

## 第十三組時間數列報告

# 目錄

前言.....	
資料敘述.....	
資料來源.....	
預測資料.....	
原始資料的時數圖、ACF與PACF圖.....	
一階差分後的時數圖、ACF與PACF圖.....	
季節性差分後的時數圖、ACF與PACF圖.....	
一階季節性差分後的時數圖、ACF與PACF圖.....	
參數模型表.....	
殘差模型.....	
程式碼.....	

## 前言

使用1949年1月到1959年12月的數據來預測1960年紐約市與其他國際目的地之間每月的航班乘客總數。

## 資料敘述

- 時間範圍:數據涵蓋了11年, 從1949年1月到1959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。

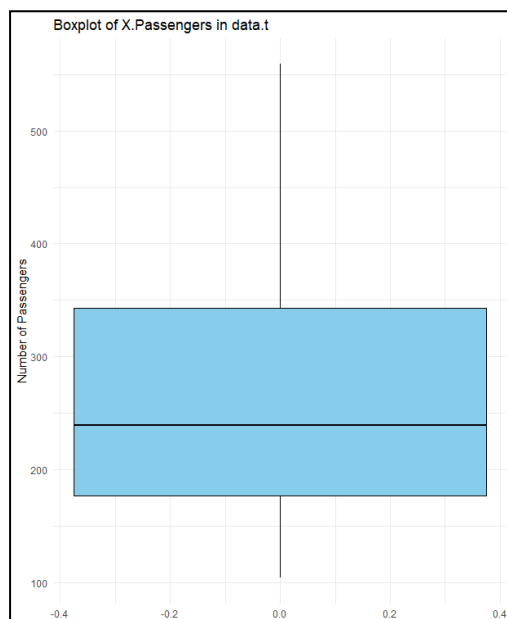
## 資料來源

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.

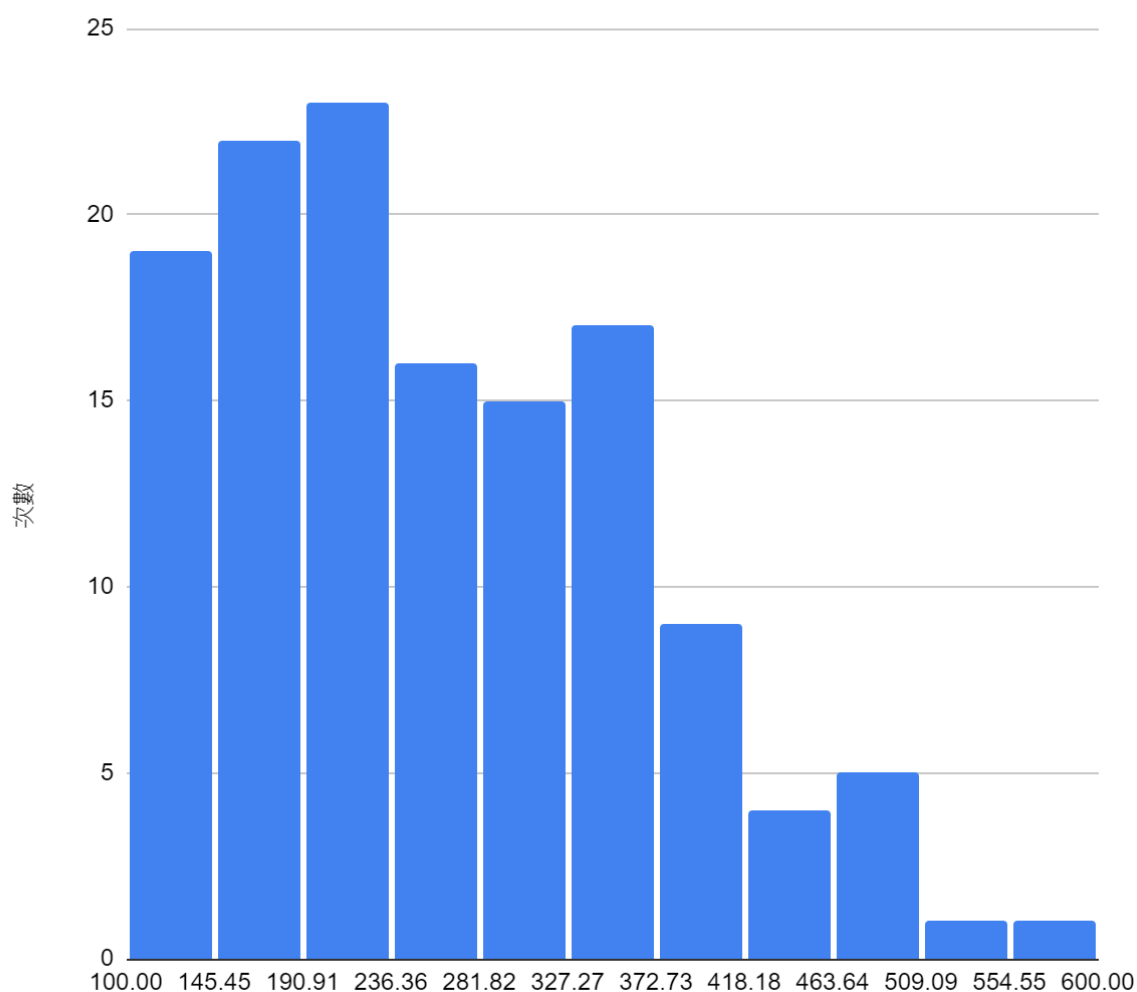
-The classic Box & Jenkins airline data.

## 千人計乘客數量箱型圖



## 千人計乘客數量直方圖

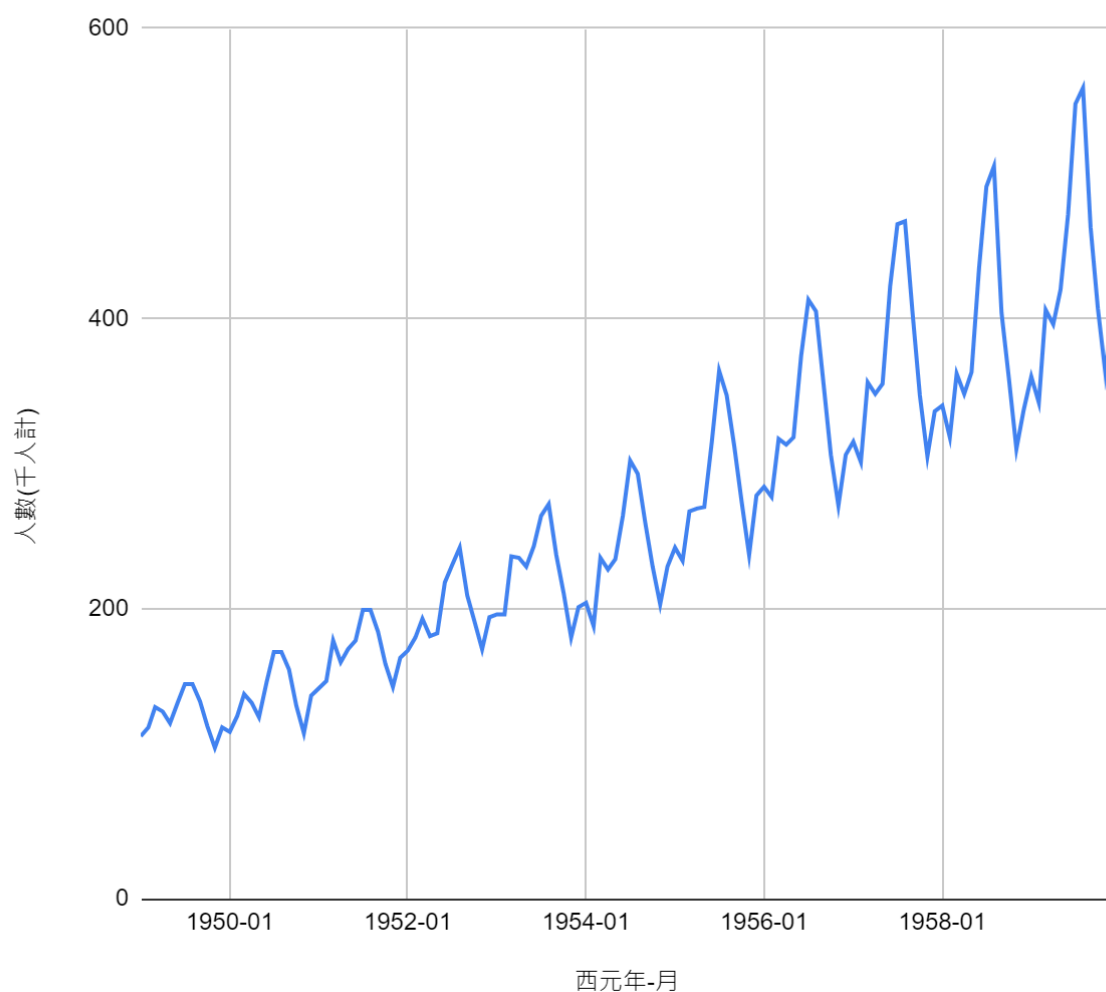
1949-1961年間每月紐約市與其他國際目的地之間的航班乘客  
(千人計)總數量的區間分布圖



## 原始資料的時數圖、ACF與PACF圖

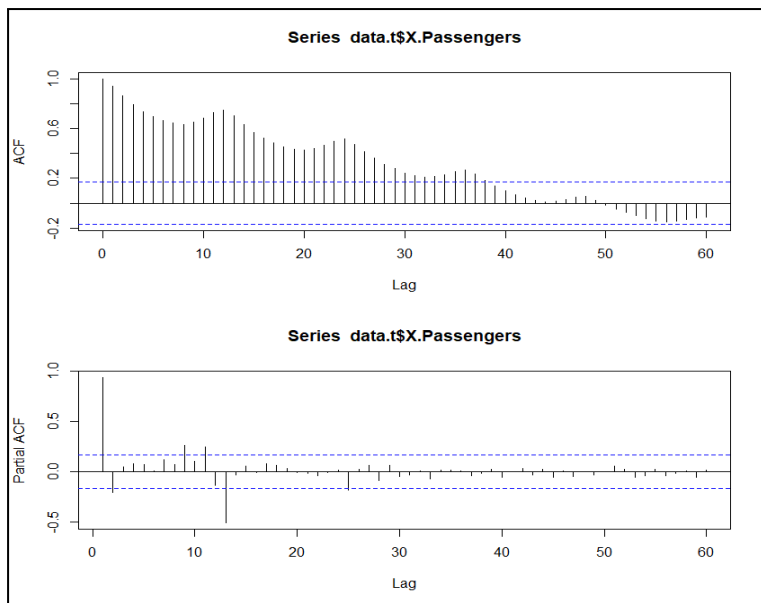
原144筆，去掉1960年分12筆資料後的訓練樣本的132筆資料(1949-01~1959-12)。

1949-1959年間每月紐約市與其他國際目的地之間的航班乘客  
總數量時間數列圖



有趨勢、季節性且震盪隨時間加劇，故不平穩。7、8月份通常是全年最高。

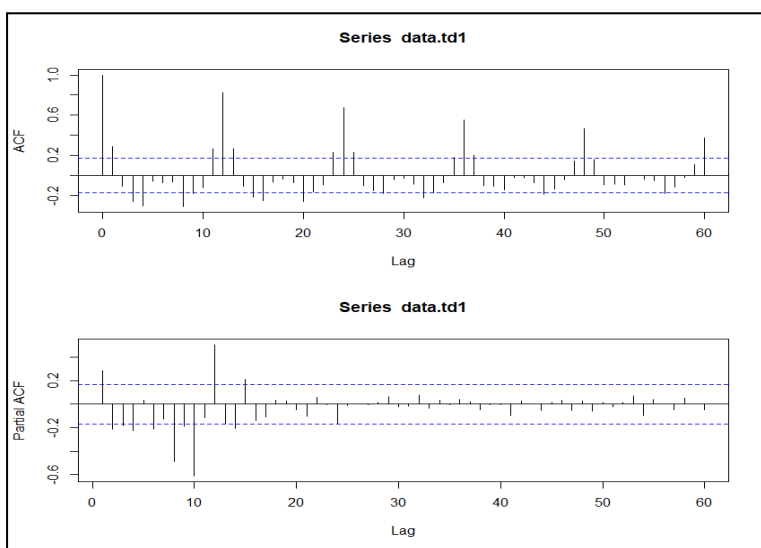
ACF圖與PACF圖



S

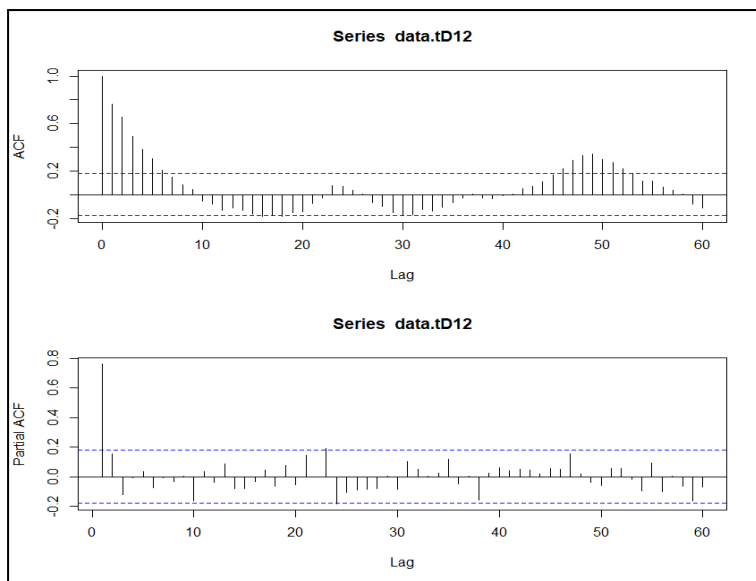
ACF下降緩慢，有趨勢，故不平穩。

## 一階差分後的時數圖、ACF與PACF圖



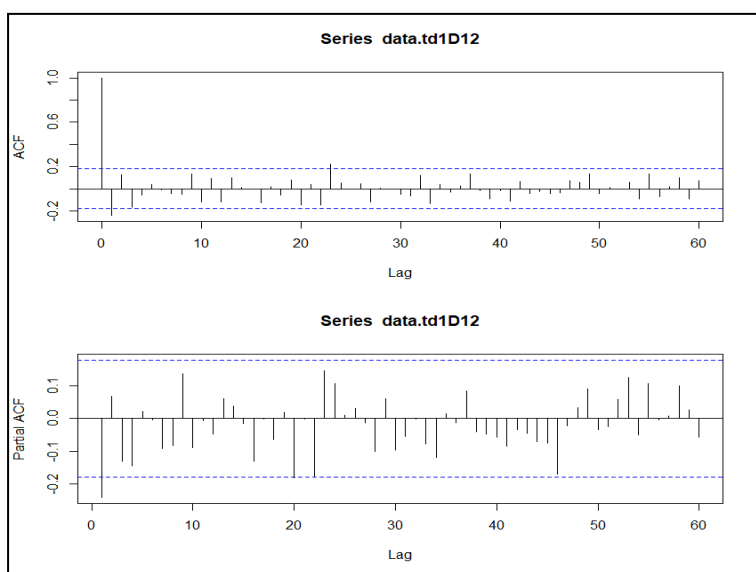
Lag=12時，ACF下降較慢，有季節性，故不平穩。

## 季節性差分後的時數圖、ACF與PACF圖



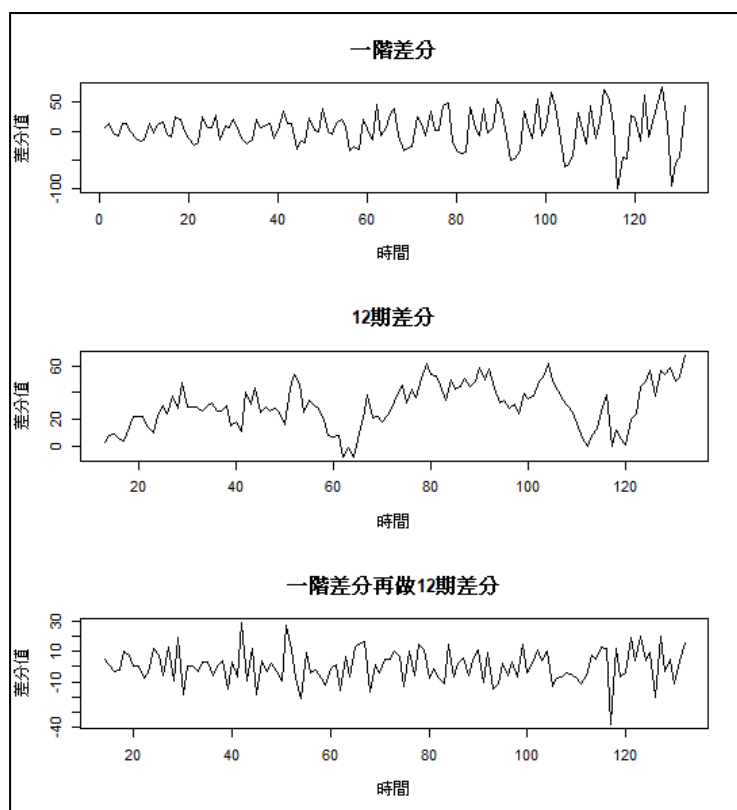
季節性差分後的ACF下降緩慢且波動範圍明顯超過-0.2與0.2之間，故不平穩。

## 一階季節性差分後的時數圖、ACF與PACF圖



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

## 各差分後時間數列圖



由上述內容所知資料data.td1D12平穩，並藉由data.td1D12的ACF及PACF配適模型，因此選擇在 $d=1$ 、 $D=1$ 、 $p \leq 1$ 、 $q \leq 2$ 中找出最佳模型。

## 參數模型表

參數模型表是指data.fitpqQ以data.fit00就是一個有季節性與一階差分的AR(1)

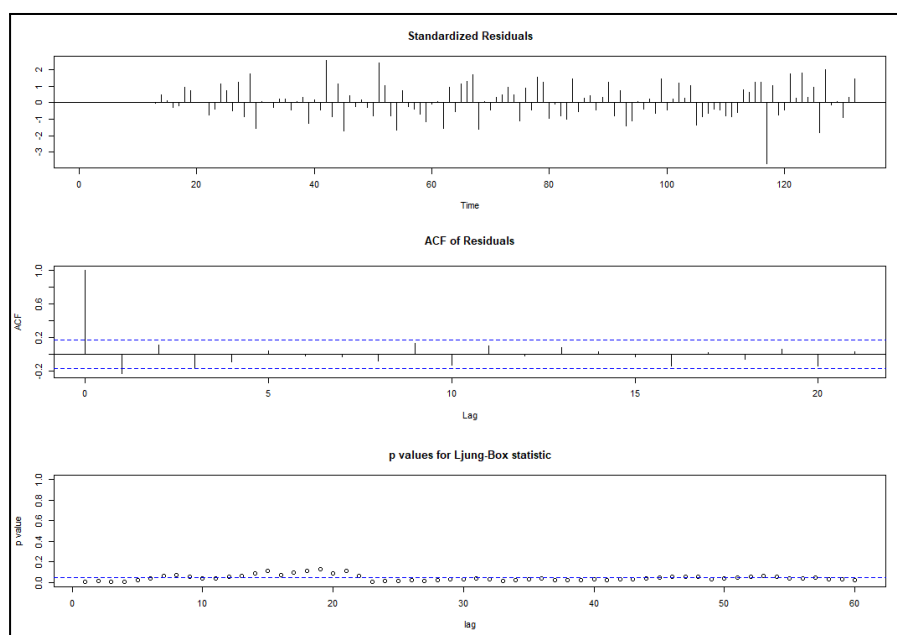
SARIMA	AR	MA	SMA
$(0, 1, 0 \times 0, 1, 1)_{12}$	AR: None	MA: None	SMA: -0.115
$(0, 1, 0 \times 0, 1, 2)_{12}$	AR: None	MA: None	SMA: -0.129, 0.069
$(0, 1, 1 \times 0, 1, 0)_{12}$	AR: None	MA: -0.226	SMA: None
$(0, 1, 1 \times 0, 1, 1)_{12}$	AR: None	MA: -0.217	SMA: -0.084
$(0, 1, 1 \times 0, 1, 2)_{12}$	AR: None	MA: -0.249	SMA: -0.115, 0.134
$(1, 1, 0 \times 0, 1, 0)_{12}$	AR: -0.243	MA: None	SMA: None
$(1, 1, 0 \times 0, 1, 1)_{12}$	AR: -0.232	MA: None	SMA: -0.079
$(1, 1, 0 \times 0, 1, 2)_{12}$	AR: -0.268	MA: None	SMA: -0.11, 0.142



$(1, 1, 1 \times 0, 1, 0)_{12}$	AR: -0.553	MA: 0.332	SMA: None
$(1, 1, 1 \times 0, 1, 1)_{12}$	AR: -0.539	MA: 0.325	SMA: -0.074
$(1, 1, 1 \times 0, 1, 2)_{12}$	AR: -0.513	MA: 0.268	SMA: -0.102, 0.135

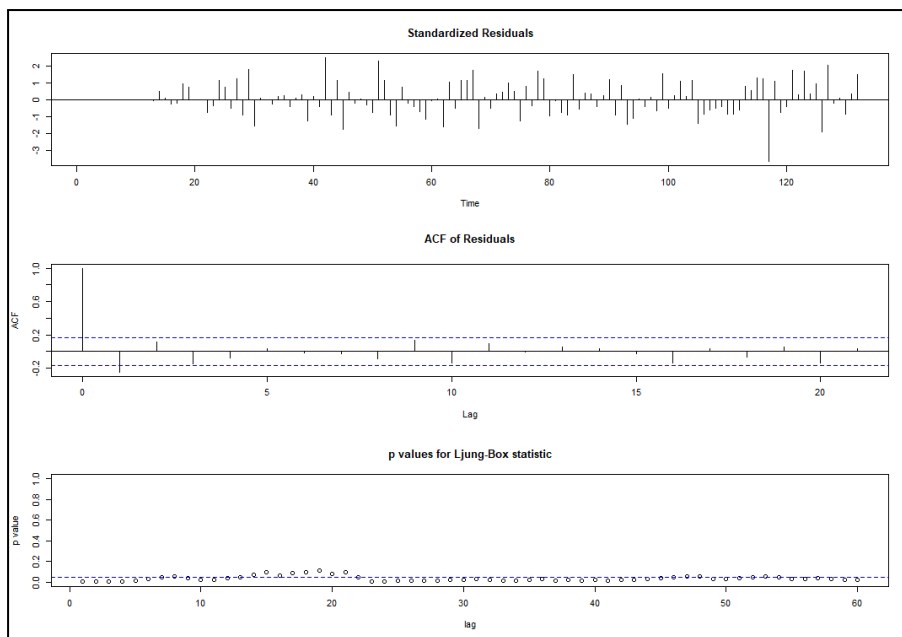
## 殘差模型

### 模型一殘差



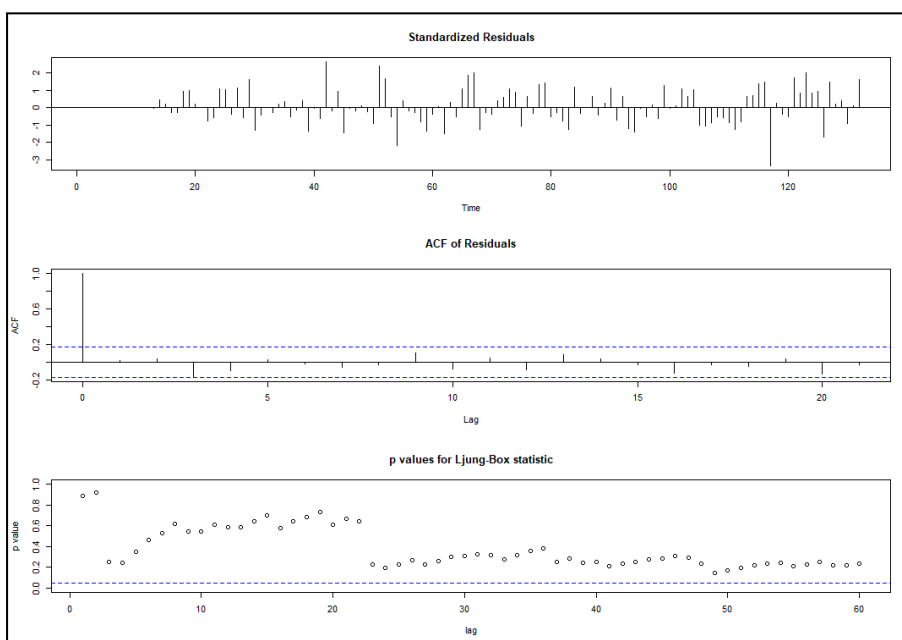
p-value  $\leq 0.05$ , 拒絕白噪音假設

### 模型二殘差



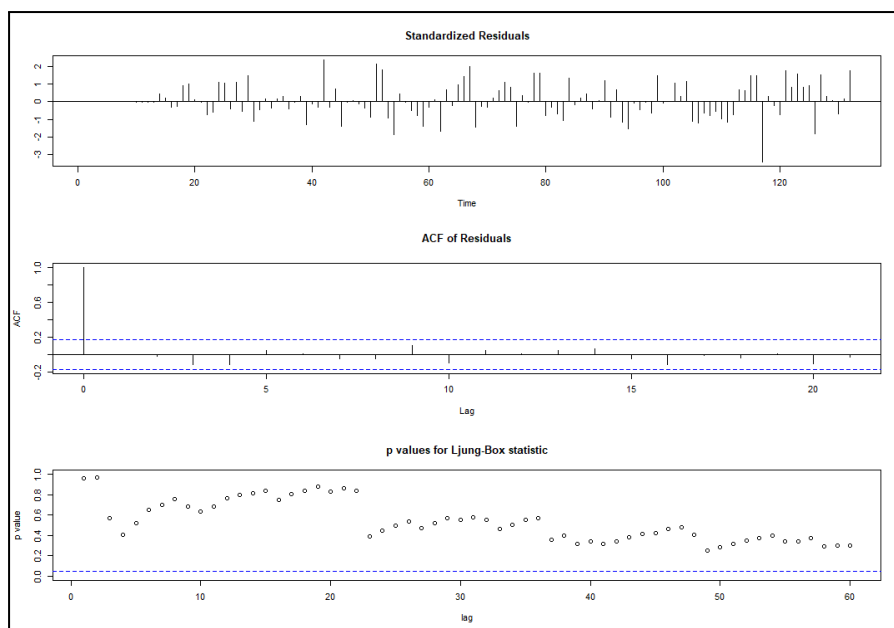
p-value $\leq$ 0.05, 拒絕白噪音假設

## 模型六殘差



p-value $>$ 0.05, 不拒絕白噪音假設

## 模型11殘差



p-value>0.05, 不拒絕白噪音假設

## 比較AIC和BIC

模型編號	SARIMA	是否滿足白噪音假設	AIC	BIC
1	$(0, 1, 0) \times (0, 1, 1)_{12}$	否	905.4189	910.9772
2	$(0, 1, 0) \times (0, 1, 2)_{12}$	否	906.8886	915.226
3	$(0, 1, 1) \times (0, 1, 0)_{12}$	是	900.6852	906.2435
4	$(0, 1, 1) \times (0, 1, 1)_{12}$	否	901.7211	910.0584
5	$(0, 1, 1) \times (0, 1, 2)_{12}$	是	901.7468	912.8633
6	$(1, 1, 0) \times (0, 1, 0)_{12}$	是	899.9021	905.4604
7	$(1, 1, 0) \times (0, 1, 1)_{12}$	否	901.0524	909.3898
8	$(1, 1, 0) \times (0, 1, 2)_{12}$	是	900.8583	911.9748
9	$(1, 1, 1) \times (0, 1, 0)_{12}$	是	900.9716	909.309
10	$(1, 1, 1) \times (0, 1, 1)_{12}$	是	902.2214	913.3379
11	$(1, 1, 1) \times (0, 1, 2)_{12}$	是	902.2385	916.1341

data.fit100是最適模型, AIC、BIC最小

data.fit100 是SARIMA(1,1,0),(0,1,0)12

方程式為

$$(1 - (-0.243B))(1 - B)(1 - B^{12})X_t = \omega_t$$

其中 $\omega_t \sim N(0, \sigma^2)$ ,  $\hat{\sigma}^2 = 108.9$

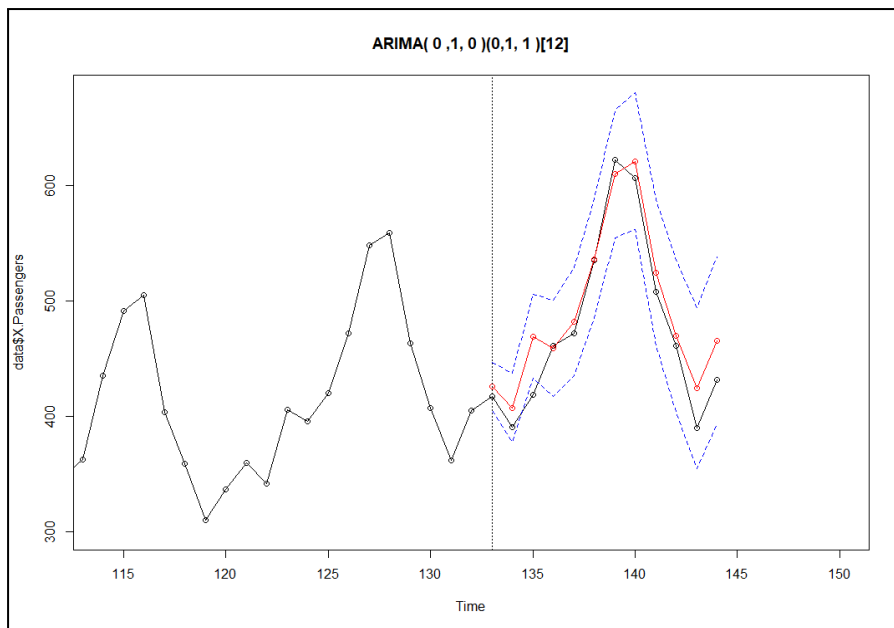
## 預測結果

## 預測資料

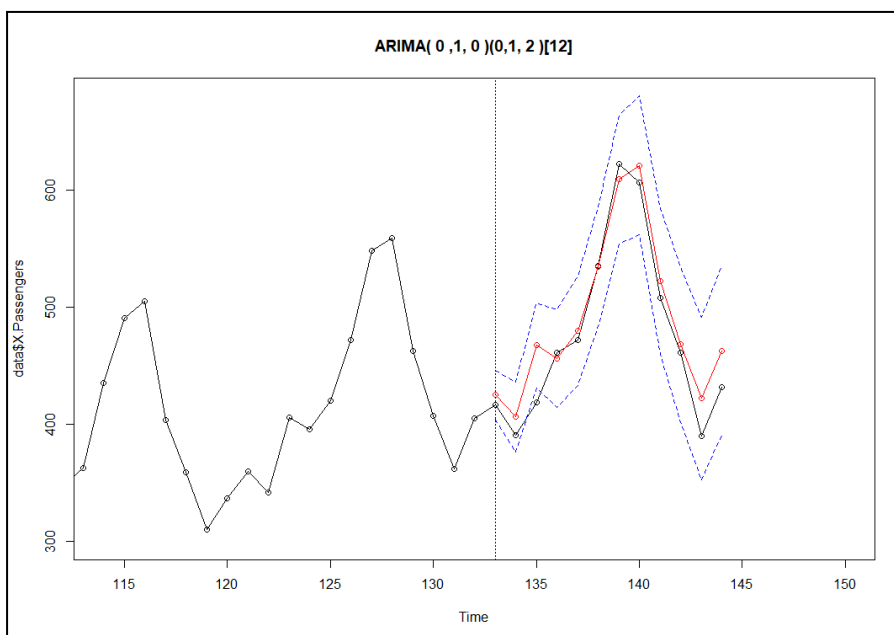
	Month	X.Passengers
133	1960-01	417
134	1960-02	391
135	1960-03	419
136	1960-04	461
137	1960-05	472
138	1960-06	535
139	1960-07	622
140	1960-08	606
141	1960-09	508
142	1960-10	461
143	1960-11	390
144	1960-12	432

預測資料一共是12筆, 時間從1960年1月到1960年12月。

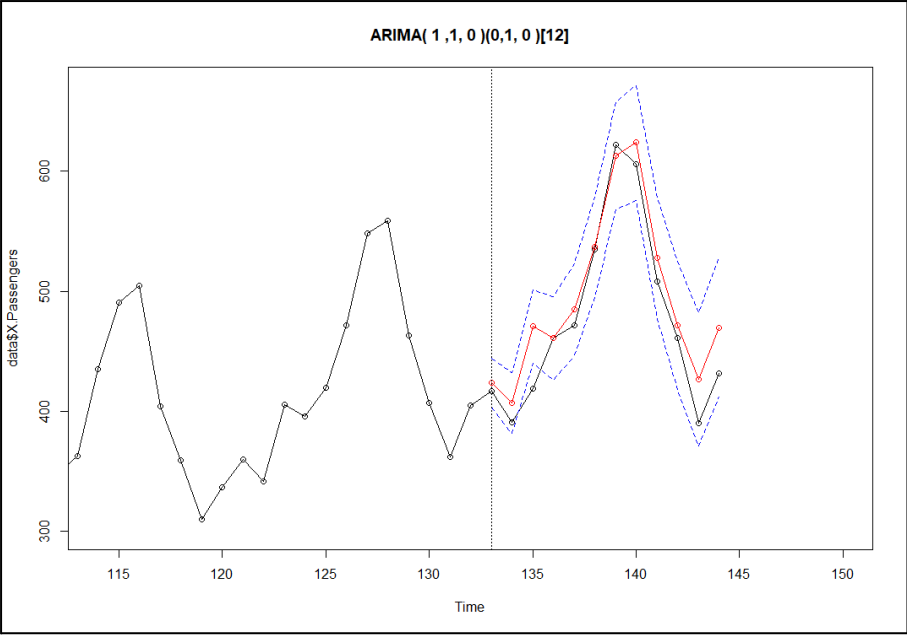
## 模型一預測結果



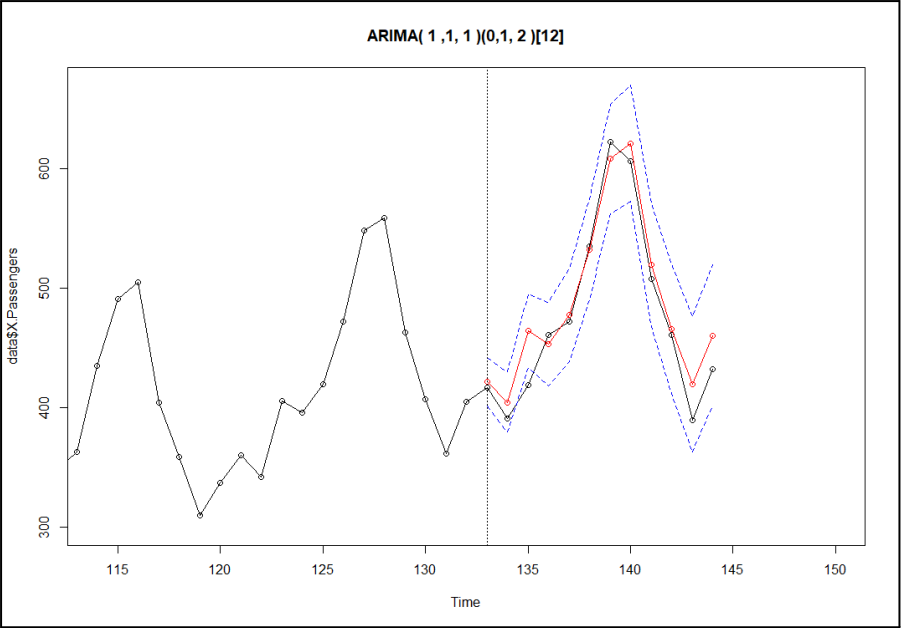
模型二預測結果



模型六預測結果



模型十一預測結果



月份	1960年一月	1960年二月	1960年三月	1960年四月	1960年五月	1960年六月
	424.1099	407.0557	470.8257	460.8817	484.8681	536.8714
	1960年七月	1960年八月	1960年九月	1960年十月	1960年十一月	1960年十二月
	612.8706	623.8708	527.8707	471.8707	426.8707	469.8707

1960年3月，紐約經歷了一場嚴重的冬季風暴（東北風暴），發生在3月3日至5日。這場風暴被評為“嚴重”級別，帶來了暴風雪、強風和大量降雪。紐約市的降雪量達到歷史高位，導致學校停課、航班取消，並給通勤和居民生活帶來了極大困難。這場惡劣天氣可能是3月份班機乘客數量較少的主要原因之一。這也導致所有模型在1960三月都沒預測好。

## 結論

透過SARIMA(110)(010)s=12的時間數列模型，除了三月以外，1960年紐約來往的國際航班數量都成功在信賴區域內，說明模型有良好的預測能力。

# 程式碼

```
#####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)

```