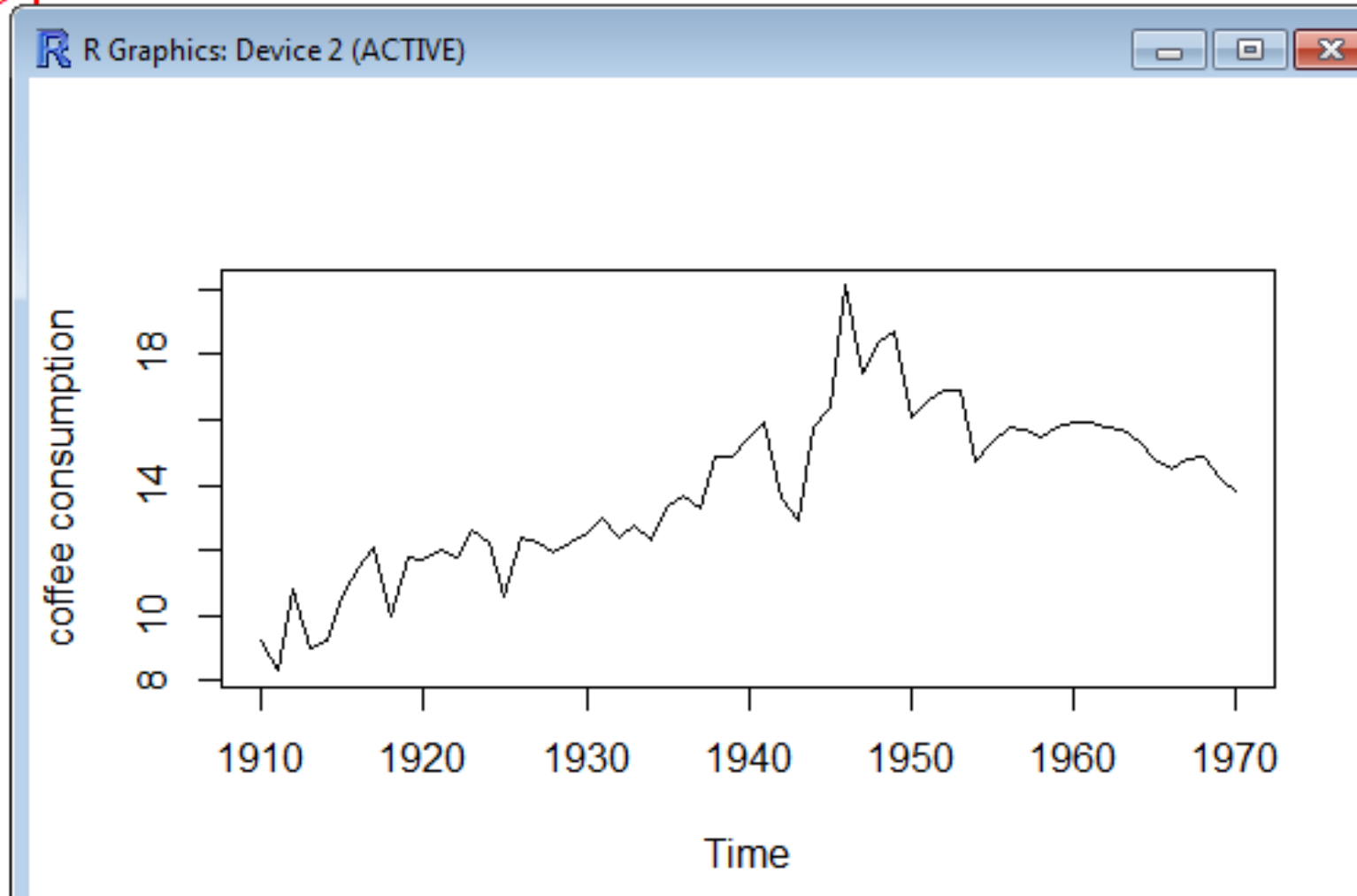
- Máme $ARIMA(p,k,q)$ model a chceme spraviť predikcie na N periód dopredu:
 - ◇ príkaz `sarima.for(x,N,p,k,q)`
 - ◇ dostaneme predikcie a intervaly spoľahlivosti
- CVIČENIE: Spravte vo vašom modeli predikcie na nasledujúcich 10 minút

Príklad 2

- **Nájdite vhodný ARMA/ARIMA model** pre tieto dáta:
 - ◇ Spotreba kávy v USA v rokoch 1910 - 1970
 - ◇ Zdroj:
<http://robjhyndman.com/tsdldata/data/coffee.dat>
 - ◇ Na stránke cvičení: [kava.txt](#)
- Poznámka k funkcii `ts`:
 - ◇ Ak máme mesačné, kvartálne alebo ročné dáta, dá sa spraviť napr.:
`x=ts(x,frequency=12,start=c(2000,1))`
pre mesačné dáta so začiatkom v januári 2000
 - ◇ Na x-ovej osi budeme mať tieto časy

Príklad 2

```
> x=ts(x,frequency=1,start=1910)
> plot(x,ylab="coffee consumption")
> |
```



Príklad 3

- Dáta: ceny kakaa (boli aj na prednáške); na stránke cvičení: [pcocoa.txt](#)
- CVIČENIE:
 - ◇ Rovnako ako doteraz: nájdite vhodný model pre logaritmy cien a vysvetlite, ako ste k nemu dospeli
 - ◇ Zadanie z prednášky: pre diferencie logaritmov sa dá použiť AR(2) aj MA(1) model - ako je to možné; porovnajte ich Woldovu reprezentáciu
 - ◇ Zadanie z prednášky: pre diferencie logaritmov porovnajte ARMA(1,2) a MA(1) model (korene ARMA modelu)
 - ◇ Testovanie jednotkového koreňa: z akej regresie sa získa štatistika, prečo takáto regresia, aký model pre dáta to znamená a ako sa z neho dostaneme k regresii z výstupu?