# 107-1 Statistics
# 第二次期中考講解

廖晧宇、吳家禎、賴冠宇

2018/11/30

**1-1.** Vehicle speeds at a certain highway location are believed to have approximately a normal distribution with mean $\mu$ = 60 mph and standard deviation $\sigma$ =6 mph. The speeds for a randomly selected sample of $n$=36 vehicles will be recorded.

**(1)** Give numerical values for the mean and standard deviation of the sampling distribution of possible sample means for randomly selected samples of $n$=36 from the population of vehicle speeds. [7%]

```
mu = 60
sigma = 6
n = 36

mean = mu; mean
```

```
## [1] 60
```

```
sd = sigma / sqrt(n); sd
```

```
## [1] 1
```

**Ans:**

**The mean of the sampling distribution is 60. The standard deviation of the sampling distribution is 1.**

**1-1.** Vehicle speeds at a certain highway location are believed to have approximately a normal distribution with mean $\mu$ = 60 mph and standard deviation $\sigma$ =6 mph. The speeds for a randomly selected sample of *n*=36 vehicles will be recorded.

**(2)** For a random sample of *n*=36 vehicles, there is about a 95% chance that the mean vehicle speed in the sample will be between *A* and *B* mph. What are the estimated values for *A* and *B*? [7%]

```
n = 36
t = qt(1-0.025, df = n-1)
A = mu - t*sd; A
```

```
## [1] 57.96989
```

```
B = mu + t*sd; B
```

```
## [1] 62.03011
```

**Ans:**
**A = 57.96989, B = 62.03011**

**1-1.** Vehicle speeds at a certain highway location are believed to have approximately a normal distribution with mean $\mu$ = 60 mph and standard deviation $\sigma$ =6 mph. The speeds for a randomly selected sample of $n$=36 vehicles will be recorded.

**(3)** Sample speeds for a random sample of 36 vehicles are measured at this location, and the sample mean is 66 mph. Given the answer to part (1), explain whether this result is consistent with the belief that the mean speed at this location is μ = 60 mph. [6%]

*From the sampling distribution of $\mu$ = 60 mph :*

```
low.3sd = mean - 3*sd; low.3sd
```

```
## [1] 57
```

```
upp.3sd = mean + 3*sd; upp.3sd
```

```
## [1] 63
```

```
x.bar = 66
```

or

*From the sample mean of 66 mph:*

```
x.bar = 66

# 99% confidence interval
t99 = qt(1-0.005, df = n-1)

low.99 = x.bar - t99*sd; low.99
```

```
## [1] 63.27619
```

```
upp.99 = x.bar + t99*sd; upp.99
```

```
## [1] 68.72381
```

# 1-1. (3)

**Cont.**

**Ans:**

**The result with sample mean = 66 mph is not consistent with the answer in part (1).**

***From the sampling distribution of*** **$\mu$ = 60 mph *:***

The probability of a sample mean greater than the value $\mu+3\sigma=63$mph is 0.15%. However, the sample mean of 66 mph is even greater than 63 mph, which infers a really small chance of having a sample mean of 66 mph.

***From the sample mean of 66 mph:***

The 99% confidence interval of the sample mean = 66 mph is 63.27619 ~ 68.72381 and does not cover $\mu$ = 60 mph. This result could infer that there's a really small chance (almost impossible) of having a sample mean of 66 mph based on the answer in part (1).

**1-2.** Students in a statistics class at National Taiwan University were asked, "Would you date someone with a great personality even though you did not find them attractive?" The results were that 61.1% of 131 women answered "yes," while 42.6% of 61 men answered "yes".

**(1)** Compute a 99% confidence interval for the difference in proportions. [10%]

```
p1.hat = 0.611
n1 = 131
p2.hat = 0.426
n2 = 61

se = sqrt(p1.hat*(1-p1.hat)/n1 + p2.hat*(1-p2.hat)/n2); se
```

```
## [1] 0.07630815
```

```
z = qnorm(1-0.005)
```

```
low.99 = (p1.hat-p2.hat) - z*se; low.99
```

```
## [1] -0.01155677
```

```
upp.99 = (p1.hat-p2.hat) + z*se; upp.99
```

```
## [1] 0.3815568
```

**Ans:**

**The 99% confidence interval is -0.01155677 ~ 0.3815568.**

**1-2.** Students in a statistics class at National Taiwan University were asked, “Would you date someone with a great personality even though you did not find them attractive?” The results were that 61.1% of 131 women answered “yes,” while 42.6% of 61 men answered “yes”.

**(2)** Would a 95% confidence interval for the difference in proportions be wider or narrower than the 99% confidence interval computed in part (1)? Explain. [5%]

**Ans:**

**Narrower. By definition, the 95% confidence interval would cover a smaller probability than the 99% confidence interval; therefore, the 95% confidence interval would be narrower.**

**1-2.** Students in a statistics class at National Taiwan University were asked, “Would you date someone with a great personality even though you did not find them attractive?” The results were that 61.1% of 131 women answered “yes,” while 42.6% of 61 men answered “yes”.

**(3)** Does the interval computed in part (1) include the value 0? What does this tell us about whether there is a difference between college men and women for this question? [5%]

**Ans:**

**Yes. The confidence interval in (1) does include 0. From the sample, we wouldn’t be able to conclude that there’s a difference between college men and women.**

**1-3.** A random sample of five college women was asked for their own heights and their mothers' heights. The researchers wanted to know whether college women are taller on average than their mothers. The results (in inches) follow:

**(1)** Defined the parameter of interest in this situation. And, find a 95% confidence interval for the parameter you defined. [10%]

| Pair | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Daughter | 66 | 64 | 64 | 69 | 66 |
| Mother | 66 | 62 | 65 | 66 | 63 |

```
daugh = c(66,64,64,69,66)
mom = c(66,62,65,66,63)
diff = daugh - mom
n = 5

d.bar = mean(diff)
d.sd = sd(diff)

se = d.sd / sqrt(n)

t = qt(1-0.025, df = n-1)

CI.low = d.bar - t*se; CI.low
```

```
## [1] -0.8555947
```

```
CI.upp = d.bar + t*se; CI.upp
```

```
## [1] 3.655595
```

**Ans:**

**The parameter of interest is $\mu d$, the mean of difference of each college woman's height and her mother's height. The 95% confidence interval is -0.8555947 ~ 3.655595.**

**1-3.** A random sample of five college women was asked for their own heights and their mothers' heights. The researchers wanted to know whether college women are taller on average than their mothers. The results (in inches) follow:

**(2)** Using the interval in part (1), write a sentence or two about the relationship between women students' heights and their mothers' heights for the population. Your explanation should be written to be understood by someone with no training in statistics. [5%]

**Ans:**
**Because the confidence interval covers 0, we can say that college women were not taller than their mother.**

**1-3.** A random sample of five college women was asked for their own heights and their mothers' heights. The researchers wanted to know whether college women are taller on average than their mothers. The results (in inches) follow:

**(3)** Explain two different things that the researchers could have done to obtain a narrower interval than the one you found in part (1). [5%]

**Ans:**

**The researcher could either choose a smaller confidence level or enlarge the sample size.**

**1-4.** In the table below are data collected in 2015 on ear piercing and tattoos for male NTU students. Assuming that these men are a random sample of college men, test the hypothesis that college men with at least one ear pierced are more likely to have a tattoo than college men with no ears pierced. Show all five steps of the hypothesis testing procedure. Use the significance level $\alpha = 0.05$ as threshold. [20%]

| Ear piercing | Tattoo? No | Yes | Total |
|---|---|---|---|
| No | 381 | 43 | 424 |
| Yes | 99 | 42 | 141 |
| Total | 480 | 85 | 565 |

Step 1. $H_0: p_1 - p_2 \leq 0$
$H_1: p_1 - p_2 > 0$ 右尾檢定

Step 2.

$$n_1 \times p_1 \geq 10$$
$$n_1 \times (1 - p_1) \geq 10$$
$$n_2 \times p_2 \geq 10$$
$$n_2 \times (1 - p_2) \geq 10$$

```
p1.hat = 42/141
p2.hat = 43/424
n1 = 141
n2 = 424

n1*p1.hat
```

```
## [1] 42
```

```
n1*(1-p1.hat)
```

```
## [1] 99
```

```
n2*p2.hat
```

```
## [1] 43
```

```
n2*(1-p2.hat)
```

```
## [1] 381
```

```
p.hat = (42+43) / (141+424)

se = sqrt(p.hat*(1-p.hat)*(1/n1 + 1/n2))

z = (p1.hat - p2.hat) / se
z
```

```
## [1] 5.652684
```

```
z.star = qnorm(0.05, lower.tail = F)
z.star
```

```
## [1] 1.644854
```

**1-4.** **Cont.**

Step 3.

```r
p.value = pnorm(z, lower.tail = F)
p.value
```

```
## [1] 7.898095e-09
```

Step 4.

```r
if (abs(z) >= abs(z.star)) {
  print("Reject H0. ")
} else {
  print("Do not reject H0.")
}
```

```
## [1] "Reject H0. "
```

Step 5.

**Ans:**

**Yes, college men with at least one ear pierced are more likely to have a tattoo.**

**1-5.** In a study, the mean number of hangover symptoms was compared for students whose parents have alcohol problems and students whose parents do not. Researchers are interested in knowing if the mean number of hangover symptoms is higher for the population of students whose parents have alcohol problems than for the population whose parents do not. The sample statistics are as follows. Carry out the five steps of hypothesis testing using the unpooled procedure. Use the significance level $\alpha = 0.05$ as threshold. [20%]

| Group | Mean | Standard deviation |
| --- | --- | --- |
| Parental alcohol problems ($n_1$=282) | $\overline{x_1}$=5.9 | $s_1$=3.6 |
| No parental alcohol problems ($n_2$=945) | $\overline{x_2}$ =4.9 | $s_2$=3.4 |

Step 1. $H_0: \mu_1 - \mu_2 \leq 0$
$H_1: \mu_1 - \mu_2 > 0$ 右尾檢定

Step 2.

$n_1 \geq 30$
$n_2 \geq 30$

```
x1.bar = 5.9
x2.bar = 4.9
s1 = 3.6
s2 = 3.4
n1 = 282
n2 = 945

se = sqrt((s1^2)/n1 + (s2^2)/n2)
```

```
t = (x1.bar - x2.bar) / se
t
```

```
## [1] 4.145481
```

```
t.star = qt(0.05, df = n1-1, lower.tail = F)
```

# 1-5.

**Cont.**

Step 3.

```
p.value = pt(t, df = n1-1, lower.tail = F)
p.value
```

```
## [1] 2.246921e-05
```

Step 4.

```
if (abs(t) >= abs(t.star)) {
  print("Reject H0. ")
} else {
  print("Do not reject H0.")
}
```

```
## [1] "Reject H0. "
```

Step 5.

**Ans:**

**Yes, the number of hangover symptoms is higher for the population of students whose parents have alcohol problems.**

2.

假設現在是2020年。金門縣「公車捷運藍線(以下簡稱BRT-B)」在2019年7月通車，縣政府在2019年8月對沿線所有居民進行調查，沿線居民通勤使用**BRT-B的比例為0.20**。

(1)

BRT-B服務半年後，縣長認為若居民通勤使用BRT-B的比例沒有比上次調查結果還高的話，應該中止BRT-B服務。於是縣府交通局在2020年1月隨機取樣調查沿線**200位居民**，發現其中**有50位**居民通勤使用BRT-B。在顯著水準**$\alpha$=0.05**的條件下，交通局應該建議縣長中止BRT-B服務嗎？[7%]

p1=50/200=0.25

n1=200

**Step1: 設立假說**

$H_0$: p1 ≤ 0.2

$H_1$: p1 > 0.2

α=0.05 –> 單尾檢定

$$z = \frac{\text{sample estimate} - \text{null value}}{\text{null standard error}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

**Step2: 計算z-statistics**

必要條件:np、n(1-p)皆大於10

```
z=(p1-0.2)/sqrt(p0*(1-p0)/n1);z
```

z=1.767767

**Step3: 計算P值**

```
p_value=pnorm(z,lower.tail = F);p_value
```

p_value = 0.03854994

**Step4: 檢定結果**

p.value < α ，拒絕Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，交通局不應該建議縣長中止BRT-B服務

(2)

交通局也規劃了「公車捷運橘線(以下簡稱BRT-O)」，BRT-O沿線居民要求實現這個服務，縣長認為，若BRT-O沿線居民表達通勤會使用BRT-O的比例高過BRT-B沿線居民在2020年1月的使用比例，便實現這個服務。於是交通局在2020年2月隨機取樣調查BRT-O沿線**200位居民**，發現其中有**60位**居民表示通勤將使用BRT-O。假設路線沿線居民在兩條路線之間互相獨立，在顯著水準**$\alpha=0.05$**的條件下，交通局應該建議縣長實現BRT-O嗎？[7%]

$p_2=60/200=0.3$

$n_2=200$

**Step1: 設立假說**

$H_0$: p2-p1 ≤ 0

$H_1$: p2-p1 > 0

α=0.05 –> 單尾檢定

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$z = \frac{\text{sample statistic} - \text{null value}}{\text{null standard error}} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

**Step2: 計算z-statistics**

必要條件:np、n(1-p)皆大於10

```
p_hat=(p1*200+p2*200)/(200+200);p_hat
```

p_hat= 0.275

```
z=(p2-p1-0)/sqrt(p_hat*(1-p_hat)*((1/n1)+(1/n2)));z
```

z= 1.119785

**Step3: 計算P值**

```
p_value=pnorm(z,lower.tail = F);p_value
```

p_value = 0.1314027

**Step4: 檢定結果**

p_value > α ，接受Ho。

**Step5: 結論**

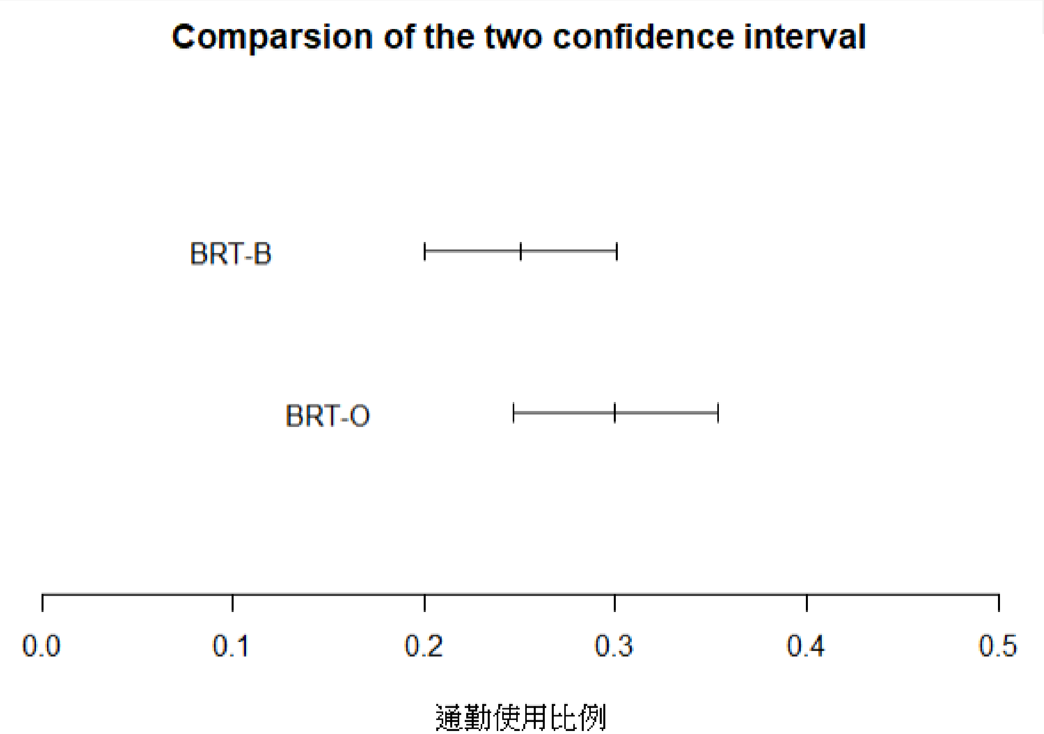
根據0.05顯著水準下的統計檢定，交通局應該建議縣長終止BRT-O

(3)

交通局需要以圖表向BRT-O沿線居民說明，請利用繪製信賴區間的方式，繪製2020年1月BRT-B沿線居民調查和2020年2月BRT-O沿線居民調查的使用比例信賴區間於同一張圖表，並與(2)小題答案作比較和說明。[6%]

```
#(3)畫圖
#先算BRT-B和BRT-O的信賴區間
#這題沒寫信心水準要多少，所以自己假設，令信心水準為90%

p_B=50/200
p_O=60/200

alpha=1-0.9
half.alpha=alpha/2

z=qnorm(half.alpha,lower.tail = F);z

p_B.upper=p_B+z*sqrt(p_B*(1-p_B)/200)
p_B.lower=p_B-z*sqrt(p_B*(1-p_B)/200)

p_O.upper=p_O+z*sqrt(p_O*(1-p_O)/200)
p_O.lower=p_O-z*sqrt(p_O*(1-p_O)/200)
```

```
plot(c(0,3),type='n',xlim=c(0,0.5),xlab='通勤使用比例',ylab='',axes=FALSE,main='Comparsion
     of the two confidence interval')
lines(c(p_B.lower,p_B.upper),c(2,2))
text(0.1,2,labels='BRT-B')
lines(c(p_B.lower,p_B.lower),c(1.95,2.05))
lines(c(p_B,p_B),c(1.95,2.05))
lines(c(p_B.upper,p_B.upper),c(1.95,2.05))

lines(c(p_O.lower,p_O.upper),c(1,1))
text(0.15,1,labels='BRT-O')
lines(c(p_O.lower,p_O.lower),c(0.95,1.05))
lines(c(p_O,p_O),c(0.95,1.05))
lines(c(p_O.upper,p_O.upper),c(0.95,1.05))

axis(1)
```




從圖片可知，在90%信心水準下，BRT-B跟BRT-O的信賴區間有許多重疊的部分，進一步支持第(2)小題的結論，BRT-O沒有顯著高於BRT-B的使用率

3.

台北市交通局隨機取樣調查「使用大眾運輸」通勤居民以及「自行開車」通勤居民各100位，記錄每位居民通勤時間，假設使用兩種通勤方式居民之間互相獨立，調查統計結果如下：

| 通勤方式 | 樣本數 | 通勤時間平均值 | 通勤時間標準差 |
|---|---|---|---|
| 使用大眾運輸 | 100 | 35 分鐘 | 20 分鐘 |
| 自行開車 | 100 | 30 分鐘 | 10 分鐘 |

mu1=35

mu2=30

n1=100

n2=100

sd1=20

Sd2=10

(1)

請問在95%信心水準下，交通局是否可以宣稱在台北市使用大眾運輸通勤時間高於自行開車通勤時間？[6%]

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

**Step1: 設立假說**

Ho: mu1-mu2 ≤ 0

H1: mu1-mu2 > 0

α=0.05 –> 單尾檢定

**Step2: 計算t-statistics**

必要條件:n大於30

```
t=(mu1-mu2-0)/sqrt(((sd1^2)/n1)+((sd2^2)/n2));t
```

t = 2.236068

**Step3: 計算P值**

```
p_value=pt(t,df=n1-1,lower.tail = F);p_value     #unpooled的df=小的n-1
```

p_value = 0.01379615

**Step4: 檢定結果**

p_value < α ，拒絕Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，交通局可以宣稱在台北市使用大眾運輸通勤時間高於自行開車通勤時間

(2)

為鼓勵居民使用大眾運輸系統，以達到城市永續發展目標，市政府推動大眾運輸改善計畫，計畫完成後，交通局再對之前調查的**同樣那100位「使用大眾運輸」通勤居民**，記錄每位居民通勤時間在計畫實施後相對於實施前的差距，發現差距平均值為-1分鐘、差距標準差為5分鐘。請問在**95%信心水準**下，交通局是否可以宣稱該大眾運輸改善計畫成功減少使用大眾運輸通勤居民的通勤時間？[7%]

**Testing Hypotheses about Mean of Paired Differences**

d_bar = -1

sd3 = 5

n = 100

**Step1: 設立假說**

$H_0$: mud ≥ 0

$H_1$: mud < 0

α=0.05 –> 左尾檢定

$$t = \frac{\text{sample mean} - \text{null value}}{\text{standard error}} = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$$

**Step2: 計算t-statistics**

必要條件:n大於30

```
t=(d_bar-0)/(sd3/sqrt(n));t
```

t = -2

**Step3: 計算P值**

```
p_value=pt(t,df=n-1,lower.tail = T);p_value    #df=n-1
```

p_value = 0.02411985

**Step4: 檢定結果**

p_value < α ，拒絕Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，交通局可以宣稱該大眾運輸改善計畫成功減少使用大眾運輸通勤居民的通勤時間

(3)

若第(2)小題中，交通局對之前調查的同樣那100位「使用大眾運輸」通勤居民，記錄計畫實施後的通勤時間，發現通勤時間的**平均值改變為32.5**分鐘、**標準差改變為10**分鐘，請問在**95%信心水準**下，交通局是否可以宣稱該大眾運輸改善計畫成功讓使用大眾運輸通勤居民跟自行開車通勤居民之通勤時間沒有差異？請由顯著水準與效應值這兩方面進行論述。[7%]

| 通勤方式 | 樣本數 | 通勤時間平均值 | 通勤時間標準差 |
|---|---|---|---|
| 使用大眾運輸 | 100 | 32.5 分鐘 | 10 分鐘 |
| 自行開車 | 100 | 30 分鐘 | 10 分鐘 |

mu1=32.5

mu2=30

n1=100

n2=100

sd1=10

sd2=10

**Step1: 設立假說**

$H_0$: mu1-mu2 = 0

$H_1$: mu1-mu2 ≠ 0

α=0.05 –> 雙尾檢定

$$t=\frac{\text{sample mean}-\text{null value}}{\text{standard error}}=\frac{(\bar{x}_1-\bar{x}_2)-0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

**Step2: 計算t-statistics**

必要條件:n大於30

```
t=(mu1-mu2-0)/sqrt(((sd1^2)/n1)+((sd2^2)/n2));t
```

t = 1.767767

**Step3: 計算P值**

```
p_value=2*pt(t,df=n1-1,lower.tail = F);p_value    #unpooled的df=小的n-1
```

p_value = 0.08018078

**Step4: 檢定結果**

p_value > α ，接受Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，交通局可以宣稱該大眾運輸改善計畫成功讓使用大眾運輸通勤居民跟自行開車通勤居民之通勤時間沒有差異

## 計算效應值

$$effectsize = t * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

```
effect.size=1.767767*sqrt((1/n1)+(1/n2));effect.size
```

**effect size=0.25，屬於small magnitude，也就是說大眾運輸通勤時間與自行開車通勤時間的差異不大**

4.

PENN STATE UNIVERSITY校方隨機取樣調查校內205位同學運動習慣，調查結果如資料檔PENNSTATE2.CSV，欄位S為性別：F女性、M男性，欄位T為每週運動量是否達到WHO標準：Y有、N無。請運用R與該資料檔，在**95%信心水準**下，分析以下問題：

```
setwd("D:/碩班二上/統計學_助教/老師上課內容/Lab")
UQ=read.csv(file="pennstate2.csv",header=T)
```

| | s | e | t | c | h | h.1 | l | f |
|---|---|---|---|---|---|---|---|---|
| 1 | F | 2 | N | 25 | 70.00 | 70.0 | 4.0 | O |
| 2 | M | 0 | N | 47 | 77.00 | 77.0 | 6.0 | N |
| 3 | M | 2 | N | 50 | 71.00 | 71.0 | 5.0 | O |
| 4 | F | 2 | N | 50 | 65.00 | 65.0 | 4.0 | O |
| 5 | M | 4 | Y | 50 | 74.00 | 72.0 | 5.0 | N |
| 6 | F | 9 | Y | 70 | 61.00 | 65.0 | 7.0 | O |
| 7 | M | 0 | N | 162 | 76.00 | 76.0 | 4.0 | S |
| 8 | M | 0 | N | 150 | 73.00 | 76.0 | 4.0 | N |
| 9 | F | 6 | N | 50 | 65.00 | 69.0 | 4.0 | N |
| 10 | F | 4 | N | 30 | 63.00 | 66.0 | 5.0 | N |

(1)

校方宣稱該校有超過20%的同學，每周運動量達到WHO標準，請問前述調查結果是否支持該項宣稱？[10%]

$$z = \frac{\text{sample estimate} - \text{null value}}{\text{null standard error}} = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

```
XT=xtabs(~t+s,data=UQ)
XT
FN=XT[1,1];FN
MN=XT[1,2];MN
FY=XT[2,1];FY
MY=XT[2,2];MY
n=FN+MN+FY+MY
```

```
t     F   M
  N 119  55
  Y  18  13
```

```
p=0.2
p1=(FY+MY)/(FN+MN+FY+MY);p1
```

**Step1: 設立假說**

H0: p1 ≤ 0.2

H1: p1 > 0.2

α=0.05 –> 單尾檢定

**Step2: 計算z-statistics**

必要條件:np、n(1-p)皆大於10

```
z1=(p1-p)/sqrt(p*(1-p)/n);z1
```

z=-1.746076

**Step3: 計算P值**

```
pvalue1=pnorm(z1,lower.tail = T);pvalue1
```

p_value = 0.0403989

**Step4: 檢定結果**

p.value < α=0.05 ，拒絕Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，調查結果支持校方宣稱該校有超過20%的同學，每周運動量達到WHO標準

(2)

校方又宣稱該校男同學每周運動量達到WHO標準的比例較女同學為高，請問前述調查結果是否支持該項宣稱？[10%]

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$z = \frac{\text{sample statistic} - \text{null value}}{\text{null standard error}} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

```
p_male=MY/(MN+MY)
p_female=FY/(FY+FN)
p_hat=(p_male*(MN+MY)+p_female*(FY+FN))/(MY+MN+FY+FN)
```

**Step1: 設立假說**

$H_0$: p_male-p_female ≤ 0

$H_1$: p_male-p_female > 0

α=0.05 –> 單尾檢定

**Step2: 計算z-statistics**

必要條件:np、n(1-p)皆大於10

```
z=(p_male-p_female-0)/sqrt(p_hat*(1-p_hat)*((1/(MN+MY))+(1/(FY+FN))));z
```

z=1.125024

**Step3: 計算P值**

```
p_value=pnorm(z,lower.tail = F)
```

p_value = 0.1302894

**Step4: 檢定結果**

p.value > α=0.05 ，接受Ho。

**Step5: 結論**

根據0.05顯著水準下的統計檢定，調查結果不支持校方所宣稱的”該校男同學每周運動量達到WHO標準的比例較女同學為高”