108 年度 警大研究所 統計學考前猜題

考前提醒：統計學歷年考題難度中等偏高，但僅需得到 60 分以上即可符合錄取資格，因此考試時建議先將會寫的題目先完成，再回頭解較難的題型。準備部分重點章節為卜瓦松分布、卡方檢定、常態分布、比例(%)的檢定。今年要注意迴歸出題的可能，但建議用 20%的心力準備迴歸題型即可。

一、 2017年自行車使用概況抽查指出，國內12歲以上民眾每週均會騎自行車人口占比約24%。

(1)詢問50名民眾，每周均會騎自行車人數的期望值與其95%信賴區間為何?

(2)若已知母體實際比例在30%至20%之間，而此調查想控制95%信賴區間的抽樣誤差在3%以內，請問需要多少樣本數?

(3)某一社區調查36名12歲以上民眾，有9名民眾每周均會騎自行車，此一社區民眾每週騎車比率是否高於一般民眾?

(4)如果你的結論其實是錯的，那是犯了什麼錯誤? (Type I or Type II?)，有什麼方

法可以降低這類錯誤?

(5)若此社區民眾實際騎車比率為30%，請問檢定力為何?

Ans:

(1)p=24%=0.24，每周均會騎車人數期望值即平均數=n×p=50×0.24=12

因為n×p=12且n×p×(1-p)=9.12皆大於5，所以此二項分配近似一個平均數為n×p=12且標準差為$\sqrt{n\times p\times(1-p)}\doteqdot 3$的常態分布

期望值的95%信賴區間=12±1.96×3=12±5.88=(6.12 , 17.88)

(2)誤差的計算是由樣本資料來計算，所以母體實際值的範圍並不影響計算。

$$\text{誤差}d = 0.03 = 1.96\times\sqrt{\frac{0.24\times 0.76}{n}} \doteqdot \frac{0.837}{\sqrt{n}}$$

$$\text{所以}n = \left(\frac{0.837}{0.03}\right)^2 = (27.9)^2 = 778.41$$

最少需要779筆樣本

(3)此社區民眾騎車比例$\hat{p}=\frac{9}{36}=0.25$，一般民眾騎車比例$p_0=0.24$

單一樣本比例檢定

H_0:此社區民眾每週騎車比率不高於一般民眾($P_0 \leq 0.24$)

H_1:此社區民眾每週騎車比率高於一般民眾($P_0 > 0.24$)

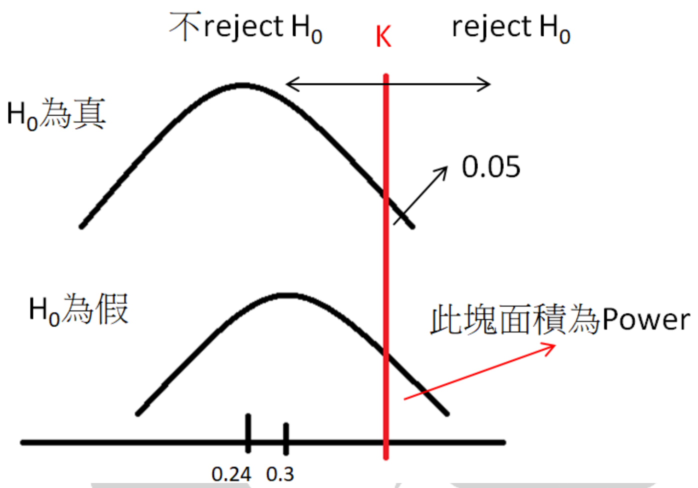
右單尾檢定，α=0.05時，拒絕域為$Z>Z_{0.05}$=1.645

$$Z=\frac{0.25-0.24}{\sqrt{\frac{0.24\times 0.76}{36}}}\fallingdotseq\frac{0.01}{0.07}\fallingdotseq 0.143$$

因為Z=0.143< $Z_{0.05}$=1.645，沒有位在拒絕域，無法拒絕H_0，認為此社區民眾每週騎車比率不高於一般民眾

(4)如果我們由(3)的結果所下的結論是錯的，代表實際上H_0為假，但我們的結果卻無法拒絕H_0，犯了型2錯誤。提高樣本數可以減少此類錯誤。

(5)此類題目請一定要畫圖輔助解題



在α=0.05，H_0為真(P=0.24)時，求K之值

$$P\left(Z>\frac{k-0.24}{\sqrt{\frac{0.24\times 0.76}{36}}}\right)=P(Z>1.645)$$

$$\text{所以}\frac{k-0.24}{\sqrt{\frac{0.24\times 0.76}{36}}}\fallingdotseq\frac{k-0.24}{0.07}=1.645\text{，得}k=1.645\times 0.07+0.24\fallingdotseq 0.355$$

從H_0為假時(P=0.3)求>K之面積即為Power

$$P(p>0.355)=P\left(Z>\frac{0.355-0.30}{\sqrt{\frac{0.355\times 0.645}{36}}}\right)\fallingdotseq P\left(Z>\frac{0.055}{0.08}\right)\fallingdotseq P(Z>0.69)$$

$=1-P(Z<0.69)=1-0.7549_{(\text{查表})}$=0.2451

檢定力為0.2451

二、 觀光地區巡迴巴士每10分鐘抵達站牌的次數如下：

1	2	0	2	0	0	2	2	0	0
2	3	2	0	0	3	2	2	0	0
2	3	3	3	0	3	3	3	0	1
1	2	3	0	3	2	1	1	1	2

(1)計算每10分鐘內巴士抵達次數的平均數、中位數、眾數、標準差、變異係數

(2)若抵達站牌次數符合卜瓦松分布，請問每10分鐘、每半小時與每小時預期會有幾台車抵達?

(3)10分鐘內都沒有車抵達，或是超過2台車抵達的機率各是多少?

(4)若抵達站牌時看到巴士剛離開，請問下一班車最少還要等多久?

Ans:

(1)建議可先把數據整理成以下形式

抵達台數	0	1	2	3
次數	12	6	12	10

$$平均數=\frac{\sum nx_i}{n}=\frac{12\times0+6\times1+12\times2+10\times3}{12+6+12+10}=\frac{60}{40}=1.5台/每10分鐘$$

因為n為偶數，所以中位數位於$\frac{n}{2}$與$\frac{n}{2}+1$位置之間，資料由小到大排列後，第20與第21個數據都是2，所以中位數為$\frac{2+2}{2}=2$

眾數為0跟2（各出現12次最多）

$$標準差=\frac{\sum n(x_i-\bar{x})^2}{n-1}$$

$$=\sqrt{\frac{12\times(0-1.5)^2+6\times(1-1.5)^2+12\times(2-1.5)^2+10\times(3-1.5)^2}{12+6+12+10-1}}$$

$$=\sqrt{\frac{27+1.5+3+22.5}{59}}=\sqrt{\frac{54}{59}}\fallingdotseq0.96$$

$$變異係數=\frac{標準差}{平均數}=\frac{0.96}{1.5}=0.64$$

(2)期望值即不同時間長度下的λ

每10分鐘：期望值=λ=1.5台/每10分鐘

每半小時：期望值=3×λ=4.5台/每半小時

每小時：期望值=6×λ=9台/每小時

(3)觀察時間10分鐘，λ=1.5台/每10分鐘

由查表可知$P(X=0)=\frac{e^{-1.5}\times 1.5^{0}}{0!}$=0.223（附表$\lambda$=1.5且c=0之數據）

10分鐘內都沒有車抵達的機率為0.223

$P(X>2)=1-P(X\leqq 2)=1-[P(X=0)+P(X=1)+P(X=2)]=1-[\frac{e^{-1.5}\times 1.5^{0}}{0!}+\frac{e^{-1.5}\times 1.5^{0}}{0!}+\frac{e^{-1.5}\times 1.5^{0}}{0!}]=1-0.809$（附表$\lambda$=1.5且c=2之數據)=0.191

10分鐘內超過2台車抵達的機率為0.191

(4)因為λ=1.5台/每10分鐘，所以兩台車之間的抵達時間間隔為$\frac{1}{\lambda}=\frac{10}{1.5}\fallingdotseq 6.67$分/每台，所以下一班車抵達至少還要等6.67分。

三、 107年法務部統計，第四級毒品案件新收人數較多的前五名城市資料如下：

	男性	女性
新北	64	31
桃園	96	12
台中	77	24
台南	66	65
高雄	127	42

請檢視α=0.05時，城市與性別這兩個因素是否有關聯性？需列出假說、計算過程與結論。

$X^2_{0.05,1}=3.84$，$X^2_{0.05,2}=5.99$，$X^2_{0.05,3}=7.82$，$X^2_{0.05,4}=9.49$，$X^2_{0.05,5}=11.07$

Ans:

H_0:城市與性別無關聯性

H_1:城市與性別有關聯性

卡方自由度為(5-1)×(2-1)=4，α=0.05，右單尾拒絕域為$X^2>X^2_{0.05,4}=9.49$

	男性	女性	邊際次數
新北	64(68)	31(27)	95
桃園	96(77)	12(31)	108
台中	77(72)	24(29)	101
台南	66(93)	65(38)	131
高雄	127(120)	42(49)	169
邊際次數	430	174	604

(因研究所考試不能用計算機，建議期望值四捨五入至整數或題目指定小數位數)

$$X^2=\frac{(64-68)^2}{68}+\frac{(96-77)^2}{77}+\frac{(77-72)^2}{72}+\frac{(66-93)^2}{93}+\frac{(127-120)^2}{120}+\frac{(31-27)^2}{27}+\frac{(12-31)^2}{31}+\frac{(24-29)^2}{29}+\frac{(65-38)^2}{38}+\frac{(42-49)^2}{49}$$

=0.24+4.69+0.35+7.84+0.41+0.59+11.65+0.86+19.18+1=46.81

因為$X^2=46.81>X^2_{0.05,4}=9.49$，位在拒絕域中，所以拒絕$H_0$，認為城市與性別有關聯性