統計學與實習上
第二次作業

1. (Text, p.62) The demand for the video games provided by Mid-Tech Video Games Inc. has exploded in the last several years. Hence, the owner needs to hire several new technical people to keep up with the demand. Mid-Tech gives each applicant a special test that Dr. McGraw, the designer of the test, believes is closely related to the ability to create video games. For the general population, the mean on this test is 100. Below are the scores on this test for the applicants.

95 105 120 81 90 115 99 100 130 10

The president is interested in the overall quality of the job applicants based on this test. Compute the mean and the median scores for the 10 applicants. What would you report to the president? Does it seem that the applicants are better than the general population? **(by hand) (0.5 points)**

Mean = 94.5; Median = 99.5; Applicants are no better than regular people.

2. (Text p.92) The following frequency distribution reports the electricity cost for a sample of 50 two-bedroom apartments in Albuquerque, New Mexico, during the month of May last year.

Electricity Cost	Frequency
$ 80 up to $100	3
$ 100 up to $120	8
$ 120 up to $140	12
$ 140 up to $160	16
$ 160 up to $180	7
$ 180 up to $200	4
Total	50

a. Estimate the mean cost. **(by hand) (0.3 points)**
b. Estimate the standard deviation. **(by hand) (0.3 points)**
c. Use the Empirical Rule to estimate the proportion of costs within two standard deviations of the mean. What are these limits? Use 1-2 sentences to explain what you find. **(by hand) (0.6 points)**

Class	M	f	$f(M)$	$M-\mu$	$(M-\mu)^2$	$f(M-\mu)^2$
80 up to 100	90	3	270	–51.2	2621.44	7864.3
100 up to 120	110	8	880	–31.2	973.44	7787.5
120 up to 140	130	12	1560	–11.2	125.44	1505.3
140 up to 160	150	16	2400	8.8	77.44	1239.0
160 up to 180	170	7	1190	28.8	829.44	5806.1
180 up to 200	190	4	760	48.8	2381.44	9525.8
			7060			33728

a. The mean is $141.20, found by 7060/50.
b. The standard deviation is $26.24, found by the square root of 33728/49.
c. The empirical rule says that 95 percent of the costs are between $88.73 and 193.67, found by 141.20 ± 2(26.24).

3. 利用以下的程式碼做為開始，使用 ifelse 判斷式並遵從下圖的條件將結果呈現出來。weather <- sample(c("sunny", "rainy", "cloudy"), size = 1) **(0.6 points)**


```
1.  weather <- sample(c("sunny", "rainy", "cloudy"), size = 1)
2.  if(weather == "sunny"){
    print("騎單車(Go cycling)")
3.  }else if(weather == "rainy"){
    print("去健身房(Go Gym)")
4.  }else{
    print("待在家(Stay at home)")
5.  }
```

```
[1] "待在家(Stay at home)"
```

4. 某機構針對部分成年人進行健康調查並記錄了各項健康指標，變數名稱分別為 min(分鐘數)、number(號碼)、year(年紀)、gender(性別)、height(身高)、weight(體重)、BMI(身體質量指數)、temperature(體溫)、body.fat(體脂肪)、SBP (收縮壓) 和 DBP (舒張壓)。請使用資料檔 health_examination.txt 回答以下問題。

   a. 計算 SBP 的第四十百分位數。**(0.3 points)**

   b. 呈上，請說明這個數值代表的意涵。**(0.2 points)**

   c. 請依照不同性別繪製 BMI 的盒鬚圖(boxplot)。**(0.4 points)**

   d. 請根據你所繪製的盒鬚圖說明不同性別的 BMI 的**中位數**（median）、**四分位距**（interquartile range），以及**偏態**（skewness）分佈情形。**(1 points)**

   e. 以身高作為 x 軸，體重作為 y 軸，繪製一個散佈圖(scatter plot)，並在圖上添加兩條直線，分別為以平均身高作為垂直線和以平均體重作為水平線。(圖需有 x 及 y 軸名稱，以及圖的標題，可自由發揮) **(0.6 points)**

   f. 請根據你所繪製的散佈圖說明你的發現。**(0.2 points)**

a.

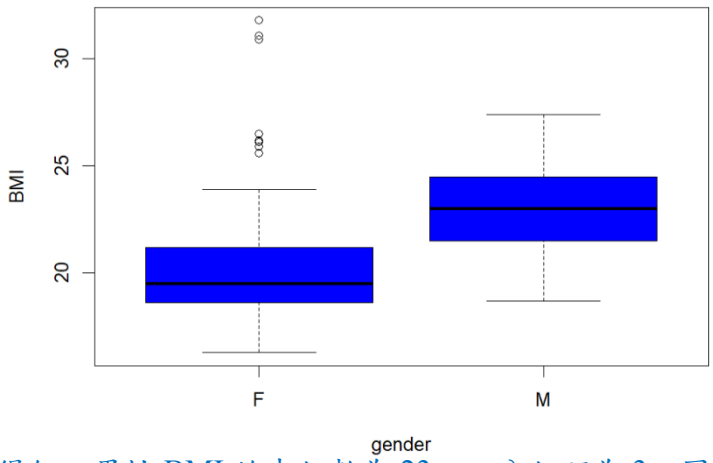
```
1.  data = read.table(file = "C:/Users/lab205/Desktop/health_examination.txt", header
    = T)
2.  SBP = data$SBP
3.  quantile(SBP, probs=0.4, type=6)
```

```
> quantile(SBP, probs=0.4, type=6)
 40%
93.4
```

b. 第四十分百分位數為 93.4，表示至少有 40%的觀測值小於或等於這個數，而同時也有 60%的資料大於或等於這個觀測值。

c.

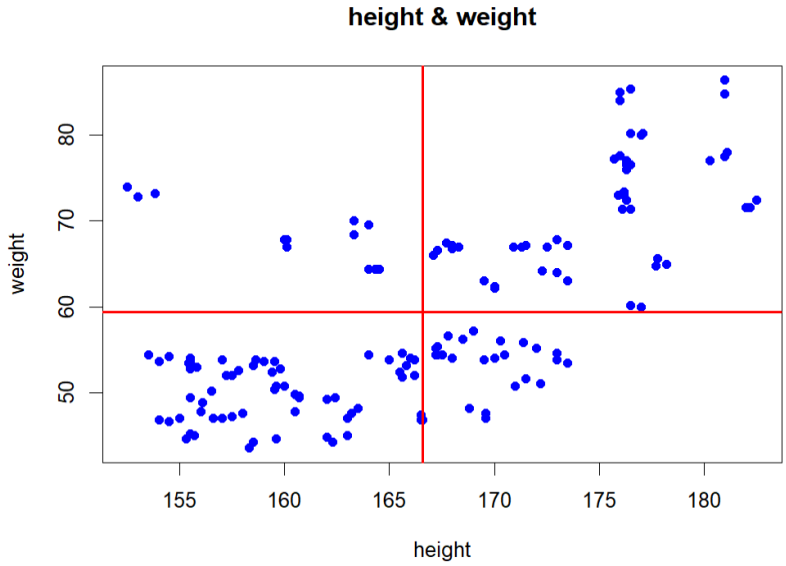
1. boxplot(formula = BMI ~ gender , # Y ~ X (代表 X 和 Y 軸要放的數值)
   data = data, # 資料
   xlab = "gender", # X 軸名稱
   ylab = "BMI", # Y 軸名稱
   col ="blue")
2. male = data[data[["gender"]] == "M",7]
3. summary(male)
4. female = data[data[["gender"]] == "F",7]
5. summary(female)



d. 從圖中可以得知，男性 BMI 的中位數為 23，四分位距為 3，圖形分佈趨近對稱；女性 BMI 的中位數為 19.5，四分位距為 2.6，偏態分佈為正偏態。

e.

1. plot(x = data$height,
2. y = data$weight,
3. main = "height & weight",
4. xlab = "height",
5. ylab = "weight",
6. pch = 16, # 點的圖形
7. col = "blue" # 顏色
8.)
9. # 添加垂直線
10. abline(h = mean(data$weight),
    a. v = mean(data$height),
    b. lwd = 2, # 線條類型
    c. col = "red" # 顏色
11.



f. 大部分觀測值落在右上方與左下方，代表調查樣本可大致分為兩組，一組為高於平均體重和平均身高，另一組則是低於平均體重及平均身高。