

Time Series Analysis

shake_it

第一周

方法

- 描述性时序分析
- 统计时序分析 (频域分析方法 + 时域分析方法)

生成数据

从 2005 年 1 月开始的月度数据。start 指定起始读入时间, frequency 指定序列每年读入的数据频率。

```
price <- c(101, 82, 66, 35, 31, 7)

price <- ts(price, start = c(2005, 1), frequency = 12)
```

例 1-1

读入 1884-1939 年英格兰和威尔士小麦平均亩产量数据 file1.csv

```
x <- read.csv("D:/Documents/UIBE/6/TimeSeries/file1.csv")
head(x)
```

```
##   year yield
## 1 1884  15.2
## 2 1885  16.9
## 3 1886  15.3
## 4 1887  14.9
## 5 1888  15.7
## 6 1889  15.1
```

截取 1925 年之后的数据 subset

```
z <- subset(x, year > 1925, select = yield)
head(z)
```

```
##   yield
## 43  16.0
## 44  16.4
## 45  17.2
## 46  17.8
## 47  14.4
## 48  15.0
```

对 yield 序列进行对数变换, 并将对数序列和原序列值导出, 保存为数据文件 yield.csv。

```
ln_yield <- log(x$yield)
x_new <- data.frame(x, ln_yield) # 新数据框
write.csv(x_new, file = "D:/Documents/UIBE/6/TimeSeries/yield.csv", row.names = F)
```

缺失值插值

R 中缺失值用 NA 表示。常用的插值方法：线性插值和样条插值

```
library(zoo)
a <- 1:7
a[4] <- NA
cat("a: ", a)

## a:  1 2 3 NA 5 6 7

y1 <- na.approx(a)
y2 <- na.spline(a)
cat(" ", y1, "\n ", y2)

##  1 2 3 4 5 6 7
##  1 2 3 4 5 6 7
```

第二周时间序列数据的预处理

统计性质

平稳时间序列自协方差函数和自相关系数只依赖于时间的平移长度而与时间的起止点无关

- 延迟 k 自协方差函数

$$\gamma(k) = \gamma(t, t+k), \quad \forall k \in \mathbb{N}$$

- 延迟 k 自相关系数

$$\rho_k = \frac{\gamma(t, t+k)}{\sqrt{DX_t \cdot DX_{t+1}}} = \frac{\gamma(k)}{\gamma(0)}$$

- 估计均值函数

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- 估计延迟 k 自相关系数

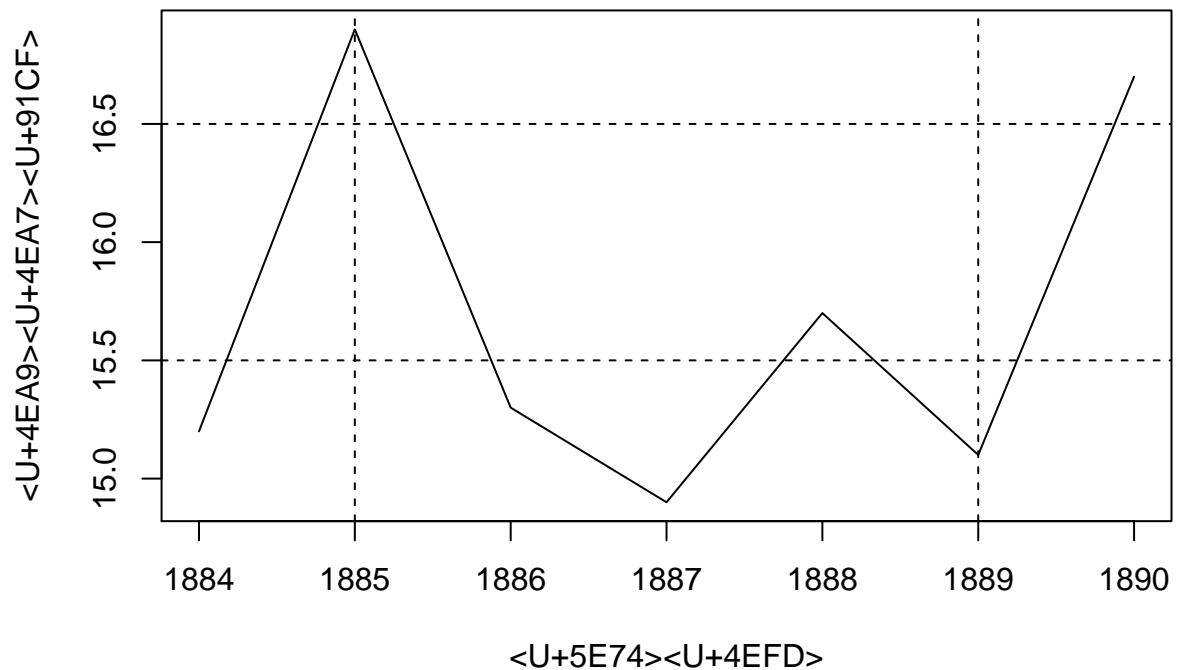
$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \forall 0 < k < n$$

时序图与自相关图

时序图

```
yield <- c(15.2,16.9,15.3,14.9,15.7,15.1,16.7)
yield <- ts(yield, start = 1884)
plot(yield, main = "1884-1890 年英格兰和威尔士地区小麦平均亩产量",
      xlab = " 年份", ylab=" 亩产量")
abline(v = c(1885,1889),
       h = c(15.5,16.5),
       lty = 2)
```

548C><U+5A01><U+5C14><U+58EB><U+5730><U+533A><U+5C0F><U



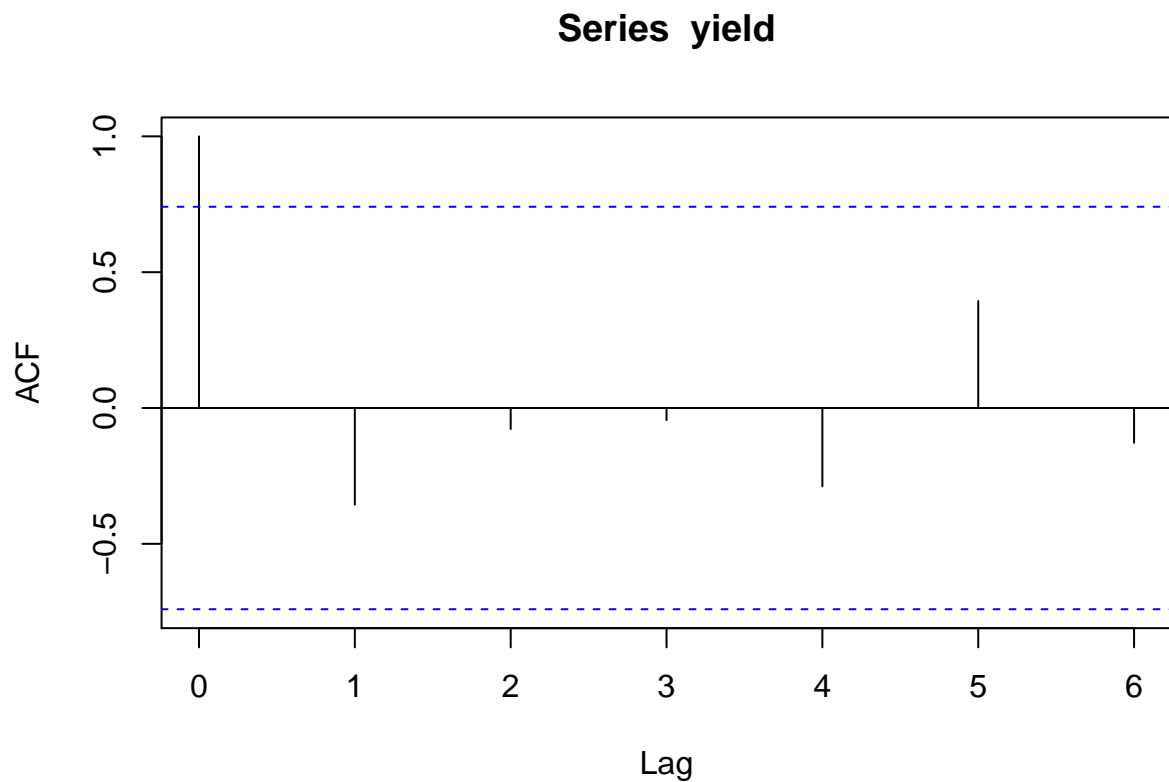
plot 各项参数:

```
if(FALSE) {
  type = "p" # 点
        "l" # 线
        "b" # 点连线
        "o" # 线穿过点
        "h" # 悬垂线
        "s" # 阶梯线
  pch = 17 # 点的符号
  lty = 2  # 连线的类型
  lwd = 2  # 连线的宽度 (默认宽度的 2 倍)
  col = 1  # col = "black"
  col = 2  # col = "red"
  col = 3  # col = "green"
  col = 4  # col = "blue"
  xlim = c(1886,1890)
  ylim = c(15,16) # 指定坐标轴范围
}
```

自相关图

自相关图是一个平面悬垂线图，横坐标表示延迟期数，纵坐标表示自相关系数，悬垂线表示自相关系数的大小。

```
acf(yield) # 虚线为自相关系数 2 倍标准差位置
```



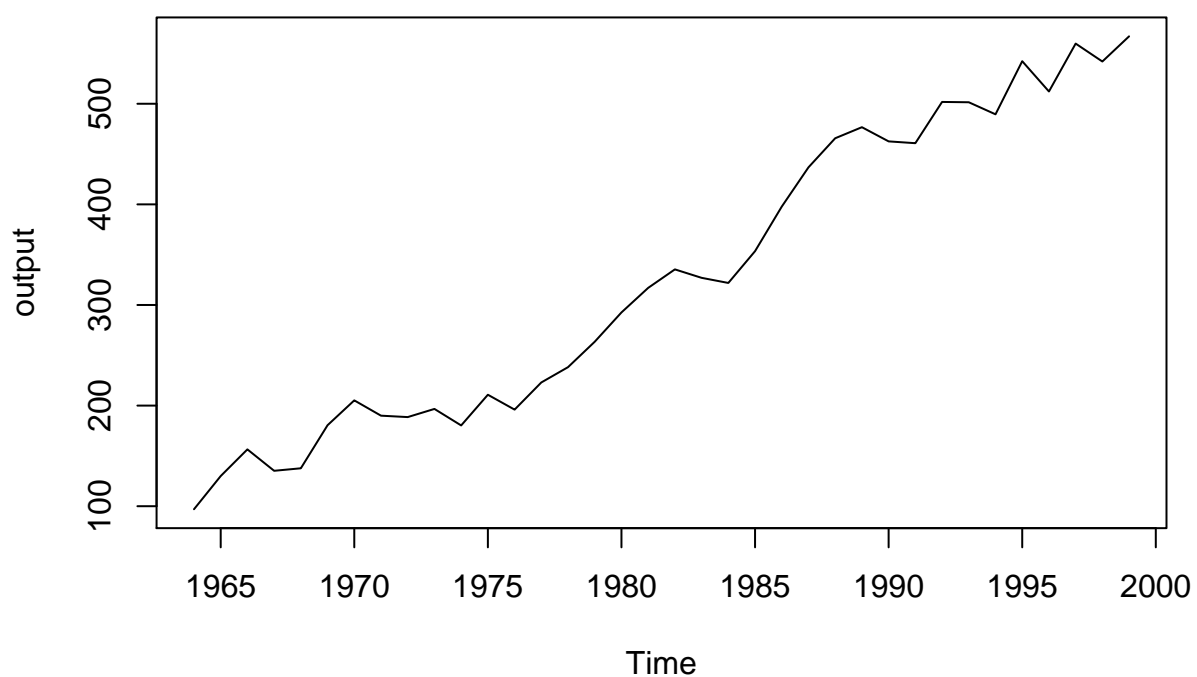
平稳性的检验（图检验方法）

- 时序图检验：始终在一个常数值附近随机波动，而且波动的范围有界、无明显趋势及周期特征。
- 自相关图检验：平稳序列通常具有短期相关性。该性质用自相关系数来描述就是随着延迟期数的增加，平稳序列的自相关系数会很快地衰减向零。

例 2.1 检验 1964 年-1999 年中国纱年产量序列的平稳性

```
library(readr)
sha <- read_csv("timeseries_data/file4.csv")
output <- ts(sha$output, start = 1964)
plot(output, main = " 时序图")
```

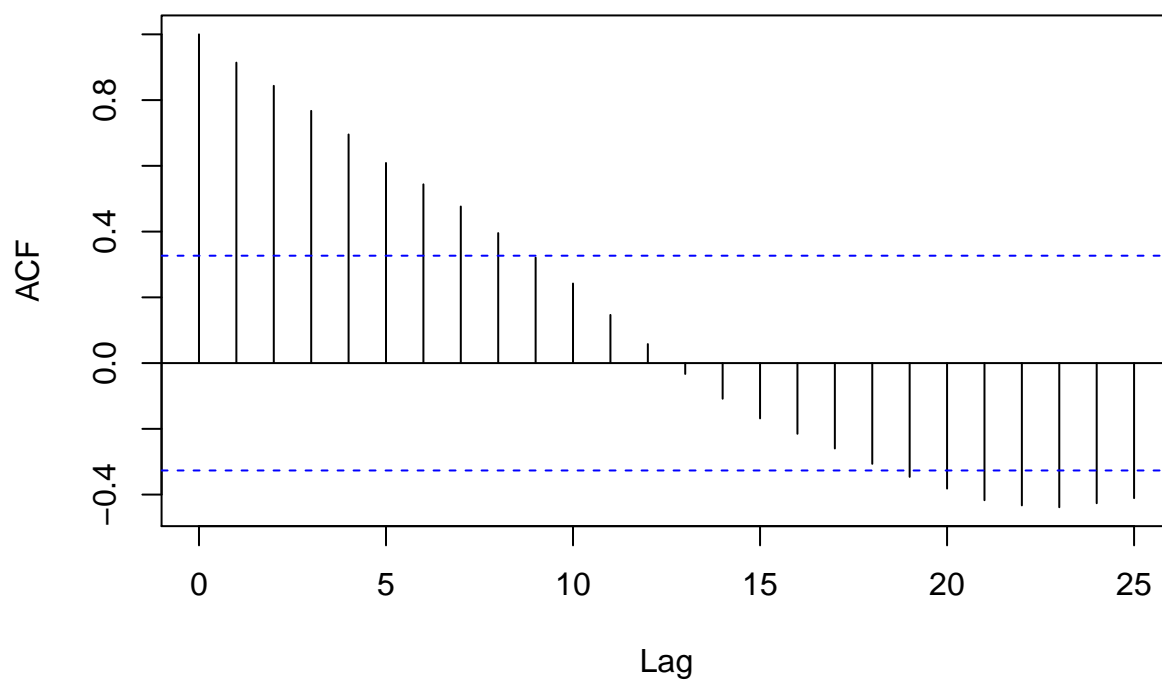
<U+65F6><U+5E8F><U+56FE>



时序图有递增趋势，非平稳

```
acf(output, lag = 25,  
      main = " 自相关图")
```

<U+81EA><U+76F8><U+5173><U+56FE>



自相关图衰减缓慢，非平稳。