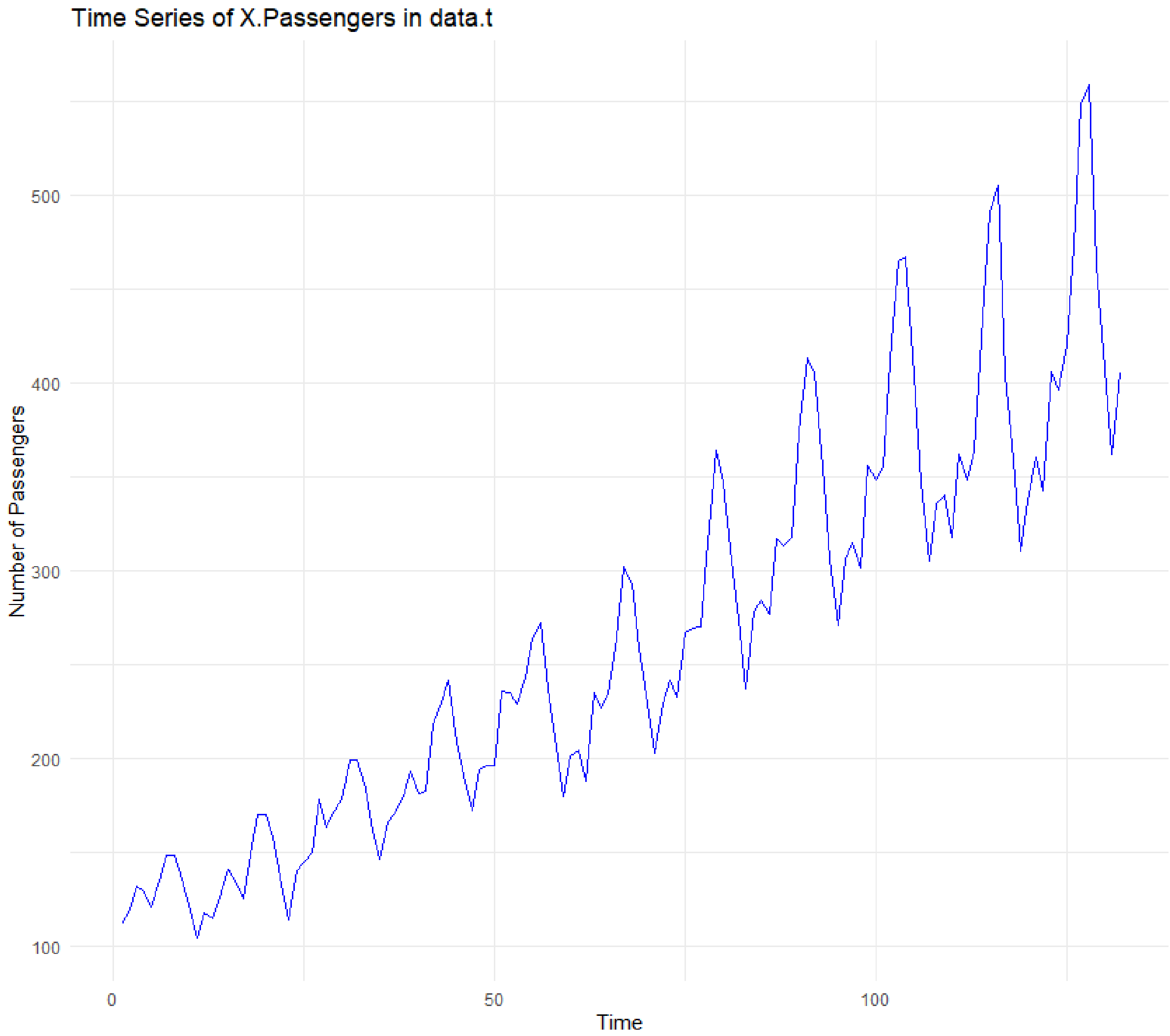
期末報告預備1

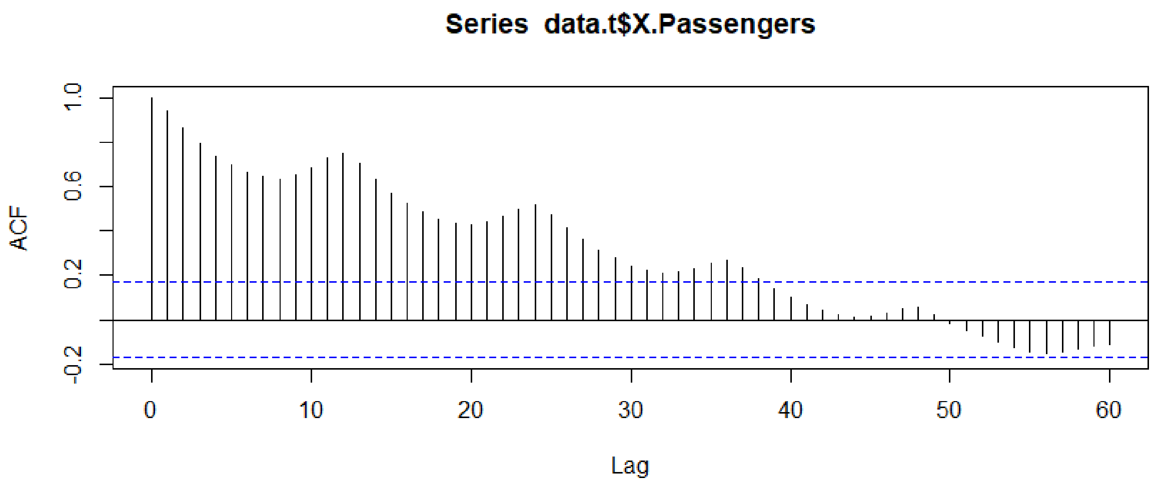
1.原始資料的時數圖、ACF與PACF圖。判別資料平穩嗎?

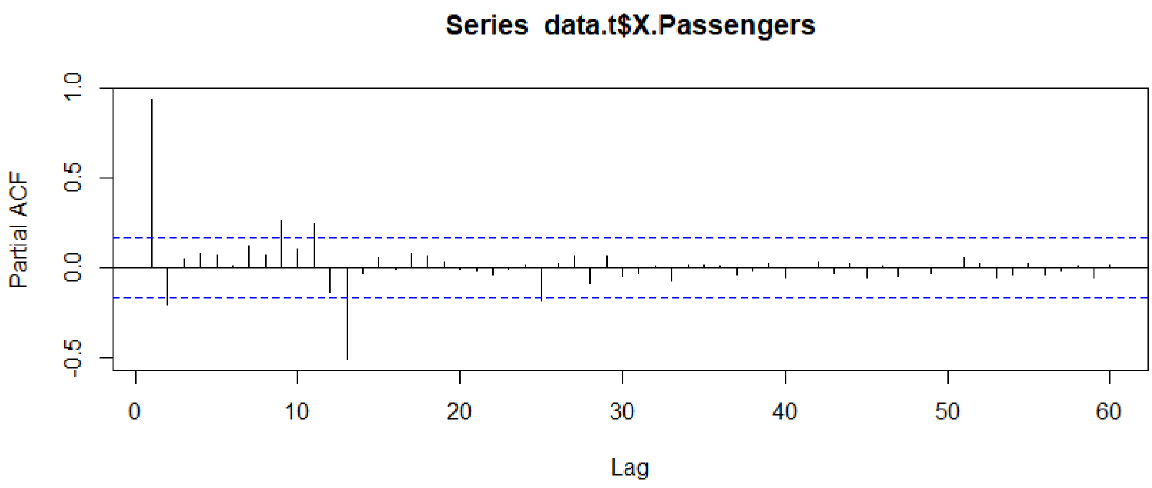
原144筆，去掉1960年分12筆資料後的訓練樣本的132筆資料(1949-01~1959-12)。



不平穩，有趨勢、季節性且震盪隨時間加劇。7、8月份通常是全年最高

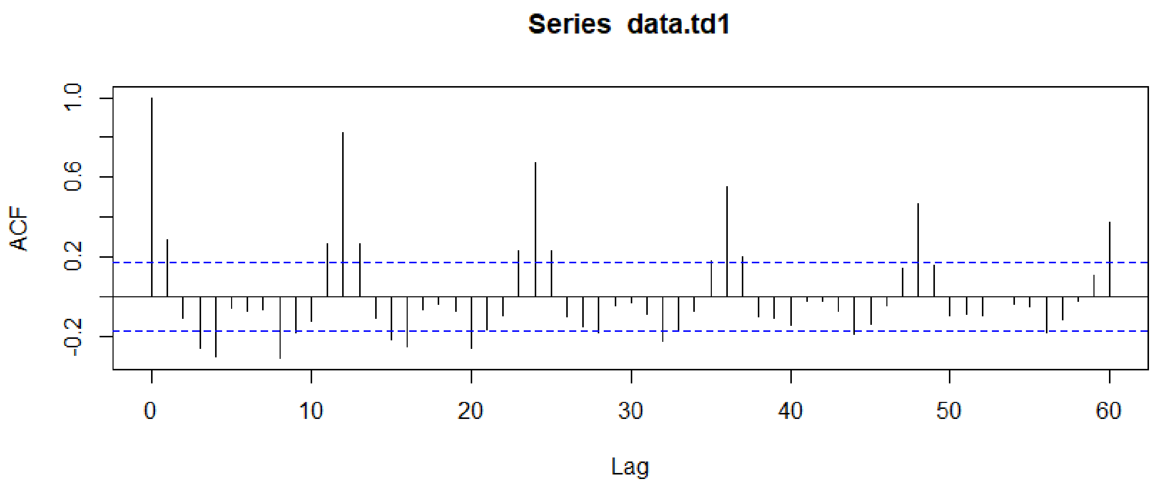
ACF圖與PACF圖

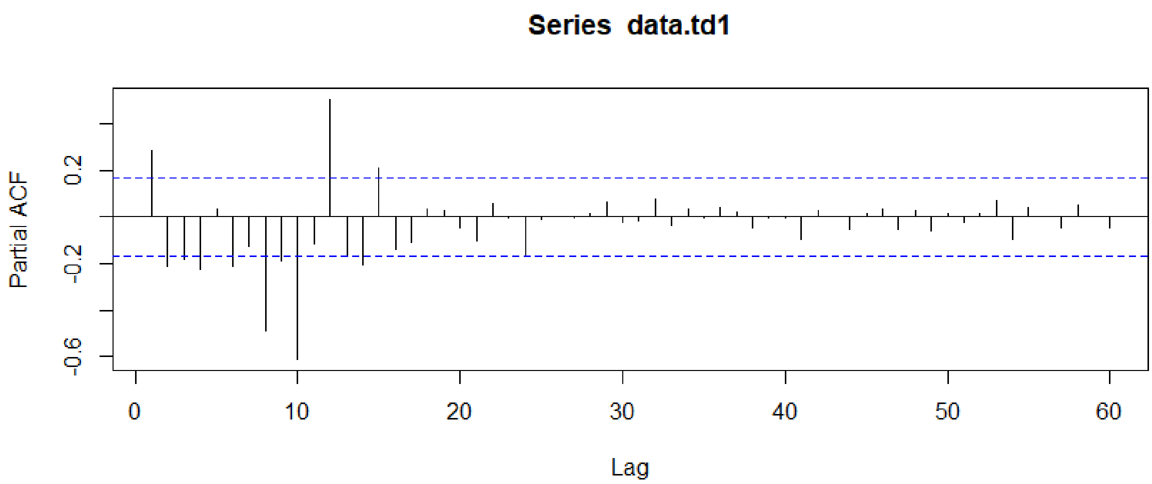




ACF下降緩慢，有趨勢;lag=12時，PACF下降比較慢，有季節性，不平穩。

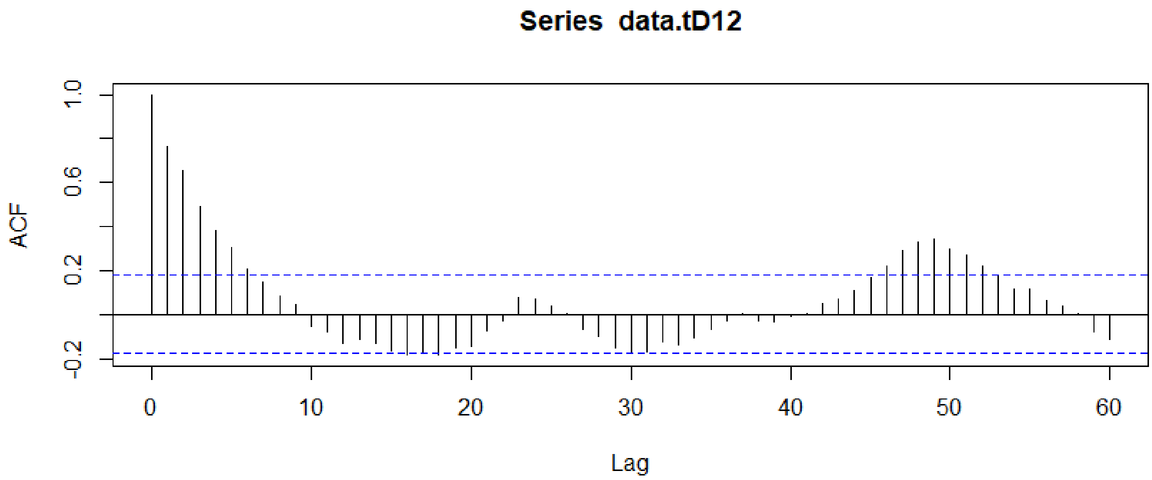
2.一階差分後的時數圖、ACF與PACF圖。判別資料平穩嗎?

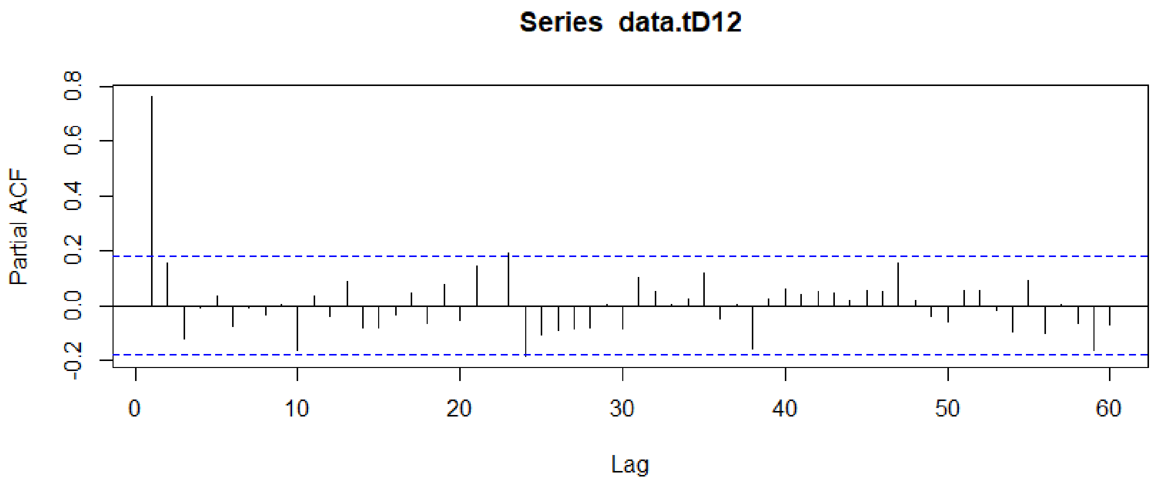




結論:Lag=12時，ACF下降較慢，有季節性，不平穩

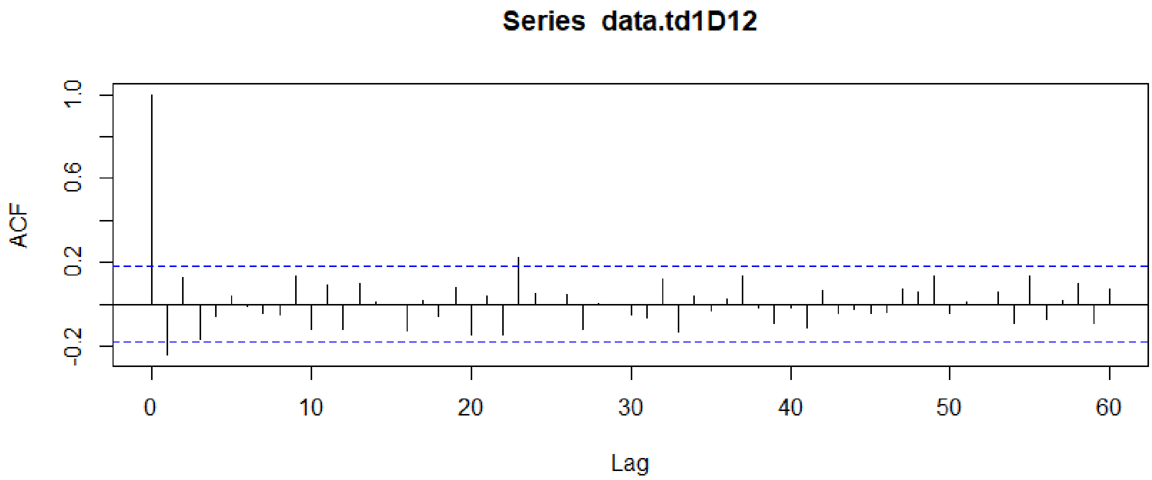
3.季節性差分後的時數圖、ACF與PACF圖。判別資料平穩嗎?

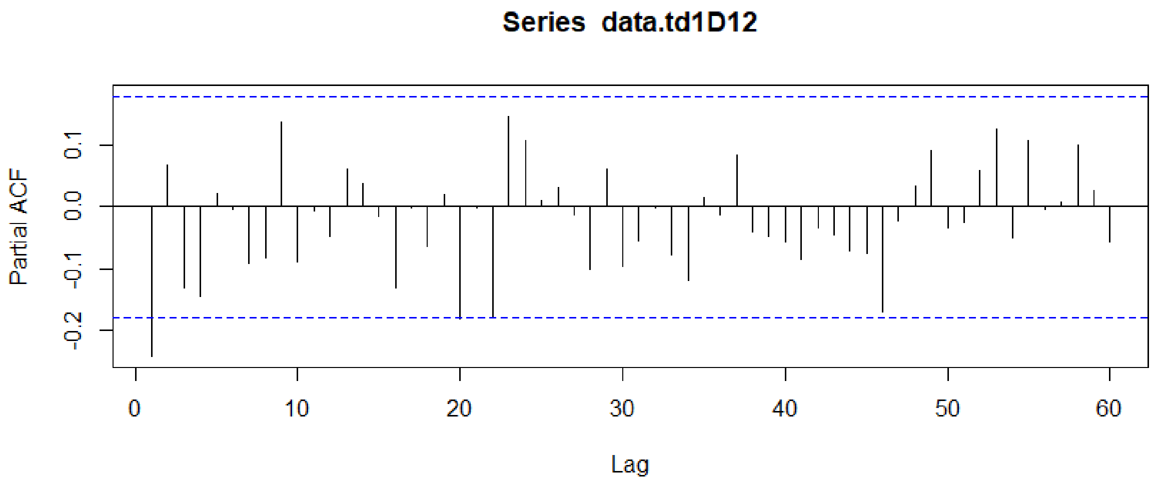




結論:季節性差分後的ACF下降緩慢波動範圍，明顯超過-0.2與0.2之間⇒不平穩

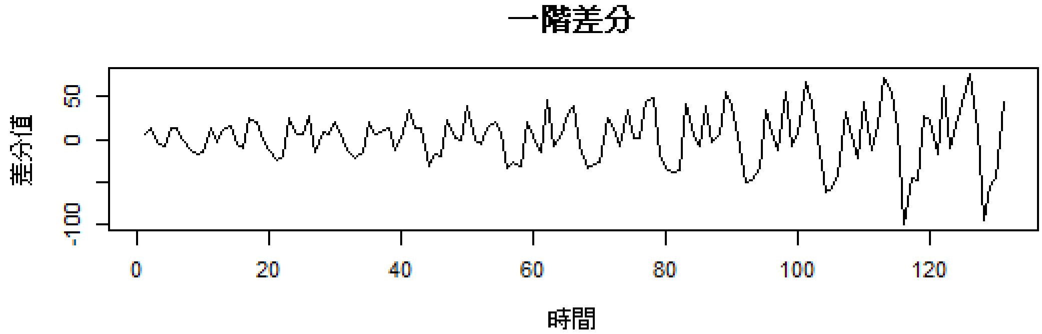
4..一階季節性差分後的時數圖、ACF與PACF圖。判別資料平穩嗎?

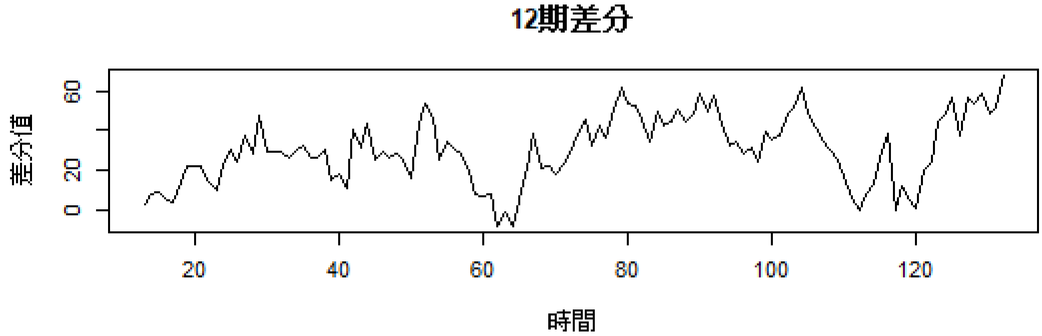


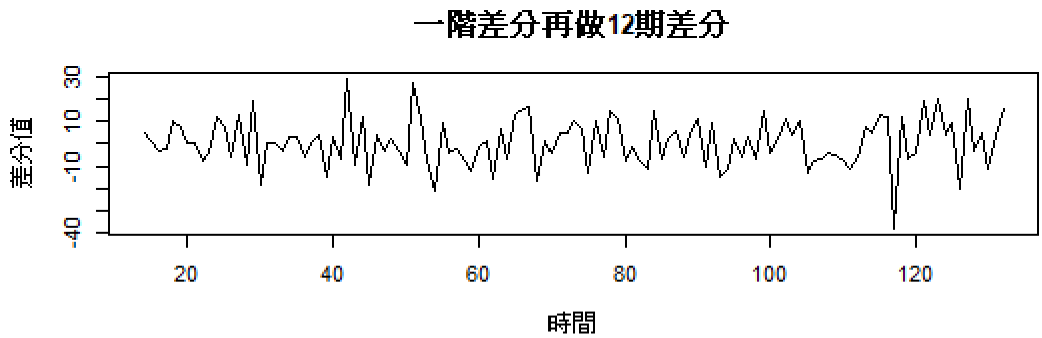


ACF在lag=0後切斷，PACF在lag=1後切斷，沒有緩慢下降情形，資料平穩。

各差分後時間數列圖:







5.由以上訊息,選d與D

由以上訊息發現只有資料data.td1D12平穩，因此選d=1、D=1

選擇模型$p,q\le 1, Q\le 2, d=1, D=1$

⇒ 最佳模型

資料來源:(1)

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) *Time Series Analysis, Forecasting and Control.* Third Edition. Holden-Day. Series G.

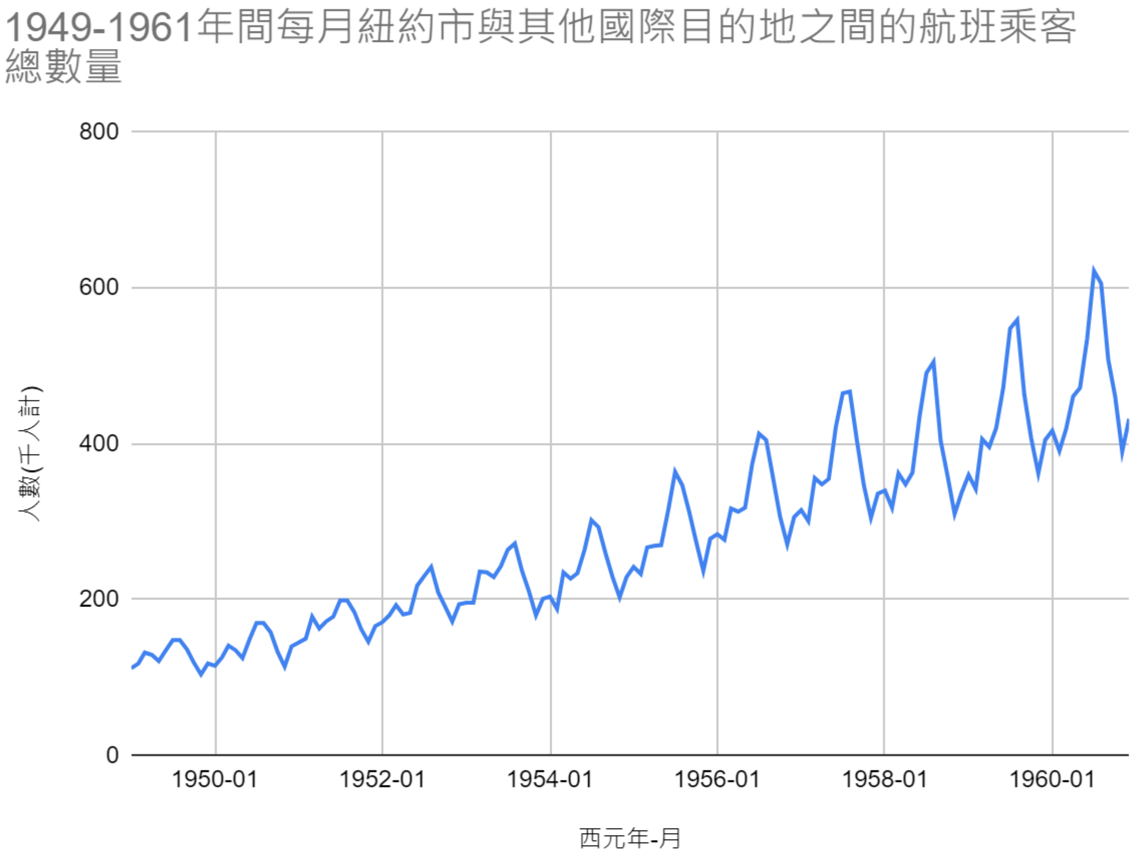
-The classic Box & Jenkins airline data.

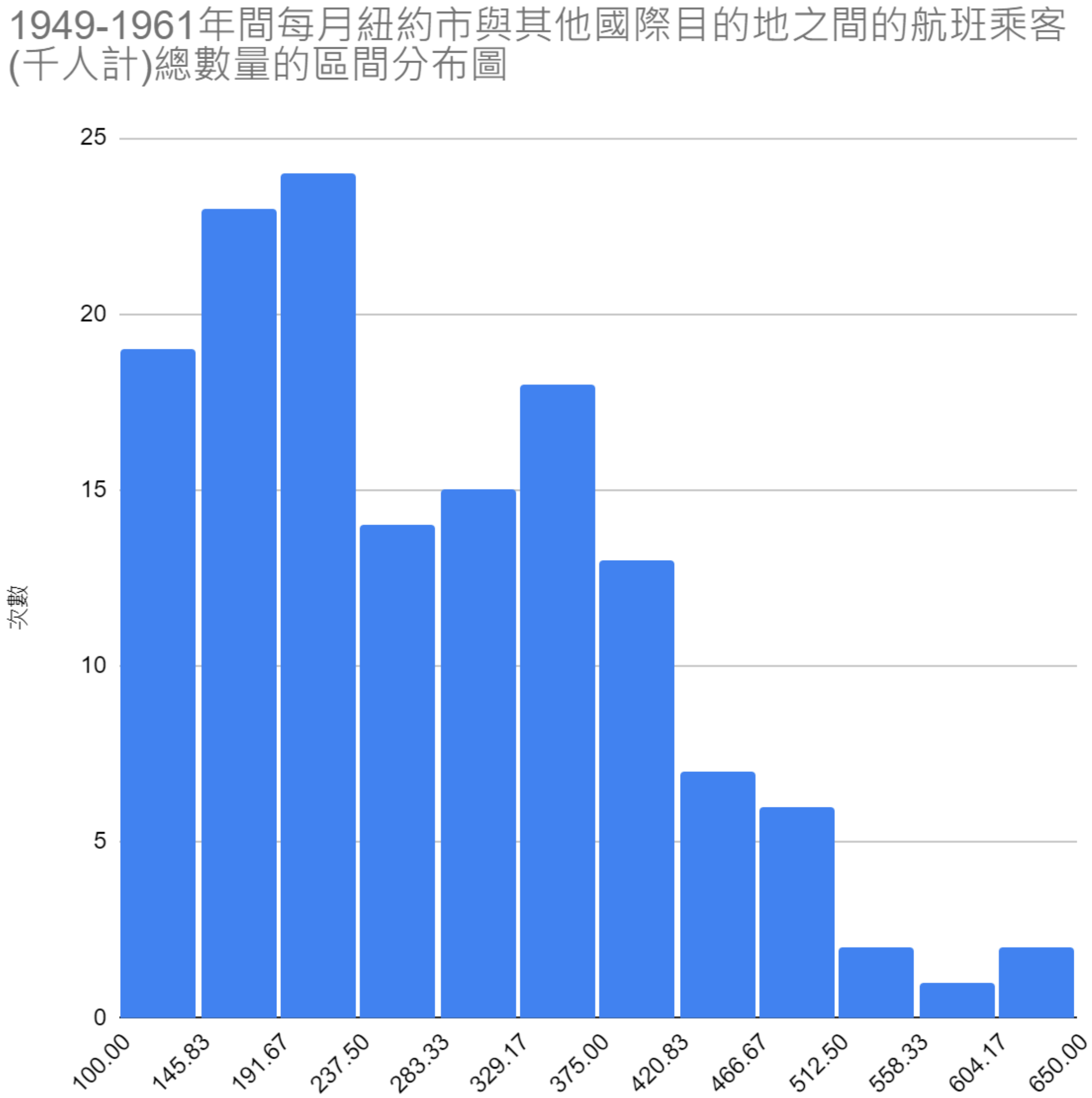
時間範圍:數據涵蓋了12年，從1949年1月到1960年12月。
頻率:按月記錄。
測量內容:每月紐約市與其他國際目的地之間的航班乘客總數量，以千人為單位。
地區:數據主要代表了紐約市與其他國際目的地之間的航班。。
季節性:數據顯示出明顯的季節性變化，乘客數量通常在夏季達到高峰，在冬季下降。
趨勢:有上升趨勢。

	Model	AIC	BIC
1	data.fit001	905.4189	910.9772
2	data.fit002	906.8886	915.2260
3	data.fit010	900.6852	906.2435
4	data.fit011	901.7211	910.0584
5	data.fit012	901.7468	912.8633
6	data.fit100	899.9021	905.4604
7	data.fit101	901.0524	909.3898
8	data.fit102	900.8583	911.9748
9	data.fit110	900.9716	909.3090
10	data.fit111	902.2214	913.3379
11	data.fit112	902.2385	916.1341

model00是最適模型，AIC、BIC最小
預測資料

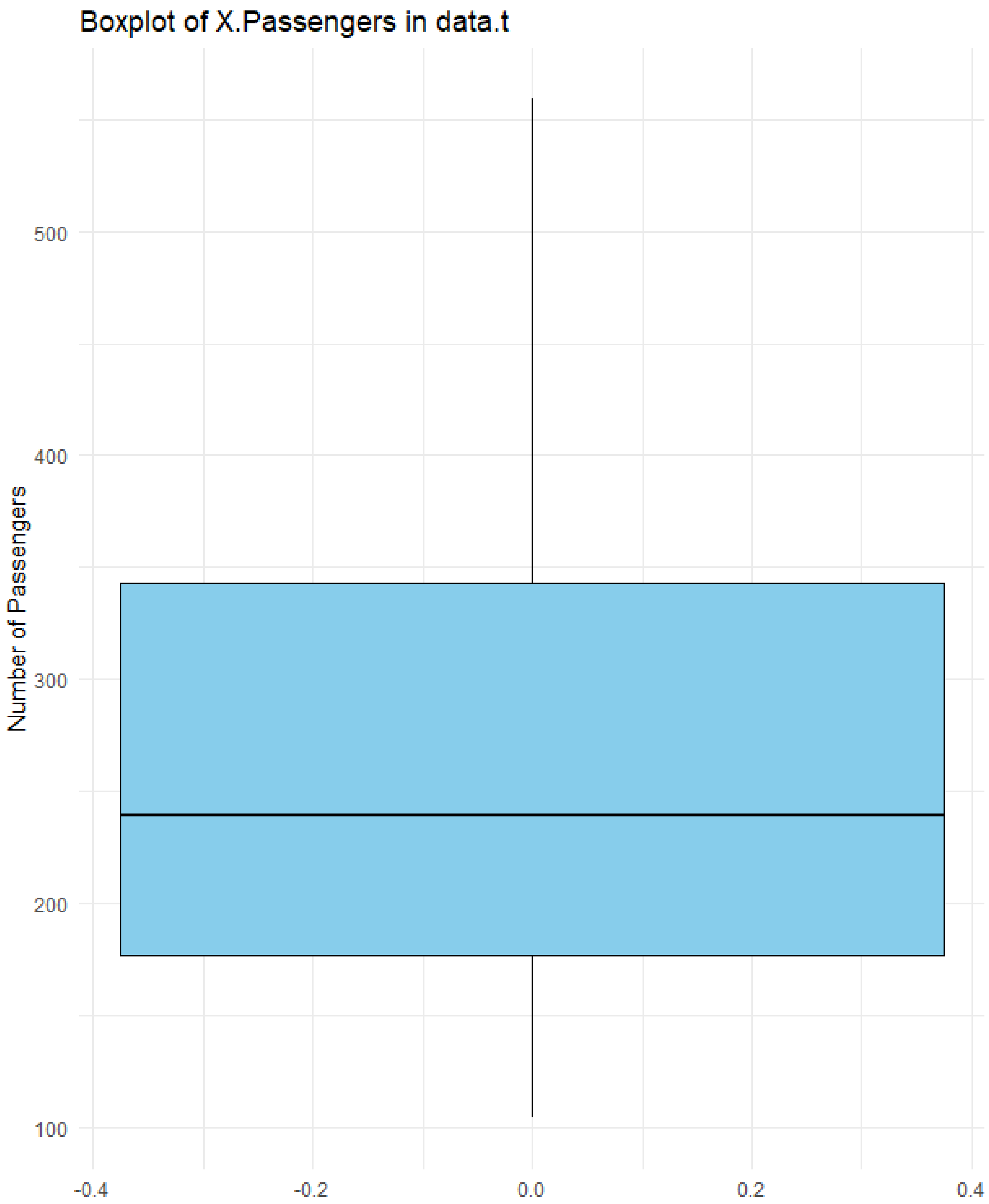
	Month	X.Passengers
133	1960-01	417
134	1960-02	391
135	1960-03	419
136	1960-04	461
137	1960-05	472
138	1960-06	535
139	1960-07	622
140	1960-08	606
141	1960-09	508
142	1960-10	461
143	1960-11	390
144	1960-12	432


1949-1961年間每月紐約市與其他國際目的地之間的航班乘客總數量
800
600
400
200
0
人數(千人計)
1950-01
1952-01
1954-01
1956-01
1958-01
1960-01
西元年-月


1949-1961年間每月紐約市與其他國際目的地之間的航班乘客(千人計)總數量的區間分布圖
25
20
15
10
5
0
次數
100.00
145.83
191.67
237.50
283.33
329.17
375.00
420.83
466.67
512.50
558.33
604.17
650.00

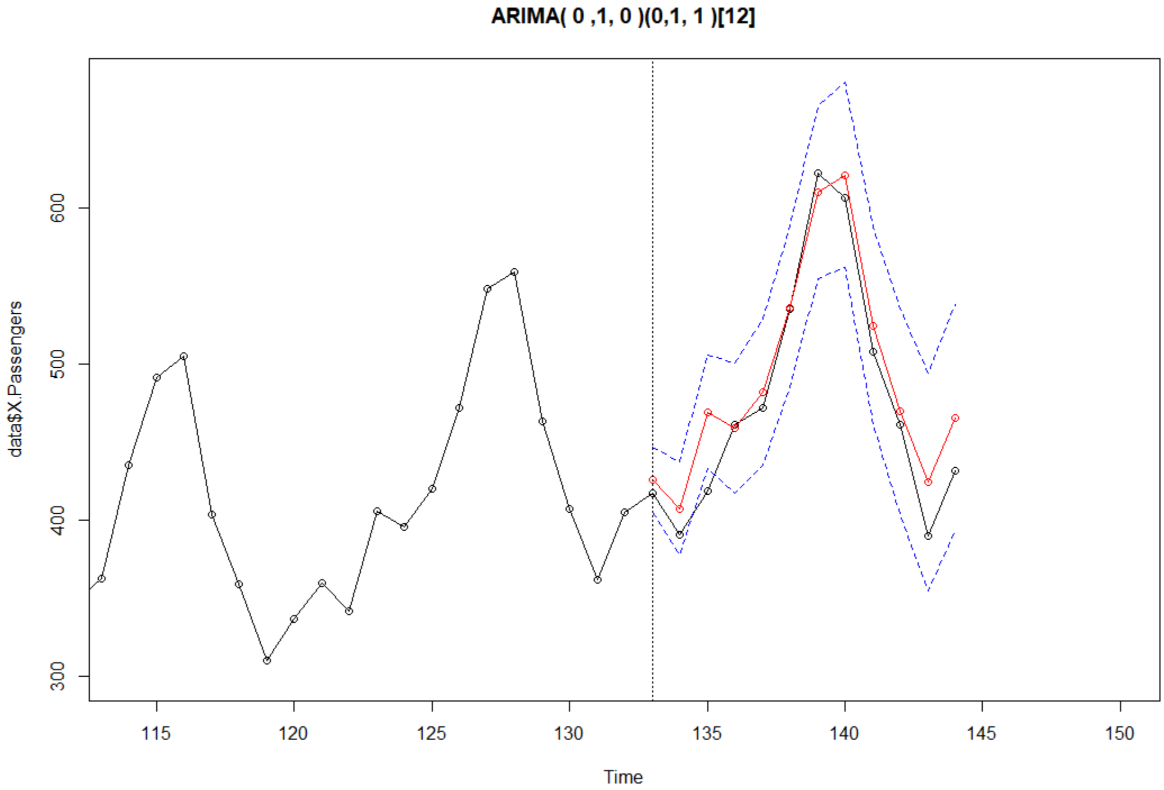
.

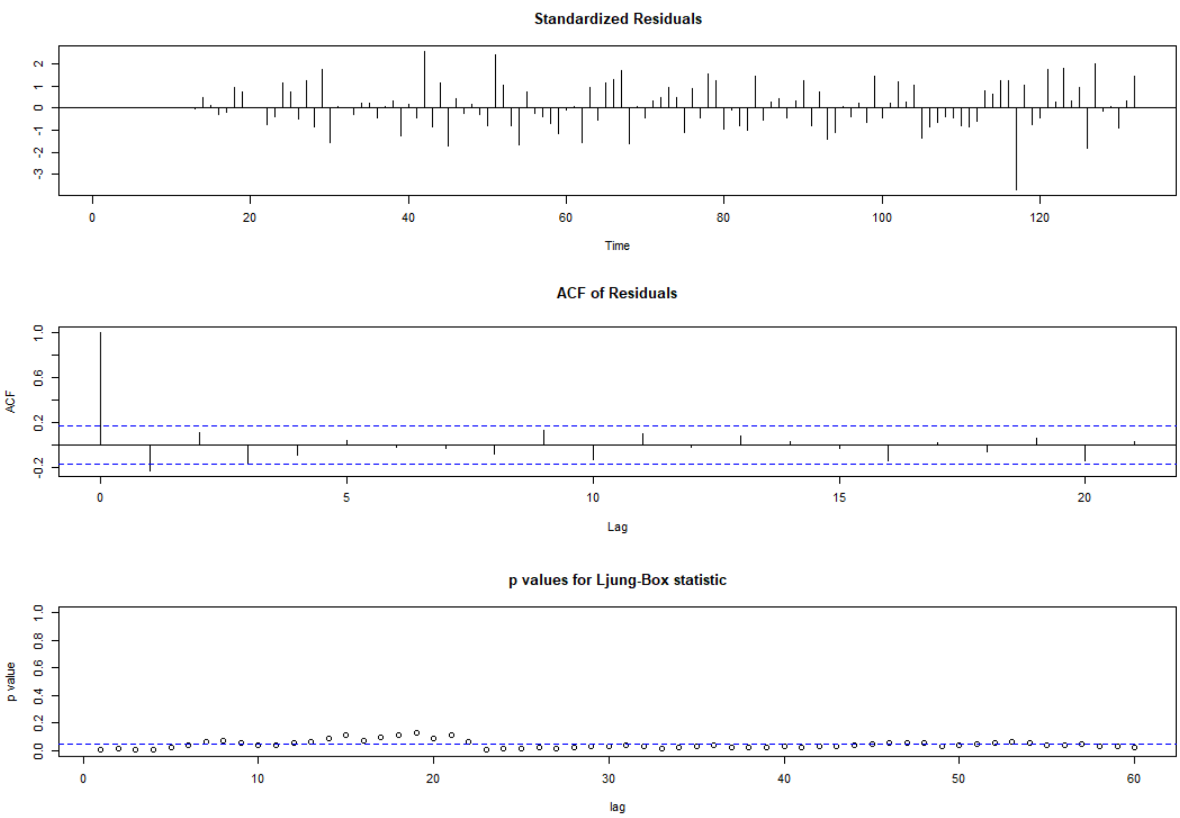
千人計乘客數量箱型圖



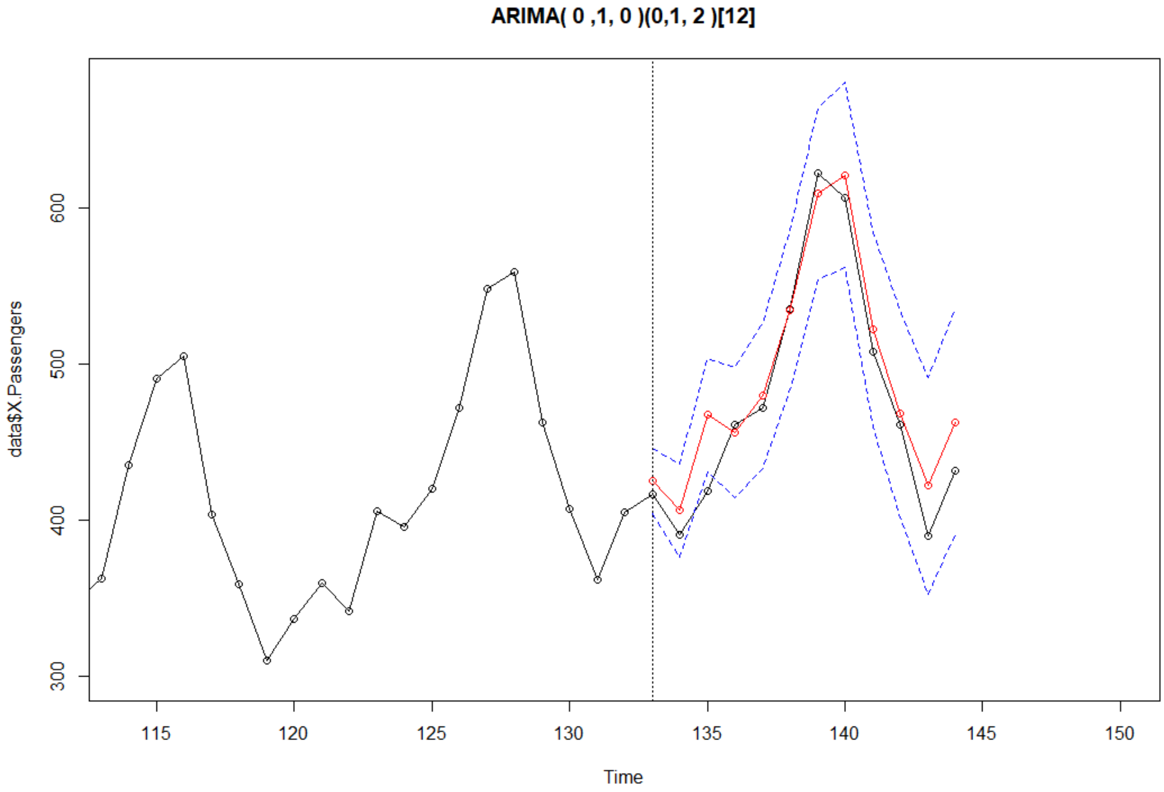
殘差與預測

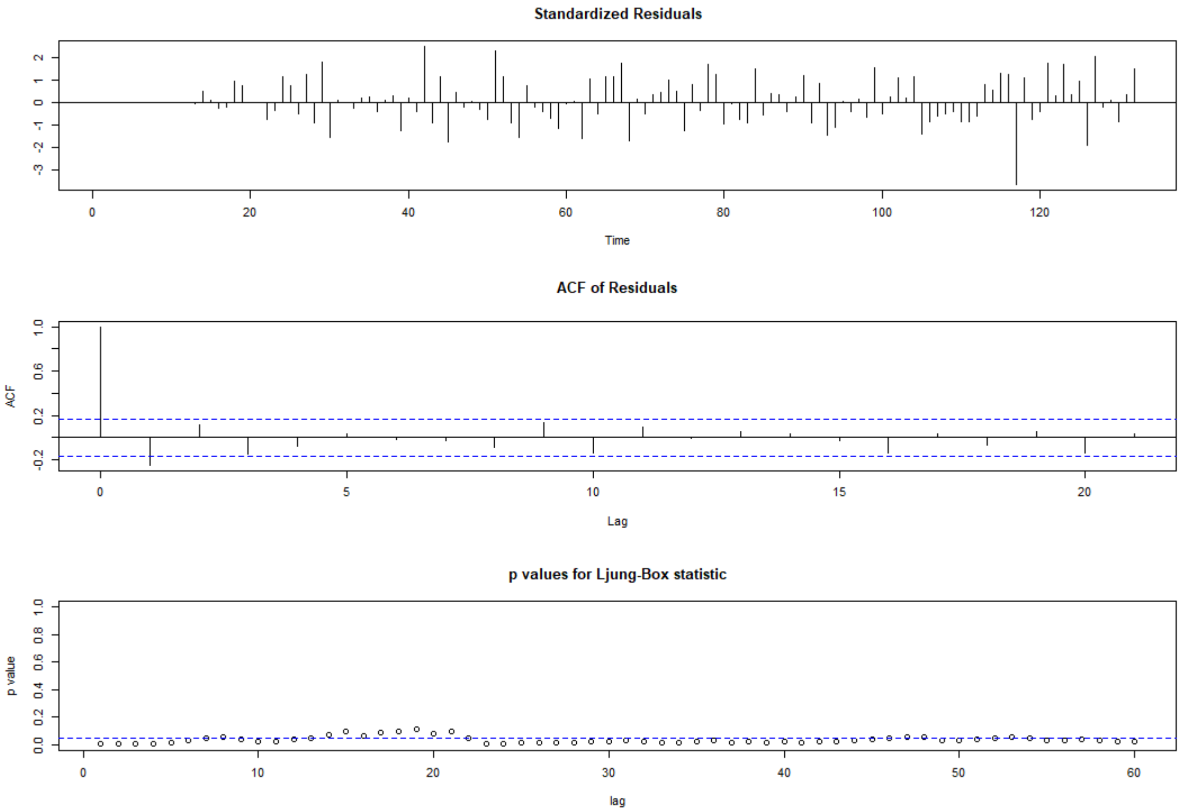
模型1


ARIMA(0 ,1, 0)(0,1, 1)[12]
data$X.Passengers
Time
300
400
500
600
115
120
125
130
135
140
145
150

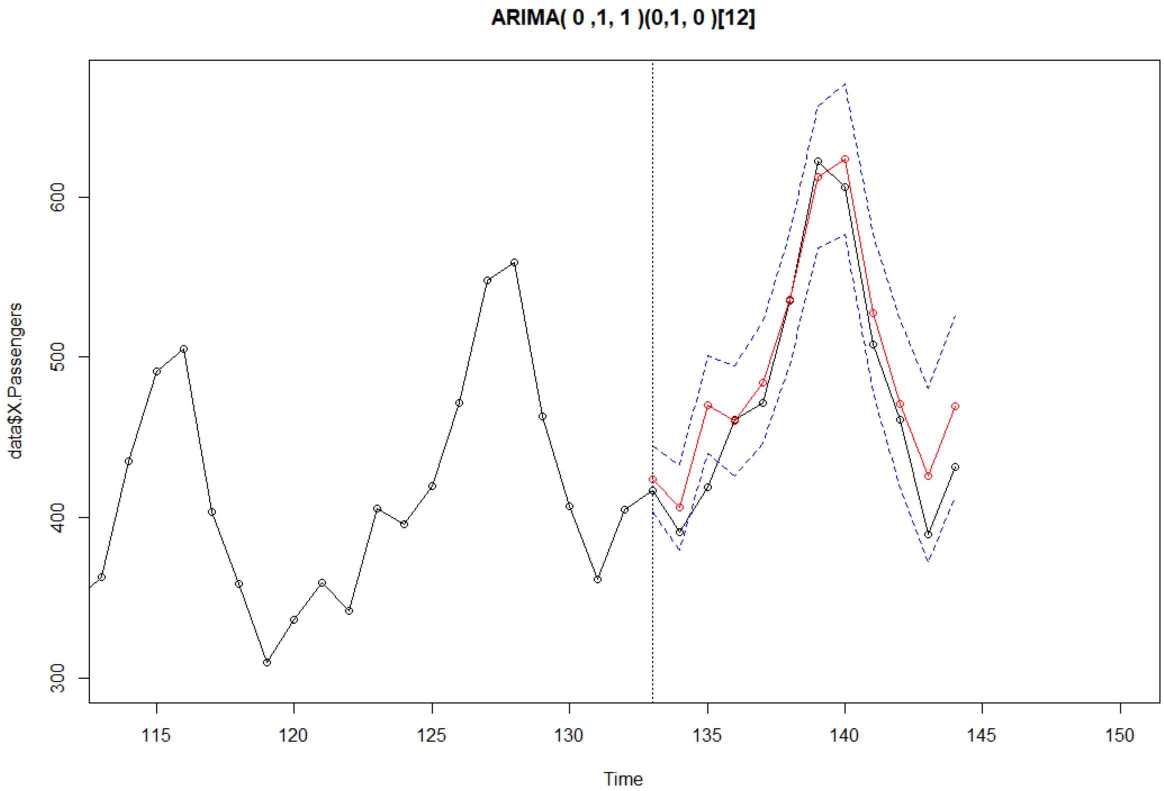

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

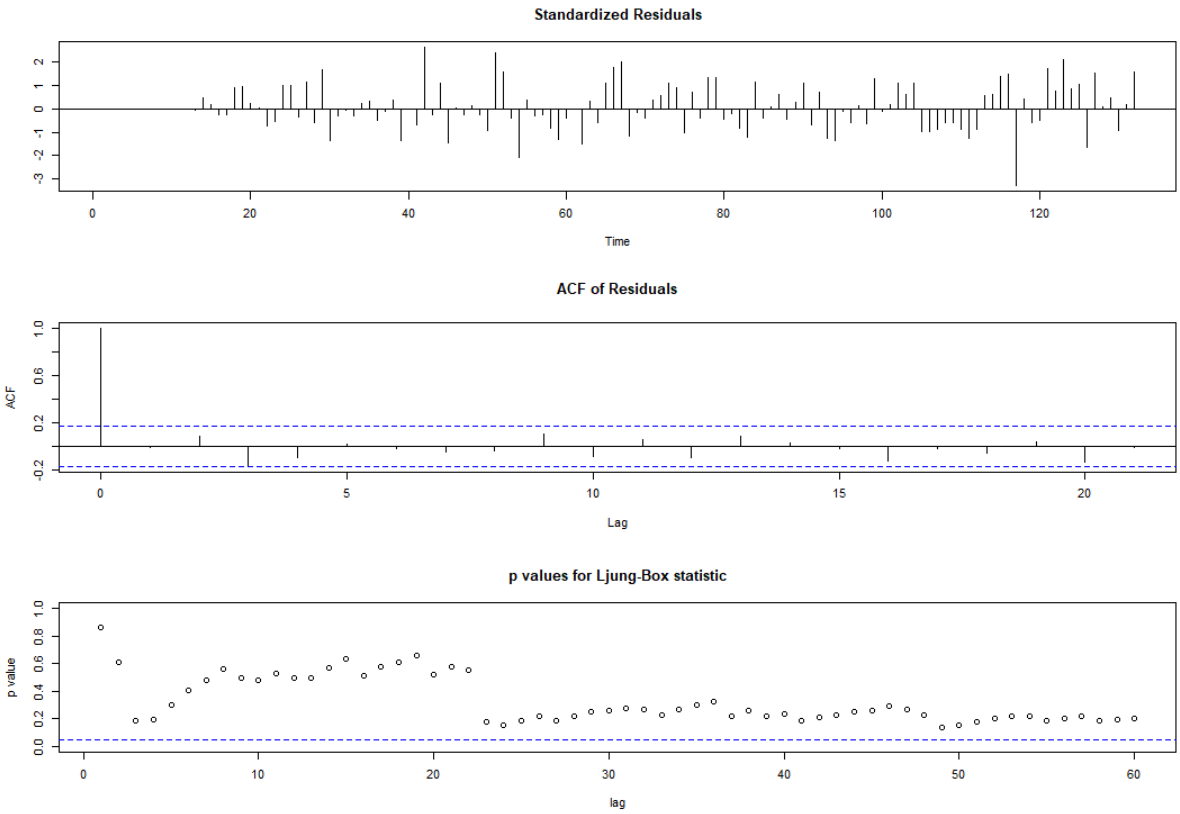
模型2


ARIMA(0 ,1, 0)(0,1, 2)[12]
data$X.Passengers
Time

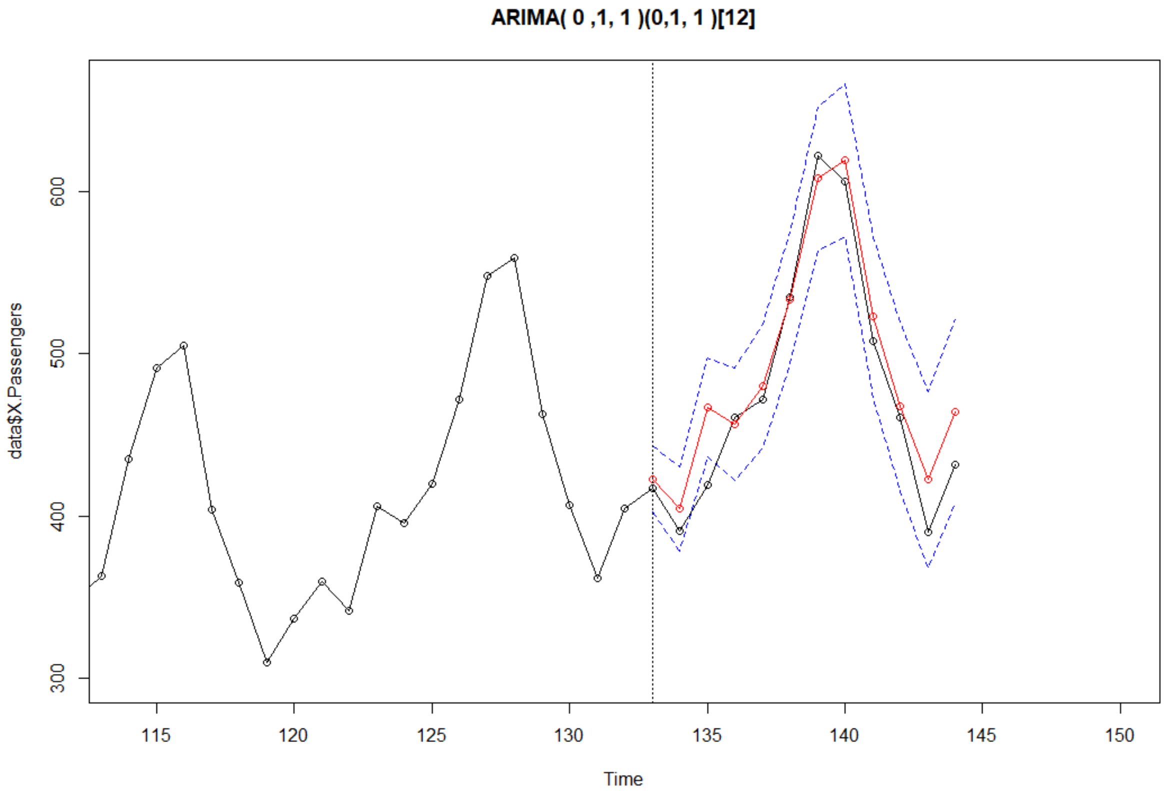

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

模型3

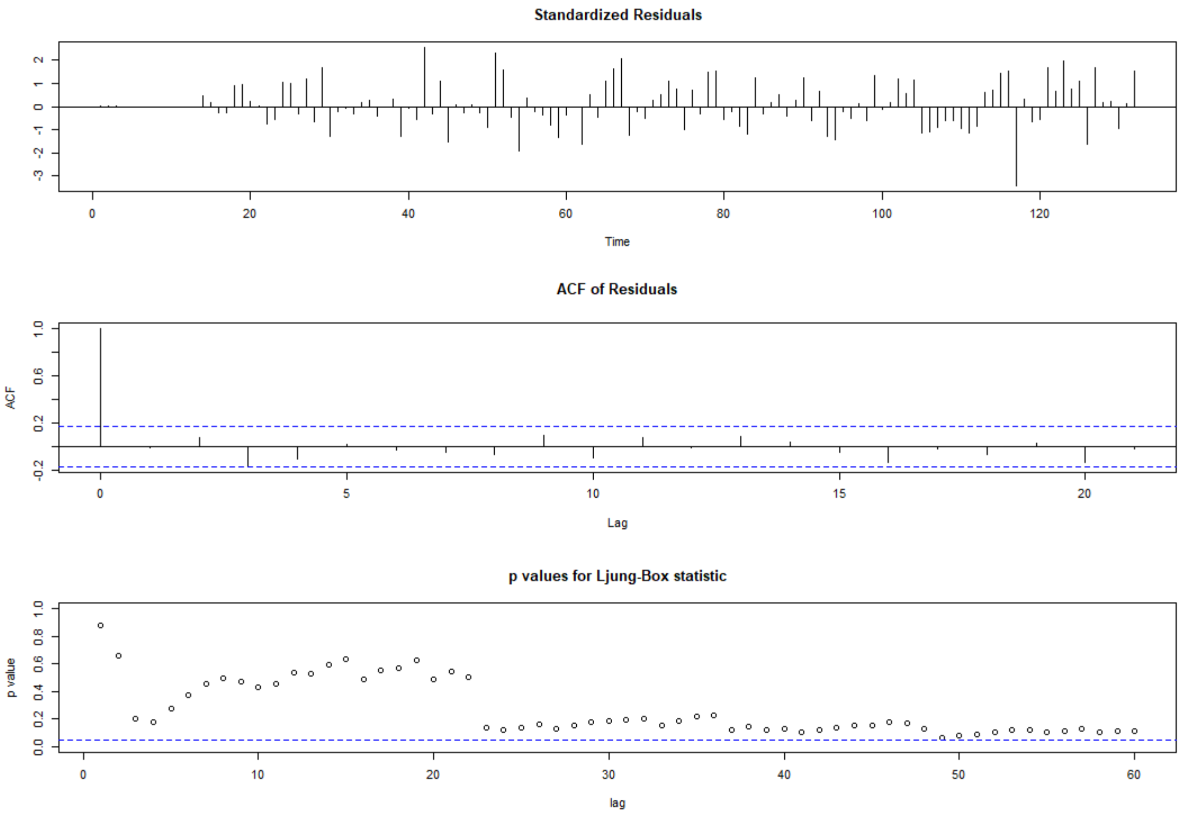

ARIMA(0 ,1, 1)(0,1, 0)[12]
data$X.Passengers
300
400
500
600
115
120
125
130
135
140
145
150
Time


Standardized Residuals
-3
-2
-1
0
1
2
0
20
40
60
80
100
120
Time
ACF of Residuals
ACF
-0.2
0.2
0.6
1.0
0
5
10
15
20
Lag
p values for Ljung-Box statistic
p value
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
60
lag

模型4

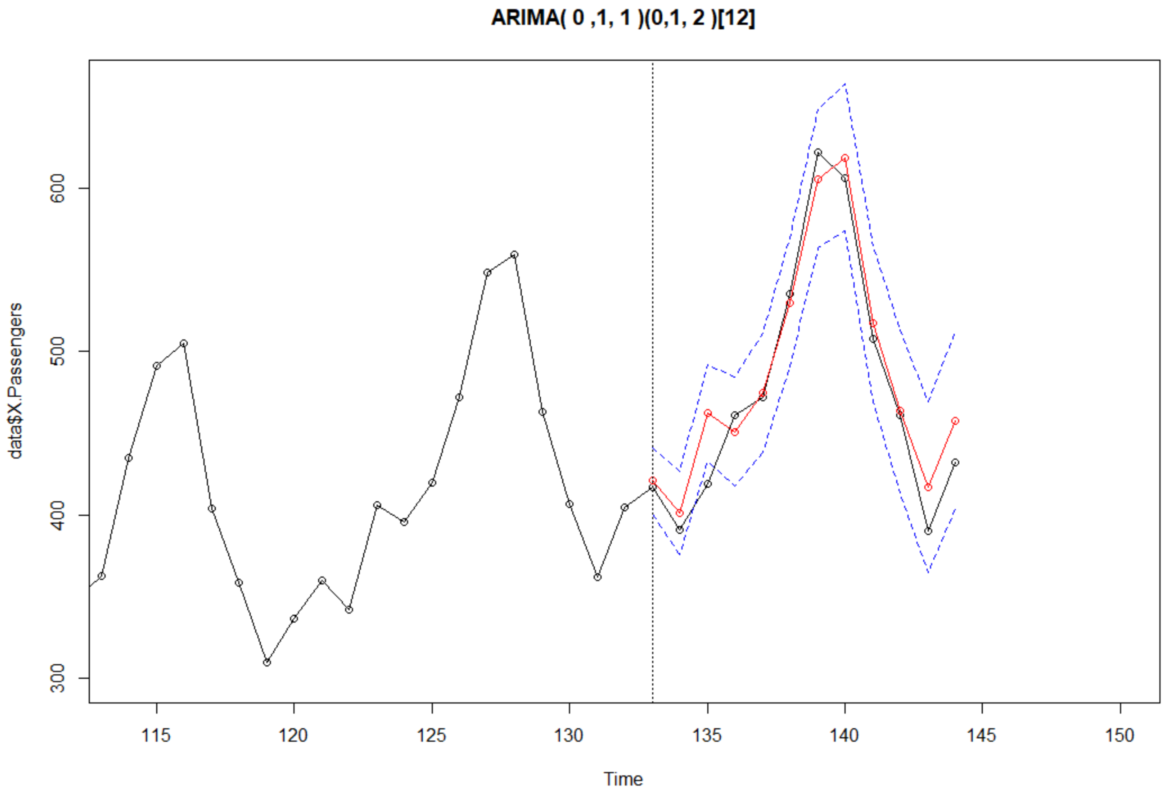


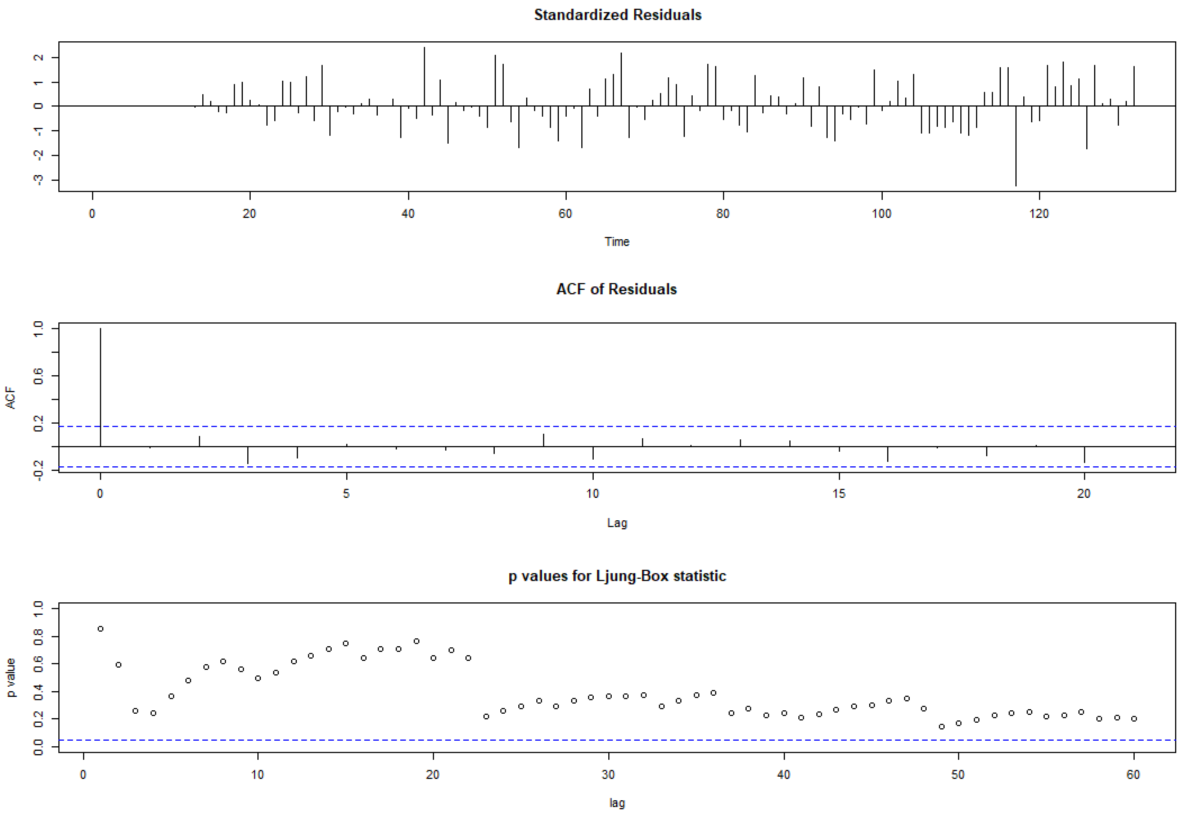
ARIMA(0 ,1, 1)(0,1, 1)[12]
data$X.Passengers
Time



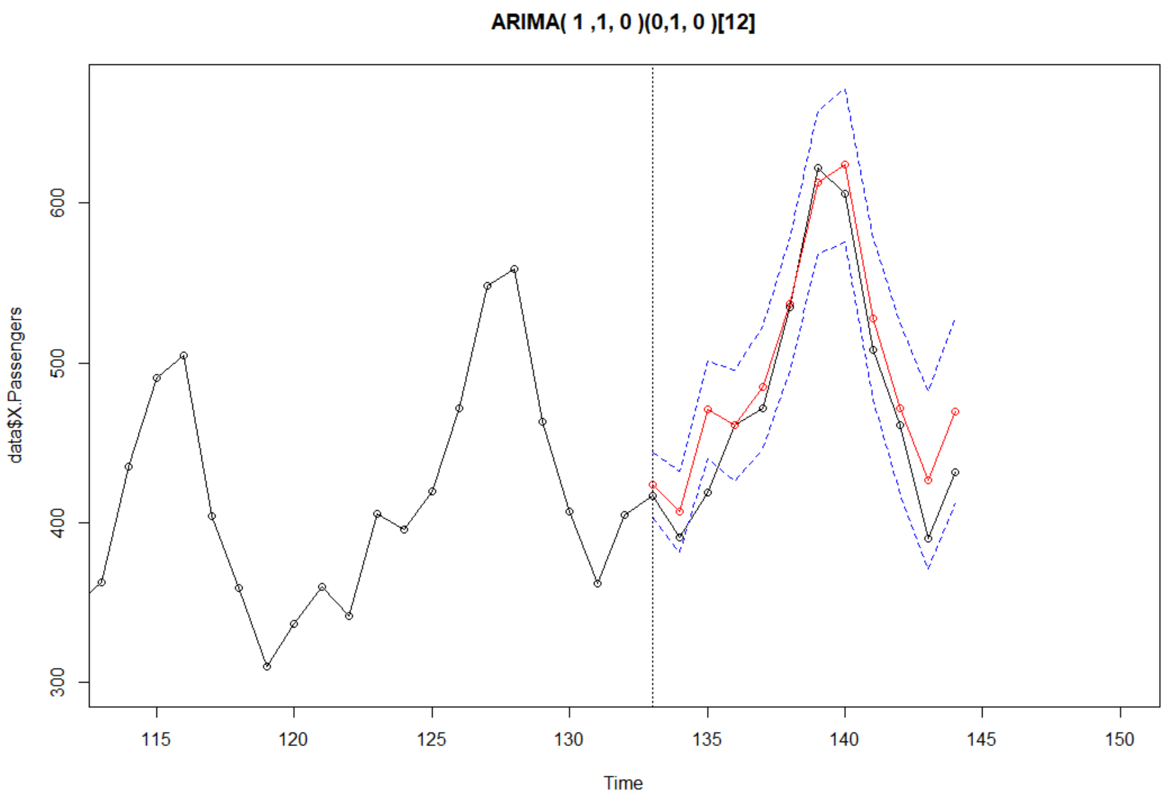
Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

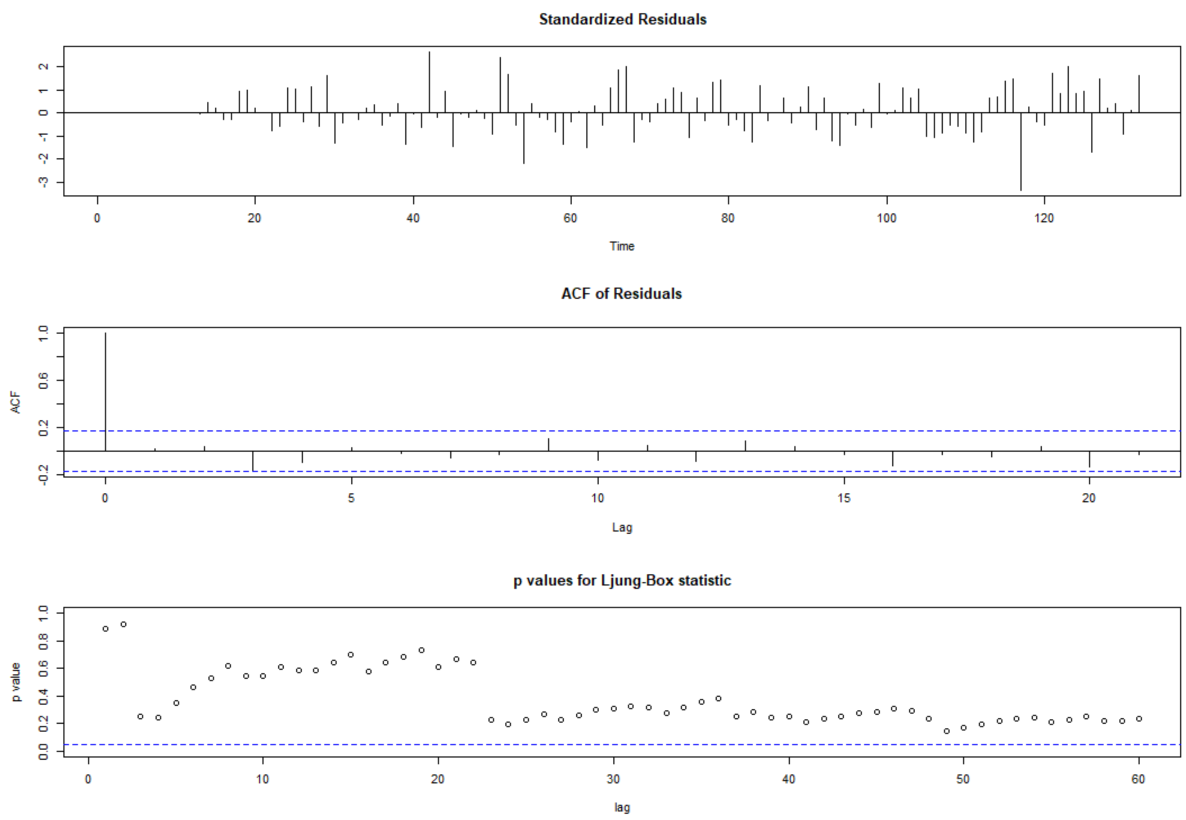
模型5


ARIMA(0 ,1, 1)(0,1, 2)[12]
data$X.Passengers
300
400
500
600
115
120
125
130
135
140
145
150
Time

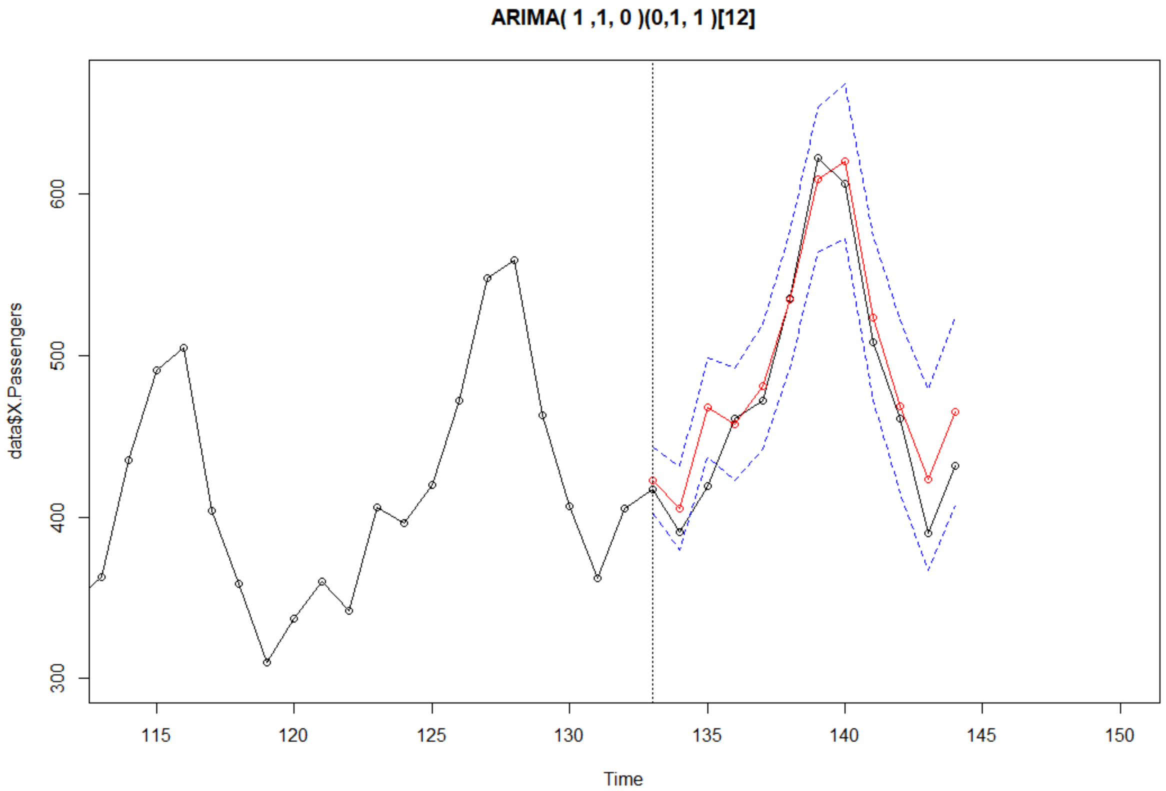

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

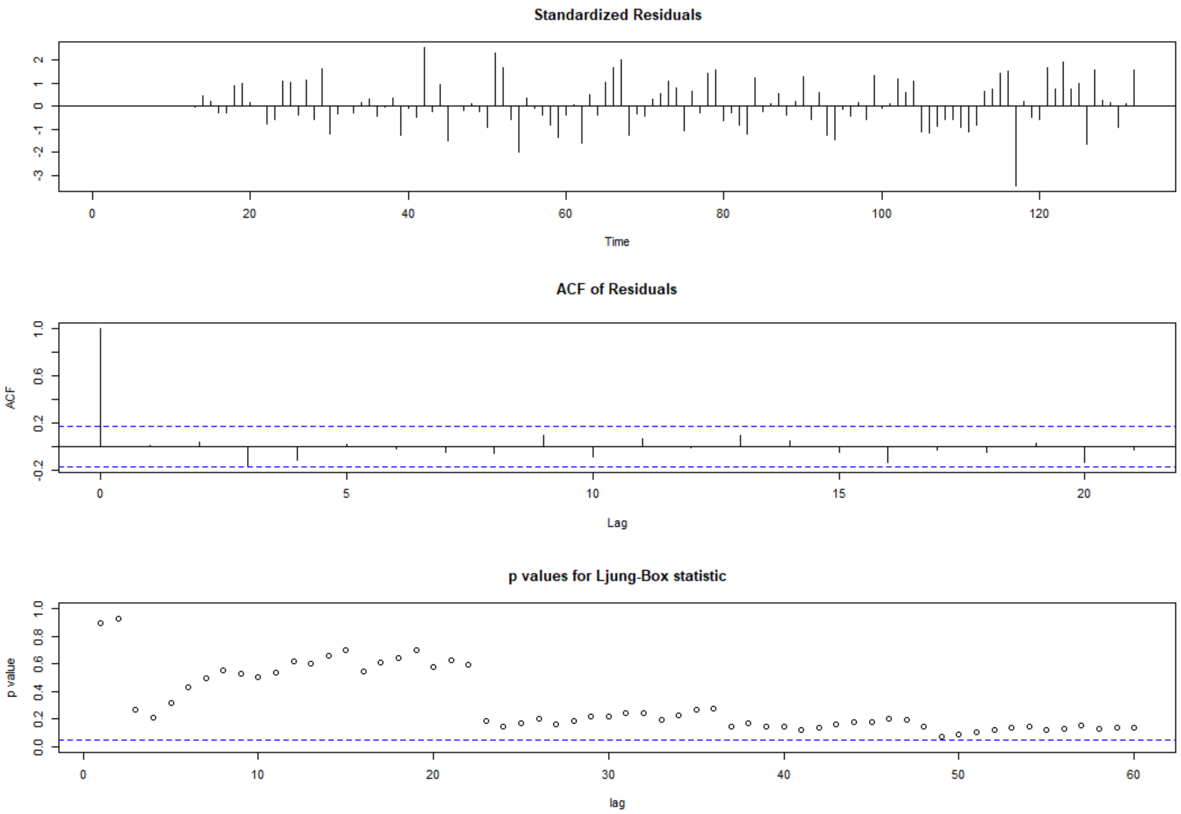
模型6


ARIMA(1 ,1, 0)(0,1, 0)[12]
data$X.Passengers
Time

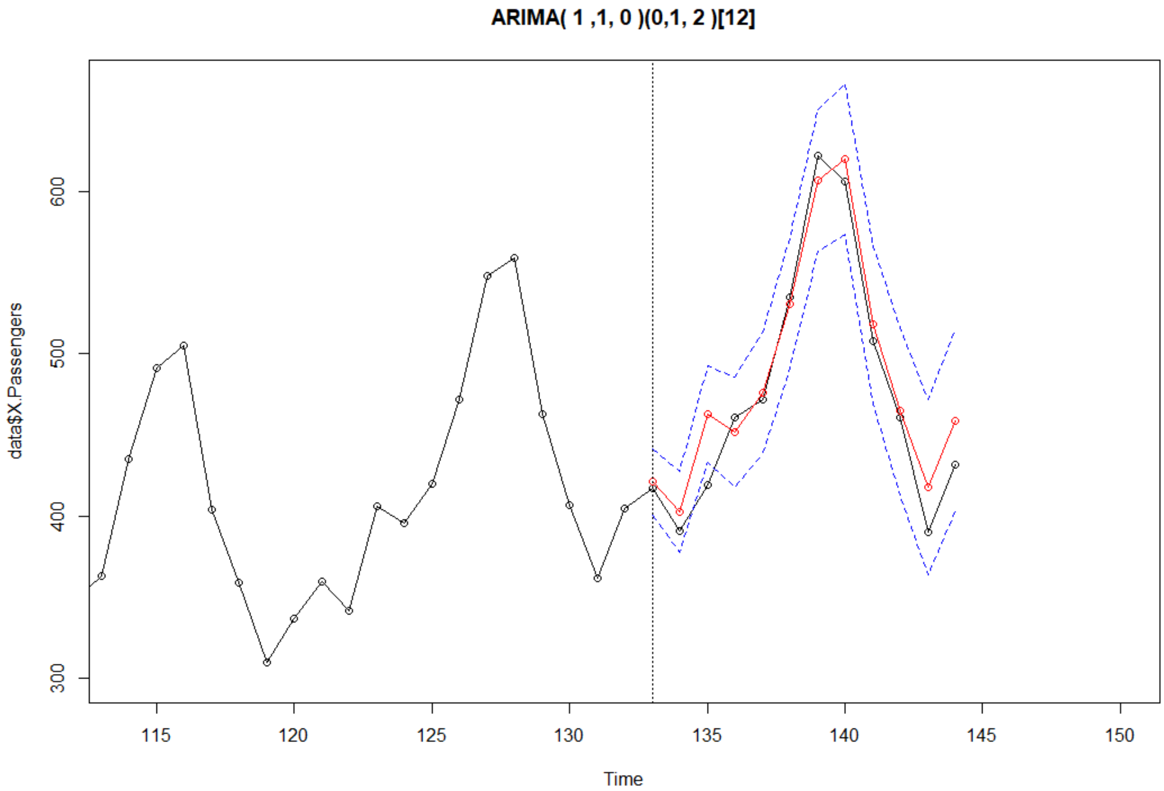

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

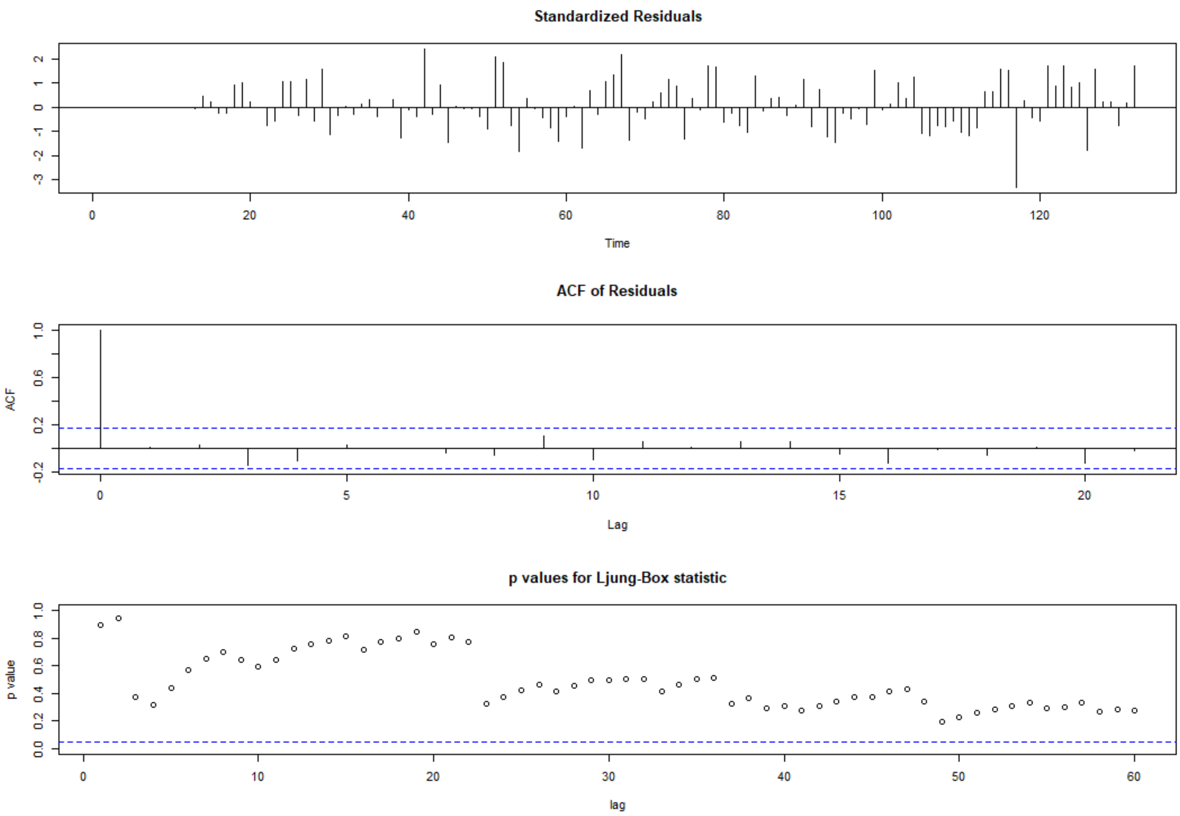
模型7


ARIMA(1 ,1, 0)(0,1, 1)[12]
data$X.Passengers
Time

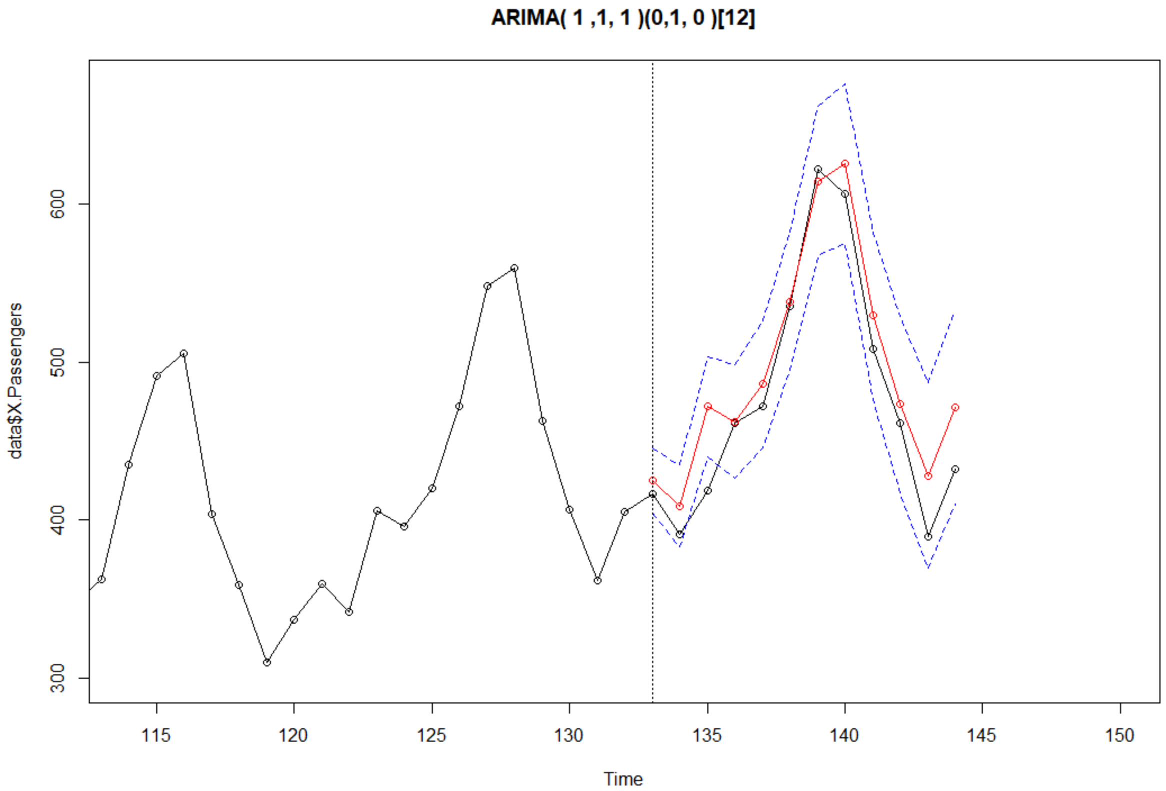

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

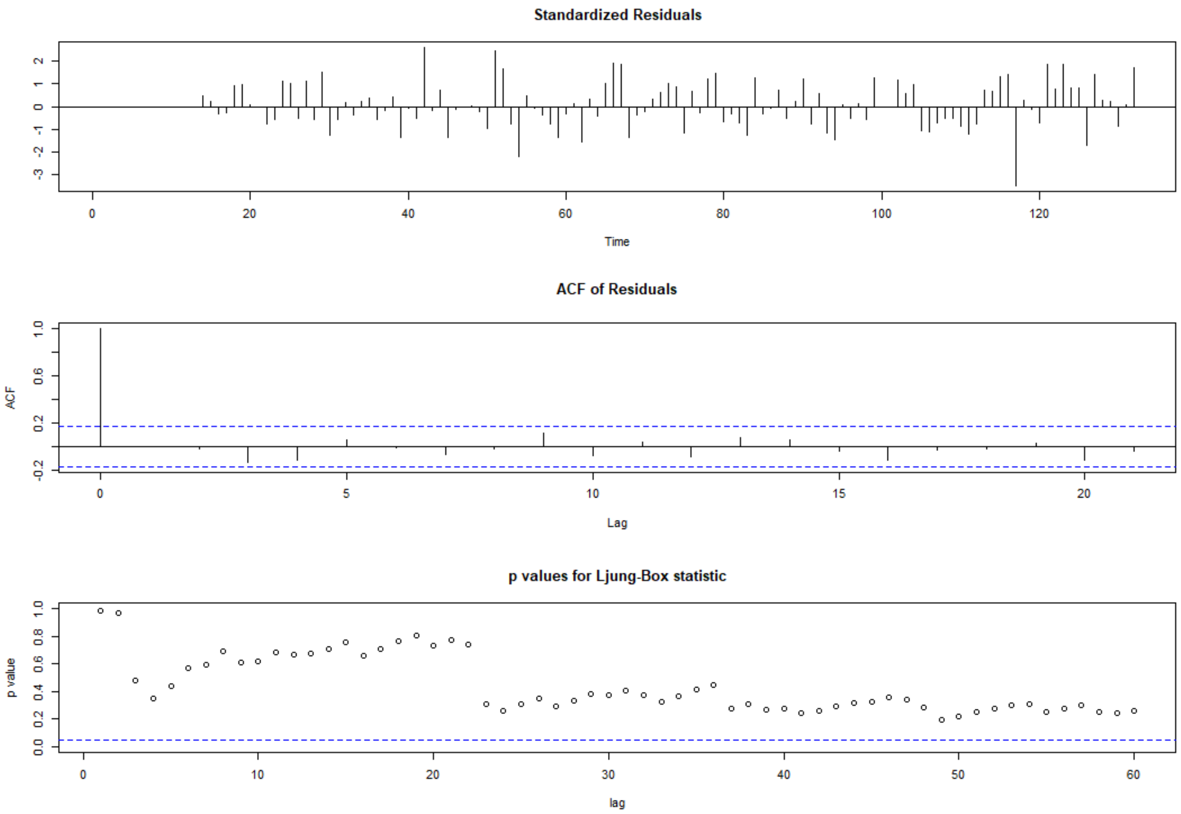
模型8


ARIMA(1 ,1, 0)(0,1, 2)[12]
data$X.Passengers
Time

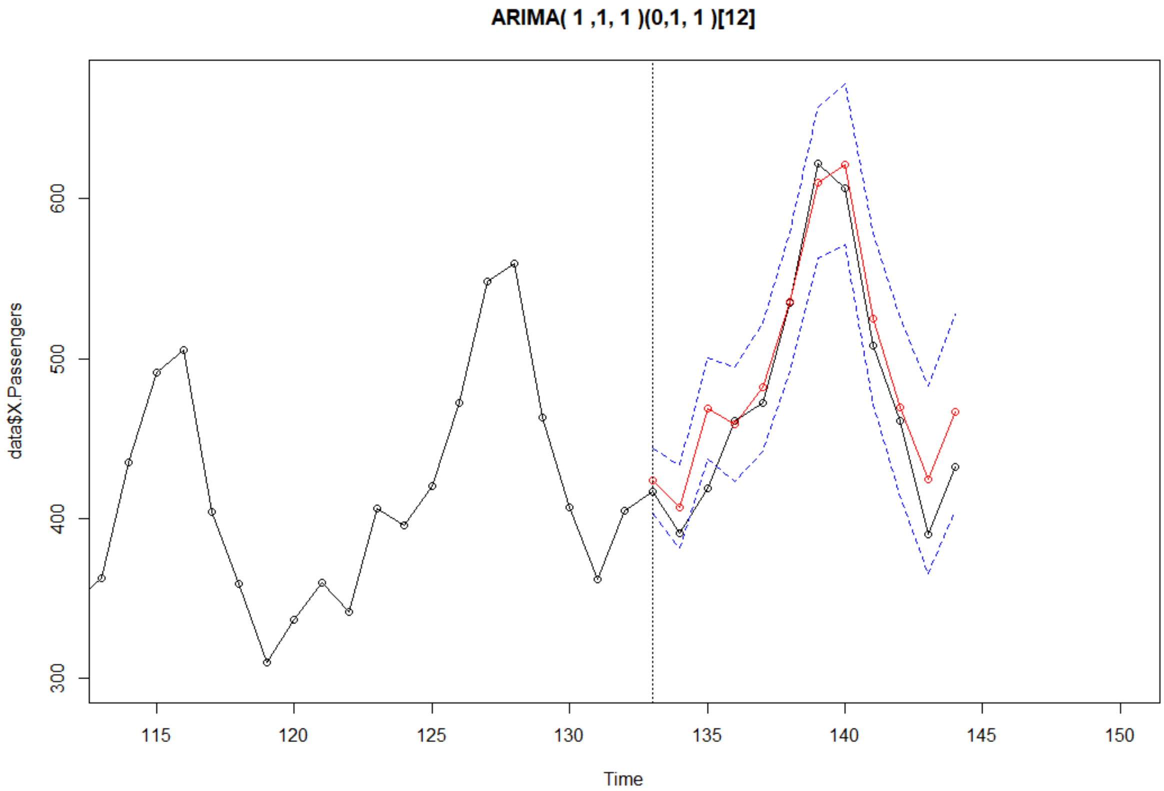

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

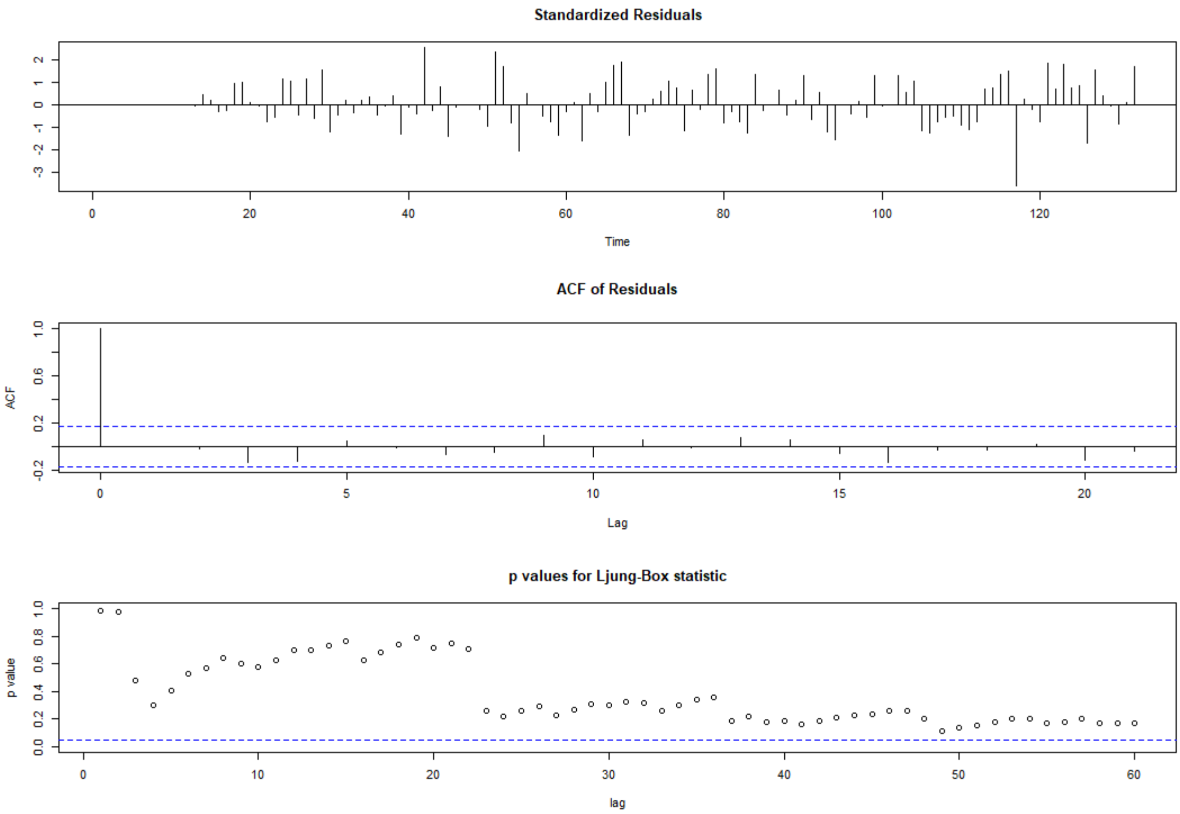
模型9


ARIMA(1 ,1, 1)(0,1, 0)[12]
data$X.Passengers
Time

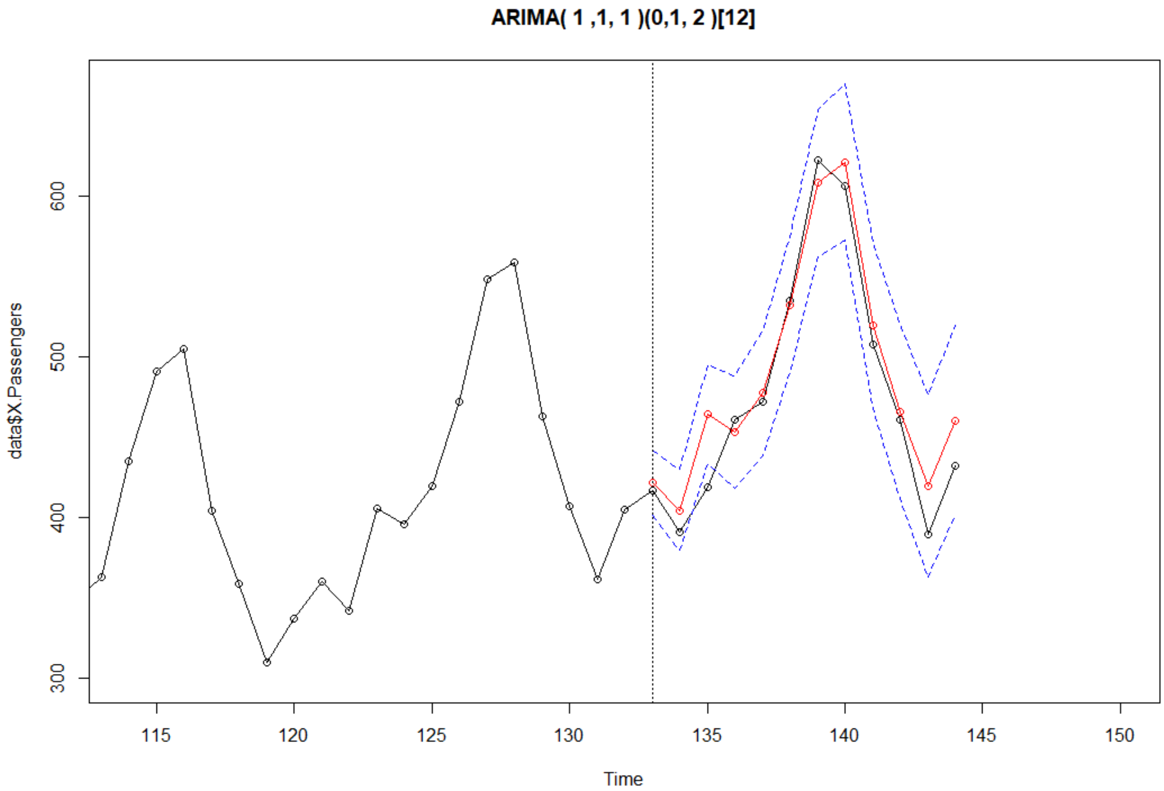

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

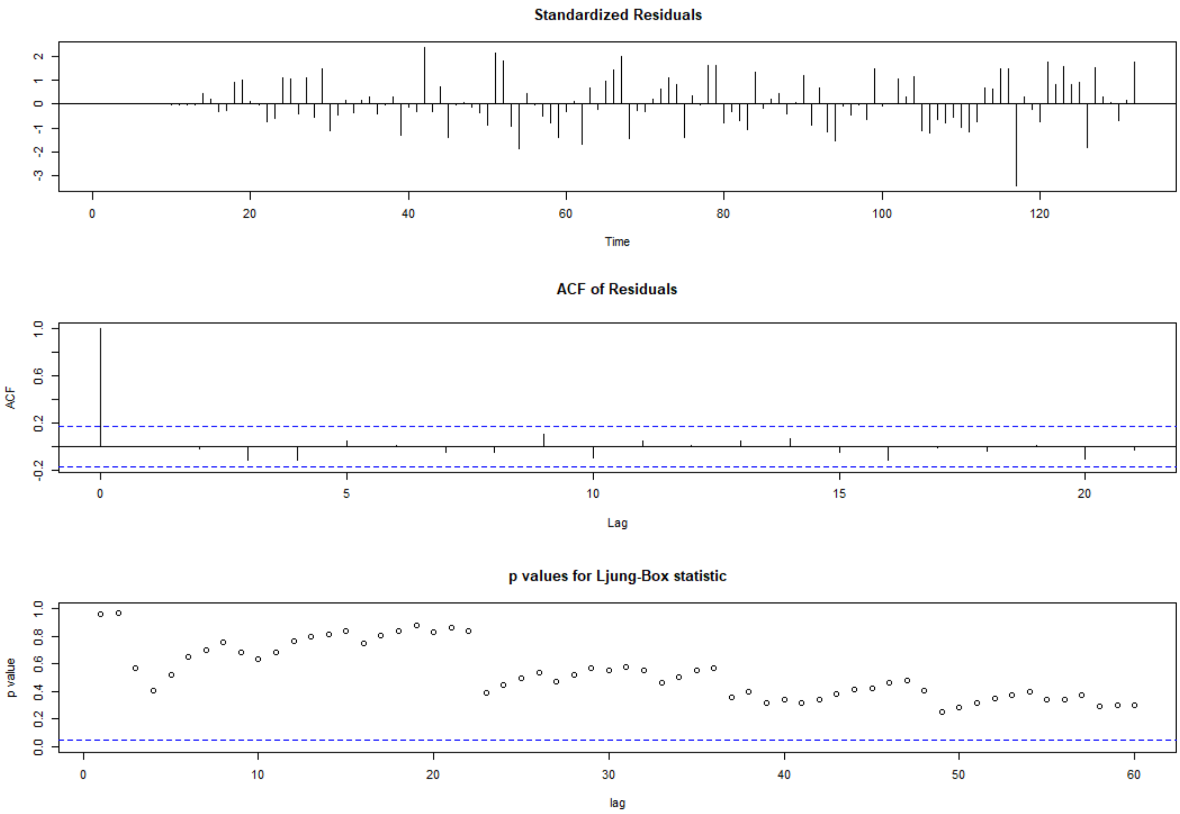
模型10


ARIMA(1 ,1, 1)(0,1, 1)[12]
data$X.Passengers
300
400
500
600
115
120
125
130
135
140
145
150
Time


Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung-Box statistic
p value
lag

模型11





EXCEL連結

https://docs.google.com/spreadsheets/d/1vnZssn69WA4-W8RlS8jAxj93jUtrjqPJ12P_1s6niMk/edit?usp=sharing

```
############ts_Final_report#####

file_path <- "C:/Users/User/Desktop/學/時間數列/AirPassengers.csv"

###讀CSV文件並檢視###
data <- read.csv(file_path)
head(data)
str(data)
summary(data)
##拆，以1949-1959預測1960年##
n=length(data$X.Passengers)
data.t <- data[1:(n-12), ]
data.p <- data[(n-12+1):n, ]
head(data.t)
tail(data.t)
data.p

########以data.t分析##############
install.packages("ggplot2")
library(ggplot2)
##########直方(90-110,110-130......)##################
ggplot(data.t, aes(x = X.Passengers)) +
  geom_histogram(binwidth = 20, fill = "skyblue", color = "black") +
  labs(title = "Histogram of X.Passengers in data.t",
       x = "Number of Passengers",
       y = "Frequency") +
  theme_minimal()

#############箱型######################
ggplot(data.t, aes(y = X.Passengers)) +
  geom_boxplot(fill = "skyblue", color = "black") +
  labs(title = "Boxplot of X.Passengers in data.t",
       y = "Number of Passengers") +
  theme_minimal()

###########ts圖#############
Time <- 1:nrow(data.t)
ggplot(data.t, aes(x = Time, y = X.Passengers)) +
```

```
  geom_line(color = "blue") +
  labs(title = "Time Series of X.Passengers in data.t",
       x = "Time",
       y = "Number of Passengers") +
  theme_minimal()
#######這也可以#########
ts.plot(data.t$X.Passengers)
############未差分APACF###################
par(mfrow=c(2,1))
acf(data.t$X.Passengers,60)
pacf(data.t$X.Passengers,60)
################差分######################
data.td1=diff(data.t$X.Passengers)
data.tD12=diff(data.t$X.Passengers,12)
data.td1D12=diff(data.td1,12)

################
# 加載需要的庫
library(ggplot2)

# 計算一階差分
data.td1 <- diff(data.t$X.Passengers)

# 計算12期差分
data.tD12 <- diff(data.t$X.Passengers, lag = 12)

# 計算一階差分再做12期差分
data.td1D12 <- diff(data.td1, lag = 12)

# 創建時間序列對象
ts_data.td1 <- ts(data.td1, start = start(data.t$X.Passengers), frequency =
frequency(data.t$X.Passengers))
ts_data.tD12 <- ts(data.tD12, start = start(data.t$X.Passengers) + c(0, 12), frequency
= frequency(data.t$X.Passengers))
ts_data.td1D12 <- ts(data.td1D12, start = start(data.t$X.Passengers) + c(0, 12 + 1),
frequency = frequency(data.t$X.Passengers))

# 繪製圖形
par(mfrow = c(3, 1)) # 設置圖形排列方式
```

```
# 繪製一階差分
plot(ts_data.td1, main = "一階差分", ylab = "差分值", xlab = "時間")

# 繪製12期差分
plot(ts_data.tD12, main = "12期差分", ylab = "差分值", xlab = "時間")

# 繪製一階差分再做12期差分
plot(ts_data.td1D12, main = "一階差分再做12期差分", ylab = "差分值", xlab = "時間")

############差分APACF######################
##########td1###############
par(mfrow=c(2,1))
acf(data.td1,60)
pacf(data.td1,60)
##########tD12##############
par(mfrow=c(2,1))
acf(data.tD12,60)
pacf(data.tD12,60)
##########td1D12############
par(mfrow=c(2,1))
acf(data.td1D12,60)
pacf(data.td1D12,60)

#########sarima傻瓜法######################
#p1############################################################

data.fit <- arima(data$X.Passengers[1:(length(data$X.Passengers)-12)], c(1, 1, 0),
seasonal = list(order = c(0, 1, 0), period = 12))
tsdiag(data.fit,60)
data.fit
names(data.fit)
data.fit$coef
data.fit$sigma2
data.fit$aic
tsdiag(data.fit,60)
AIC(data.fit)
BIC(data.fit)
data.pre <- predict(data.fit, n.ahead = 12)
```

```
names(data.pre)
U <- data.pre$pred + 1.96 * data.pre$se
L <- data.pre$pred - 1.96 * data.pre$se
plot.ts(data$X.Passengers, xlim = c(120, length(data$X.Passengers)), ylim = c(300, 700),type = "o")
lines(data.pre$pred, col = "red", type = "o")
lines(U, col = "blue", lty = "dashed")
lines(L, col = "blue", lty = "dashed")
abline(v = (length(data$X.Passengers) - 11), lty = "dotted")



##################################################
#####################巨集測試####################
##################################################
# Define a function to fit the model, generate predictions, plot the results, and store AIC/BIC values
fit_and_plot <- function(p, q, Q, data, model_name, aic_bic_table) {
  # Fit the ARIMA model
  data.fit <- arima(data$X.Passengers[1:(length(data$X.Passengers) - 12)],
                    order = c(p, 1, q),
                    seasonal = list(order = c(0, 1, Q), period = 12))

  # Assign the fitted model to a dynamically named variable
  assign(model_name, data.fit, envir = .GlobalEnv)

  # Print model details
  cat("\nModel:", model_name, "\n")
  print(data.fit)
  print(names(data.fit))
  print(data.fit$coef)
  print(data.fit$sigma2)
  print(data.fit$aic)
  print(AIC(data.fit))
  print(BIC(data.fit))

  # Append AIC and BIC to the table
  aic_bic_table <<- rbind(aic_bic_table, data.frame(Model = model_name, AIC = AIC(data.fit), BIC = BIC(data.fit)))
```

```
  # Predict the next 12 periods
  data.pre <- predict(data.fit, n.ahead = 12)
  U <- data.pre$pred + 1.96 * data.pre$se
  L <- data.pre$pred - 1.96 * data.pre$se

  # Plot the original data and the forecast
  plot.ts(data$X.Passengers, xlim = c(114, length(data$X.Passengers) + 12), ylim =
c(300, max(data$X.Passengers, U)), type = "o",
          main = paste("ARIMA(", p, ",1,", q, ")(0,1,", Q, ")[12]"))
  lines(data.pre$pred, col = "red", type = "o")
  lines(U, col = "blue", lty = "dashed")
  lines(L, col = "blue", lty = "dashed")
  abline(v = (length(data$X.Passengers) - 11), lty = "dotted")

  # Plot diagnostic plots
  tsdiag(data.fit, 60)
}

# Initialize the table to store AIC and BIC values
aic_bic_table <- data.frame(Model = character(), AIC = numeric(), BIC = numeric(),
stringsAsFactors = FALSE)

# Loop through the values of p, q, and Q and fit the models
for (p in 0:1) {
  for (q in 0:1) {
    for (Q in 0:2) {
      if (p == 0 && q == 0 && Q == 0) next  # Skip the case where p = q = Q = 0

      # Construct model name
      model_name <- paste0("data.fit", p, q, Q)

      # Fit and plot the model
      fit_and_plot(p, q, Q, data, model_name, aic_bic_table)
    }
  }
}

# Print the AIC and BIC table
```

```
print(aic_bic_table)
```