

科學計算軟體

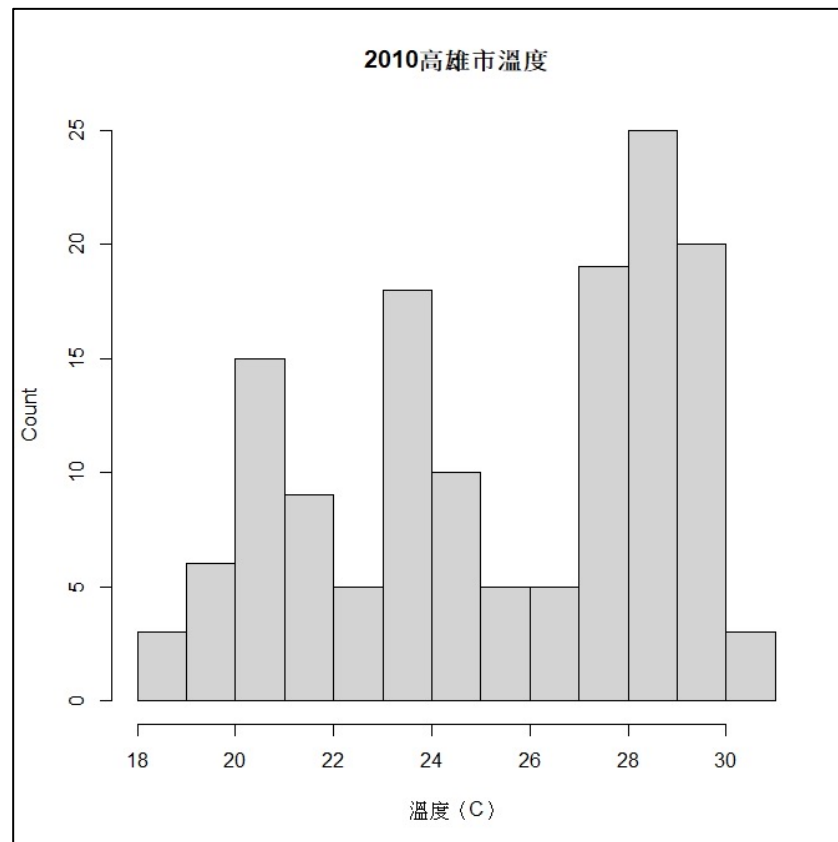
W10課堂練習

黃薇庭F64101032

第一題 (1) -1

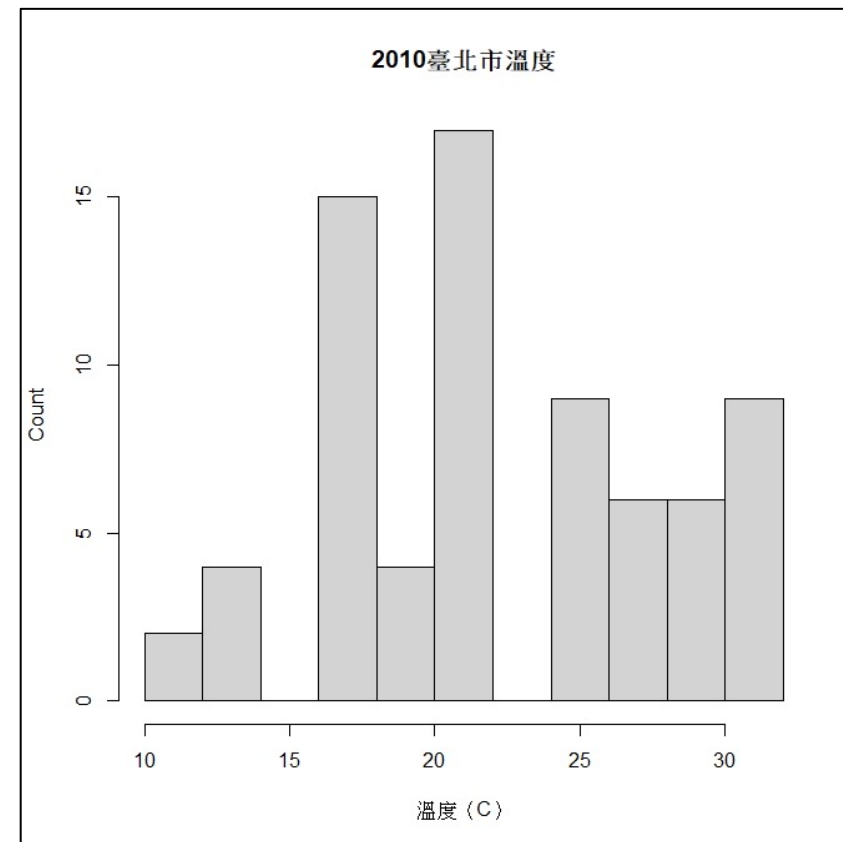
高雄市 2010

最大值	30.47
最小值	18.41
平均值	25.36
標準差	3.45
四分位數	22.57 ,26.02 ,28.46



·台北市 2010

最大值	31.09
最小值	11.06
平均值	22.30
標準差	5.44
四分位數	17.90 ,21.34 ,26.60



第一題 (1) -2

- 樣本數>50，屬於大樣本，因此使用lillie.test
- 台北的p-value<0.05，為非常態分佈。左偏分布，低闊峰，比較平緩
- 高雄的P值<0.05，為非常態分佈。左偏分布（比台北嚴重），低闊峰，也就是比較平緩。

```
> skew(taipei$溫度)
[1] -0.01425896
```

```
> skew(KC$溫度)
[1] -0.3235191
```

```
> kurtosi(taipei$溫度)
[1] -0.9319399
```

```
> kurtosi(KC$溫度)
[1] -0.9768056
```

```
> lillie.test(taipei$溫度)

      Lilliefors (Kolmogorov-Smirnov) normality test

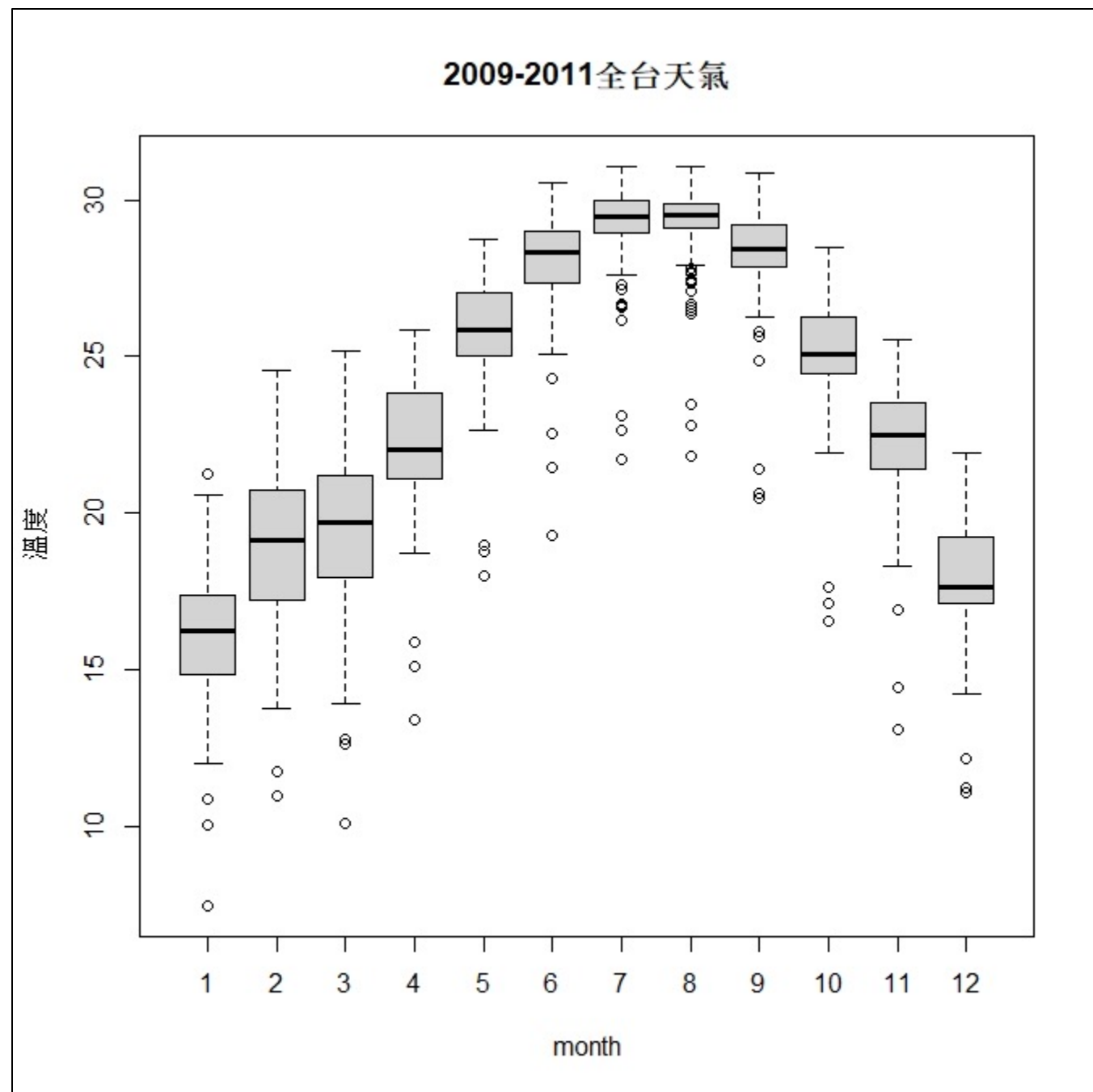
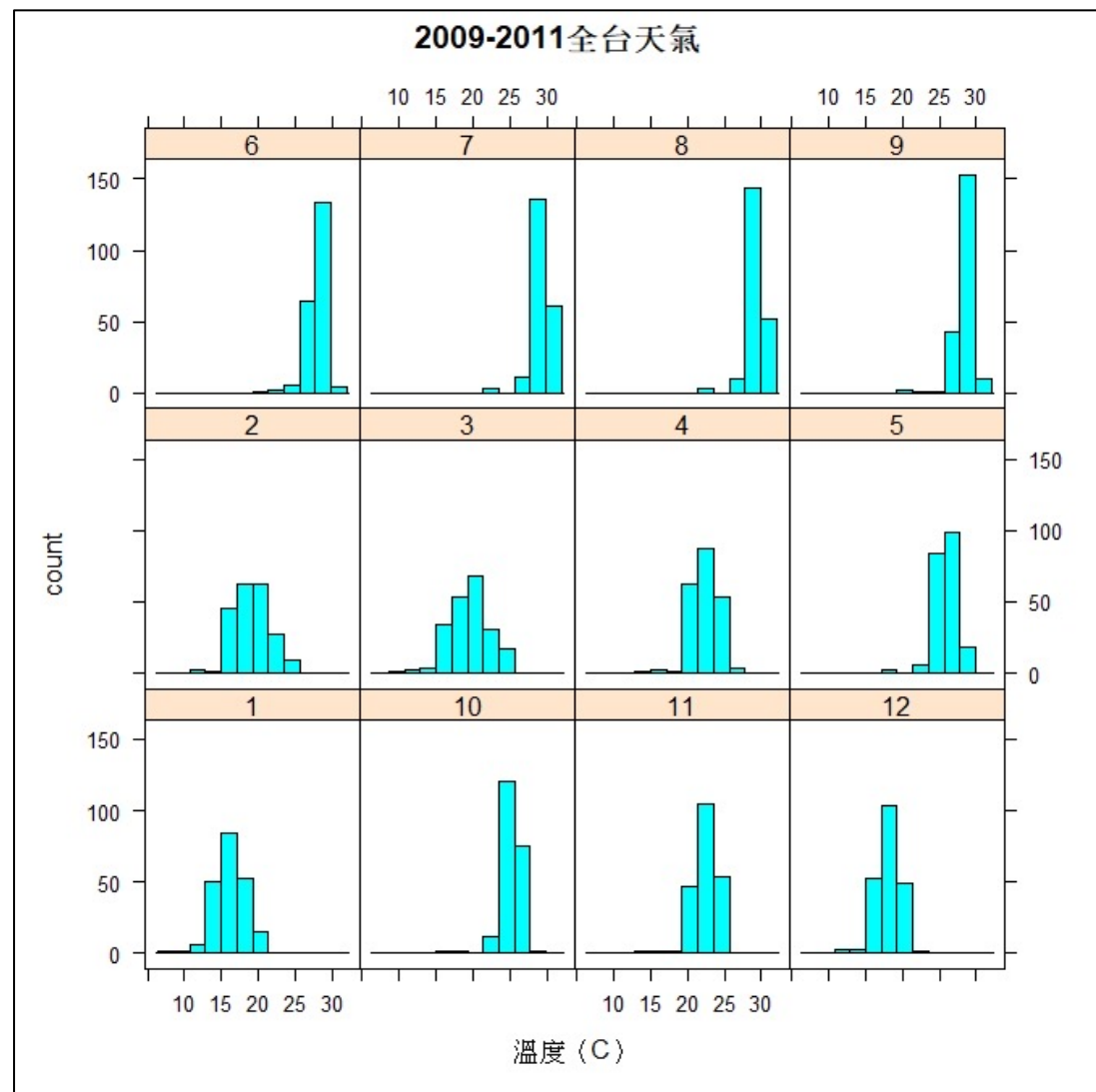
data:  taipei$溫度
D = 0.1176, p-value = 0.01515

> lillie.test(KC$溫度)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  KC$溫度
D = 0.15952, p-value = 1.567e-09
```

第一題 (2)



第二題 (1)

- 一年級：因為 $p\text{-value}=0.2151>0.05$ ，因此不拒絕虛無假設，因此與全台灣平均體重相同。
- 二年級：因為 $p\text{-value}=0.213>0.05$ ，因此不拒絕虛無假設，因此與全台灣平均體重相同。

```
> t.test(dataset1$weight_1, mu =20.80)
```

One Sample t-test

```
data: dataset1$weight_1
t = -1.2673, df = 29, p-value = 0.2151
alternative hypothesis: true mean is not equal to 20.8
95 percent confidence interval:
 19.59668 21.08258
sample estimates:
mean of x
 20.33963
```

```
> t.test(dataset1$weight_2, mu =23.35)
```

One Sample t-test

```
data: dataset1$weight_2
t = -1.2734, df = 29, p-value = 0.213
alternative hypothesis: true mean is not equal to 23.35
95 percent confidence interval:
 22.14404 23.63048
sample estimates:
mean of x
 22.88726
```

第二題 (2)

- 設 H_0 :一二年級時成績相同， H_a :一二年級時成績不同，因為 $p\text{-value} = 0.02773 < 0.05$ ，因此拒絕虛無假設，因此一二年級時的體育成績不同。

```
> ##Two-sample paired T-test  
> #read dataset2  
> t.test(dataset1$score_1, dataset1$score_2,paired=TRUE)
```

Paired t-test

```
data: dataset1$score_1 and dataset1$score_2  
t = 2.3176, df = 29, p-value = 0.02773  
alternative hypothesis: true mean difference is not equal to 0  
95 percent confidence interval:  
 0.148883 2.384450  
sample estimates:  
mean difference  
 1.266667
```

第三題 (1)

- 樣本數皆大於50，因此使用 `lillie.test` 檢驗是否為常態分佈，由下圖可以看出兩直轄市的 `p-value` 皆 < 0.05 ，因此兩縣市溫度都非常態分佈。
- 皆非常態分佈，因此選用 `leveneTest`，由下圖可以發現 `p-value = 4.742e-7 < 0.05，因此兩者變異數有差異。`
- 因變異數有差異，因此 `var.equal = default`，`p-value = 3.195e-5 < 0.05，因此拒絕 H_0 ，兩縣市氣溫有差異。`

```
> lillie.test(kaohsiung$溫度)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  kaohsiung$溫度
D = 0.15952, p-value = 1.567e-09

> lillie.test(Taipai$溫度)

      Lilliefors (Kolmogorov-Smirnov) normality test

data:  Taipei$溫度
D = 0.1176, p-value = 0.01515

> leveneTest(all$溫度, all$city, center=mean)
Levene's Test for Homogeneity of Variance (center = mean)
      Df F value    Pr(>F)
group  1  27.005 4.742e-07 ***
      213
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> t.test(all$溫度[all$city=="高雄市"], all$溫度[all$city=="臺北市"])
```

```
Welch Two Sample t-test

data:  all$溫度[all$city == "高雄市"] and all$溫度[all$city == "臺北市"]
t = 4.3562, df = 100.53, p-value = 3.195e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.668152 4.458114
sample estimates:
mean of x mean of y
25.36119  22.29806
```

第三題 (2)

- 台北樣本數大於50，因此使用 `lillie.test` 檢驗是否為常態分佈，由下圖可以看出 $p\text{-value}=0.01515<0.05$ ，因此兩縣市溫度都非常態分佈。
- 新北樣本數 <50 ，選用 `shapiro.test` 檢驗是否為常態分佈，由下圖可以看出 $p\text{-value}=2.385e-6<0.05$ ，因此非常態分佈。
- 因為兩者皆非常態分佈，因此選用 `leveneTest`，由下圖可以發現 $p\text{-value}=0.4586>0.05$ ，因此兩者變異數相同。
- 因變異數有差異，因此 `var.equal = true`， $p\text{-value}=0.2039>0.05$ ，因此不拒絕 H_0 ，兩縣市氣溫相同。

```
> lillie.test(Taipai$溫度)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data: Taipai$溫度
```

```
D = 0.1176, p-value = 0.01515
```

```
> shapiro.test(NewTaipai$溫度)
```

Shapiro-Wilk normality test

```
data: NewTaipai$溫度
```

```
W = 0.91993, p-value = 2.385e-06
```

```
> leveneTest(all2$溫度,all$city, center=mean)
```

Levene's Test for Homogeneity of Variance (center = mean)

	Df	F value	Pr(>F)
group	1	0.5516	0.4586
	190		

...

```
> t.test(all2$溫度[all2$city=="新北市"], all2$溫度[all2$city=="臺北市"],var.equal = TRUE)
```

Two Sample t-test

```
data: all2$溫度[all2$city == "新北市"] and all2$溫度[all2$city == "臺北市"]
```

```
t = 1.2748, df = 190, p-value = 0.2039
```

```
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.5288962  2.4614518
```

```
sample estimates:
```

```
mean of x mean of y
```

```
23.26433  22.29806
```


第四題-1

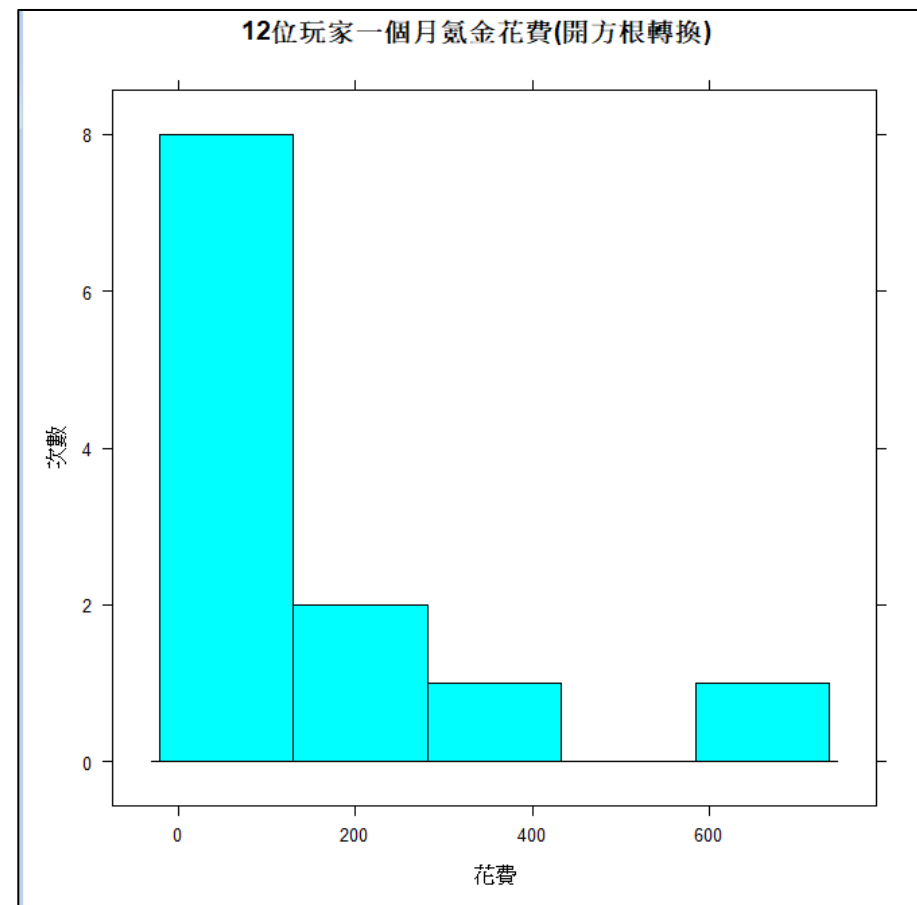
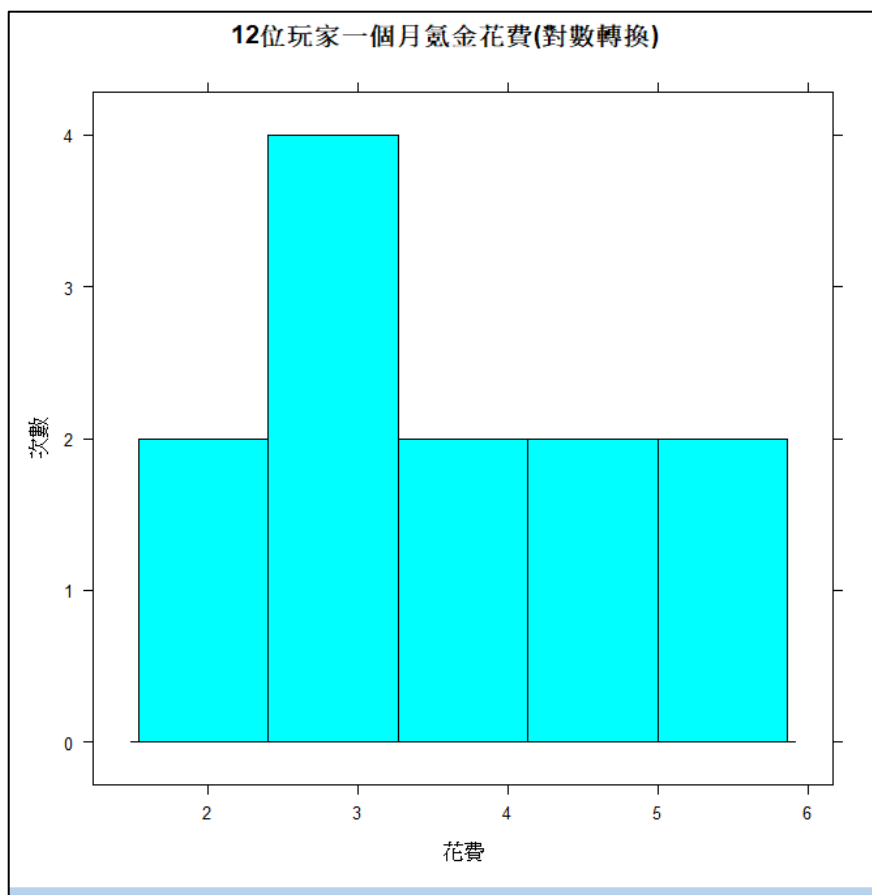
- P 值 <0.05 ，
因此為非常態分佈。

```
> shapiro.test(Q4$NT.dollar)
```

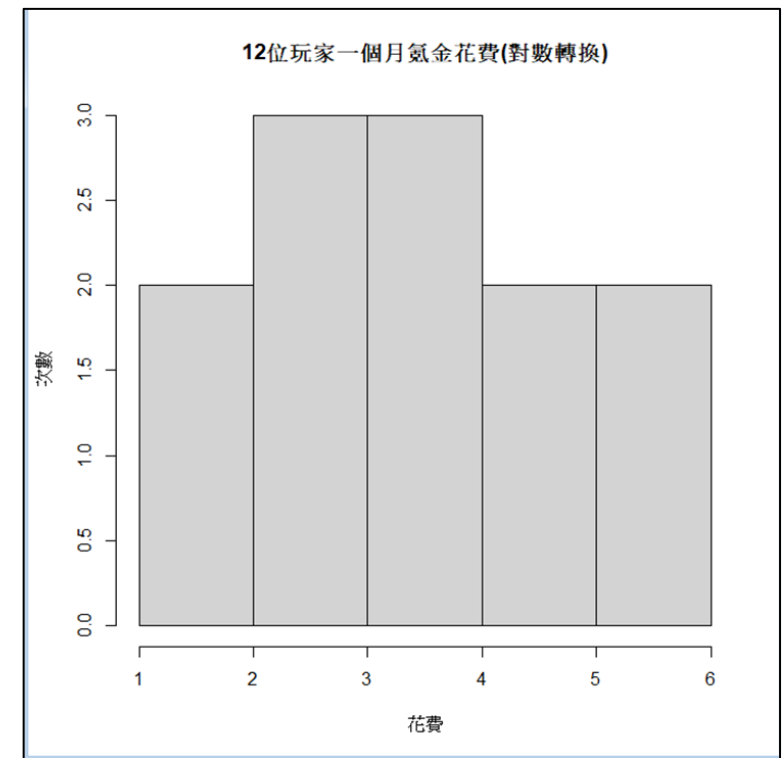
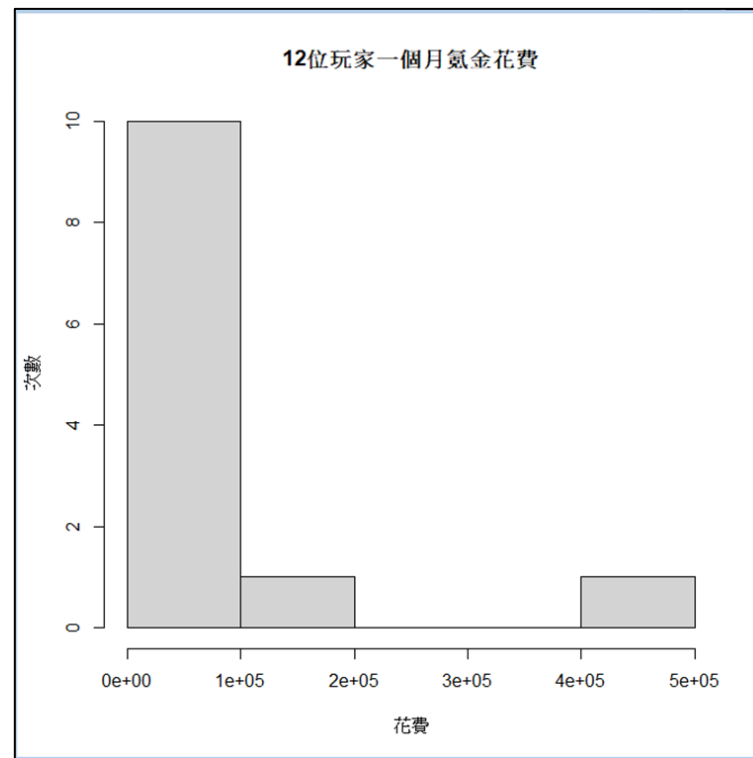
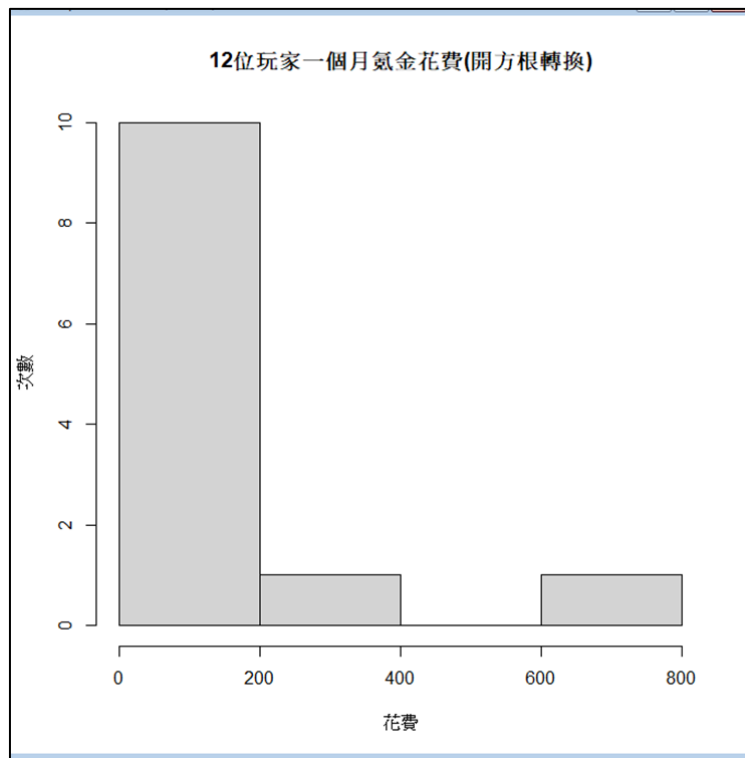
Shapiro-Wilk normality test

data: Q4\$NT.dollar

W = 0.48475, p-value = 1.367e-05



第四題-2



第五題 (1)

- P值 <0.05 ，所以資料有差別要進行事後檢定。從事後檢定可知三年的氣溫互相有差異。

```
> summary(aov)
              Df Sum Sq Mean Sq F value Pr(>F)
factor(yearid)  2   1978    988.9   43.87 <2e-16 ***
Residuals      2032  45809     22.5
---
```

```
> leveneTest(Q5_1$溫度, Q5_1$yearid, center=mean)
Levene's Test for Homogeneity of Variance (center = mean)
              Df F value    Pr(>F)
group         2  55.657 < 2.2e-16 ***
              2032
```

```
> res<-gamesHowellTest(aov)
> summary(res)

Pairwise comparisons using Games-Howell test

data: 溫度 by factor(yearid)
alternative hypothesis: two.sided
P value adjustment method: none
H0
              q value    Pr(>|q|)
2 - 1 == 0 -16.055 2.1971e-13 ***
3 - 1 == 0  -3.820  0.01911  *
3 - 2 == 0  11.786 < 2.22e-16 ***
---
```

第五題 (2)

- P值 <0.05 ，所以資料有差別需要進行事後檢定。從事後檢定可知台南市分別與其他兩個縣市有差別。

```
> summary(aov2)
              Df Sum Sq Mean Sq F value    Pr(>F)
factor(cityid)  2     447   223.67    8.387 0.000267 ***
Residuals      429   11441    26.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

```
> leveneTest(Q5_2$溫度, Q5_2$cityid, center=mean)
Levene's Test for Homogeneity of Variance (center = mean)
              Df F value    Pr(>F)
group        2   6.3864 0.001849 **
            429
---
```

```
> summary(res2)

Pairwise comparisons using Games-Howell test

data: 溫度 by factor(cityid)
alternative hypothesis: two.sided
P value adjustment method: none
H0
              q value    Pr(>|q|)
2 - 1 == 0    0.782    0.845274
3 - 1 == 0    6.007 8.2173e-05 ***
3 - 2 == 0    3.663    0.028814 *
```

第六題

- P值小於0.05，所以抽菸對罹癌有影響

```
7. 又抽菸 10 90  
> Q6 <- data.frame(row.names=c("抽菸", "沒抽菸"), 罹癌= c(30,10), 沒罹癌 = c(70,90))  
> Q6  
      罹癌 沒罹癌  
抽菸    30    70  
沒抽菸   10    90  
> chisq.test(Q6)  
  
      Pearson's Chi-squared test with Yates' continuity correction  
  
data:  Q6  
X-squared = 11.281, df = 1, p-value = 0.0007829
```