

Time Series HomeWork (1)

钟瑜 222018314210044

2020 年 9 月 24 日

1. 附录 B 中的 B6 是 1973 至 1978 年美国在意外事故中的死亡人数。

利用至少两种方法对该时间序列进行分解。要求如下：

- (1) 画出数据图，给出数据的周期 T ；
- (2) 给出趋势项、季节项和随机项的计算公式；
- (3) 画出趋势项、季节项和随即项的数据图；
- (4) 对 1979 年的意外死亡人数做出预测。

解. 数据图如下：

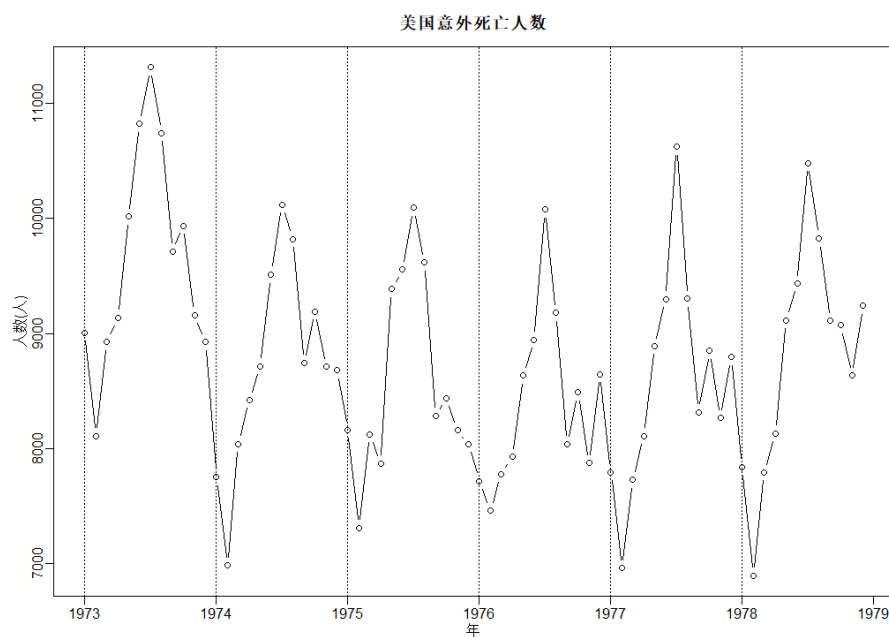


图 1: 美国意外死亡人数

显然如图可知，数据的周期为 12 个月。

- 方法一：分段趋势

趋势项 $\{T_t\}$ 定义成年平均值, 公式如下

$$T_t = T(t) = \begin{cases} \frac{\sum_{j=1}^{12} c_j}{12}, t = 1, \dots, 12 \\ \frac{\sum_{j=13}^{24} c_j}{12}, t = 13, \dots, 24 \\ \dots \\ \frac{\sum_{j=61}^{72} c_j}{12}, t = 61, \dots, 72 \end{cases} \quad (1)$$

$$= \begin{cases} 8982.083333, t = 1, \dots, 12 \\ 8723.75, t = 13, \dots, 24 \\ 8585.833333, t = 25, \dots, 36 \\ 8396.75, t = 37, \dots, 48 \\ 8576.833333, t = 49, \dots, 60 \\ 8796.75, t = 61, \dots, 72 \end{cases} \quad (2)$$

其中 c_j 表示第 j 个月的死亡人数; 用第 k 个月的平均值作为季节项 $S(k), 1 \leq k \leq 12$ 的估计. 用 $x_{j,k}, T_{j,k}$ 分别表示第 j 年 第 k 个月的数据和趋势项, 则时刻 (j,k) 的时间次序指标为 $k + 12(j - 1)$

$$\begin{aligned} S(k) &= \frac{1}{6} \sum_{j=1}^6 (x_{j,k} - T_{j,k}) \\ &= \frac{1}{6} \sum_{j=0}^5 (x_{k+12j} - T_{k+12j}), 1 \leq k \leq 12 \end{aligned} \quad (3)$$

计算结果如下

$S(k)$
-744.61111
-1504.77778
-724.77778
-523.77778
337.55556
806.72222
1664.22222
960.55556
-88.27778
206.55556
-321.44444
-67.94444

下图为原始序列、趋势、拟合 (包括趋势与季节项):

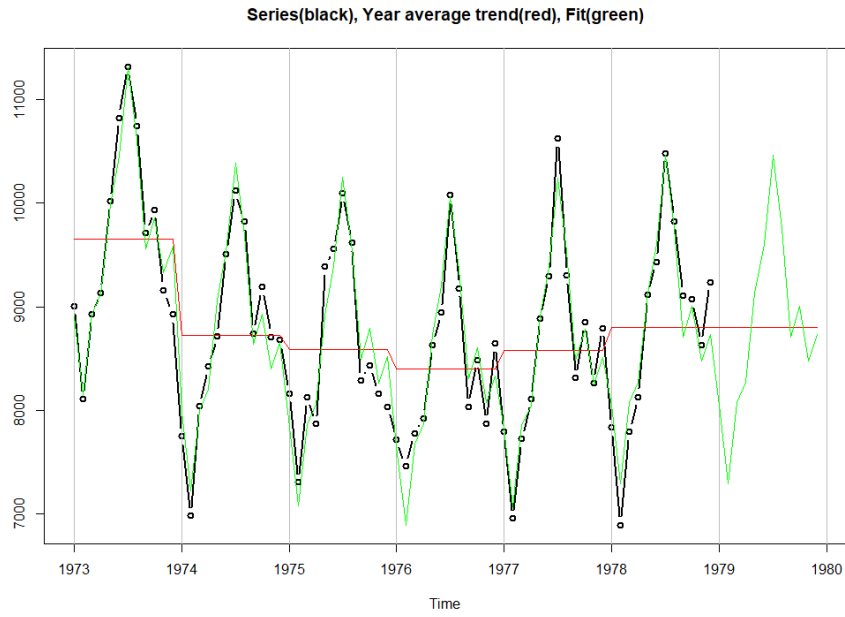


图 2: 原始序列、趋势、拟合（包括趋势与季节项）

最后，随机项的计算公式

$$R_t = x_t - T_t - S_t, 1 \leq t \leq 72 \quad (4)$$

下图为去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项：

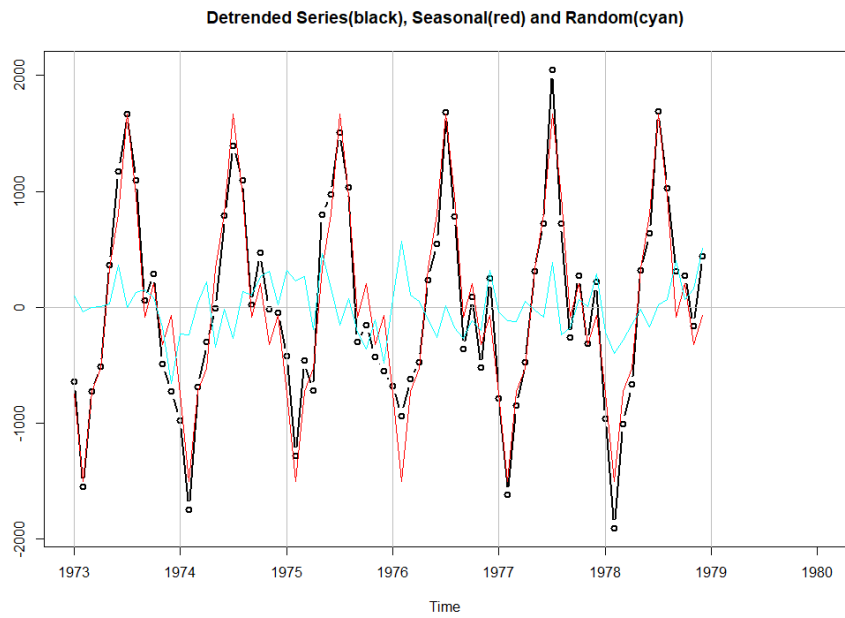


图 3: 去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项

- 方法二：回归直线趋势

趋势项用回归直线表示，这时认为 (x_t, t) 满足一元线性回归模型

$$x_t = a + bt + \epsilon_t, t = 1, \dots, 72 \quad (5)$$

定义

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_{72} \end{pmatrix}^T,$$

$$\mathbf{Y} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 2 & \dots & 72 \end{pmatrix},$$

$$\epsilon = \begin{pmatrix} \epsilon_1 & \epsilon_2 & \dots & \epsilon_{72} \end{pmatrix}^T,$$

$(a, b)^T$ 的最小二乘估计由公式

$$(a, b)^T = (Y Y^T)^{-1} Y X$$

决定, 用 r 计算得

$$a = 9099.17$$

$$b = -8.51$$

回归方程为

$$x_t = 9099.17 + -8.51t \quad (6)$$

趋势项 $\{T_t\}$ 的估计值是回归直线

$$T_t = 9099.17 + -8.51t$$

下图绘制了原始数据、估计的趋势、拟合值（包括趋势与季节项）：

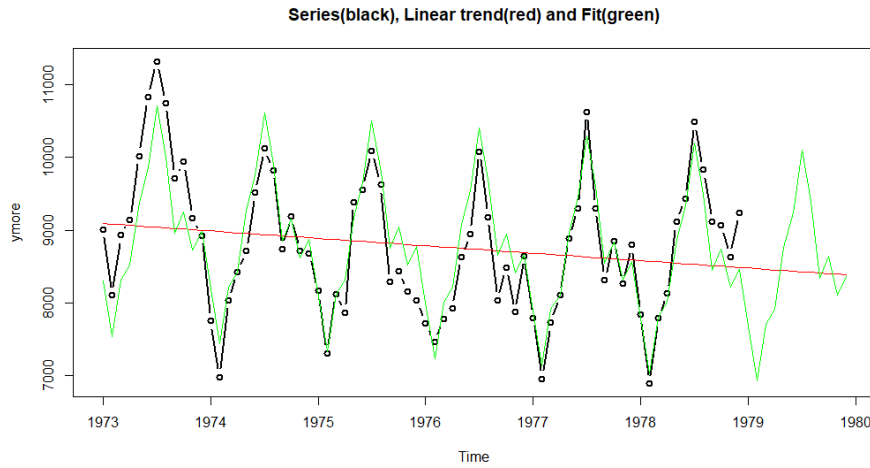


图 4: 原始数据、估计的趋势、拟合值（包括趋势与季节项）

利用原始数据 $\{x_t\}$ 减去趋势项的估计 $\{T_t\}$ 后得到的数据基本只含有季节项和随机项。我仍用第 k 月的平均值作为季节项 $S(k)$ 的估计。利用方法一的公式计算结果如下

$S(k)$
-791.40798
-1543.06612
-754.55760
-545.04908
324.79277
802.46796
1668.47648
973.31834
-67.00647
236.33538
-283.15610
-21.14758

最后，随机项的计算公式

$$R_t = x_t - T_t - S_t, 1 \leq t \leq 72$$

(7)

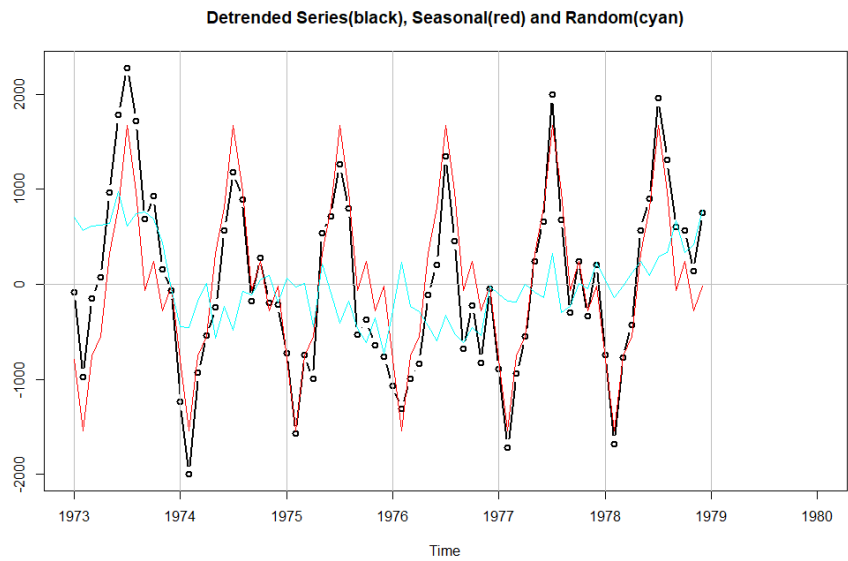


图 5: 去掉了趋势后的序列、季节项、取掉了趋势与季节项后的随机项

附录

A 数据图

```
1 B6.siwangrenshu <-
2 ts(c(
3 1973,9007,8106,892,9137,10017,10826,11317,10744,9713,9938,9161,8927,
4 1974,7750,6981,8038,8422,8714,9512,10120,9823,8743,9192,8710,8680,
```

```

5 1975,8162,7306,8124,7870,9387,9556,10093,9620,8285,8433,8160,8034,
6 1976,7717,7461,7776,7925,8634,8945,10078,9179,8037,8488,7874,8647,
7 1977,7792,6957,7726,8106,8890,9299,10625,9302,8314,8850,8265,8796,
8 1978,7836,6892,7791,8129,9115,9434,10484,9827,9110,9070,8633,9240),
9 frequency=12, start=c(1973,1))
10
11 demo.siwang.data <- function(){
12 opar <- par(mar=c(3,3,3,1), mgp=c(1.5,0.5,0))
13 on.exit(par(opar))
14 plot(B6.siwangrenshu, lty=1, type='b',
15 main='美国意外死亡人数',
16 xlab='年', ylab='人数(人)')
17 abline(v=1973:1978, lty=3)
18 }
19 demo.siwang.data()

```

B 方法一的代码

```

1
2 y <- B6.siwangrenshu
3 ymore <- ts(c(y, rep(NA,12)), start=start(y), frequency=12)
4 ymat <- matrix(c(y), byrow=TRUE, nrow=6, ncol=12)
5
6 cols <- rainbow(20)
7 ic <- 1
8
9 ## 用同季度的值平均得到12个季节项
10 get.season <- function(yd){ # input: Detrended series
11     ymat <- matrix(c(yd), byrow=TRUE, ncol=12)
12
13     ## season
14     seas0 <- apply(ymat, 2, mean, na.rm=TRUE)
15
16     seas0
17 }
18
19 ## 画去除趋势后的序列、季节项和随机项
20 plot.season <- function(yd, seas0){ # input: Detrended series
21     ## season
22     seas <- rep(seas0, 6)
23     seas <- ts(seas, start=c(1973,1), frequency=12)

```

```

24
25     ## Error
26     r <- c(yd) - seas
27     r <- ts(r, frequency=12, start=c(1973,1))
28
29     plot(yd, type='b', lwd=2,
30          main="Detrended□Series (black)□□Seasonal (red)□and□Random (cyan)",
31          xlim=c(1973,1980), ylab="")
32     abline(v=1973:1979, col="gray")
33     abline(h=0, col="gray")
34     lines(seas, type="l", col="red")
35     lines(r, type="l", col="cyan")
36 }
37
38 plot.season(y.detrended, seas0)

```

C 方法二的代码

```

1
2 yy <- c(y)
3 tt <- seq(length(y))
4 lmr <- lm(yy ~ tt)
5 tr.more <- ts(predict(lmr, newdata=list(tt=seq(length(y)+12))),
6 frequency=12, start=c(1973,1))
7
8 ## season
9 y.detrended <- y - tr.more[1:length(y)]
10 seas0 <- get.season(y.detrended)
11 seas.more <- ts(rep(seas0, 7),
12 start=start(y), frequency=12)
13 y.pred <- tr.more + seas.more
14
15 plot(ymore, main="Series (black)□□Linear□trend (red)□and□Fit (green)",
16 lwd=2,
17 type="b", col="black")
18 lines(tr.more, col="red", type="l")
19 lines(y.pred, col="green", type="l")
20
21 plot.season(y.detrended, seas0)

```