时间序列分析作业与第三次实验报告

姓名：康江睿

学号：2018213779

指导老师： 张晓飞

2020 年 10 月 23 日

1 使用R语言实现对部分确定趋势时间序列模型的简单回归分析

1.1 函数DTTSM(Deterministic Trend Time Series Model)简介

函数DTTSM实现了对六种主要的确定趋势模型（常均值趋势，线性趋势，二次趋势，季节性均值趋势，季节性线性趋势，余弦趋势）进行简单回归分析的功能。其中，为了保证数值稳定，避免直接求设计矩阵的逆而造成的舍入误差，这里使用了QR 分解的方法来实现最小二乘估计问题的求解。代码如下：

```
DTTSM <- function(Data,Model,Intercept){
  Time <- time(Data);n <- length(Data);
  if (length(grep(Model,"Constant"))==1){
    X <- cbind(Intercept = rep(1,n));
    Para <- qr.solve(X,Data);
  }else if (length(grep(Model,"Linear"))==1){
    X <- cbind(Intercept = rep(1,n),Time);
    Para <- qr.solve(X,Data);
  }else if (length(grep(Model,"Quadratic"))==1){
    Time <- time(Data);n <- length(Data);
    X <- cbind(Intercept = rep(1,n),Time,Time^2);
    Para <- qr.solve(X,Data);
  }else if (length(grep(Model,"Cosine"))==1){
    X <- cbind(Intercept = rep(1,n),Cos = cos(2*pi*Time),Sin = sin(2*pi*Time));
    Para <- qr.solve(X,Data);
```

```
16  }else if (length(grep(Model,"SeasonC"))==1){
17    t <- Time;Time <- integer(0);FirstMonth <- (t[1]-floor(t[1]))*12+1;
18    m1 <- FirstMonth-12;m2 <- floor((n+m1-1)/12);
19    if (m2 == -1){
20      Time <- FirstMonth:(FirstMonth+n-1);
21    }else if (FirstMonth == 1){
22      Time <- c(FirstMonth:12,rep(1:12,m2));
23    }else{
24      Time <- c(FirstMonth:12,rep(1:12,m2),1:(n+FirstMonth-13-12*m2));
25    }
26    Time <- round(Time);
27    dimnames = list(rep("0",n),c("Jan","Feb","Mar","Apr",
28                                 "May","June","July","Aug",
29                                 "Sept","Oct","Nov","Dec"))
30    X <- matrix(0,nrow = n,ncol = 12,dimnames = dimnames);
31    if (Intercept){
32      X[,1] <- rep(1,n);
33      for (i in 2:12){
34        idx <- which(Time == i);
35        u <- rep(0,n);
36        u[idx] <- 1;
37        X[,i] <- u;
38      }
39    }else{
40      for (i in 1:12){
41        idx <- which(Time == i);
42        u <- rep(0,n);
43        u[idx] <- 1;
44        X[,i] <- u;
45      }
46    }
47    Para <- qr.solve(X,Data);
48  }else if (length(grep(Model,"SeasonL"))==1){
49    t <- Time;Time <- integer(0);FirstMonth <- (t[1]-floor(t[1]))*12+1;
50    m1 <- FirstMonth-12;m2 <- floor((n+m1-1)/12);
51    if (m2 == -1){
52      Time <- FirstMonth:(FirstMonth+n-1);
53    }else if (FirstMonth == 1){
54      Time <- c(FirstMonth:12,rep(1:12,m2));
55    }else{
56      Time <- c(FirstMonth:12,rep(1:12,m2),1:(n+FirstMonth-13-12*m2));
57    }
58    Time <- round(Time);
59    dimnames = list(rep("0",n),c("Jan","Feb","Mar","Apr",
60                                 "May","June","July","Aug",
61                                 "Sept","Oct","Nov","Dec",
62                                 "Slope"))
63    X <- matrix(0,nrow = n,ncol = 13,dimnames = dimnames);
```

```
64    if (Intercept){
65      X[,1] <- rep(1,n);
66      for (i in 2:12){
67        idx <- which(Time == i);
68        u <- rep(0,n);
69        u[idx] <- 1;
70        X[,i] <- u;
71      }
72    }else{
73      for (i in 1:12){
74        idx <- which(Time == i);
75        u <- rep(0,n);
76        u[idx] <- 1;
77        X[,i] <- u;
78      }
79    }
80    X[,13] <- t;
81    Para <- qr.solve(X,Data);
82  }
83  qrr <-qr(X);
84  R <- qr.R(qrr);
85  invXtX <- solve(R)%*%solve(t(R));
86  Fitval <- X%*%Para;Res <- Data-Fitval;Fitval <- Data-Res;m <- ncol(X);
87  SE <- sqrt((t(Res)%*%Res)/(n-m));SEPara <- sqrt(as.vector((SE^2))*diag(invXtX));
88  if (length(grep(Model,"SeasonC"))==1 || length(grep(Model,"SeasonL"))==1 && Intercept==FALSE){
89     tStatistic <- as.vector(Para/SEPara);t <- qt(p = 0.975,df = n-m);
90    SSR <- sum(Fitval^2);SSE <-SE^2*(n-m);SST <- SSR+SSE;
91     FStatistic <- as.vector(SSR*(n-m)/SSE/m);F <-qf(0.95,m,n-12);
92    tTest <- cbind(tStatistic,t);FTest <- cbind(FStatistic,F);R2 <- as.vector(SSR/SST);
93    Hat <- X%*%invXtX%*%t(X);StuRes <- Res/as.vector(SE)/sqrt(diag(diag(1,n)-Hat));
94    EStuRes <- StuRes*sqrt((n-m-1)/(n-m-StuRes^2));
95  }else if (m!=1){
96     tStatistic <- as.vector(Para/SEPara);t <- qt(p = 0.975,df = n-m);
97    MeanData <- mean(Data);
98    SSR <- sum((Fitval-MeanData)^2);SSE <-SE^2*(n-m);SST <- SSR+SSE;
99     FStatistic <- as.vector(SSR*(n-m)/SSE/(m-1));F <-qf(0.95,m-1,n-m);
100   tTest <- cbind(tStatistic,t);FTest <- cbind(FStatistic,F);R2 <- as.vector(SSR/SST);
101   Hat <- X%*%invXtX%*%t(X);StuRes <- Res/as.vector(SE)/sqrt(diag(diag(1,n)-Hat));
102   EStuRes <- StuRes*sqrt((n-m-1)/(n-m-StuRes^2));
103 }else{
104    tStatistic <- as.vector(Para/SEPara);t <- qt(p = 0.975,df = n-m);
105   tTest <- cbind(tStatistic,t);FTest <- integer(0);R2 <- integer(0);EStuRes <- integer(0);
106 }
107 mdl <- list(Coefficients = Para,Fitted_Value = Fitval,Designed_Matrix = X,Residuals = Res,
108             External_Studentized_Residuals = EStuRes,Standard_Error = SE,
109              Coefficients _Standard_Error = SEPara,t_Test_Result = tTest,
110             F_Test_Result = FTest,Rsquare = R2,Model_Type = Model,Data = Data)
111 return(mdl)
```

```
112 }
```

1.2 函数DTTSM的使用说明

(1)输入参数Data是用于建模和参数估计的数据集；要求是时间序列格式。

(2)输入参数Model是回归模型选项；要求是字符串向量。可提供的模型选项如下：

Model参数的选项	参数对应模型的趋势类型
”Constant”	常数均值趋势
”Linear”	线性趋势
”Quadratic”	二次趋势
”SeasonC”	季节性均值趋势
”SeasonL”	季节性线性趋势
”Cosine”	余弦趋势

表 1: Model参数的选项与对应的含义

(3)”Season”类选项只能用于研究月度数据的季节性均值或者线性趋势模型，不可用于研究季度或者半年度数据。”Season”类选项要求的Data包含年份和月份，例如，数据格式应与TSA程辑包中的tempdub 数据集一致。Intercept是选择Model=”SeasonC”或者”SeasonL”后，用于选择“模型是否包括截距项”的参数；要求是布尔型变量。

(4)输出参数mdl为一个列表。具体内容如下：

列表mdl包含的成分	含义
Coefficients	回归系数的最小二乘估计
Fitted_Value	拟合值
Designed_Matrix	设计矩阵
Residuals	残差
External_Studentized_Residuals	残差
Standard_Error	模型的回归标准误差
Coefficients_Standard_Error	参数的回归标准误差
t_Test_Result	回归系数的t检验结果
F_Test_Result	回归方程的F检验结果
Rsquare	决定系数
Model_Type	模型类别
Data	训练数据集

表 2: 列表mdl包含的成分与对应的含义

1.3 例：使用函数DTTSM对rwalk数据进行基于线性趋势的回归

以下为示例代码；summary(mdl)展示了列表mdl的内部结构。

```
> library("TSA");data("rwalk");
> source('D:/R Files/TSAcourse/DTTSM.R')
> mdl <- DTTSM(rwalk,"Linear")
> summary(mdl)
                              Length Class  Mode
 Coefficients                 2      -none- numeric
Fitted_Value                  60     ts     numeric
Designed_Matrix               120    mts    numeric
Residuals                     60     ts     numeric
External_Studentized_Residuals 60    ts     numeric
Standard_Error                1      -none- numeric
 Coefficients _Standard_Error 2      -none- numeric
t_Test_Result                 4      -none- numeric
F_Test_Result                 2      -none- numeric
Rsquare                       1      -none- numeric
Model_Type                    1      -none- character
Data                          60     ts     numeric
```

2 使用TSA中的数据集检验算法的准确性

载入TSA程辑包，并获取数据集rwalk和tempdub；载入函数DTTSM至工作环境。

```
> library("TSA")
> data("tempdub")
> data("rwalk")
> source('D:/R Files/TSAcourse/DTTSM.R')
```

2.1 常均值趋势模型

选取rwalk数据集作为常均值趋势模型的训练集。首先使用lm函数进行回归分析：

```
> mdl=lm(rwalk~1);summary(mdl)

Call:
lm(formula = rwalk ~ 1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.1194 -1.9395 -0.0851  2.1335  5.3278

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.0818     0.3355   9.185 5.62e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.599 on 59 degrees of freedom
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(rwalk,"Constant")
> mdl$Coefficients
                    y
Intercept  3.081771
> mdl$Standard_Error
          y
y 2.598897
> mdl$Coefficients_Standard_Error
Intercept
0.3355161
> mdl$t_Test_Result
      tStatistic           t
```

```
[1,]  9.185165 2.000995
```

二者结果一致。

2.2 线性趋势模型

选取rwalk数据集作为线性趋势模型的训练集。首先使用lm函数进行回归分析：

```
> mdl=lm(rwalk~1+time(rwalk))
> summary(mdl)

Call:
lm(formula = rwalk ~ 1 + time(rwalk))

Residuals:
     Min       1Q   Median       3Q      Max
−2.70045 −0.79782 0.06391  0.63064  2.22128

 Coefficients :
                Estimate Std. Error t value Pr(>|t|)
(Intercept) −1.007888   0.297245  −3.391  0.00126 **
time(rwalk)  0.134087   0.008475  15.822  < 2e−16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’ . 0.1 ‘’ 1

Residual standard error: 1.137 on 58 degrees of freedom
Multiple R−squared: 0.8119,   Adjusted R−squared: 0.8086
F−statistic: 250.3 on 1 and 58 DF, p−value: < 2.2e−16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"Linear")
> mdl$Coefficients
                [,1]
Intercept  46.26597
> mdl$Standard_Error
          y
y 1.136865
> mdl$Coefficients_Standard_Error
  Intercept       Time
0.297245218 0.008474867
> mdl$t_Test_Result
      tStatistic        t
[1,]  −3.390763 2.001717
[2,]  15.821744 2.001717
> mdl$F_Test_Result
```

```
      FStatistic         F
[1,]   250.3276 4.006873
> mdl$Rsquare
          y
y 0.8118884
```

二者结果一致。

2.3 二次趋势模型

选取rwalk数据集作为二次趋势模型的训练集。首先使用lm函数进行回归分析：

```
> t <- time(rwalk);tt <- t^2;mdl=lm(rwalk~1+t+tt);summary(mdl)

Call:
lm(formula = rwalk ~ 1 + t + tt)

Residuals:
     Min       1Q   Median       3Q      Max
-2.69623 -0.76802  0.00826  0.85337  2.34468

Coefficients :
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.4272911  0.4534893  -3.147  0.00262 **
t            0.1746746  0.0343028   5.092 4.16e-06 ***
tt          -0.0006654  0.0005451  -1.221  0.22721
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' . 0.1 ' '  1

Residual standard error: 1.132 on 57 degrees of freedom
Multiple R-squared: 0.8167,   Adjusted R-squared: 0.8102
F-statistic:   127 on 2 and 57 DF,  p-value: < 2.2e-16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(rwalk,"Quadratic")
> mdl$Coefficients
                          y
Intercept  -1.4272911210
Time        0.1746745556
Time^2     -0.0006653669
> mdl$Standard_Error
          y
y 1.132091
> mdl$Coefficients_Standard_Error
   Intercept          Time        Time^2
```

```
0.4534893010 0.0343027541 0.0005450561
> mdl$t_Test_Result
      tStatistic          t
[1,]  −3.147353 2.002465
[2,]   5.092144 2.002465
[3,]  −1.220731 2.002465
> mdl$F_Test_Result
      FStatistic          F
[1,]   126.9667 3.158843
> mdl$Rsquare
           y
y 0.816681
```

二者结果一致。

2.4 季节性均值趋势模型（不带截距项）

选取tempdub数据集作为季节性均值趋势模型的训练集。首先使用lm函数进行回归分析：

```
> month.=season(tempdub);mdl=lm(tempdub~month.−1);summary(mdl)

Call:
lm(formula = tempdub ~ month. − 1)

Residuals:
    Min      1Q  Median      3Q     Max
−8.2750 −2.2479 0.1125  1.8896  9.8250

 Coefficients :
                  Estimate Std. Error t value Pr(>|t|)
month.January       16.608      0.987   16.83   <2e−16 ***
month.February      20.650      0.987   20.92   <2e−16 ***
month.March         32.475      0.987   32.90   <2e−16 ***
month.April         46.525      0.987   47.14   <2e−16 ***
month.May           58.092      0.987   58.86   <2e−16 ***
month.June          67.500      0.987   68.39   <2e−16 ***
month.July          71.717      0.987   72.66   <2e−16 ***
month.August        69.333      0.987   70.25   <2e−16 ***
month.September     61.025      0.987   61.83   <2e−16 ***
month.October       50.975      0.987   51.65   <2e−16 ***
month.November      36.650      0.987   37.13   <2e−16 ***
month.December      23.642      0.987   23.95   <2e−16 ***
−−−
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’ . 0.1 ‘’ 1

Residual standard error: 3.419 on 132 degrees of freedom
```

```
Multiple R−squared: 0.9957,   Adjusted R−squared: 0.9953
F−statistic:  2569 on 12 and 132 DF, p−value: < 2.2e−16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"SeasonC",FALSE)
> mdl$Coefficients
                [,1]
Month1 16.60833
Month2 20.65000
Month3 32.47500
Month4 46.52500
Month5 58.09167
Month6 67.50000
Month7 71.71667
Month8 69.33333
Month9 61.02500
Month10 50.97500
Month11 36.65000
Month12 23.64167
> mdl$Standard_Error
          [,1]
[1,] 3.418932
> mdl$Coefficients_Standard_Error
    Month1    Month2    Month3    Month4    Month5    Month6    Month7    Month8    Month9   Month10
        Month11
0.9869607 0.9869607 0.9869607 0.9869607 0.9869607 0.9869607 0.9869607 0.9869607 0.9869607 0.9869607
      0.9869607
   Month12
0.9869607
> mdl$t_Test_Result
      tStatistic        t
 [1,]   16.82776 1.978099
 [2,]   20.92282 1.978099
 [3,]   32.90405 1.978099
 [4,]   47.13967 1.978099
 [5,]   58.85915 1.978099
 [6,]   68.39178 1.978099
 [7,]   72.66416 1.978099
 [8,]   70.24934 1.978099
 [9,]   61.83124 1.978099
[10,]   51.64846 1.978099
[11,]   37.13420 1.978099
[12,]   23.95401 1.978099
> mdl$F_Test_Result
     FStatistic        F
[1,]   2568.838 1.826197
> mdl$Rsquare
```

```
              [,1]
[1,] 0.9957231
```

二者结果一致。

2.5 季节性均值趋势模型（带截距项）

选取tempdub数据集作为季节性均值趋势模型的训练集。首先使用lm函数进行回归分析：

```
> month.=season(tempdub);mdl=lm(tempdub~month.);summary(mdl)

Call:
lm(formula = tempdub ~ month.)

Residuals:
    Min      1Q  Median      3Q     Max
−8.2750 −2.2479 0.1125  1.8896  9.8250

Coefficients :
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         16.608      0.987  16.828  < 2e−16 ***
month.February       4.042      1.396   2.896  0.00443 **
month.March         15.867      1.396  11.368  < 2e−16 ***
month.April         29.917      1.396  21.434  < 2e−16 ***
month.May           41.483      1.396  29.721  < 2e−16 ***
month.June          50.892      1.396  36.461  < 2e−16 ***
month.July          55.108      1.396  39.482  < 2e−16 ***
month.August        52.725      1.396  37.775  < 2e−16 ***
month.September     44.417      1.396  31.822  < 2e−16 ***
month.October       34.367      1.396  24.622  < 2e−16 ***
month.November      20.042      1.396  14.359  < 2e−16 ***
month.December       7.033      1.396   5.039 1.51e−06 ***
−−−
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’ . 0.1 ‘’ 1

Residual standard error: 3.419 on 132 degrees of freedom
Multiple R−squared: 0.9712,   Adjusted R−squared: 0.9688
F−statistic: 405.1 on 11 and 132 DF, p−value: < 2.2e−16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"SeasonC",TRUE)
> mdl$Coefficients
              [,1]
Month1 16.608333
Month2  4.041667
```

```
Month3 15.866667
Month4 29.916667
Month5 41.483333
Month6 50.891667
Month7 55.108333
Month8 52.725000
Month9 44.416667
Month10 34.366667
Month11 20.041667
Month12 7.033333
> mdl$Standard_Error
          [,1]
[1,] 3.418932
> mdl$Coefficients_Standard_Error
    Month1    Month2    Month3    Month4    Month5    Month6    Month7    Month8    Month9   Month10
        Month11
0.9869607 1.3957732 1.3957732 1.3957732 1.3957732 1.3957732 1.3957732 1.3957732 1.3957732 1.3957732
      1.3957732
  Month12
1.3957732
> mdl$t_Test_Result
       tStatistic          t
 [1,]  16.827755 1.978099
 [2,]   2.895647 1.978099
 [3,]  11.367654 1.978099
 [4,]  21.433759 1.978099
 [5,]  29.720683 1.978099
 [6,]  36.461272 1.978099
 [7,]  39.482297 1.978099
 [8,]  37.774761 1.978099
 [9,]  31.822266 1.978099
[10,]  24.621956 1.978099
[11,]  14.358827 1.978099
[12,]   5.039023 1.978099
> mdl$F_Test_Result
                        F
[1,] 405.1247 1.861868
> mdl$Rsquare
          [,1]
[1,] 0.9712316
```

二者结果一致。

2.6 季节性线性趋势模型（不带截距项）

选取tempdub数据集作为季节性线性趋势模型的训练集。首先使用lm函数进

行回归分析：

```
> month.=season(tempdub);mdl=lm(tempdub~month.+time(tempdub)-1);summary(mdl)

Call:
lm(formula = tempdub ~ month. + time(tempdub) - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-8.2809  -2.2459   0.1151   1.8922   9.8296

 Coefficients :
                    Estimate Std. Error t value Pr(>|t|)
month.January     19.190720 163.172413   0.118    0.907
month.February    23.232496 163.179317   0.142    0.887
month.March       35.057605 163.186221   0.215    0.830
month.April       49.107714 163.193125   0.301    0.764
month.May         60.674490 163.200029   0.372    0.711
month.June        70.082933 163.206933   0.429    0.668
month.July        74.299709 163.213837   0.455    0.650
month.August      71.916485 163.220740   0.441    0.660
month.September   63.608260 163.227644   0.390    0.697
month.October     53.558370 163.234548   0.328    0.743
month.November    39.233479 163.241452   0.240    0.810
month.December    26.225255 163.248356   0.161    0.873
time(tempdub)     -0.001311   0.082848  -0.016    0.987

Residual standard error: 3.432 on 131 degrees of freedom
Multiple R-squared: 0.9957,   Adjusted R-squared: 0.9953
F-statistic:  2353 on 13 and 131 DF, p-value: < 2.2e-16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"SeasonL",FALSE)
> mdl$Coefficients
          Jan           Feb           Mar           Apr           May          June          July           Aug
         Sept
19.190719697 23.232495629 35.057604895 49.107714161 60.674490093 70.082932692 74.299708625 71.916484557
      63.608260489
          Oct           Nov           Dec         Slope
53.558369755 39.233479021 26.225254953 -0.001311189
> mdl$Standard_Error
           [,1]
[1,]  3.431953
> mdl$Coefficients_Standard_Error
          Jan           Feb           Mar           Apr           May          June          July           Aug
         Sept
163.17241323 163.17931711 163.18622100 163.19312488 163.20002876 163.20693265 163.21383653 163.22074042
      163.22764430
```

```
          Oct           Nov           Dec         Slope
163.23454819 163.24145207 163.24835595 0.08284814
> mdl$t_Test_Result
      tStatistic         t
 [1,]  0.11761007 1.978239
 [2,]  0.14237402 1.978239
 [3,]  0.21483189 1.978239
 [4,]  0.30091779 1.978239
 [5,]  0.37177990 1.978239
 [6,]  0.42941149 1.978239
 [7,]  0.45522923 1.978239
 [8,]  0.44060874 1.978239
 [9,]  0.38969049 1.978239
[10,]  0.32810683 1.978239
[11,]  0.24034017 1.978239
[12,]  0.16064636 1.978239
[13,] -0.01582641 1.978239
> mdl$F_Test_Result
     FStatistic       F
[1,]    2353.276 1.79498
> mdl$Rsquare
[1] 0.9957362
```

二者结果一致。

2.7 季节性线性趋势模型（带截距项）

选取tempdub数据集作为季节性线性趋势模型的训练集。首先使用lm函数进行回归分析：

```
> month.=season(tempdub);mdl=lm(tempdub~month.+time(tempdub));summary(mdl)

Call:
lm(formula = tempdub ~ month. + time(tempdub))

Residuals:
    Min      1Q  Median      3Q     Max
-8.2809 -2.2459  0.1151  1.8922  9.8296

Coefficients :
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      19.190720 163.172413   0.118  0.90656
month.February    4.041776   1.401106   2.885  0.00458 **
month.March      15.866885   1.401157  11.324  < 2e-16 ***
month.April      29.916994   1.401242  21.350  < 2e-16 ***
month.May        41.483770   1.401361  29.602  < 2e-16 ***
month.June       50.892213   1.401514  36.312  < 2e-16 ***
```

```
month.July       55.108989   1.401701  39.316  < 2e-16 ***
month.August     52.725765   1.401922  37.610  < 2e-16 ***
month.September 44.417541   1.402177  31.678  < 2e-16 ***
month.October    34.367650   1.402466  24.505  < 2e-16 ***
month.November 20.042759   1.402789  14.288  < 2e-16 ***
month.December   7.034535   1.403146   5.013  1.7e-06 ***
time(tempdub)   -0.001311   0.082848  -0.016  0.98740
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 ' ' . 0.1 ' '  1

Residual standard error: 3.432 on 131 degrees of freedom
Multiple R-squared: 0.9712,   Adjusted R-squared: 0.9686
F-statistic: 368.6 on 12 and 131 DF, p-value: < 2.2e-16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"SeasonL",TRUE)
> mdl$Coefficients
          Jan           Feb           Mar           Apr           May          June          July           Aug
         Sept
19.190719697  4.041775932 15.866885198 29.916994464 41.483770396 50.892212995 55.108988928 52.725764860
     44.417540793
          Oct           Nov           Dec         Slope
34.367650058 20.042759324 7.034535256 -0.001311189
> mdl$Standard_Error
          [,1]
[1,]  3.431953
> mdl$Coefficients_Standard_Error
          Jan           Feb           Mar           Apr           May          June          July           Aug
         Sept
163.17241323   1.40110614   1.40115717   1.40124221   1.40136126   1.40151432   1.40170136   1.40192237
     1.40217735
          Oct           Nov           Dec         Slope
  1.40246627   1.40278911   1.40314584   0.08284814
> mdl$t_Test_Result
       tStatistic         t
 [1,]  0.11761007 1.978239
 [2,]  2.88470361 1.978239
 [3,] 11.32412950 1.978239
 [4,] 21.35033776 1.978239
 [5,] 29.60248118 1.978239
 [6,] 36.31230337 1.978239
 [7,] 39.31578480 1.978239
 [8,] 37.60961791 1.978239
 [9,] 31.67754830 1.978239
[10,] 24.50515268 1.978239
[11,] 14.28779224 1.978239
[12,]  5.01340278 1.978239
```

```
[13,] −0.01582641 1.978239
> mdl$F_Test_Result
      FStatistic          F
[1,]   368.5517 1.826769
> mdl$Rsquare
[1] 0.9712317
```

二者结果一致。

2.8 余弦趋势模型

选取tempdub数据集作为余弦趋势模型的训练集。首先使用lm函数进行回归分析：

```
> har.=harmonic(tempdub,1);mdl=lm(tempdub~har.);summary(mdl)

Call:
lm(formula = tempdub ~ har.)

Residuals:
     Min       1Q   Median       3Q      Max
−11.1580 −2.2756 −0.1457  2.3754  11.2671

 Coefficients :
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          46.2660     0.3088 149.816  < 2e−16 ***
har.cos(2*pi*t) −26.7079     0.4367 −61.154  < 2e−16 ***
har.sin(2*pi*t)  −2.1697     0.4367  −4.968 1.93e−06 ***
−−−
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘’ . 0.1 ‘’ 1

Residual standard error: 3.706 on 141 degrees of freedom
Multiple R−squared: 0.9639,  Adjusted R−squared: 0.9634
F−statistic:  1882 on 2 and 141 DF, p−value: < 2.2e−16
```

接下来使用DTTSM函数进行回归分析：

```
> mdl=DTTSM(tempdub,"Cosine")
> mdl$Coefficients
                [,1]
Intercept   46.26597
Cos        −26.70793
Sin         −2.16975
> mdl$Standard_Error
          [,1]
[1,] 3.705826
> mdl$Coefficients_Standard_Error
```

```
Intercept        Cos        Sin
0.3088188 0.4367358 0.4367358
> mdl$t_Test_Result
      tStatistic          t
[1,]  149.815907 1.976931
[2,]  −61.153529 1.976931
[3,]   −4.968106 1.976931
> mdl\$F_Test_Result
                        F
[1,] 1882.218 3.060292
> mdl$Rsquare
           [,1]
[1,] 0.9638965
```

二者结果一致。

2.9 计算学生化外残差

选取tempdub数据集和拟合的余弦趋势模型作为残差计算的对象。首先使用rstudent函数计算学生化外残差。

```
> har.=harmonic(tempdub,1);mdl=lm(tempdub~har.);rstudent(mdl)
          1           2           3           4           5           6           7           8           9
 1.40708173  0.99495526 −0.11765154  0.92775484  1.41184369  0.05139963  0.19733065 −0.70246578 −0.10842494
         10          11          12          13          14          15          16          17          18
 0.01746649  1.31478575 −1.15243607 −0.94263780 −0.80382503 −1.88002370  0.35443722  0.97037433
         −0.49249136
         19          20          21          22          23          24          25          26          27
−0.23751618 −0.56601145 −0.35313656  0.67069873  0.71020799  1.89299022 −2.54546949 −0.12265254
          1.74893150
         28          29          30          31          32          33          34          35          36
 0.16408153 −1.24066827 −0.08446912  0.19733065 −0.62056312 −0.21714984  0.47980525  0.65558962  0.10296375
         37          38          39          40          41          42          43          44          45
 0.52821902 −2.02669030  1.08247782  1.14766260 −1.02029007 −0.03012062 −0.91956345 −1.30694252
         −0.18996454
         46          47          48          49          50          51          52          53          54
 0.18052681 −0.43263218  0.48378885 −0.12447077 −0.39459742  2.54809106  1.61919834 −0.66431050
         −0.16600297
         55          56          57          58          59          60          61          62          63
−0.61873770 −0.04907414 −0.24433858  0.61611091  0.27423413 −0.95993948 −1.52275417  0.55730175
         −0.44403673
         64          65          66          67          68          69          70          71          72
 0.68148451  0.23349584 −1.67711137 −0.23751618  0.33148140  0.10897001 −0.22713437 −0.16062585
         −1.04233660
         73          74          75          76          77          78          79          80          81
−3.13673302 −0.83119308  0.09974360  1.25804592  1.05278988 −0.05729441 −0.21032834  0.03244607  0.27207662
         82          83          84          85          86          87          88          89          90
```

```
21 1.16433860  0.51921869 −0.11443128 −2.31417996 −0.55803710 −0.38958895 0.98263356 −0.52792453
       1.33699526
22          91          92          93          94          95          96          97          98          99
23 −1.36047544 −0.92139791 0.92699396 2.40240312 0.76487162  0.94836137 −1.69032654 −1.32652657
       −0.06330018
24         100         101         102         103         104         105         106         107         108
25 −0.10767093 1.24570655 −0.51973921 −0.75529370 0.30427873 0.16332739 −0.66290202 −0.56886571
       −1.90490214
26         109         110         111         112         113         114         115         116         117
27 0.80125658  0.99495526  3.16956659  0.30002827 −0.60972745 0.16010669 −0.18314383 0.49482286 0.27207662
28         118         119         120         121         122         123         124         125         126
29 1.97366272  0.87434890 −1.04233660 −0.53260097 −0.42181919 0.86287566 1.39648670 −0.80096094
       −1.23237289
30         127         128         129         130         131         132         133         134         135
31 0.27890033 −0.92139791 −1.03626143 0.64339970 0.54647464 0.34766884  0.22883079 −0.66716846 −1.76743346
32         136         137         138         139         140         141         142         143         144
33 −0.76139534 1.10781744  0.05139963 −0.26470778 0.16832909 −1.14634533 1.10923780 1.59271156 0.53828758
```

接下来使用DTTSM函数计算学生化外残差：

```
1  > mdl=DTTSM(tempdub,"Cosine");mdl$External_Studentized_Residuals
2                Jan          Feb          Mar          Apr          May          Jun          Jul          Aug
                     Sep
3  1964  1.40708173  0.99495526 −0.11765154 0.92775484 1.41184369  0.05139963  0.19733065 −0.70246578
       −0.10842494
4  1965 −0.94263780 −0.80382503 −1.88002370 0.35443722 0.97037433 −0.49249136 −0.23751618 −0.56601145
       −0.35313656
5  1966 −2.54546949 −0.12265254 1.74893150 0.16408153 −1.24066827 −0.08446912 0.19733065 −0.62056312
       −0.21714984
6  1967  0.52821902 −2.02669030 1.08247782 1.14766260 −1.02029007 −0.03012062 −0.91956345 −1.30694252
       −0.18996454
7  1968 −0.12447077 −0.39459742 2.54809106 1.61919834 −0.66431050 −0.16600297 −0.61873770 −0.04907414
       −0.24433858
8  1969 −1.52275417 0.55730175 −0.44403673 0.68148451 0.23349584 −1.67711137 −0.23751618 0.33148140
       0.10897001
9  1970 −3.13673302 −0.83119308 0.09974360 1.25804592 1.05278988 −0.05729441 −0.21032834 0.03244607
       0.27207662
10 1971 −2.31417996 −0.55803710 −0.38958895 0.98263356 −0.52792453 1.33699526 −1.36047544 −0.92139791
       0.92699396
11 1972 −1.69032654 −1.32652657 −0.06330018 −0.10767093 1.24570655 −0.51973921 −0.75529370 0.30427873
       0.16332739
12 1973  0.80125658  0.99495526  3.16956659  0.30002827 −0.60972745 0.16010669 −0.18314383 0.49482286
       0.27207662
13 1974 −0.53260097 −0.42181919 0.86287566 1.39648670 −0.80096094 −1.23237289 0.27890033 −0.92139791
       −1.03626143
14 1975  0.22883079 −0.66716846 −1.76743346 −0.76139534 1.10781744 0.05139963 −0.26470778 0.16832909
       −1.14634533
15               Oct          Nov          Dec
16 1964  0.01746649  1.31478575 −1.15243607
```

```
1965  0.67069873  0.71020799  1.89299022
1966  0.47980525  0.65558962  0.10296375
1967  0.18052681 −0.43263218 0.48378885
1968  0.61611091  0.27423413 −0.95993948
1969 −0.22713437 −0.16062585 −1.04233660
1970  1.16433860  0.51921869 −0.11443128
1971  2.40240312  0.76487162  0.94836137
1972 −0.66290202 −0.56886571 −1.90490214
1973  1.97366272  0.87434890 −1.04233660
1974  0.64339970  0.54647464  0.34766884
1975  1.10923780  1.59271156  0.53828758
                                        $
```

二者结果一致。

2.10 结论

依照函数DTTSM在TSA中的数据集rwalk和tempdub上的表现看来，算法是正确的。

3 应用函数DTTSM解决习题（3.5，3.7，3.11，3.13）

3.1 习题3.5

实现习题要求的代码如下（如果要运行这段代码，请先下载程辑包”TSA”和”lawstat”）：

```
library("TSA");library("lawstat");
source('D:/R Files/TSAcourse/DTTSM.R');
data(wages);
plot.ts(wages,type = "o");
lmdl = DTTSM(wages,"Linear");
lCoef = lmdl$Coefficients;
ltTest = lmdl$t_Test_Result;lFTest = lmdl$F_Test_Result;
lR2 = lmdl$Rsquare;
lESRE = lmdl$External_Studentized_Residuals;
plot.ts(lESRE,type = "o",ylab = "Studentized Residuals");abline(h = c(−3,3));

qmdl = DTTSM(wages,"Quadratic");
qCoef = qmdl$Coefficients;
qtTest = qmdl$t_Test_Result;qFTest = qmdl$F_Test_Result;
```

```
qR2 = qmdl$Rsquare;
qESRE = qmdl$External_Studentized_Residuals
plot.ts(qESRE,type = "o",ylab = "Studentized Residuals");abline(h = c(-3,3));
qRun = runs.test(qESRE);
acf(qESRE);hist(qESRE);qqnorm(qESRE);qqline(qESRE);                    $
```

(a) 绘制的时间序列图像如下：

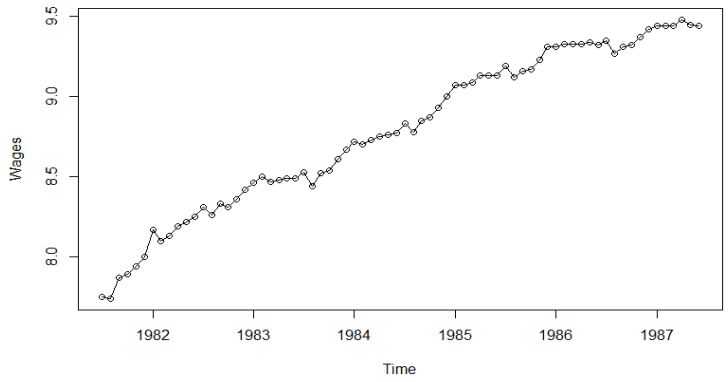

图 1: 工资时间序列图

从图像可以看出，时间序列有较为明显的线性趋势，可以尝试用线性趋势或者二次趋势拟合时间序列。

(b) 用最小二乘法拟合线性趋势，得到列表对象lmdl。其回归系数如下：

```
> lCoef
                Wages
Intercept -549.0060630
Time         0.2810805
```

查看回归模型的显著性检验结果：

```
> ltTest
     tStatistic         t
[1,]  -49.23962 1.994437
[2,]   50.02768 1.994437
> lFTest
     FStatistic         F
[1,]   2502.768 3.977779
```

可以得出结论：回归方程和回归系数都是显著的。再查看决定系数：

```
> lR2
[1] 0.972792
```

可以看出，回归直线的拟合程度较好。

(c) 绘制的来自lmdl的学生化外残差的时间序列图像如下：

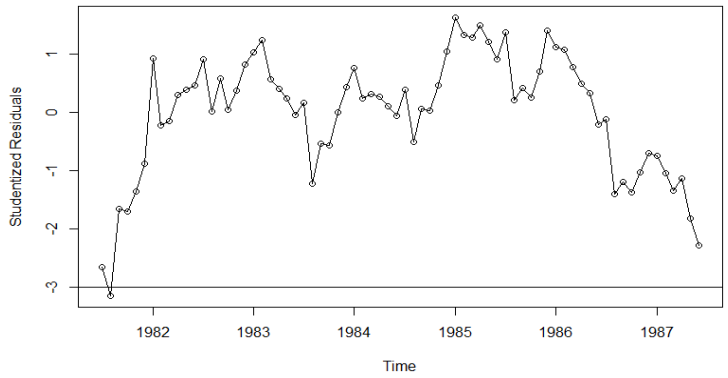

图 2: 线性趋势模型的学生化外残差的时间序列图

发现有一个点的残差在[-3,3]以外；故认为这个残差属于异常值，而且线性趋势并不是一个好的模型。此外，从图像可以看出，图像较平滑，连续时间点的残差可能存在一定相关性。

(d) 用最小二乘法拟合二次趋势，得到列表对象qmdl。其回归系数如下：

```
> qCoef
                    Wages
Intercept  −8.494973e+04
Time        8.534287e+01
Time^2     −2.143199e−02
```

查看回归模型的显著性检验结果：

```
> qtTest
      tStatistic         t
[1,]   −8.335556 1.994945
[2,]    8.309054 1.994945
[3,]   −8.281688 1.994945
> qFTest
      FStatistic         F
[1,]    2493.913 3.129644
```

可以得出结论：回归方程和回归系数都是显著的。查看决定系数：

```
> qR2
[1] 0.9863551
```

可以看出，回归直线的拟合度较好，且较线性趋势的拟合效果更好。

(e) 绘制的来自qmdl的学生化外残差的时间序列图像如下：

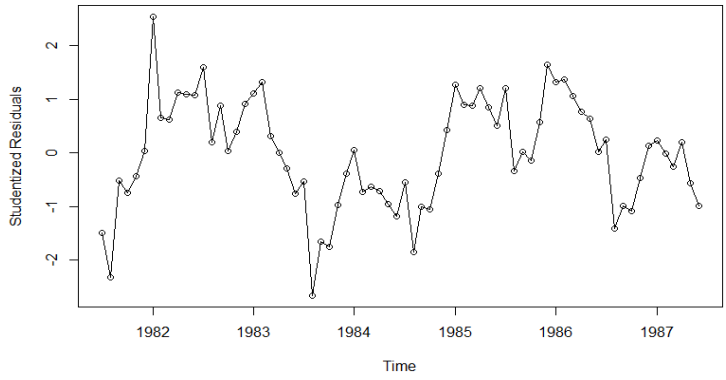

图 3: 二次趋势模型的学生化外残差的时间序列图

发现所有残差都在[-3,3]内；故认为没有异常值，而且二次模型在残差的初步分析上的表现比较好。此外，从图像可以看出，图像平滑性较线性趋势的学生化外残差图更低，且连续时间点的残差相关性也更不明显。

3.2 习题3.11

(a) 二次趋势的学生化外残差储存在时间序列格式变量qESRE中。

(b) 对学生化外残差的时间序列进行游程检验（游程检验的函数runs.test来自"lawstat"包）的结果如下：

```
> qRun

  Runs Test - Two sided

data:  qESRE
Standardized Runs Statistic = -5.2224, p-value = 1.767e-07
```

这个结果说明，应该拒绝“学生化外残差序列是随机的”这一假设。

(c) 使用acf函数计算学生化外残差的样本自相关，并可视化：

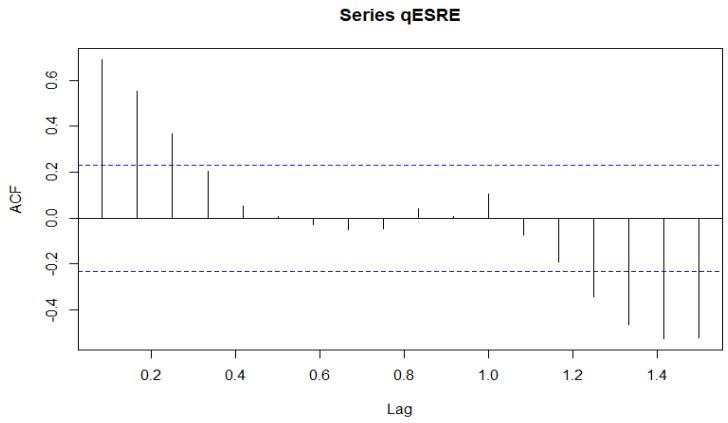

图 4: 学生化外残差的样本自相关图

样本自相关图说明，学生化外残差序列并不是白噪声过程；回归的基本假设没有得到满足。

(d) 学生化外残差的直方图和Q-Q图如下：

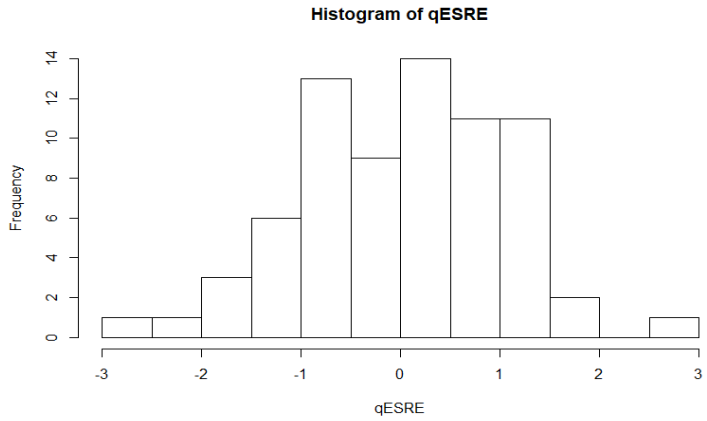

图 5: 学生化外残差的直方图

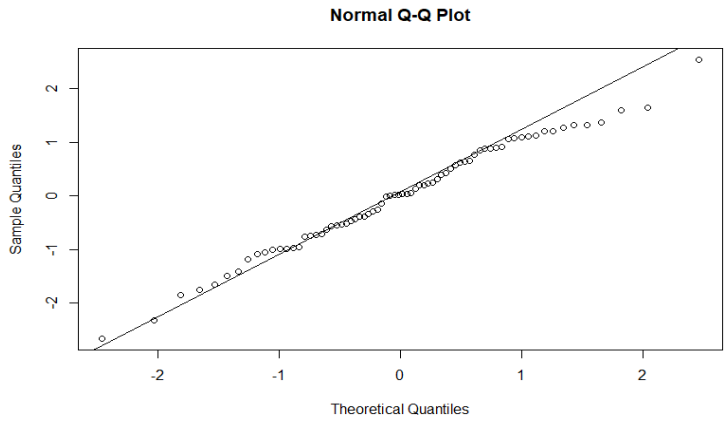

图 6: 学生化外残差的Q-Q图

直方图并没有显现出正态性。图的右侧没有从高值到低值逐渐减小的特征，左侧甚至还出现了和正态分布相反的单调区间。Q-Q图表明，靠近均值的点的分位数和正态分布分位数非常接近，但是远离均值的点的分位数和正态分布分位数差距较大。到这里，可以认为残差很可能不是正态的；为了比较严格的验证这个想法，使用Shapiro-Wilk检验方法来对残差时间序列进行正态性检验：

```
> qSW

  Shapiro−Wilk normality test

data:  qESRE
W = 0.98868, p−value = 0.7693
```

从检验结果来看，目前的证据不足以拒绝残差时间序列的正态性假设。综上，我们仍然肯定残差的正态性。

3.3 习题3.7

实现习题要求的代码如下（如果要运行这段代码，请先下载程辑包”TSA”和”lawstat”）：

```
library("TSA");library("lawstat");
data("winnebago");
source('D:/R Files/TSAcourse/DTTSM.R')
plot.ts(winnebago,type = "o");
lmdl = DTTSM(winnebago,"Linear");
lCoef = lmdl$Coefficients;
ltTest = lmdl$t_Test_Result;lFTest = lmdl$F_Test_Result;
lR2 = lmdl$Rsquare;
```

```
lESRE = lmdl$External_Studentized_Residuals;
plot.ts(lESRE,type = "o",ylab = "Studentized Residuals");abline(h = c(-3,3));

plot.ts(log(winnebago),type = "o");
lglmdl = DTTSM(log(winnebago),"Linear");
lglCoef = lglmdl$Coefficients;
lgltTest = lglmdl$t_Test_Result;lglFTest = lglmdl$F_Test_Result;
lglR2 = lglmdl$Rsquare;
lglESRE = lglmdl$External_Studentized_Residuals;
plot.ts(lglESRE,type = "o",ylab = "Studentized Residuals");abline(h = c(-3,3));

lgslmdl = DTTSM(log(winnebago),"SeasonL",TRUE);
lgslCoef = lgslmdl$Coefficients;
lgsltTest = lgslmdl$t_Test_Result;lgslFTest = lgslmdl$F_Test_Result;
lgslR2 = lgslmdl$Rsquare;
lgslESRE = lgslmdl$External_Studentized_Residuals;
plot.ts(lgslESRE,type = "o",ylab = "Studentized Residuals");abline(h = c(-3,3));
lgslRun = runs.test(lgslESRE);
acf(lgslESRE);hist(lgslESRE);qqnorm(lgslESRE);qqline(lgslESRE);
lgslSW = shapiro.test(lgslESRE);
idx = abs(lgslESRE)<3;
lgslDmdl = DTTSM(log(winnebago[idx]),"SeasonL",TRUE);
lgslDSW = shapiro.test(lgslDmdl$External_Studentized_Residuals)
```

(a) 绘制的时间序列图像如下：

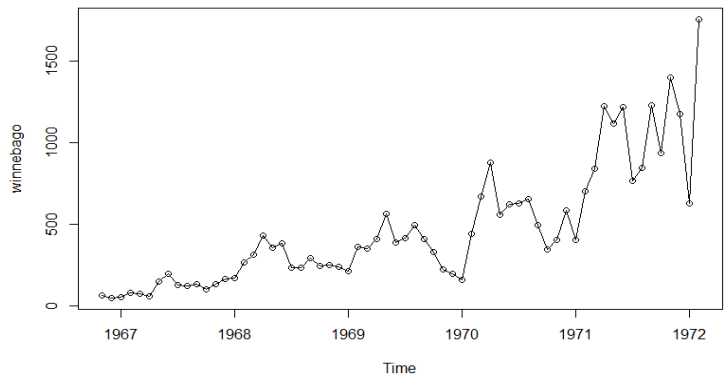

图 7: 销量时间序列图

从图像可以看出，时间序列有较为明显的季节性差异，而数据总体呈现值增大的趋势，因此可以尝试用线性趋势或者季节性趋势拟合时间序列。

(b) 用最小二乘法拟合线性趋势，得到列表对象lmdl。其回归系数如下：

```
> lCoef
   Intercept            Time
```

```
3 −394885.6849     200.7418
```

查看回归模型的显著性检验结果：

```
1 > ltTest
2       tStatistic          t
3 [1,]   −11.77366 1.998972
4 [2,]    11.78758 1.998972
5 > lFTest
6       FStatistic          F
7 [1,]     138.947 3.995887
```

可以得出结论：回归方程和回归系数都是显著的。再查看决定系数：

```
1 > lR2
2 [1] 0.6914609
```

可以看出，回归直线的拟合程度一般。接下来绘制来自lmdl的学生化外残差的时间序列图像，如下所示：

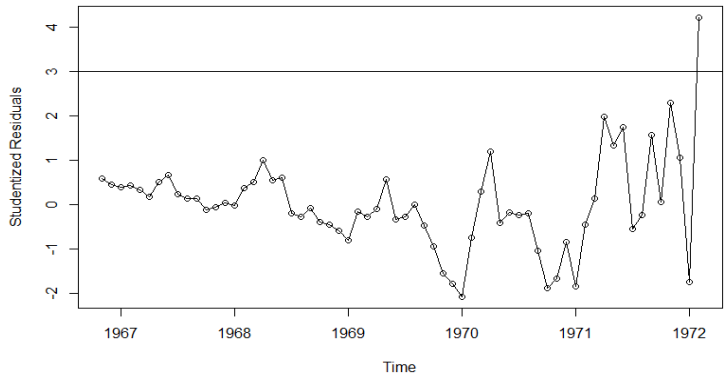

图 8: 线性趋势模型的学生化外残差的时间序列图

发现有一个点的残差在[-3,3]以外；故认为这个残差属于异常值；结合R^2的情况以及时间序列图象，认为线性趋势并不是一个好的模型。

(c)对销量取对数后绘制的时间序列图像如下：

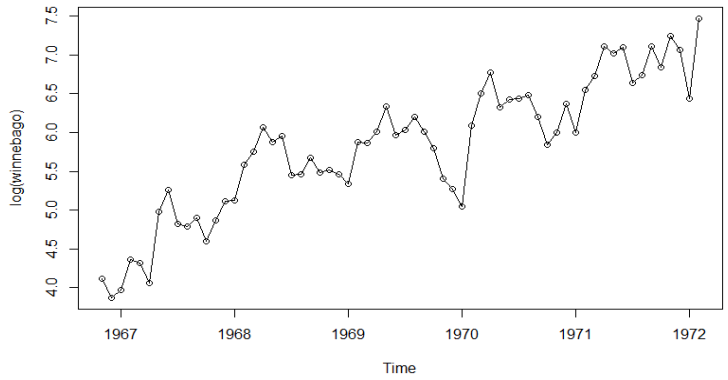

图 9: 对数销量的时间序列图

可以看出,log变换将一些本来量级差距较大的数据间的差距变小，有益于对时间序列数据进行进一步处理。同时，log变换保留了数据中季节性差异的特性。

(d) 用最小二乘法对对数变换下的销量拟合线性趋势，得到列表对象lglmdl。拟合的学生会外残差的时间序列图如下所示：

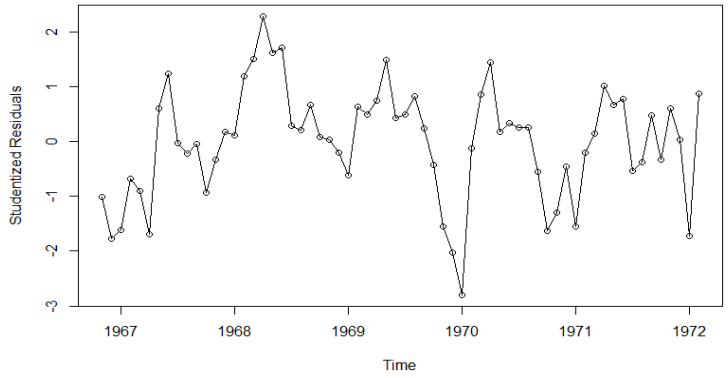

图 10: 对数变换后拟合线性趋势模型的学生化外残差的时间序列图

发现所有残差都在[-3,3]内；故认为没有异常值点，而且基于对数变换的线性趋势模型在残差的初步分析上的表现比较好。但是，可以看出图像仍有一定的平滑性，连续时间点的残差有一定相关性。

(e) 用最小二乘法对对数变换下的销量拟合季节均值+线性时间趋势，得到列表对象lgslmdl。回归系数如下：

```
1 > lgslCoef
```

```
     Jan       Feb       Mar       Apr       May      June      July       Aug
                Sept
−997.3306126 0.6244477 0.6821971 0.8095882 0.8695252 0.8630875 0.5539177 0.5698855
   0.5757167
     Oct       Nov       Dec     Slope
0.2634863 0.2868224 0.2480223 0.5090896
```

查看回归系数的显著性检验结果：

```
> lgsltTest
      tStatistic         t
 [1,] −19.694542 2.007584
 [2,]   3.434354 2.007584
 [3,]   3.573999 2.007584
 [4,]   4.243267 2.007584
 [5,]   4.558850 2.007584
 [6,]   4.525954 2.007584
 [7,]   2.904879 2.007584
 [8,]   2.988429 2.007584
 [9,]   3.018436 2.007584
[10,]   1.381002 2.007584
[11,]   1.577145 2.007584
[12,]   1.364080 2.007584
[13,]  19.799687 2.007584
```

可以发现，截距项（也是一月份对应的回归系数）和斜率非常显著，10 – 12月的系数不显著，其他月份的系数一般显著。

(f) (e)中模型学生化外残差的时间序列图如下：

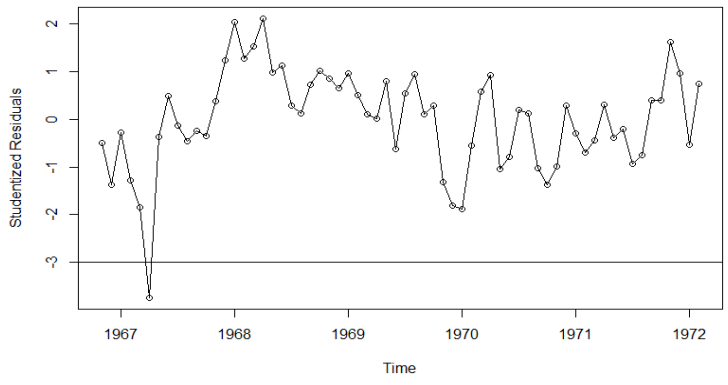

图 11: 对数变换后拟合季节均值+线性时间趋势的学生化外残差的时间序列图

发现有一个点的残差在[-3,3]以外，故认为这个残差属于异常值；但是其他的残差在图像上的表现非常好，不平滑性非常明显，有理由相信模型本身并没

有问题，异常值点只是意外产生的；如果允许的话，可以在之后的分析中删除这个异常值点，这样可能可以得出一个非常好的模型。

3.4 习题3.13

(a) 3.7(e)拟合的学生化外残差储存在时间序列格式变量lgslESRE中。

(b) 对学生化外残差的时间序列进行游程检验（游程检验的函数runs.test来自"lawstat"包）的结果如下：

```
> lgslRun

  Runs Test - Two sided

data:  lgslESRE
Standardized Runs Statistic = -2.7721, p-value = 0.00557
```

这个结果说明，应该拒绝“学生化外残差序列是随机的”这一假设。

(c) 使用acf函数计算学生化外残差的样本自相关，并可视化： 样本自相关图说

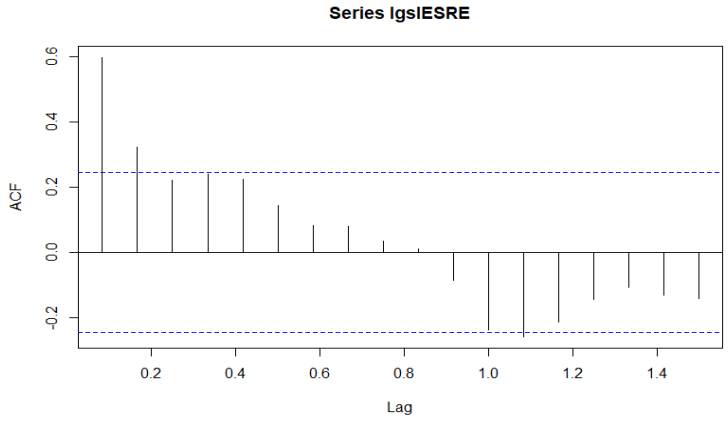

图 12: 学生化外残差的样本自相关图

明，学生化外残差序列并不是白噪声过程；回归的基本假设没有得到满足。

(d) 学生化外残差的直方图和Q-Q图如下：

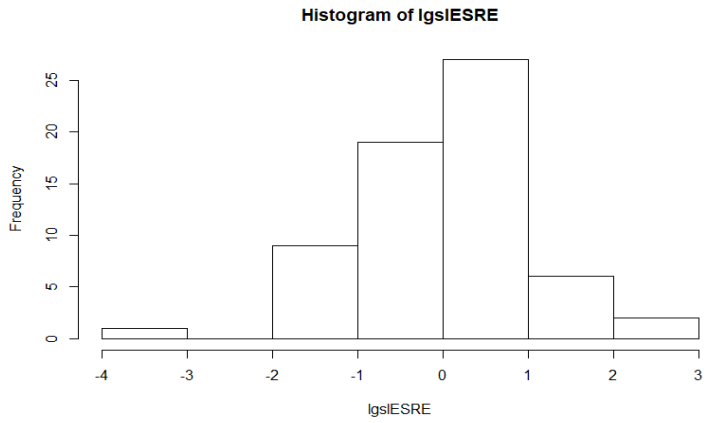

图 13: 学生化外残差的直方图

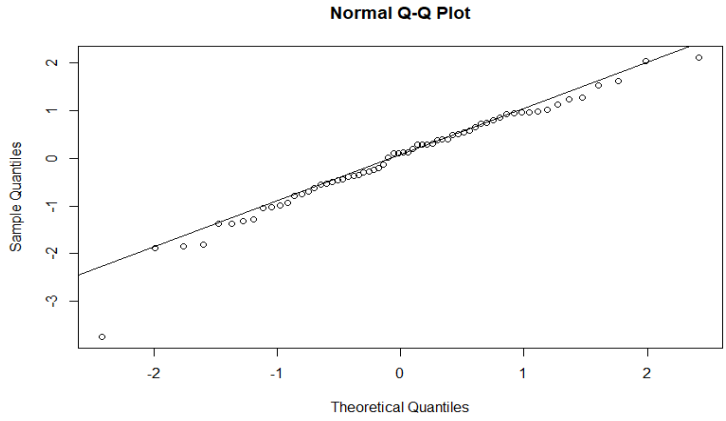

图 14: 学生化外残差的Q-Q图

直方图显现出了一定的正态性，频数分布的单调性和正态分布几乎一致。Q-Q图表明，仅有不超过3个样本点不满足正态性——而这并不能直接推翻残差正态的假设。于是进行Shapiro-Wilk 检验，从比较严谨的方面验证模型残差的正态性。

```
> lgslSW

  Shapiro−Wilk normality test

data:  lgslESRE
W = 0.97035, p−value = 0.1262
```

从检验结果来看，可以接受残差时间序列的正态性假设，但是残差时间序列的正态性并不是很显著。因此，考虑删除残差中的异常值点后，重新进行回归分析，再对新模型的学生化外残差进行Shapiro-Wilk检验。

```
> lgslDSW

  Shapiro−Wilk normality test

data:  lgslDmdl$External_Studentized_Residuals
W = 0.98333, p−value = 0.551
```

此时，残差时间序列的正态性非常显著；同时，这也说明模型效果非常好，在3.7中根据残差对模型和异常值点的推断是没有问题的。综上，我们肯定残差的正态性。