

## Cvičenie 2

### Časové rady

Časové rady sú dátové množiny pri ktorých sú jednotlivé hodnoty usporiadané sekvenčne v čase. Sú definované počiatočným a koncovým časom (napr. dáta od roku 2016 do 2017) a frekvenciou vzorkovania (napr. s frekvenciou 12 pre mesačné dáta). V prostredí R sa používajú na reprezentáciu časových radov objekty typu `ts`, ktoré môžete priamo načítať ako dátovú množinu. Funkcia `summary(ts)` zobrazuje len základné charakteristiky číselný hodnôt, ktoré tvoria časový rad. Počiatočný a koncový čas a frekvenciu vzorkovania môžete vypísať funkciami `start(ts)`, `end(ts)` a `frequency(ts)`:

```
data("AirPassengers")
ap <- AirPassengers
summary(ap)
start(ap); end(ap); frequency(ap);
```

Množina `AirPassengers` udáva nárast cestujúcich v leteckej dobe v USA a obsahuje hodnoty od januára 1949 (1949,1) do decembra 1960 (1949,12). Vzorkovacia frekvencia je 12 hodnôt na rok (tzn. mesačné dáta). Časové rady si môžete priamo vykresliť pomocou funkcie `plot`:

```
plot(ap, type="o")
```

Pomocou funkcie `window` je možné vybrať podmnožinu časového radu resp. zmeniť vzorkovaciu frekvenciu, napr. ak chceme vybrať hodnoty iba za rok 1950 (od januára do decembra) môžeme zadať:

```
window(ap, start=c(1950,1), end=c(1950,12))
```

Dáta za jednotlivé roky (hlavné časové jednotky) je možné agregovať pomocou funkcie `aggregate`, napr. priemerné hodnoty za rok je možné vypočítať ako:

```
aggregate(ap, FUN=mean)
```

Dáta je možné agregovať aj s inou frekvenciou, napr. sumu za pol roka (1/2 hlavnej jednotky) a za štvrt' rok (1/4 hlavnej jednotky) môžete vypočítať ako:

```
aggregate(ap, FUN=sum, nfrequency=2)
aggregate(ap, FUN=sum, nfrequency=4)
```

Všetky hodnoty je možné agregovať priamo, napr. priemer môžeme určiť funkciou `mean`:

```
mean(ap)
```

Objekt časového radu je možné vytvoriť aj z vektora číselných hodnôt. Napr. pre priebeh funkcie  $0.25\sin(10x) + 0.5x + 1$  na intervale 0..5 s krokom 0.1:

```
x <- seq(from=0, to=5, by=0.1)
y <- 0.25*sin(10*x) + 0.5*x + 1
```

Objekt vytvoríme funkciou `ts`, kde nastavíme počiatočný čas a frekvenciu vzorkovania, napr.:

```
t <- ts(y, start=c(2017,1), frequency=12)
plot(t)
```

Pomocou funkcie `ts` môžeme vytvoriť časový rad aj z tabuľkových dát. Dáta môžeme stiahnuť priamo z internetu, napr. nasledujúci súbor `Maine.dat` udáva mesačnú mieru nezamestnanosti v štáte Maine za obdobie 1996 do 2006:

```
data <- read.table("http://web.tuke.sk/fei-
cit/butka/ie/Maine.dat", header=TRUE)
maine <- ts(data, start=c(1996,1), frequency=12)
```

## Základné charakteristiky časových radov

Z priebehu dát `AirPassengers` vidno, že sa počet pasažierov postupne systematicky zvyšuje, tzn. dáta majú rastúci trend. Okrem trendu vidno cyklické zmeny, ktoré sa periodicky opakujú – ide o tzv. sezónnosť. Časové rady môžeme modelovať rôznou kombináciou trendu  $t$ , sezónnosti  $s$  a náhodnej zložky  $z$ . Najjednoduchším modelom je aditívny model, kde sa hodnoty radu rovnajú súčtu zložiek:  $x = t + s + z$ . Medzi ďalšie modely patrí napr. multiplikatívny model:  $x = t \cdot s \cdot z$ .

Najjednoduchší spôsob extrahovania trendu je pomocou centrovaného kĺzavého priemeru v okolí každej hodnoty  $x_t$ , ktorý je možné vypočítať pomocou funkcie `filter`. Centrovaný kĺzavý priemer je vážený priemer, kde prvá a posledná hodnota má váhu  $1/2$ , tzn. napr. pre 12 mesiacov sú váhy  $(0.5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5)/12$ :

```
ap.t = filter(ap, c(0.5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0.5)/12)
plot(ap.t)
```

Pre aditívny model je možné zvyšné zložky extrahovať odpočítaním od pôvodného radu:

```
ap.sz = ap - ap.t
plot(ap.sz)
```

Z priebehu vidno, že sa cyklické zložky postupne zvyšujú, čo môže indikovať, že je pre tieto dáta vhodnejší multiplikatívny model. Všetky zložky časového radu pre rôzne typy modelov je možné vypočítať pomocou funkcie `decompose`, napr. pre multiplikatívny model:

```
ap.decomp = decompose(ap, type="multiplicative")
plot(ap.decomp)
```

Jednotlivé zložky modelu sú dostupné vo vlastnostiach `seasonal`, `trend` a `random`. Napr. nasledujúci príkaz vypočíta základné štatistiky pre sezónnu zložku:

```
summary(ap.decomp$seasonal)
```

## Úlohy na cvičení

1. Načítajte súbor <http://web.tuke.sk/fei-cit/butka/ie/Maine.dat> a vytvorte časový rad pre mesačné dáta od januára 1996. Vykreslite priebeh. Vypočítajte priemernú hodnotu pre prvé tri a posledné tri mesiace v roku 1996. Zistite o koľko % sa tieto priemery odlišujú od priemeru počas celého roku 1996.
2. Pre dátovú množinu Maine vypočítajte agregovaný časový rad pre priemerné štvrtročné hodnoty a zobrazte jeho priebeh.
3. Načítajte súbor <http://web.tuke.sk/fei-cit/butka/ie/CBE.dat>. Dáta obsahujú dátovú množinu so stĺpcami určujúcimi mesačnú produkciu čokolády (stĺpec 1), piva (2) a elektrickej energie (3) od roku 1958. Načítajte dáta ako časové rady (pre výber stĺpca môžete použiť zápis CBE[,1], CBE[,2], CBE[,3]). Zobrazte všetky tri časové rady v jednom grafe pomocou funkcie plot a cbind.
4. Vypočítajte koreláciu medzi produkciou čokolády a piva.
5. Na dátovej množine AirPassengers postupne aplikujte kĺzavý priemer s šírkou 4 mesiace, 6 mesiacov a 12 mesiacov (tzn. váhy pre 4 mesiace budú  $(0.5, 1, 1, 0.5)/4$ ). Vyextrahujte trendovú zložku pre každý filter podľa aditívneho modelu a zobrazte jej priebeh.
6. Pomocou funkcie decompose vygenerujte a zobrazte zložky pre aditívny a multiplikatívny model pre dátovú množinu AirPassengers
7. Na intervale  $[0..5]$  s krokom 0.1 vygenerujte priebeh funkcií  $y = 0.5x + 1$  a  $z = 0.25\sin(10x)$ . Vygenerujte rovnako veľký náhodný vektor  $n$  podľa normálneho rozdelenia s 0 strednou hodnotou a štandardnou odchýlkou 0.5. Vykreslite priebehy funkcií na jednom zloženom grafe pomocou funkcie plot a cbind.
8. Zo zložiek s predchádzajúceho príkladu vytvorte časový rad s aditívnym modelom  $y + z + n$ , a multiplikatívnym modelom  $y \cdot z \cdot n$ .
9. Dekomponujte časové rady z predchádzajúceho priebehu pomocou funkcie decompose. Vytvorte pre oba priebehy obe dekompozície aj pre aditívny aj multiplikatívny model.
10. Nahraďte lineárny trend logaritmickým  $y = \log(x + 1)$ , vytvorte aditívny a multiplikatívny časový rad a dekomponujte ich zložky.