Tutorial 8

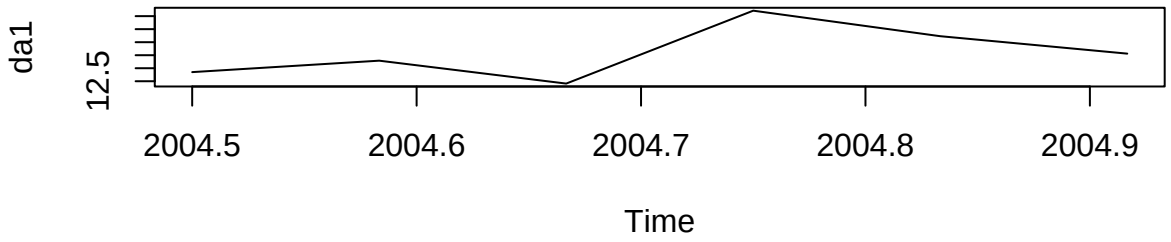
Dayu

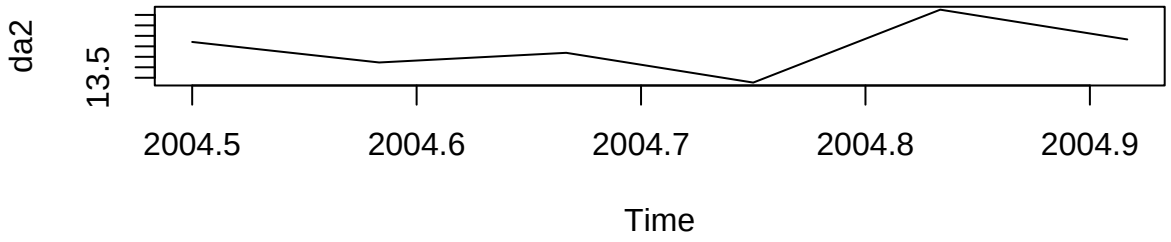
2022-05-21

案例 1, P35, 绘制时序图

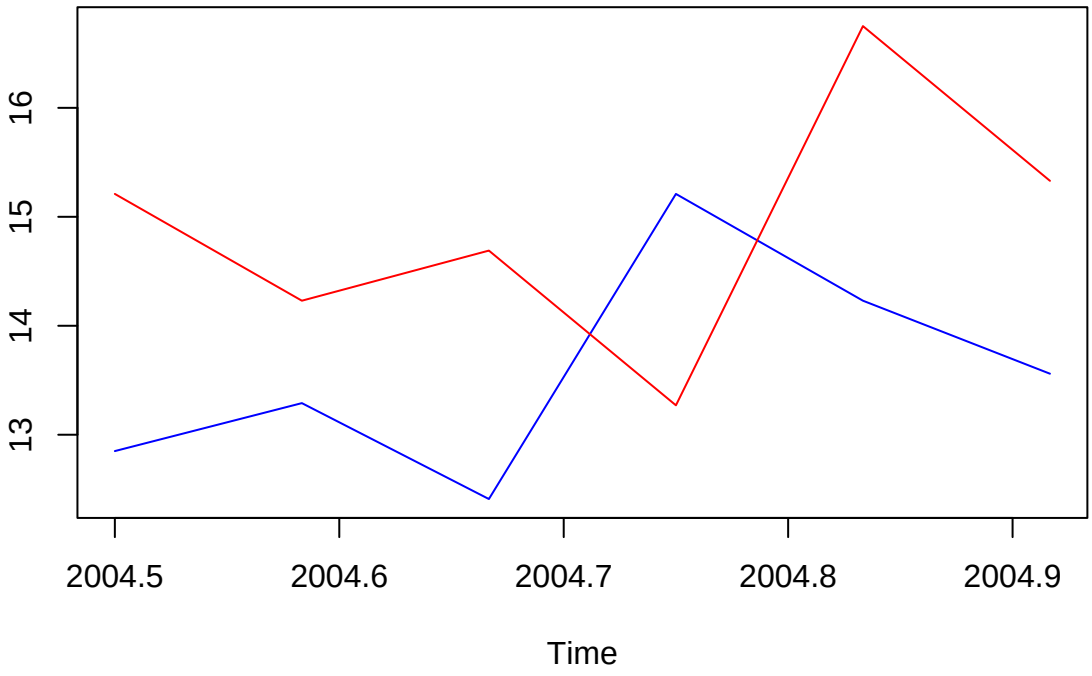
```
x1 <- c(12.85,13.29,12.41,15.21,14.23,13.56)
x2 <- c(15.21,14.23,14.69,13.27,16.75,15.33)
da <- data.frame(x1, x2)
# ts 函数：是将数值型数据向量转化成一个时间序列对象。
# 即每个数据带上一个日期标签
da1 <- ts(da[,1],frequency = 12, start = c(2004,7))
da2 <- ts(da[,2],frequency = 12, start = c(2004,7))
# frequency = 1(年), 12(月), 4(季度)

# 将两张图放在一页
par(mfcol = c(2,1))
plot(da1,type = 'l')  #'l'表示画线
plot(da2,type = 'l')
```






```
# 将两个数据放在一张图上
par(mfcol = c(1,1))
ts.plot(da1,da2,gpars=list(col=c("blue","red")))
```



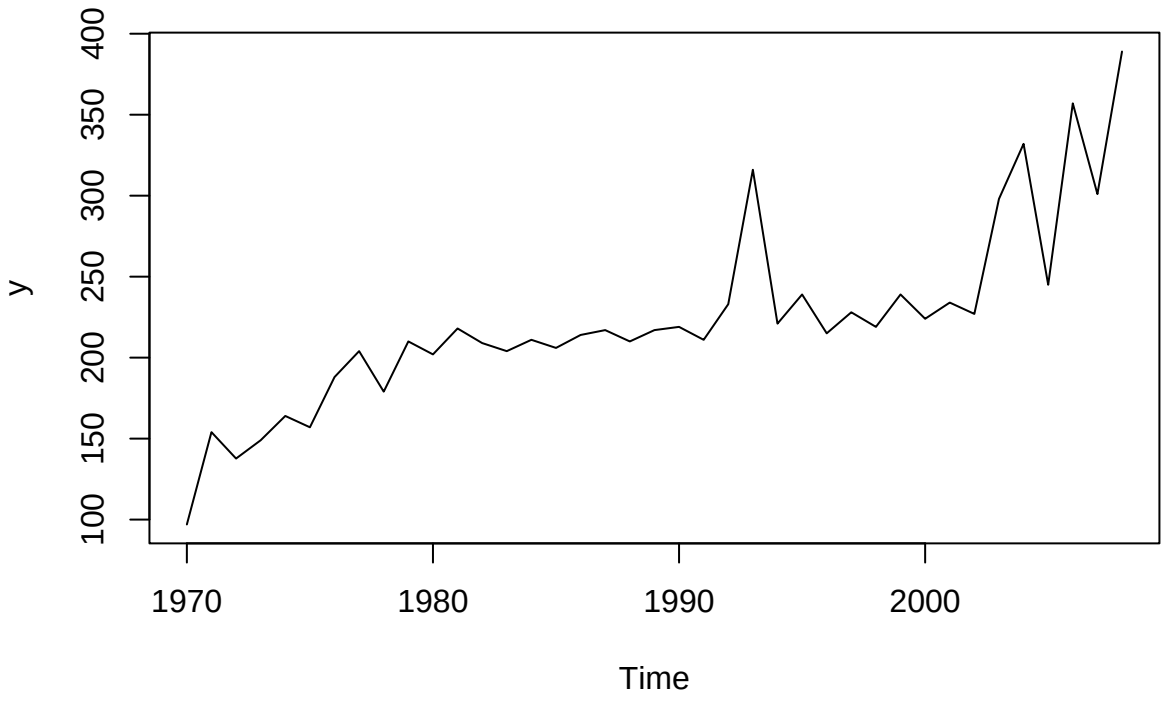
案例 2, P37, 平稳性检验

```r
x <- c(97,154,137.7,149,164,157,188,204,179,210,202,218,209,
       204,211,206,214,217,210,217,219,211,233,316,221,239,
       215,228,219,239,224,234,227,298,332,245,357,301,389)
y <- ts(x, frequency = 1, start = 1970)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

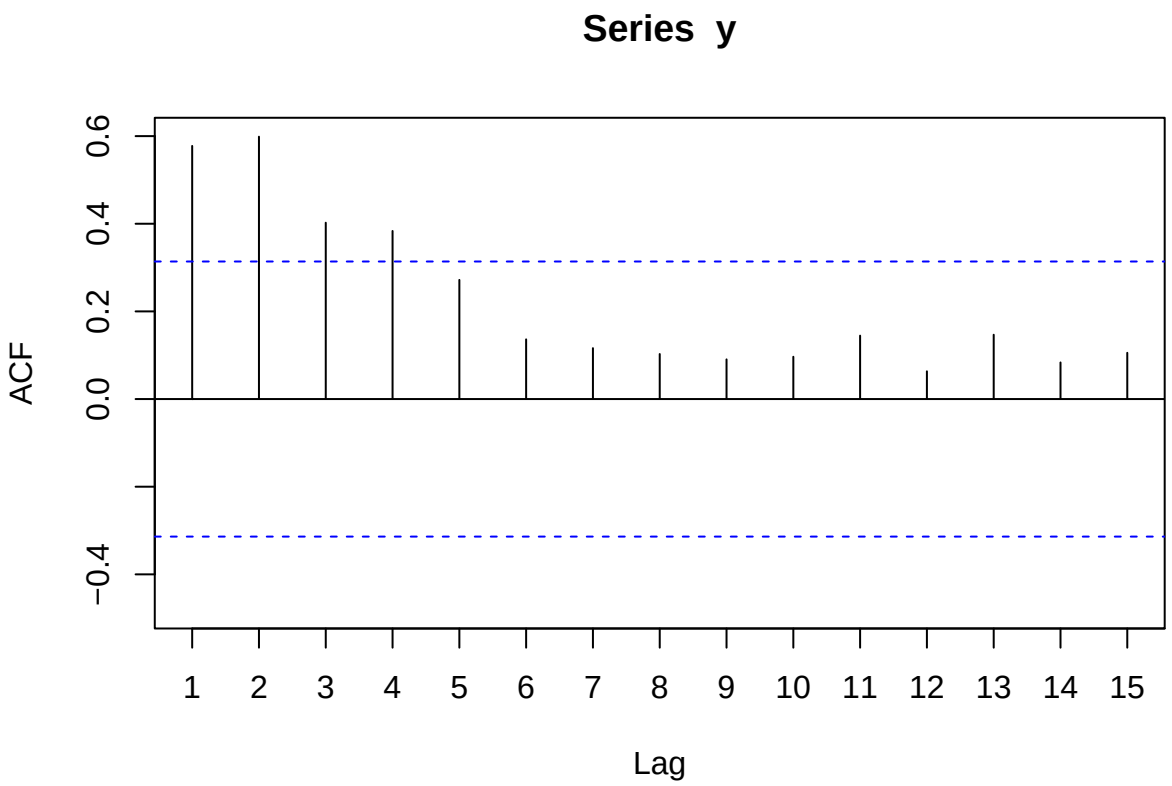
```r
library(tseries)

# 步骤 1, 平稳性检验
# 时序图检验
par(mfcol = c(1,1))
plot(y)
```



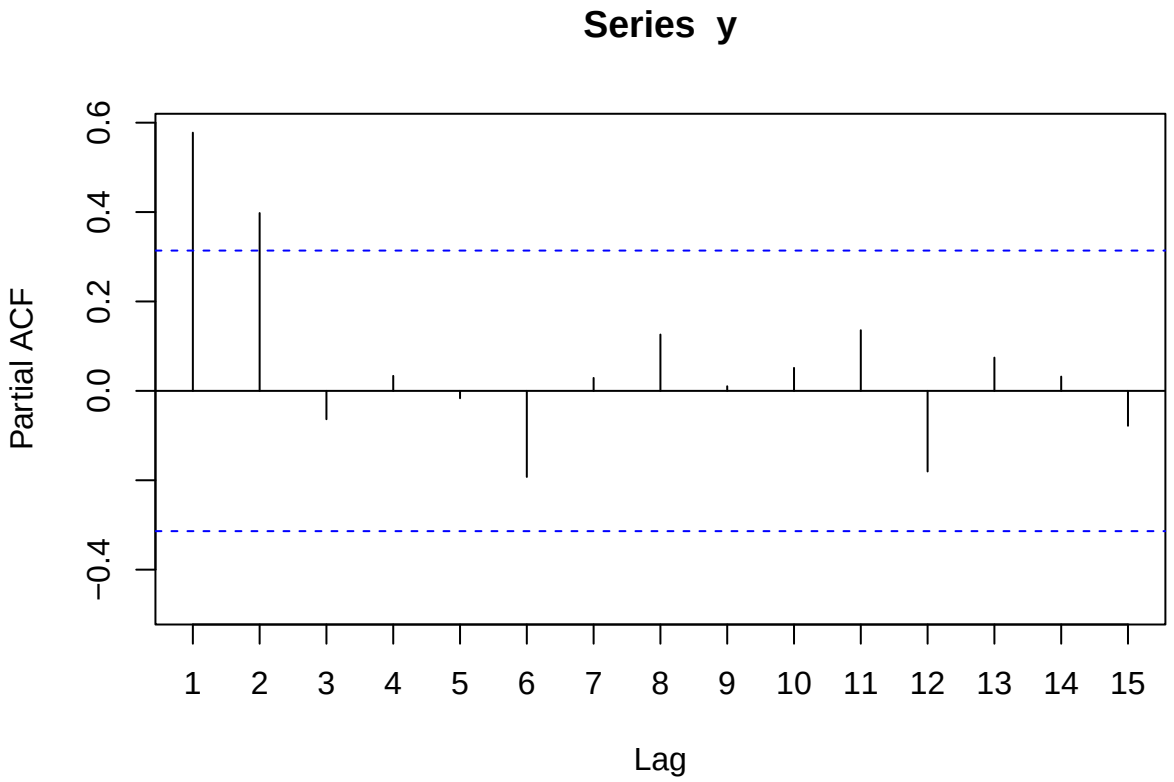

```r
# 不平稳

#par(mfcol = c(2,1))
# 自相关图
Acf(y)
```




```
#MA(q) 的截尾性， i=0 当 i>q 时。
#acf(y) ACF 时 acf 的改进，主要时 Acf 没有在滞后 0 阶的时候绘制峰值

# 偏自相关图
Pacf(y)
```




```
# 平稳的 AR(P) 模型一般有 p 阶截尾性，即 Φii=0 当 i>p 时。
```

```
# 随机性检验
LBQ <- Box.test(y,type = "Ljung-Box")
# 因为这里样本量很小，用 box-ljung 检验比较准确 P31
LBQ
```

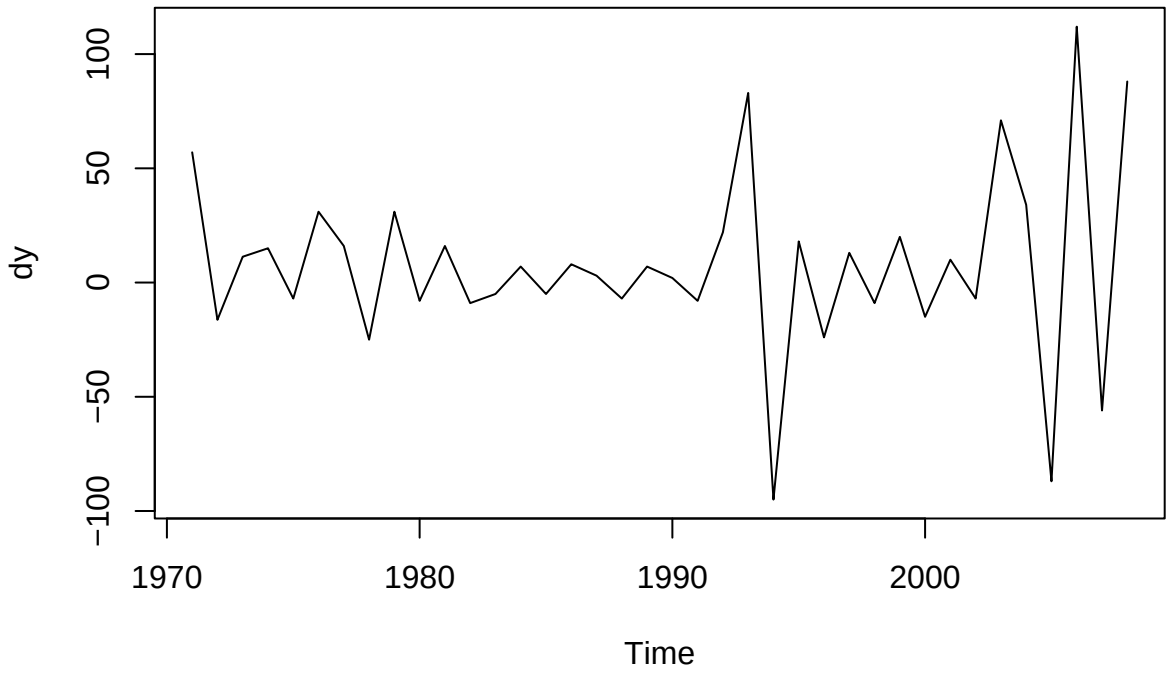
```
##
##  Box-Ljung test
##
## data:  y
## X-squared = 14.037, df = 1, p-value = 0.0001792
```

```
# 拒绝原假设，不随机，值得研究

# 差分次数
ndiffs(y)
```

```
## [1] 1
```

```
# 显示要差分一次
dy <- diff(y)
plot(dy)
```




```
#adf.test(dy)
```

```
# 平稳序列建模步骤 P66
#1. 判断序列是否为平稳非白噪声序列（时序图，随机性检验）
#2. 自相关图和偏自相关图， ARIMA(p,q) 定阶
#3. 进行模型拟合，估计参数值
#4. 有效性检验
#5. 模型改进优化
#6. 模型预测
```

P68 1950-2008 年我国邮政及农村投递线路每年新增里程数序列

```
x <- c(15.71,24.43,18.23,22.50,12.53,9.94,7.19,41.13,79.03,119.32,-12.10,-89.71,-52.26,20.01,19.92
da <- ts(x,frequency =1, start = c(1950))
```

1. 随机性检验

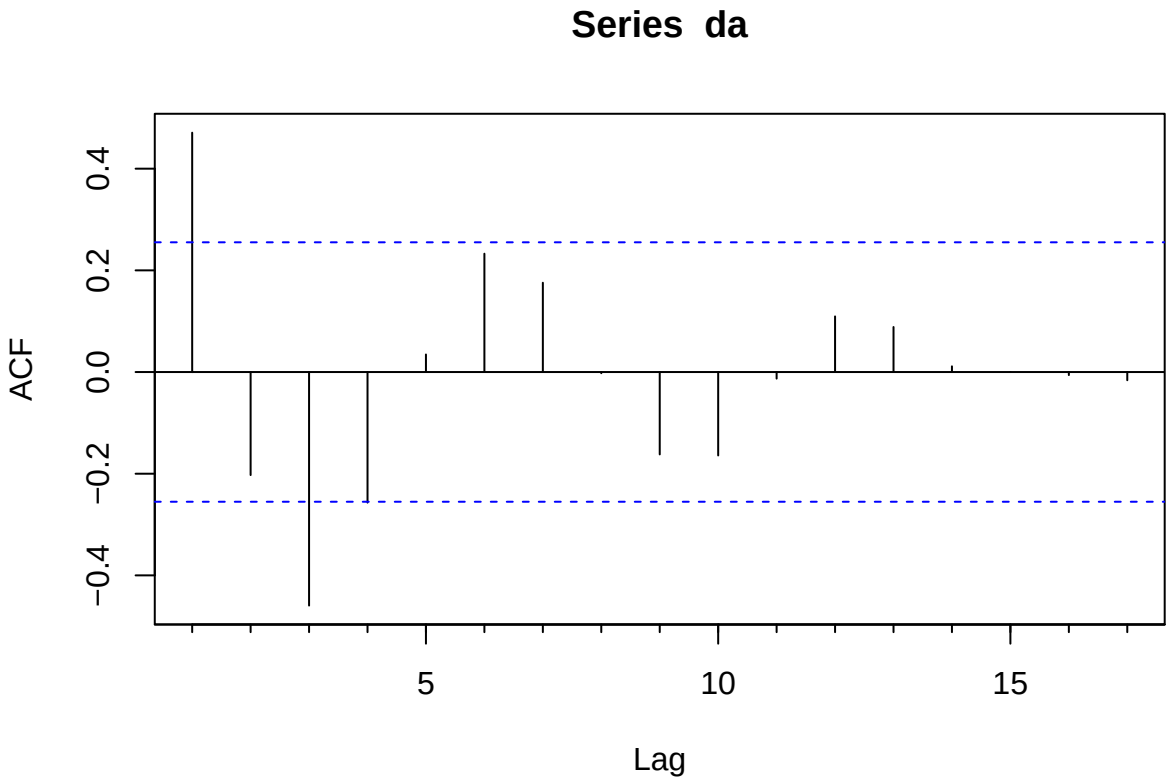
```
plot(da,type = 'l')  #'l'表示画线
```


```
# 由图显示，序列没有显著的非平稳特征。
# 随机性检验
LBQ <- Box.test(da,type = "Ljung-Box")
# 因为这里样本量很小，用 box-ljung 检验比较准确 P31
LBQ  # 拒绝原假设，不随机，值得研究
```

```
##
##  Box-Ljung test
##
## data:  da
## X-squared = 13.753, df = 1, p-value = 0.0002085
```

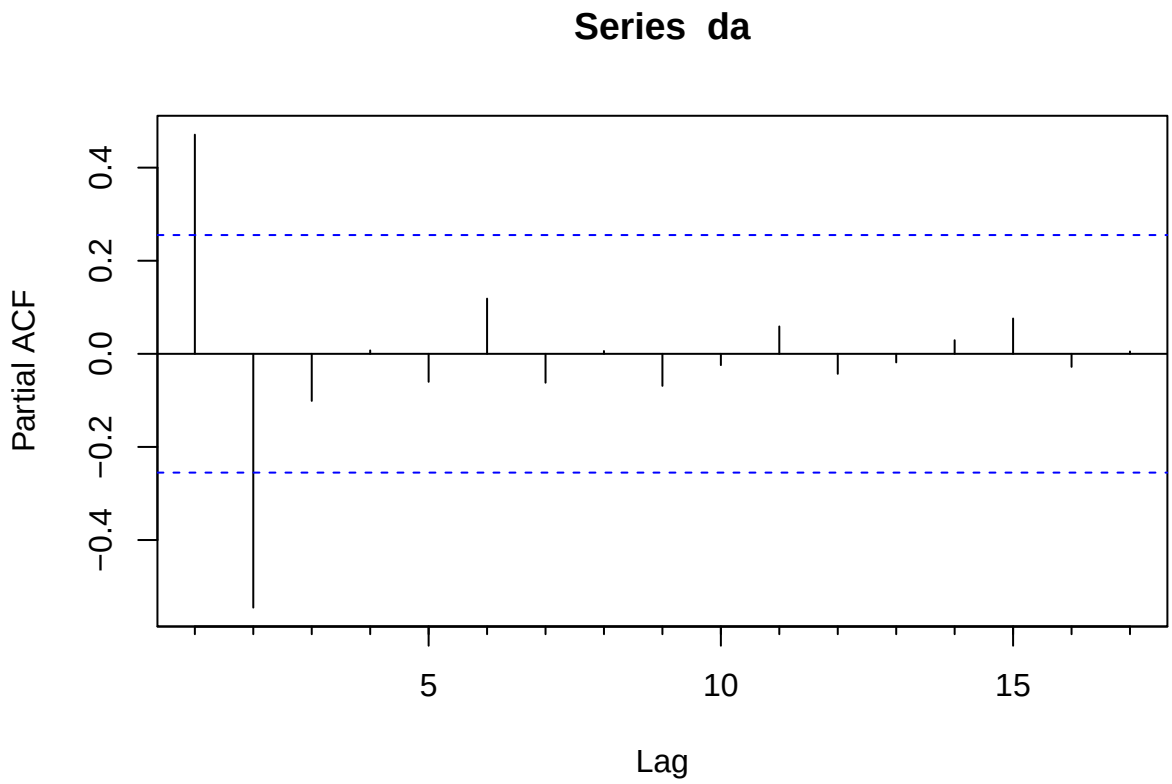
2. 自相关图

```
Acf(da)
```




```
#MA(q) 的截尾性，当 i>q 时。
# 从图中可以看出，除了 1-3 阶的自相关系数在 2 倍标准差之外，其他的
# 自相关系数都在 2 倍标准差范围内波动，且衰减到零的过程存在正弦波动
# 自相关系数的拖尾典型特征。

# 偏自相关图
Pacf(da)
```




```
# 平稳的 AR(P) 模型一般有 p 阶截尾性，即 Φii=0 当 i>p 时。
# 偏自相关图，除了 1-2 阶偏自相关系数在 2 倍以外，其他都在 2 倍以内，
#2 阶截尾典型，所以可以确定为 AR(2) 模型
```

习题 1

(1) 自行模拟 P69 例 3-10 并得出书本中的结论

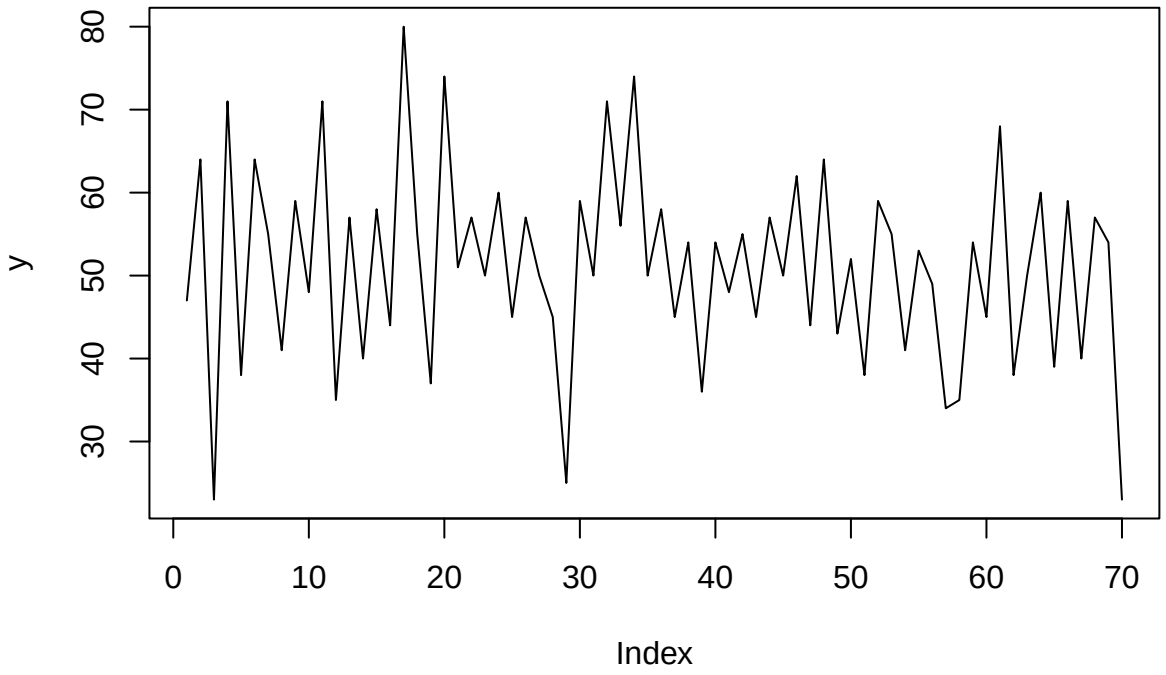
```
overshot<-c(78, -58, 53, -63, 13, -6, -16, -14, 3, -74, 89, -48, -14, 32, 56, -86, -66, 50, 26, 59
```

(2) 自行模拟 P69 例 3-11 并得出书本中的结论

```
temperature<-c(-0.40, -0.37, -0.43, -0.47, -0.72, -0.54, -0.47, -0.54, -0.39, -0.19, -0.40, -0.44,
```

P79 模型优化 (70 次化学反应过程)

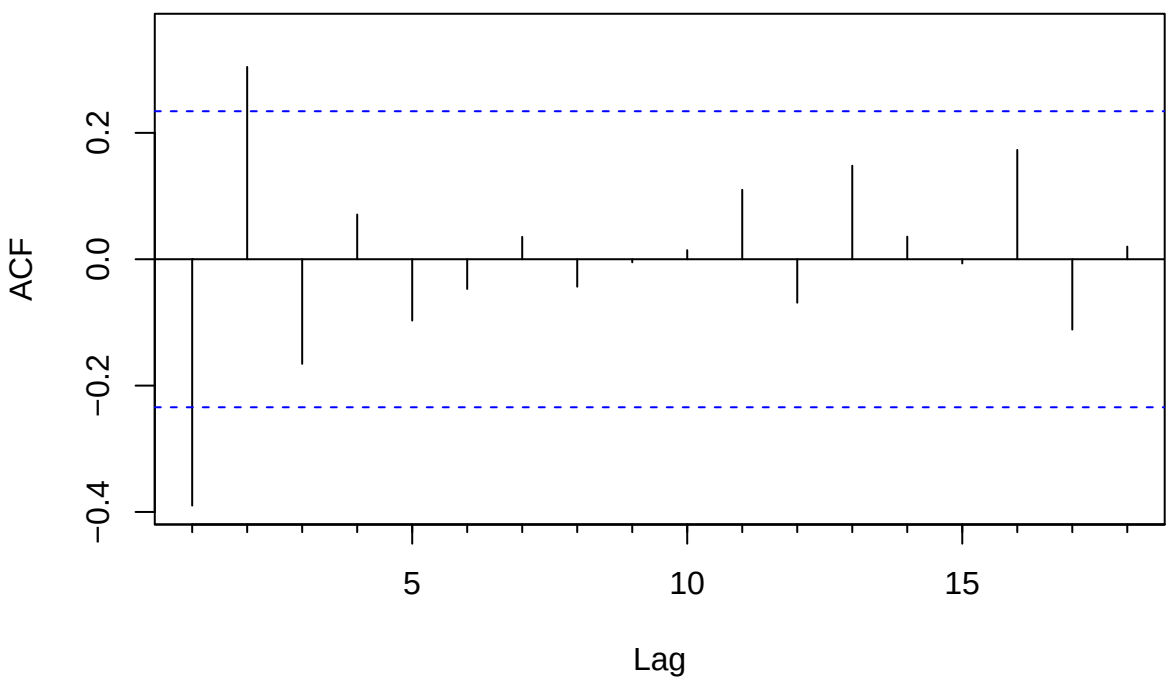
```
y <- c(47,64,23,71,38,64,55,41,59,48,71,35,57,40,
       58,44,80,55,37,74,51,57,50,60,45,57,50,45,
       25,59,50,71,56,74,50,58,45,54,36,54,48,55,
       45,57,50,62,44,64,43,52,38,59,55,41,53,49,
       34,35,54,45,68,38,50,60,39,59,40,57,54,23)
plot(y,type = 'l')
```




```
# 平稳
LBQ <- Box.test(y,type = "Ljung-Box")
LBQ
```

```
##
##  Box-Ljung test
##
## data:  y
## X-squared = 11.103, df = 1, p-value = 0.0008619
```

```
# 非白噪声
Acf(y)
```




```
#2 阶截尾 MA(2)
```

```
fit <- arima(y, order=c(0,0,2))
fit
```

```
##
## Call:
## arima(x = y, order = c(0, 0, 2))
##
## Coefficients:
##           ma1     ma2  intercept
##       -0.3194  0.3019    51.1695
## s.e.   0.1160  0.1233     1.2516
##
## sigma^2 estimated as 114.4:  log likelihood = -265.35,  aic = 538.71
```

```
#aic = 538.71


Box.test(fit$residuals,lag = 6)
```

```
##
##  Box-Pierce test
##
```

```
## data:  fit$residuals
## X-squared = 2.1105, df = 6, p-value = 0.9093
```

```
Box.test(fit$residuals,lag = 12)
```

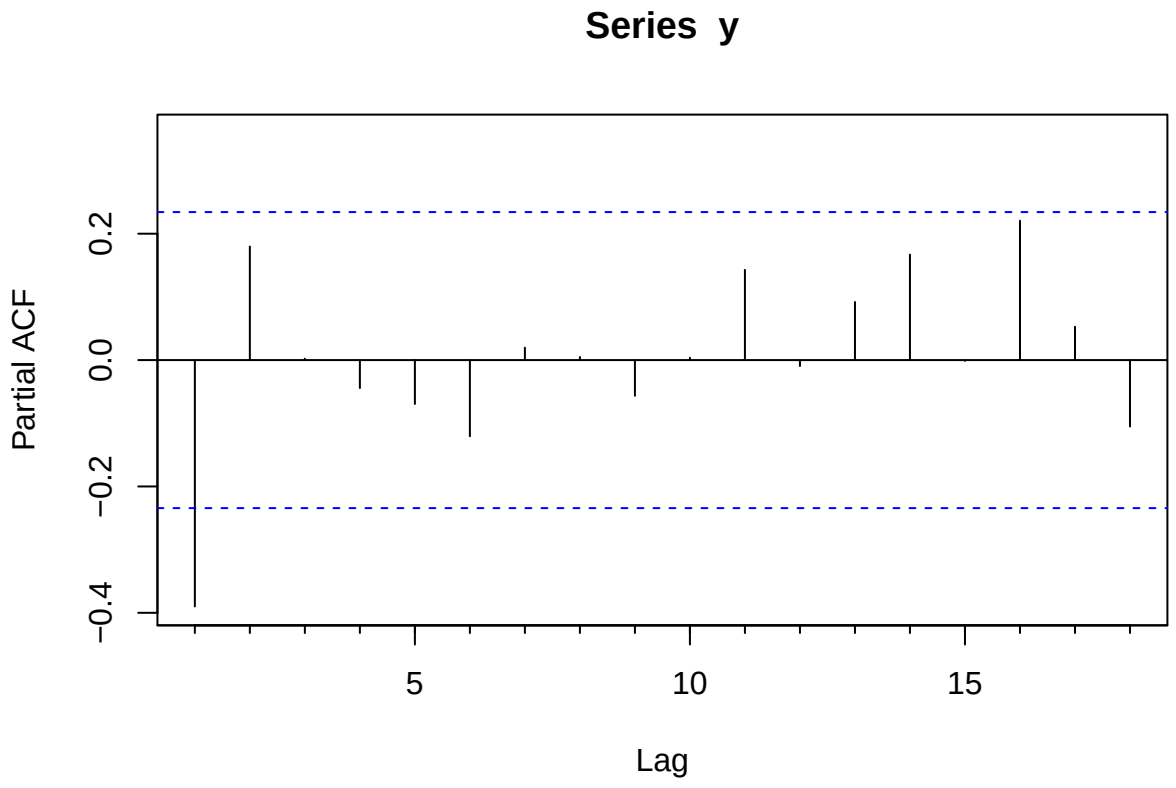
```
##
##  Box-Pierce test
##
## data:  fit$residuals
## X-squared = 3.9217, df = 12, p-value = 0.9848
```

```
Box.test(fit$residuals,lag = 18)
```

```
##
##  Box-Pierce test
##
## data:  fit$residuals
## X-squared = 8.6671, df = 18, p-value = 0.967
```

```
#P 值显著大于 0.05，残差多阶都不相关，模型有效


Pacf(y)
```



```r
#1 阶截尾倾向 AR(1)
fit2 <- arima(y, order=c(1,0,0))
fit2
```

```
##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##           ar1  intercept
##       -0.4191    51.2658
## s.e.   0.1129     0.9137
##
## sigma^2 estimated as 116.6:  log likelihood = -265.98,  aic = 537.96
```

```r
#aic = 537.96 < MA(2)

Box.test(fit2$residuals,lag = 6)
```

```
##
##  Box-Pierce test
##
## data:  fit2$residuals
## X-squared = 4.1678, df = 6, p-value = 0.654
```

```r
Box.test(fit2$residuals,lag = 12)
```

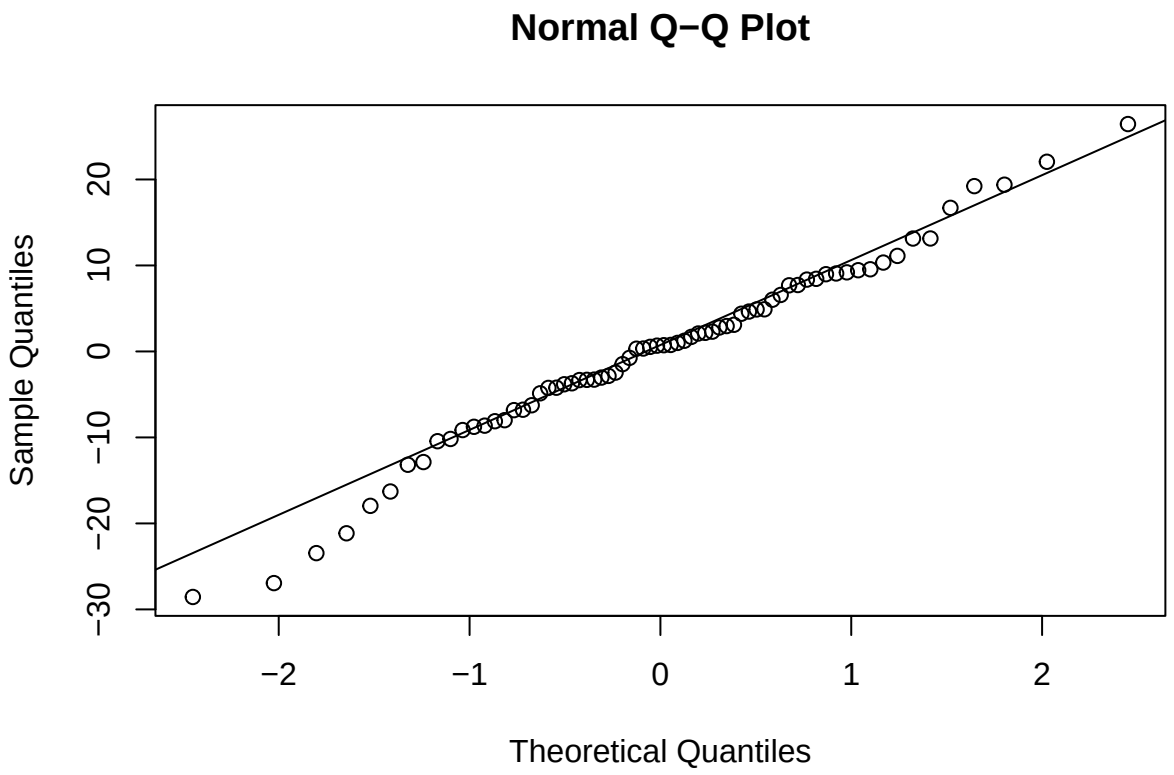
```
##
##  Box-Pierce test
##
## data:  fit2$residuals
## X-squared = 6.1411, df = 12, p-value = 0.9088
```

```r
Box.test(fit2$residuals,lag = 18)
```

```
##
##  Box-Pierce test
##
## data:  fit2$residuals
## X-squared = 11.843, df = 18, p-value = 0.8552
```

```r
###AR(1) 模型较优
```

```r
# 模型评价-qq 图检验
qqnorm(fit$residuals)
qqline(fit$residuals)
```




```r
Box.test(fit$residuals,type = "Ljung-Box")
```

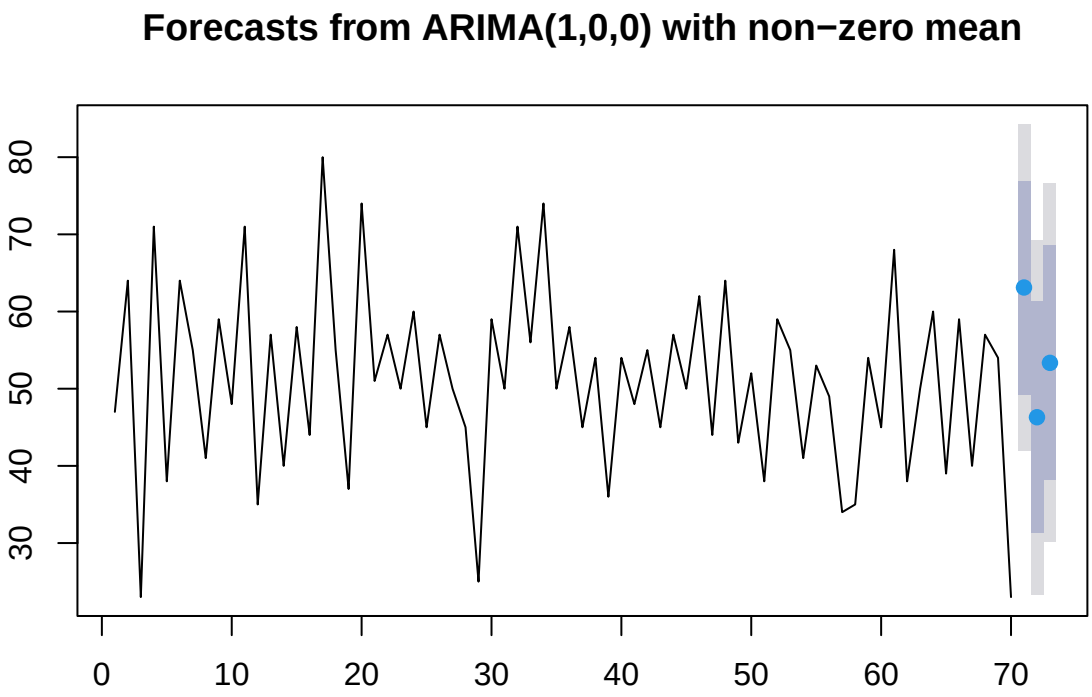
```
##
##  Box-Ljung test
##
## data:  fit$residuals
## X-squared = 0.098181, df = 1, p-value = 0.754
```

```r
### 结果都是残差不想关，有效

# 预测
forecast(fit2,3)
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 71       63.11143 49.27296 76.94990 41.94732 84.27555
## 72       46.30157 31.29702 61.30612 23.35409 69.24905
## 73       53.34623 38.14612 68.54634 30.09967 76.59280
```

```
plot(forecast(fit2,3))
```




```
# 浅灰色和深灰色分别是 80 和 95 的置信区间
```

习题 2

```
# 从习题 1 中任选一个数据完成下面的任务
# (1) 绘制样本自相关图，偏相关图
# (2) 检验平稳性，纯随机性
# (3) 模式识别：定阶，参数估计
# (4) 残差的自相关检验
# (5) 预测
```