

# 111 學年度第二學期科學計算軟體作業二

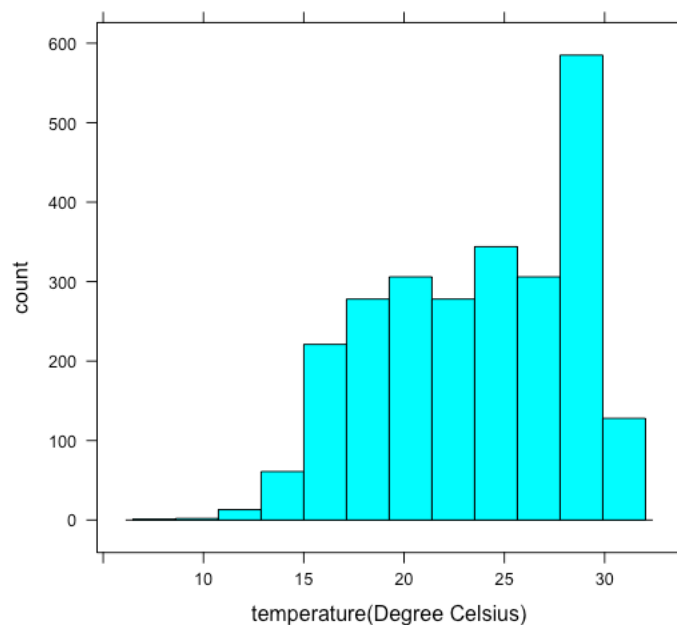
系級：114 姓名：黃薇庭 學號：F64101032

※ 書面作業格式為 pdf 檔，檔名：HW2\_學號.pdf

1. 請解答 2009-2011 年三年間**全台溫度**是否為**常態分布**(15 分)，並描述資料分布狀況(20 分)(共 35 分)。(答題提醒:1.常態分布需附上**使用的分析方法**、**p 值**，2.需提到資料分布狀況之**偏度**以及**峰度**，可畫圖輔助，例: 直方圖)。

ANS:

Temperature distribution in Taiwan during 2009-2011



- 因為此樣本數  $> 50$ ，因此選用 Lilliefors(Kolmogorov-Smirnov) Normality Test。由下圖可知  $p\text{-value} < 0.05$ ，拒絕虛無假設，就是拒絕該資料符合常態分布。

Lilliefors (Kolmogorov-Smirnov) normality test

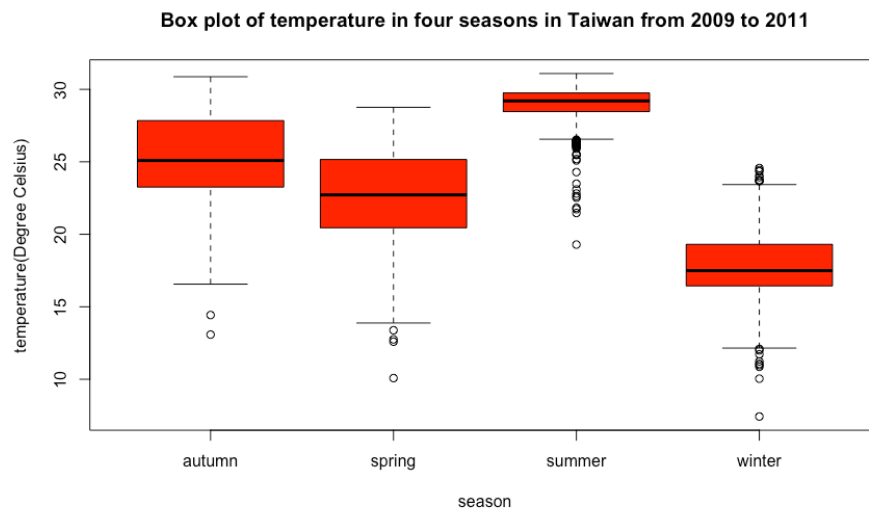
```
data: dataset2$temperature  
D = 0.093944, p-value < 2.2e-16
```

- 偏度與峰值：由下圖可以看出  $Kurtosis < 0$  (不等於 0)，代表資料分佈起伏較小，不符合常態分佈， $skew < 0$  (不等於 0)，代表資料為左偏分佈，不符合常態分佈。

```
> skew(dataset2$temperature) #Skewness
[1] -0.3315729
> kurtosi(dataset2$temperature) #Kurtosis
[1] -1.020501
```

2. 依據季節，繪製全台溫度之盒形圖(boxplot)(15 分)。(boxplot(...), x 軸為季節, y 軸為溫度, 四季結果繪製在同一張圖)

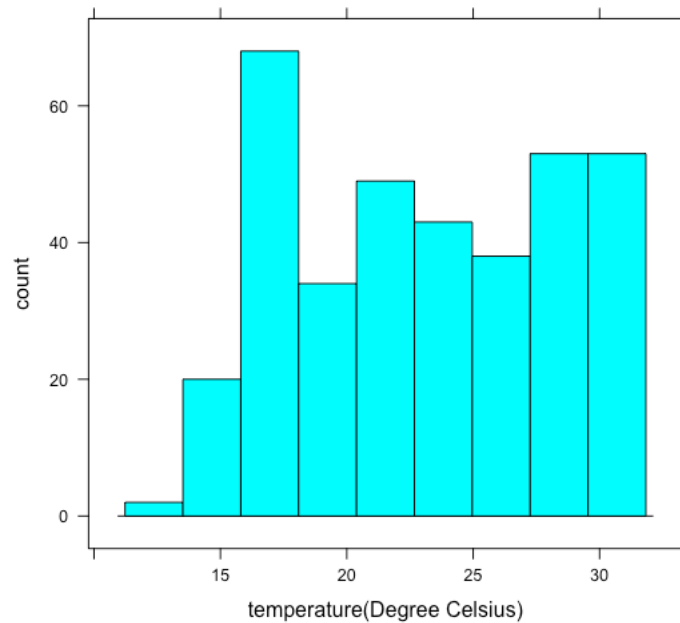
ANS:



3. 請解答 2009-2011 年三年間新北市溫度是否為常態分布 (15 分), 並描述資料分布狀況(20 分)(共 35 分)。(答題提醒: 1.可運用 `subset(dataset, 屬性欄=='特定值')` 取出特定條件資料, 例: (縣市值等於"新北市"))

ANS:

Temperature distribution in New Taipei City during 2009-2011



- 因為此樣本數  $> 50$ ，因此選用 Lilliefors(Kolmogorov-Smirnov) Normality Test，由下圖可知  $p\text{-value} < 0.05$ ，拒絕虛無假設，就是拒絕該資料符合常態分佈。

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: dataset_select$temperature
D = 0.10124, p-value = 1.737e-09
```

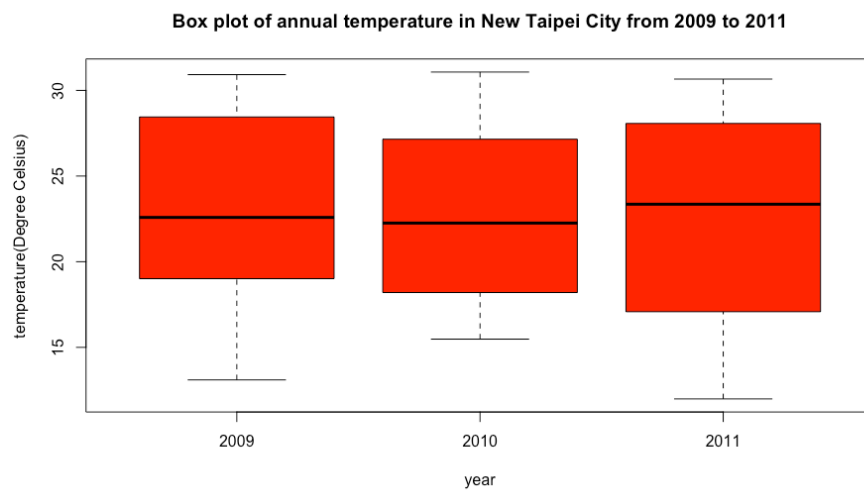
- 偏度與峰值：由下圖可以看出  $Kurtosis < 0$  (不等於 0)，代表資料分佈起伏較小，不符合常態分佈， $skew < 0$  (不等於 0)，代表資料為左偏分佈，不符合常態分佈。

```
> skew(dataset_select$temperature)
[1] -0.06518144
> kurtosi(dataset_select$temperature)
[1] -1.299147
```

4. 依據年份，畫出 2009-2011 年每一年新北市溫度之盒形圖(boxplot) (15 分)。

(boxplot(...))，x 軸為年份，y 軸為溫度，三年份結果繪製在同一張圖)

ANS:



※ 將程式碼附在作業最後

```
#####
# Set Working Directory #
#####

# Get your current working directory #
getwd() # Re-check the path for the working directory

# Change your current working directory #
setwd("/Users/huangweiting/coding/INTRODUCTION TO SCIENTIFIC COMPUTING SOFTWARE
/C3_ClassData")
getwd() # Re-check the path for the working directory

#####
# Import an example dataset #
#####

dataset <- read.csv("C3_HW.csv", fileEncoding = "utf-8")
#dataset <- read.table("C3.csv",header = TRUE, sep = ",") #Import dataset

#str(dataset)      #Check the variable format
#View(dataset)     #Check Dataset
#dim(dataset)      #Check Dataset (how many observations and variables)
#colnames(dataset) #Check variable names
#summary(dataset)  #Get the summary statistics about the object

#####
# generate a boxplot #
#####

install.packages("lattice")
library("lattice")

#boxplot(formula= Rainfall~SeasonNum ,data =dataset, xlab = "Season", ylab =
"Rainfall(mm)", col = "red")

#boxplot(formula= Rainfall~Town_name,data =dataset, xlab = "Town", ylab =
"Rainfall(mm)", col = "blue")

dataset2 <- na.omit(dataset) #刪掉有缺失的數據

#####Question#####
#Q1
#2009-2011年三年間全台灣溫度是否為常態分布，並描述資料分布狀況
#常態分布需附上使用的分析方法、p值
##Skewness & Kurtosis ##
install.packages("psych")
library(psych)
skew(dataset2$temperature) #Skewness
kurtosi(dataset2$temperature) #Kurtosis
#The Shapiro-Wilk Normality Test
#shapiro.test(dataset2$Rainfall) #小樣本
install.packages("nortest")
library(nortest)
#Lilliefors(Kolmogorov-Smirnov) Normality Test
lillie.test(dataset2$temperature) #大樣本
#需提到資料分布狀況之偏度以及峰度，可畫圖輔助。例：直方圖
#install.packages("lattice")
# How to load a package #
library("lattice")
histogram(x= ~temperature,data =dataset2, xlab = "temperature(Degree Celsius)", ylab =
"count", type = "count",main="Temperature distribution in Taiwan during 2009-2011")

#Q2 ok
boxplot(formula= temperature~season ,data =dataset2, xlab = "season", ylab =
"temperature(Degree Celsius)", col = "red",main="Box plot of temperature in four seasons

in Taiwan from 2009 to 2011")

#Q3
#請解答2009-2011年三年間新北市溫度是否為常態分布，並描述資料分布狀況
dataset_select <- subset(dataset2,city=="新北市")
lillie.test(dataset_select$temperature)
skew(dataset_select$temperature)
kurtosi(dataset_select$temperature)
histogram(x= ~temperature,data =dataset_select, xlab = "temperature(Degree Celsius)",
,ylab = "count", type = "count",main="Temperature distribution in New Taipei City during
2009-2011")

#Q4
#依據年份，畫出2009-2011年每一年新北市溫度之盒形圖(boxplot) 。
#x軸為年份，y軸為溫度，三年份結果繪製在同一張圖)
#Box plot of annual temperature in New Taipei City from 2009 to 2011"
boxplot(formula= temperature~year ,data =dataset_select, xlab = "year", ylab =
"temperature(Degree Celsius)", col = "red",main="Box plot of annual temperature in New
Taipei City from 2009 to 2011")
```