Question1

2群有高血壓 收縮壓

Consider two groups of people who receive treatments; one group receives the therapy while the other gets the placebo. We have known that the average SPB for all of them is 145 (mmHg) before they take the treatment. Moreover, according to the American Heart Association, the regular SBP would be around to 120 (mmHg). 正常人

Q1 (10 points) Can we assume each group follows the normal distribution? 是用Q-Q Plot

Q2 (10 points) Which group receives the therapy and which is not? Why? (Using the testing to explain) 哪群人得到治療

Q2 A: treat 1 病人得到治療

Q3 (5 points) Does the therapy work? 是否有功效

Q3 A=此次治療有功效

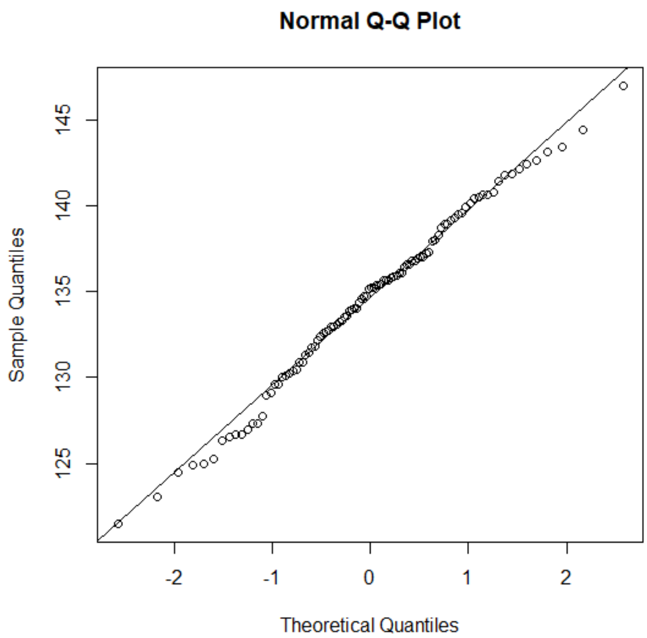
Q4 (5 points) Does the therapy lead the patient's SBP to the regular value? SBP是否能接受

hint Using the T-test.

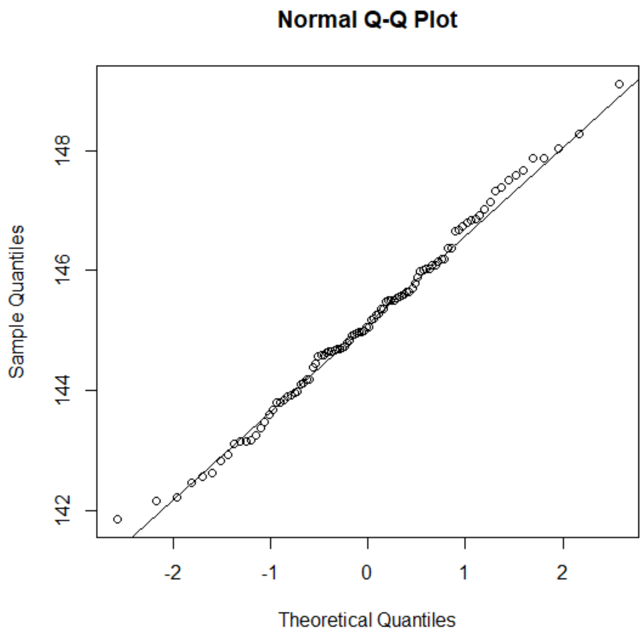
Q4: A=此次治療無法讓病人回到正常血壓

Q1

Treat1 組病人



Treat2 組病人



Comment:

1.By QQ plot , we know treat1 follow the normal distribution.(treat1 組病人的平均收縮壓服從常態分配。)

2.By QQ plot , we know treat2 follow the normal distribution. (treat2 組病人的平均收縮壓服從常態分配。)

Code:

```
#清理物件
rm(list=ls(all=TRUE))
data_1=read.table("F:\\統計方法\\data\\data1.txt",head=TRUE)#head=TRUE 宣告第一列為表頭

#顯示前 5 筆資料
head(data_1)

#Question1
##Q1
X_treat1=data_1$treat_1
X_treat1
qqnorm(X_treat1)
qqline(X_treat1)

X_treat2=data_1$treat_2
X_treat2
qqnorm(X_treat2)
qqline(X_treat2)
```

Q2

Let T_1 =the sample mean SPB of patients in treatment 1.

T_2 =the sample mean SPB of patients in treatment 2.

By the sample mean of T_1 ≈134.5

By the sample mean of T_2 ≈145.2

Comment:

(i)

Therefore, we consider H_0:$\mu_{T_1} \geq 145$(with H_1:$\mu_{T_1} < 145$)population of T_1.

And, with the t-test

By the p-value < 2.2e-16 <<< α =0.05 , which means…

Reject H_0. It is statistically significant to consider $\mu_{T_1} < 145$.

Result:

```
One Sample t-test
data:    X_treat1
t = -20.107, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is less than 145
95 percent confidence interval:
     -Inf 135.405
sample estimates:
mean of x
  134.5414
```

(ii)

Therefore, we consider H_0:$\mu_{T_2} \geq 145$(with H_1:$\mu_{T_2} < 145$)population of T_2.

And, with the t-test

By the p-value = 0.8439 >>> α =0.05 , which means…

Don't reject H_0. It is not statistically significant to consider $\mu_2 < 145$.($\therefore \mu_2 \geq 145$)

Result:

```
One Sample t-test
data:    X_treat2
t = 1.0159, df = 99, p-value = 0.8439
alternative hypothesis: true mean is less than 145
95 percent confidence interval:
      -Inf 145.4042
sample estimates:
mean of x
  145.1534
```

Code:

```
#Q2
#n=100
X_bar_1=mean(X_treat1)
X_bar_1
X_bar_2=mean(X_treat2)
X_bar_2
#H1:mu<145
t.test(X_treat1,mu=145,alternative='less',df=99)
t.test(X_treat2,mu=145,alternative='less',df=99)
```

Q3

從 Q2 可知，treatment 1 的為受到治療的病人;

treatment 2 則是使用安慰劑的病人。

所以，可執行假設檢定確認 treatment 1 病人的平均收縮壓是否顯著低於 treatment 2 病人的平均收縮壓，由此結果推論此藥物是否有療效。

Comment:

Therefore, we consider H_0:$\mu_{T_1} \geq \mu_{T_2}$(with H_1:$\mu_{T_1} < \mu_{T_2}$)population of $T_1$and T_2.

And, with two sample t-test

By the p-value < 2.2e-16 <<< α =0.05 , which means…

Reject H_0. It is statistically significant to consider $\mu_{T_1} < \mu_{T_2}$.

So, the therapy is work.(此次治療有功效)

Result:

```
        Welch Two Sample t-test

data:   X_treat1 and X_treat2
t = -19.593, df = 115.57, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -9.713959
sample estimates:
mean of x mean of y
  134.5414   145.1534
```

Code:

```
#Q3
t.test(X_treat1,X_treat2,alternative='less',df=99)
```

Q4

Comment:

(i)

Therefore, we consider H_0:$\mu_{T_1} = 120$(with H_1:$\mu_{T_1} \neq 120$)population of T_1.

And, with the t-test

By the 95 percent confidence interval: (133.5093 , 135.5735), then

$\mu_{T_1} = 120 \notin$ (133.5093 , 135.5735) , which means…

Reject H_0. It is statistically significant to consider $\mu_{T_1} \neq 120$.

Result:

```
            One Sample t-test

data:   X_treat1
t = 27.956, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 120
95 percent confidence interval:
 133.5093 135.5735
sample estimates:
mean of x
 134.5414
```

(ii)

Therefore, we consider H_0:$\mu_{T_2} = 120$(with H_1:$\mu_{T_2} \neq 120$)population of T_2.

And, with the t-test

By the 95 percent confidence interval: (144.8538 , 145.4531), then

$\mu_{T_2} = 120 \notin (144.8538 , 145.4531)$, which means…

Reject H_0. It is statistically significant to consider $\mu_{T_2} \neq 120$.

Result:

```
            One Sample t-test

data:   X_treat2
t = 166.55, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 120
95 percent confidence interval:
 144.8538 145.4531
sample estimates:
mean of x
 145.1534
```

Code:

```
#Q4
#H1:mu 不等於 120
t.test(X_treat1,mu=120,df=99)
t.test(X_treat2,mu=120,df=99)
```

Question2

The government wishes to investigate whether the average usage of the public facility has increased following the recent renovation of the activity center. Before its renovation, the average daily user count was 3.

Q1 (15 points) Calculate the 95% CI for inferring the true expected number of users under the following ideas:

- Assume the data follows the Poisson distribution. 3種方法算C.I.
- Using the central limiting theorem
- Using the T-distribution approximation

Q2 (10 points) Conduct a test to assess the impact of the renovation. 成效如何 $H_1: \mu > 3$

$\lambda = 3$

Q1

(i) Poisson:

$$\left(\hat{\lambda} \pm \frac{1.96\sqrt{\hat{\lambda}}}{\sqrt{n}}\right)$$

And $\hat{\lambda} = \overline{X} = 4.986301$, n=365, then

$$\left(\hat{\lambda} \pm \frac{1.96\sqrt{\hat{\lambda}}}{\sqrt{n}}\right) = (4.757215\,, 5.125388)$$

(ii) Central Limiting Theorem(C.L.T.):

$$\left(\overline{X} \pm \frac{1.96 \times S}{\sqrt{n}}\right)$$

And $\overline{X} = 4.986301$, S=$\sqrt{\sum_{i=1}^{n} \frac{(x_i - \hat{x})^2}{n-1}} = 2.350433$, n=365, then

$$\left(\overline{X} \pm \frac{1.96 \times S}{\sqrt{n}}\right) = (4.745168\,, 5.227435)$$

(iii) T-distribution approximation:

$$\left(\overline{X} \pm t_{n-1,\ \frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}\right)$$

And $\overline{X} = 4.986301$, S=$\sqrt{\sum_{i=1}^{n} \frac{(x_i - \hat{x})^2}{n-1}} = 2.350433$, n=365, $\frac{\alpha}{2} = 0.025$,then

$$\left(\overline{X} \pm t_{n-1,\ \frac{\alpha}{2}} \times \frac{S}{\sqrt{n}}\right) = (4.744368\,, 5.228235)$$

Result:

```
Poisson:
4.757215 5.215388
Central Limiting Theorem(C.L.T.):
4.745168 5.227435
T-distribution approximation:
                One Sample t-test

data:    x_q2
t = 0, df = 364, p-value = 1
alternative hypothesis: true mean is not equal to 4.986301
95 percent confidence interval:
  4.744368 5.228235
sample estimates:
mean of x
  4.986301
```

Code:

```
#清理物件
rm(list=ls(all=TRUE))
data_2=read.table("G:\\統計方法\\data\\data2.txt",head=TRUE)#head=TRUE 宣告第一列為表頭

#顯示前 5 筆資料
head(data_2)

#Question2
#Q1
dim(data_2)
x_q2=data_2$x
x_q2

#Poisson
la=mean(x_q2)
la
CI_la=c(la-1.96*(la/365)^0.5,la+1.96*(la/365)^0.5)
CI_la

#CLT
s_clt=sd(x_q2)
s_clt
X_bar_CI=c(la-1.96*sd(x_q2)/sqrt(365),la+1.96*sd(x_q2)/sqrt(365))
X_bar_CI

#t-teat
t.test(x_q2,mu=la,df=364)
```

Q2

Let μ =the average usage of the public facility

By the sample mean of λ =4.986301.

Before its renovation, the average daily user count was 3.

Comment:

Therefore, we consider H_0:$\mu \leq 3$(with H_1:$\mu > 3$)population of μ.

(i)And, by n=365 we can use Central Limiting Theorem(C.L.T.) with normal distribution

By the p-value = 0 <<< α =0.05, which means…

Reject H_0. It is statistically significant to consider $\mu > 3$.

Result:

```
> #假設 H0:mu<=3    H1:mu>3
> X_bar_z=(la-3)/(sd(x_q2)/sqrt(365))
> X_bar_z
[1] 16.14521
> #P-value:
> 1-pnorm(X_bar_z,0,1)
[1] 0
```

(ii)And, with the t-test

By the p-value < 2.2e-16 <<< α =0.05 , which means…

Reject H_0. It is statistically significant to consider $\mu > 3$.

Result:

```
        One Sample t-test

data:   x_q2
t = 16.145, df = 364, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 4.783423        Inf
sample estimates:
mean of x
 4.986301
```

So, by (i) and (ii) the average usage of the public facility has increased following the recent renovation of the activity center. (在翻新工程後，活動中心公共設施的平均使用率有顯著增加，表示有成效。)

Code:

```
#Q2
#假設 H0:mu<=3    H1:mu>3

#CLT
X_bar_z=(la-3)/(sd(x_q2)/sqrt(365))
X_bar_z
#P-value:
1-pnorm(X_bar_z,0,1)

#t-test
t.test(x_q2,mu=3,alternative='greater', df=364)
```

Question3

Consider that we flip a coin 150 times and then record the result as 1 for heads (otherwise, it is recorded as 0). n=150

- Q1 (5 points) Is this coin fair? $H_0 = p=0.5$, $H_1 = p \neq 0.5$ $\hat{p}=0.57333$
- Q2 (10 points) What constitutes your rejection region? hint $X \sim Bin(1,p)$ then $Var(X) = p(1-p)$. 拒絕域 Var(x)=0.246204
- Q3 (5 points) Calculate the p-values? 算P-value 用二項 p-value
- Q4 (5 points) If I claim that the true probability of the coin showing heads is 0.53, does this align with your inference from Q1? $\hat{p}=0.53$
  - (10 points) If it doesn't align with your belief, what is the reason? 是否會與Q1有矛盾
  - (10 points) Any suggestion would lead to the testing working to our belief of Q1.

n=150

Let p = the probability of the coin showing heads.

<u>Q1</u>

We want to know this coin is fair or not. If this coin is fair, then the probability of showing heads is 0.5.

Therefore, we consider H_0:p $= 0.5$(with H_1:p $\neq 0.5$)probabilty of p.

(i)And, with the t-test

By the p-value = 0.07233 >>> α =0.05 , which means…

Don't reject H_0. It is not statistically significant to consider p $\neq$ 0.5.(This coin is fair.)

Result:

```
        One Sample t-test

data:  x_q3
t = 1.8099, df = 149, p-value = 0.07233
alternative hypothesis: true mean is not equal to 0.5
95 percent confidence interval:
 0.4932680 0.6533987
sample estimates:
mean of x
0.5733333
```

(ii)And, by n=150 we can use Central Limiting Theorem(C.L.T.) with normal distribution

$\overline{X}$=0.573333, var(X)= $\sum_{i=1}^{n}\frac{(x_i-\hat{x})^2}{n-1} = 0.246264$, S=$\sqrt{\text{var(X)}}$= 0.4962499, then

$$Z^* = \frac{\overline{X}-\overline{X}_{underH_0}}{\frac{S}{\sqrt{n}}} = \frac{0.5733333-0.5}{\frac{0.4962499}{\sqrt{150}}} = 1.796292$$

By the p-value = 0.07244801 >>> α =0.05, which means…

Don't reject H_0. It is not statistically significant to consider $p \neq 0.5$.(This coin is fair.)

Result:

```
> #(ii)CLT
> p_h=mean(x_q3)
> p_h
[1] 0.5733333
> p_h_v=var(x_q3)
> p_h_v
[1] 0.246264
> p_h_z=(p_h-0.5)/(0.5/sqrt(150))
> p_h_z
[1] 1.796292
> #雙尾檢定要乘上兩倍的(1-P(X<=z*))
> #因 p_h_z 可能為負值，所以要取絕對值使用 abs(p_h_z)
> 2*(1-pnorm(abs(p_h_z),0,1))
[1] 0.07244801
```

Code:

```
#清理物件
rm(list=ls(all=TRUE))
data_3=read.table("G:\\統計方法\\data\\data3.txt",head=TRUE)#head=TRUE 宣告第一列為表頭

#顯示前 5 筆資料
head(data_3)

#Question3
#Q1
dim(data_3)
x_q3=data_3$x
x_q3

#(i)t-test
t.test(x_q3,mu=0.5, df=149)

#(ii)CLT
p_h=mean(x_q3)
p_h
p_h_v=var(x_q3)
p_h_v
p_h_z=(p_h-0.5)/(0.5/sqrt(150))
p_h_z
#雙尾檢定要乘上兩倍的(1-P(X<=z*))
#因 p_h_z 可能為負值，所以要取絕對值使用 abs(p_h_z)
2*(1-pnorm(abs(p_h_z),0,1))
```

Q2

By Q1, we consider H_0:p $= 0.5$(with H_1:p $\neq 0.5$)probabilty of p.

接受域=$\left(P_{under\ H_0} \pm 1.96 \times \frac{P_{under\ H_0}}{\sqrt{n}}\right)$

And $P_{under\ H_0} = 0.5$, n=150, then

接受域= $\left(P_{under\ H_0} \pm 1.96 \times \frac{P_{under\ H_0}}{\sqrt{n}}\right) = (0.4199833\ , 0.5800167)$

$P_{under\ H_0} = 0.5$在接受域之中，所以不拒絕H_0。

表示沒有充分證據證明此硬幣不公正。(與 Q1 的結果一致)

Result:

```
> #Q2
> p_h_Reg=c(0.5-1.96*0.5/sqrt(150),0.5+1.96*0.5/sqrt(150))
> p_h_Reg
[1] 0.4199833 0.5800167
```

Code:

```
#Q2
p_h_Reg=c(0.5-1.96*0.5/sqrt(150),0.5+1.96*0.5/sqrt(150))
p_h_Reg
```

Q3

By Q1, we consider H_0:p $= 0.5$(with H_1:p $\neq 0.5$)probabilty of p.

And, with binomial distribution

$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{150} x_i}{150}$=0.5733333,

Var(X)= $\hat{p} \times (1-\hat{p}) = 0.5733333 \times (1-0.5733333) = 0.2446222$

SD(X)=$\sqrt{\mathrm{Var(X)}} = \sqrt{0.2446222} = 0.494593$, then

$Z^* = \frac{\hat{P}-P_{underH_0}}{\frac{SD(X)}{\sqrt{n}}} = \frac{0.5733333-0.5}{\frac{0.494593}{\sqrt{150}}} = 1.81593$

By the p-value = 0.06938111 >>> α =0.05, which means…

Don't reject H_0. It is not statistically significant to consider p $\neq 0.5$.(This coin is fair.)

Result:

```
> #Q3
> #mean
> sum(x_q3)
[1] 86
> p_hat=sum(x_q3)/150
> p_hat
[1] 0.5733333
> #var(X)
> p_hat_var=p_hat*(1-p_hat)
> p_hat_var
[1] 0.2446222
> #SD(X)
> p_hat_sd=p_hat_var^0.5
> p_hat_sd
[1] 0.494593
> p_h=mean(x_q3)
> p_h
[1] 0.5733333
> p_h_z=(p_hat-0.5)/(p_hat_sd/sqrt(150))
> p_h_z
[1] 1.81593
> #雙尾檢定要乘上兩倍的(1-P(X<=z*))
> #因 p_h_z 可能為負值，所以要取絕對值使用 abs(p_h_z)
> 2*(1-pnorm(abs(p_h_z),0,1))
[1] 0.06938111
```

Code:

```
#Q3

#mean
sum(x_q3)
p_hat=sum(x_q3)/150
p_hat

#var(X)
p_hat_var=p_hat*(1-p_hat)
p_hat_var

#SD(X)
p_hat_sd=p_hat_var^0.5
p_hat_sd
p_h=mean(x_q3)
p_h
p_h_z=(p_hat-0.5)/(p_hat_sd/sqrt(150))
p_h_z
#雙尾檢定要乘上兩倍的(1-P(X<=z*))
#因 p_h_z 可能為負值，所以要取絕對值使用 abs(p_h_z)
2*(1-pnorm(abs(p_h_z),0,1))
```

Q4

當此硬幣實際出現正面的機率為 0.53 時，與 Q1 之檢定結果(為公正硬幣)相互矛盾!!!!

(i)因檢定仍有錯誤之機率(犯型 I 誤差之可能性)，其原因可能為實際出現正面之機率值與檢定機率太接近導致。

(ii)降低犯錯機率的方法為增加實驗次數(n)。