华中师范大学

实验报告书

2022 年 10 月 27 日

课程名称:	时间序列分析
专　　业:	统计学
年　　级:	2020 级
学生姓名:	陈启源
学　　号:	2020211946
指导教师:	张晓飞

华中师范大学数学与统计学学院

1 编程作业

1.1 问题重述

用程序语言实现常均值模型、线性模型、二次模型、季节趋势模型、余弦趋势模型的参数估计，并使用教才中相应数据检验算法的正确性。

1.2 问题求解

```r
library(MASS)
library(TSA)
library(matlib)

ave_model = function(x, y) {
  mu = sum(y) / length(y)
  y_hat = mu
  return(list(mu, y_hat))
}

linear_model = function(x, y) {
  ones = rep(1, length(x)) # 构造截距项
  x = cbind(ones, x) # 将截距项和原数据合并
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

quadratic_model = function(x, y) {
  ones = rep(1, length(x))
  x_quad = x ^ 2
  x = cbind(ones, x, x_quad)
  #print(dim(x))
  mu = qr.solve(x, y) # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

season_model_with_intercept = function(x, y) {
  x = model.matrix(~ x, x)
```

```r
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

season_model_without_intercept = function(x, y) {
  x = model.matrix(~ x - 1, x) # 转化为one-hot编码
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

cosine_model = function(x, y, f = 12) {
  x = as.numeric(x) - 1 #转换为数字，此处需要-1使得从0开始
  f = 1 / f
  cos_x = cos(2 * pi * f * x)
  sin_x = sin(2 * pi * f * x)
  ones = rep(1, length(x))
  x = cbind(ones, cos_x, sin_x)
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

library(TSA)
library(matlib)
data(rwalk)
data(tempdub)

x = time(rwalk)
y = rwalk
lm1 = lm(y ~ x)
linear_model(x, y) # 验证线性模型正确性

lm2 = lm(rwalk ~ time(rwalk) + I(time(rwalk) ^ 2))
quadratic_model(x, y) # 验证二次模型的正确性

x = season(tempdub)
```

```
y = tempdub

season_model_with_intercept(x, y) # 验证有截距情况
season_model_without_intercept(x, y) # 验证没有截距情况
# x=harmonic(tempdub,1) # 书本做法
cosine_model(x, y) # 验证余弦的情况
```

2 问题 3.5

2.1 问题重述

数据文件 wages 包含了 1981 年 7 月到 1987 年 6 月美国服装和纺织品行业工人的平均时薪 (以美元计) 的月度值.
(a) 画出并解释这些数据的时间序列图.
(b) 对该时间序列用最小二乘法拟合线性时间趋势. 解释回归结果. 保存拟合模型的标准残差以便进一步分析.
(c) 构造并解释来自 (b) 的标准残差的时间序列图.
(d) 对工资时间序列用最小二乘法估计二次时间趋势. 解释回归结果. 保存标准残差以便进一步分析.
(e) 绘出并解释 (d) 中标准残差的时间序列图.

2.2 问题分析与求解

2.2.1 (a)

```
library(TSA)

data("wages")

win.graph(width = 8,
          height = 8,
          pointsize = 10)
plot(
  x = zlag(wages),
  y = wages,
  xlab = "Previous Year Wages",
  ylab = "Wages"
)
win.graph(width = 12,
          height = 6,
          pointsize = 10)
plot(wages, type = 'o', pch = 21)
```

解释：通过观察图 1可以得出，前一年与后一年存在较强的相关性。通过观察 2可以得出，时薪存在明显的上升趋势。

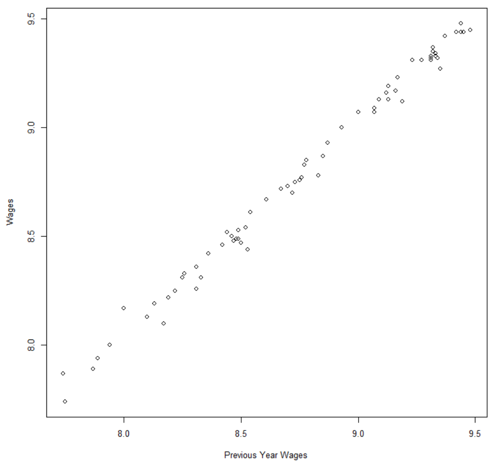

图 1: 前一年时薪与后一年时薪的散点图

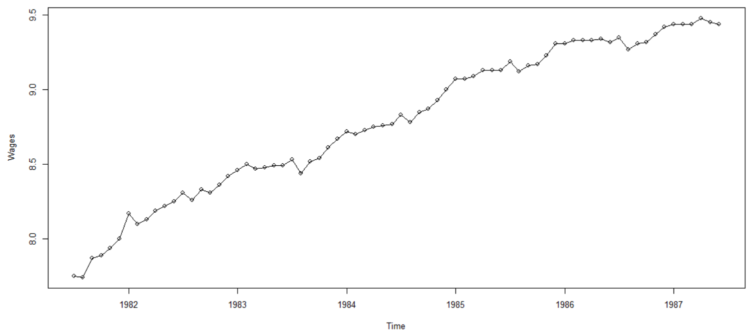

图 2: 时薪时间序列图

2.2.2 (b)

```
# b
x = time(wages)
y = wages

linear_model = function(x, y) {
  ones = rep(1, length(x)) # 构造截距项
  x = cbind(ones, x) # 将截距项和原数据合并
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
```

```
}

lm1 = lm(y ~ x) # 对比lm的结果

result = linear_model(x, y)
beta1 = result[[1]] # 自己计算的beta
y_hat = result[[2]]
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat
mean(e)
var(e)

e = scale(e)
```

其中回归的系数分别为-549.006 和 0.281。残差的均值为-1.701508e-09，残差的方差为 0.006721254。

解释： 通过查看残差的均值和方差，可以发现残差的均值和方差数值均较小，可以推测模型拟合效果较好。

2.2.3 (c)

```
# c
# 残差的标准化在b题中已经完成
win.graph(12, 6, pointsize = 10)
plot(ts(e, start = start(wages), frequency = 12),
     type = 'o',
     pch = 21)
```

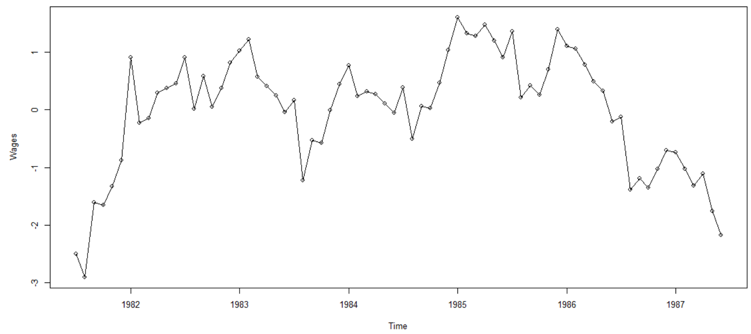

图 3: 残差时间序列图

解释： 残差随时间变化幅度较大，有一定季节性趋势，总体先上升后下降。

2.2.4 (d)

```
quadratic_model = function(x, y) {
  ones = rep(1, length(x))
  x_quad = x ^ 2
  x = cbind(ones, x, x_quad)

  mu = qr.solve(x,y) # 使用qr分解来求解矩阵的逆

  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}

x = time(wages)
y = wages

lm2 = lm(y ~ x + I(x ^ 2))

result = quadratic_model(x, y)
beta = result[[1]] # 对比lm的结果
y_hat = result[[2]]
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat

mean(e)
var(e)

e = scale(e)
```

其中回归的系数分别是-8.49e+4，85.3，-2.14e-2。残差的均值为-4.500586e-11，方差为 0.00337073。

解释： 拟合后的残差均值和方差相比线性模型残差的均值和方差都有所降低，说明二次模型对于数据的拟合效果更佳。

2.2.5 (e)

```
# e
win.graph(12, 6, pointsize = 10)
```

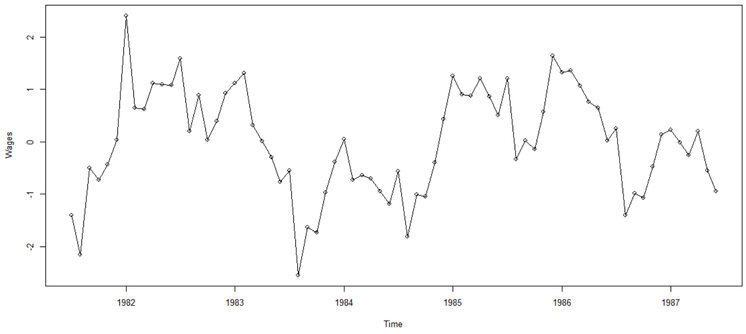

图 4: 二次模型残差时间序列图

```
plot(ts(e, start = start(wages), frequency = 12),
        type = 'o',
        pch = 21)
```

解释： 对比 3和 4，可以发现，使用二次模型拟合后残差的随机性提高了，季节性趋势变弱。说明二次模型更加适合。但是依旧存在较强的相关性。

3 问题 3.7

3.1 问题重述

数据文件 winebago 包含了 1966 年 11 月到 1972 年 2 月的 Winnebago 有限公司娱乐设备的单位销量.
(a) 画出并解释这些数据的时间序列图.
(b) 对这些数据用最小二乘法拟合直线. 解释回归结果. 画出来自拟合的标准残差的时间序列图，给出解释.
(c) 现在对月度售价数据取自然对数. 画出并解释变换后数据的时间序列图.
(d) 对对数数据用最小二乘法拟合直线. 画出并解释该拟合的标准残差的时间序列图.
(e) 现在对对数售价时间序列用最小二乘法拟合季节均值十线性时间趋势. 保存标准残差以便进一步分析. 检验该模型各回归系数的统计显著性.
(f) 画出 (e) 中标准残差的时间序列图. 给出解释.

3.2 问题分析与求解

3.2.1 (a)

```
data(”winnebago”)
win.graph(width = 8,
          height = 8,
          pointsize = 10)
plot(
  x = zlag(winnebago),
  y = winnebago,
  xlab = ”Previous␣Year␣Wages”,
  ylab = ”Wages”
) # 绘制前一年和后一年的散点图
win.graph(width = 12,
          height = 6,
          pointsize = 10)
plot(winnebago, type = ’o’, pch = 21) # 时间序列图
```

解释： 通过观察图 5和 6，可以发现，前一年和后一年存在较强的线性关系，这个也可以通过6绘制均值线得出类似结论。同时观察每一年中不同月份的走势，可以发现数据有一定的季节性变化。从总体上而言，整体上呈现上升趋势，而且变化幅度加大。

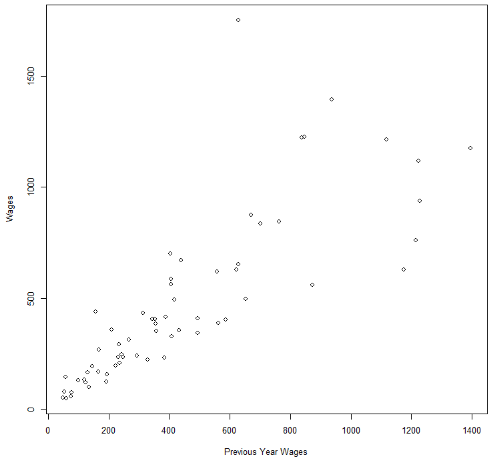

图 5: 前一年与后一年的收入散点图

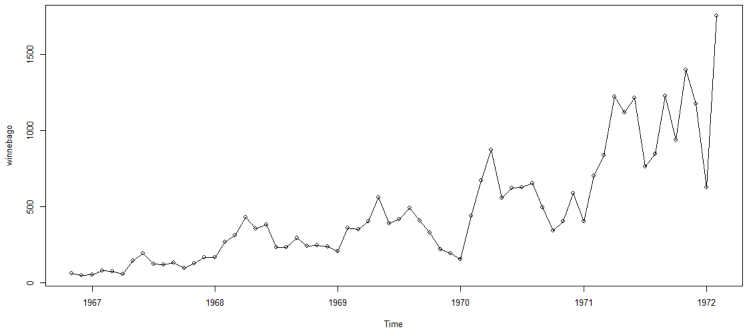

图 6: 时间序列图

3.2.2 (b)

```
linear_model = function(x, y) {
  ones = rep(1, length(x)) # 构造截距项
  x = cbind(ones, x) # 将截距项和原数据合并
  mu = solve(t(x) %*% x) %*% t(x) %*% y # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}
x = time(winnebago)
y = winnebago
lm1 = lm(y ~ x)
```

```
result = linear_model(x, y)
beta1 = result[[1]] # 对比结果
y_hat = result[[2]]
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat

e = scale(e)

win.graph(12, 6, pointsize = 10)
plot(ts(e, start = start(winnebago), frequency = 12),
     type = 'o',
     pch = 21)
```

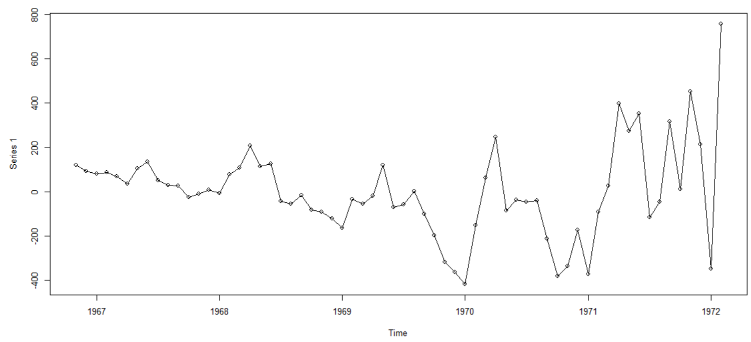

图 7: 残差时间序列图

解释：通过观察图 7可以发现，残差依旧存在季节性趋势，而且残差随着时间的推移变化幅度增大（残差方差增大）。

3.2.3 (c)

```
win.graph(width = 12,height = 6,pointsize = 10)
plot(log(winnebago),type = 'o',pch = 21)
```

解释：类似回归分析当中的 BOX-COX 变换，可以消除异方差性。所以观察 8可以发现，类似不取对数的情况，季节性变化与上升趋势依旧存在，但是整体的变化幅度（方差）差距变小。

3.2.4 (d)

```
x = time(winnebago)
```

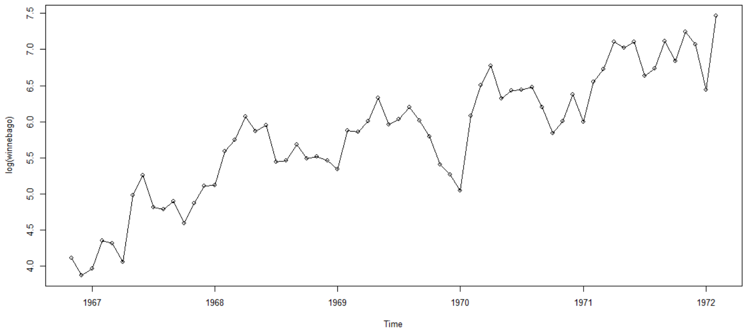

图 8: 取对数后时间序列图

```
y = log(winnebago) # 添加对数

lm1 = lm(y ~ x)

result = linear_model(x, y)
y_hat = result[[2]]
beta1 = result[[1]]
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat

e = scale(e)

win.graph(12, 6, pointsize = 10)
plot(ts(e, start = start(winnebago), frequency = 12),
      type = 'o',
      pch = 21)
```

解释： 同样的，残差的季节性并没有改变，但是残差的季节性变化幅度变小。

3.2.5 (e)

```
# e
x = season(winnebago)
x = model.matrix( ~ x, x) # 将类别转化为设计矩阵
x = cbind(x, time(log(winnebago))) # 添加时间
y = log(winnebago) # 添加对数

```

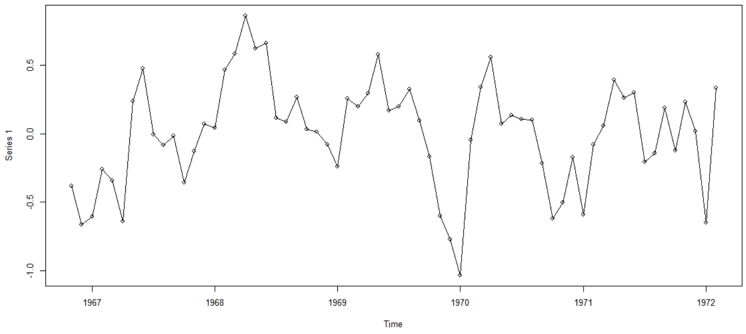

图 9: 取对数后残差时间序列图

```
lm1 = lm(y ~ x)

mu = solve(t(x) %*% x) %*% t(x) %*% y # 线性模型拟合
y_hat = x %*% mu # 预测
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat

# 假设检验
sigma_hat = as.vector(sqrt(t(e) %*% e / (length(e) - length(mu))))
c_i = as.vector(sqrt(diag(solve(t(
  x
) %*% x))))
t_value = mu / (sigma_hat * c_i)
p_value = 2 * pt(-abs(t_value), df = length(e) - length(mu))

e = scale(e)
```

其中回归的参数如下：x.(Intercept)=-997.3306122，x.xFebruary=0.6244477，x.xMarch=0.6821971，x.xApril=0.8095882，x.xMay=0.8695252，x.xJune=0.8630875，x.xJuly=0.5539177，x.xAugust=0.5698855，x.xSeptember=0.5757167，x.xOctober=0.2634863，x.xNovember=0.2868224，x.xDecember=0.2480223，time(log(winnebago))=0.5090896

对于回归参数的检验，采用 t 检验，其中原假设为对应系数为 0。

检验得到的 p 值如下：x.(Intercept)=1.718416e-25,x.xFebruary=1.188163e-03,x.xMarch=7.792694e-04,x.xApril=9.300612e-05,x.xMay=3.246420e-05,x.xJune=3.627462e-05,x.xJuly=5.420294e-03,x.xAugust=4.305322e-03,x.xSeptember=3.960057e-03,x.xOctober=1.733001e-01,x.xNovember=1.209455e-01,x.xDecember=1.785325e-01,time(log(winnebago))=1.350759e-25

其中 p 值小于 0.05 的则拒绝原假设，认为原来的系数具有统计显著性。

3.2.6 (f)

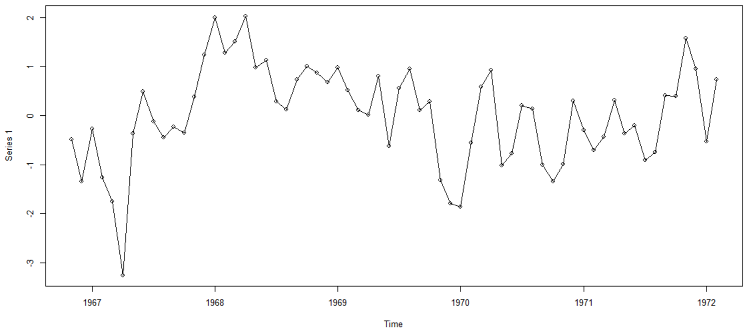

图 10: 残差时间序列图

解释：该残差图展现出一定的随机性，没有明显的趋势与自相关性。是比较好的情况。

4 问题 3.11

4.1 问题重述

3.11 (继续习题 3.5) 回到 wages 序列.
(a) 考虑用最小二乘法拟合二次时间趋势得到的残差.
(b) 对标准残差进行游程检验，并解释结果.
(c) 计算并解释标准残差的样本自相关.
(d) 研究标准残差 (误差项) 的正态性. 考虑直方图和正态图. 解释图形.

4.2 问题分析与求解

4.2.1 (a)

```
library(TSA)
library(matlib)
data(wages)
quadratic_model = function(x, y) {
  ones = rep(1, length(x))
  x_quad = x ^ 2
  x = cbind(ones, x, x_quad)
  #print(dim(x))
  mu = qr.solve(x, y) # 求参数
  y_hat = x %*% mu # 预测
  return(list(mu, y_hat))
}
x = time(wages)
y = wages
lm2=lm(y~x+I(x^2))
result = quadratic_model(x, y)
beta2 = result[[1]]
y_hat = result[[2]]
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat

mean(e)
var(e)
```

其中回归的系数为：(残差)-8.494973e+04，(一次项)8.534287e+01，(二次项)-2.143199e-02。

4.2.2 (b)

```
e=scale(e) # 标准化
runs(e)

# 结果如下

$pvalue
[1] 1.56e-07

$observed.runs
[1] 15

$expected.runs
[1] 36.75

$n1
[1] 33

$n2
[1] 39

$k
[1] 0
```

结果展示在代码框中。其中 p 值小于 0.05，我们显著的拒绝数据是随机性的原假设，而认为数据是非随机的。同时也可以对比观察游程和期望游程之间的差距，观察的游程显著小于期望，所以也可以认为数据是非随机的。

4.2.3 (c)

```
acf_result=acf(e)
acf_result

Autocorrelations of series 'e', by lag

     1      2      3      4      5      6      7      8      9
 0.700  0.563  0.376  0.211  0.062  0.014 -0.028 -0.049 -0.044

```

```
      10      11      12      13      14      15      16      17      18
   0.042   0.009   0.106  -0.076  -0.197  -0.348  -0.473  -0.532  -0.529
```

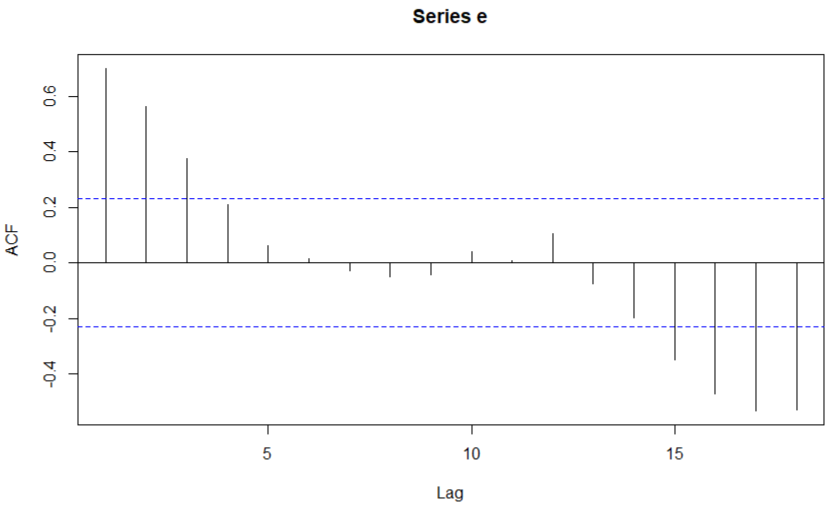

图 11: 自相关检验图

解释：通过查看残差的自相关检验结果和自相关检验图可以发现，残差存在较强的自相关性。其中超过蓝线的部分为显著的延迟。此处 e 的延迟为 1，2，3，15，16，17，18。

4.2.4 (d)

```
# d
# 此处的残差已经经过标准化
#正态性检验，原假设为符合正态分布
shapiro.test(e)
win.graph(width = 8,
          height = 8,
          pointsize = 10)
qqnorm(e)
qqline(e)
win.graph(width = 8,
          height = 8,
          pointsize = 10)
hist(e)
# 正态性检验输出
        Shapiro-Wilk normality test
```

```

data:  e
W = 0.98856, p-value = 0.7622
```

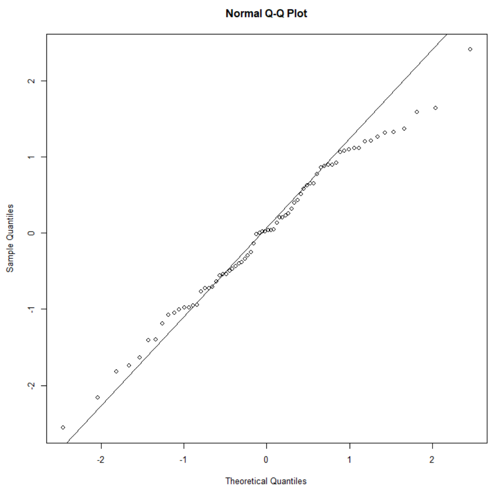

图 12: QQ 图

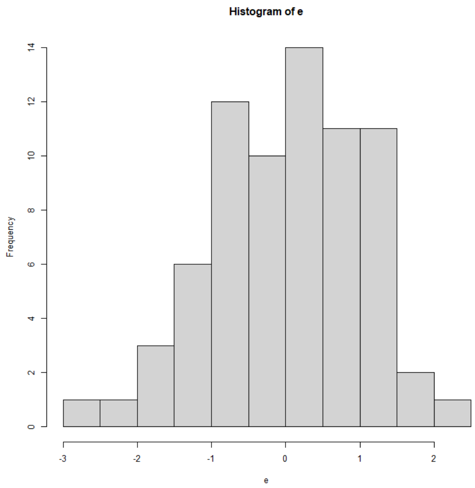

图 13: 直方图

解释：通过观察 12和 13，可以发现残差并不完全服从正态分布，尤其是直方图可以看到这个分布并不是对称的。但由正态性检验可以得到 p 值大于 0.05，其中原假设为服从正态分布，也就是不能拒绝原假设，换言之就是不能拒绝正态性假设。

5 问题 3.14

5.1 问题重述

(继续习题 3.8) 数据文件 retail 包含了 U.K. 月度零售数据.
(a) 获得用最小二乘拟合季节均值 + 线性时间趋势模型的残差.
(b) 对标准残差进行游程检验，并解释结果.
(c) 计算并解释标准残差的样本自相关.
(d) 研究标准残差 (误差项) 的正态性. 考虑直方图和正态图，解释图形.

5.2 问题分析与求解

5.2.1 (a)

```
library(TSA)
data("retail")
x = season(retail)
x = model.matrix(~ x, x) # 将类别转化为设计矩阵
x = cbind(x, time(retail)) # 添加时间
y = retail
lm1=lm(y~x)
mu = solve(t(x) %*% x) %*% t(x) %*% y # 线性模型拟合
y_hat = x %*% mu # 预测
y_num = as.numeric(y) # 转换为数值
e = y_num - y_hat
e = scale(e) # 对残差进行标准化
e=scale(e) # 对残差进行标准化
```

系数如下：x.(Intercept)=-7.249383e+03，x.xFebruary=-3.014926e+00，x.xMarch=7.469271e-02，x.xApril=3.447160e+00，x.xMay=3.107991e+00，x.xJune=3.073584e+00，x.xJuly=6.053463e+00，x.xAugust=3.138104e+00，x.xSeptember=3.427507e+00，x.xOctober=8.555004e+00，x.xNovember=2.081584e+01，x.xDecember=5.254333e+01，time(retail)=3.670026e+00。

5.2.2 (b)

```
# b
runs(e)

$pvalue
```

```
[1] 9.19e-23

$observed.runs
[1] 52

$expected.runs
[1] 127.9333

$n1
[1] 136

$n2
[1] 119

$k
[1] 0
```

结果展示在代码框中。其中 p 值小于 0.05，我们显著的拒绝数据是随机性的原假设，而认为数据是非随机的。同时也可以对比观察游程和期望游程之间的差距，观察的游程显著小于期望，所以也可以认为数据是非随机的。

5.2.3 (c)

```
# c
acf_result=acf(e)
acf_result

Autocorrelations of series 'e', by lag

     1      2      3      4      5      6      7      8
 0.400  0.112  0.042  0.146  0.228  0.212  0.211  0.100
     9     10     11     12     13     14     15     16
 0.000  0.043  0.299  0.718  0.252  0.009 -0.066  0.019
    17     18     19     20     21     22     23     24
 0.083  0.083  0.080 -0.006 -0.075 -0.045  0.175  0.485
```

解释： 通过查看残差的自相关检验结果和自相关检验图可以发现，残差存在较强的自相关性。其中超过蓝线的部分为显著的延迟。此处 e 的延迟为 1，4，5，6，7，11，12，13，23，24。

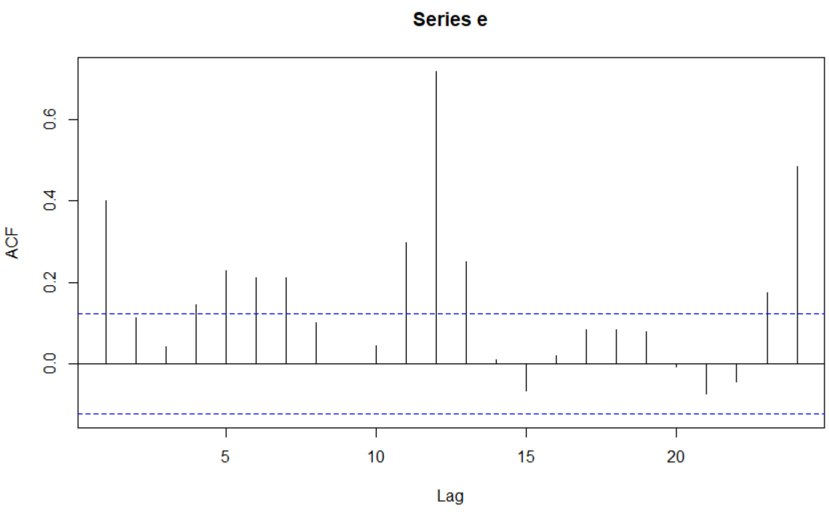

图 14: 自相关检测图

5.2.4 (d)

```
# 正态性检验，原假设为符合正态分布
shapiro.test(e)
win.graph(width = 8,
          height = 8,
          pointsize = 10)
qqnorm(e)
qqline(e)
win.graph(width = 8,
          height = 8,
          pointsize = 10)
hist(e)

        Shapiro-Wilk normality test

data:  e
W = 0.94529, p-value = 3.624e-08
```

解释：通过观察图 15和 16，观察得出残差并不服从正态分布。同时结合正态检验结果，发现 p 值小于 0.05，拒绝原假设，认为残差不服从正态分布。

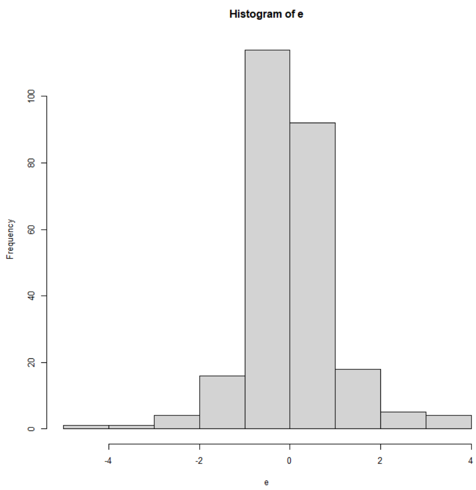

图 15: 残差直方图

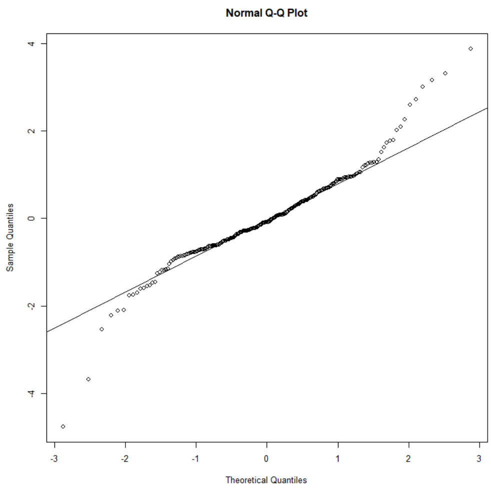

图 16: 残差 QQ 图