金門大學 資訊工程系　103 學年度上學期　　機率統計　　期末考　　出題者：陳鍾誠
考生姓名：　　　　　　　　　　　　　學號：　　　　　　　　　　得分：

基本資料與公式：

常態分布的公式為 $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}$ 已知常態分布有下列特性： $P[-\sigma < N(\mu,\sigma) < \sigma] = 0.68$ $P[-1.96\sigma < N(\mu,\sigma) < 1.96\sigma] = 0.95$	**T** 分布公式為 $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 已知自由度為 **19** 的 **t** 分布有下列特性： $P[-1.7291 < T(df=19) < -1.7291] = 0.90$ $P[-2.093 < T(df=19) < -2.093] = 0.95$ $P[-2.5395 < T(df=19) < -2.5395] = 0.98$

(所有題目均必須寫出完整的推論過程)

簡答題 (30%)

(1) 何謂敘述統計？何謂推論統計？兩者有何不同？

「敘述統計」純粹計算像是「平均值、變異數」這些數字出來，而推論統計則根據這些數字進行母體模型的推論。

(2) 何謂點估計？何謂區間估計？兩者有何不同？
點估計用來估計單一個數值點，通常會採用不偏估計式，而區間估計則是估計一個區域，像是信賴區間就是常用的區間估計概念。

(3) 何謂中央極限定理？
從任意母體抽出 **n** 個獨立樣本，其平均值的分布會符合常態分布

$$\frac{x_1+x_2+...+x_n}{n} = \overline{x} \to N(\mu, \sigma/\sqrt{n})$$

(4) 請說明中央極限定理是如何被用在平均值的區間估計上面的? (兩者有何關係)
當樣本數 **n** 足夠大時 (通常 **20** 個以上就夠大了)， **n** 個樣本的平均值 會趨近於常態分布

$$\frac{\overline{x}-\mu}{\sigma/\sqrt{n}} \to Z$$

於是我們可以根據常態分佈來計算平均值的信賴區間，像是下列算式就可以用來計算 **95%** 信賴區間。

$$P[-1.96\sigma < N(\mu,\sigma) < 1.96\sigma] = 0.95$$

(5) 請說明變異數分析 **ANOVA** 的用途為何？
變異數分析是用一次檢定來分析 **k** 組樣本的平均值是否相等的方法，也就是檢定下列算式：

$$H0{:}\mu_1 = \mu_2 = ... = \mu_k$$

(6) 請說明迴歸分析的功能與用途為何？
當我們想知道兩個或多的變數之間是否符合某些方程式，我們就可以使用迴歸分析，例如我們可以用線性回歸找出變數 **x,y** 之間的線性關係，像是 **y=3x+5**

估計：(20%)

經由隨機抽樣我們取得下列互相獨立的樣本 **X**，且我們已計算出樣本變異數 **var(X)** 與平均數 **mean(X)**：

> X = c(2, 4, 5, 2, 3, 5, 5, 3, 4, 2, 3, 2, 3, 3, 2, 3, 6, 5, 3, 2)
> var(X)
[1] 1.607895
> mean(X)
[1] 3.35
> sd(X)
[1] 1.268028

請根據此一公式進行下列計算與推估：

(1) 假如已知母體標準差為 **1.2**，請估計母體平均數的 **95%** 信賴區間。

母體標準差已知，應採用常態分佈估計，**95%** 信賴區間為 $\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}+1.96\frac{\sigma}{\sqrt{n}}$ ，在本題為

$$3.35-1.96\frac{1.2}{\sqrt{20}} \le \mu \le 3.35+1.96\frac{1.2}{\sqrt{20}}$$

(2) 假如已知母體標準差為 **1.5**，請估計母體平均數的 **95%** 信賴區間。

母體標準差已知，應採用常態分佈估計，**95%** 信賴區間為 $\bar{X}-1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}+1.96\frac{\sigma}{\sqrt{n}}$ ，在本題為

$$3.35-1.96\frac{1.5}{\sqrt{20}} \le \mu \le 3.35+1.96\frac{1.5}{\sqrt{20}}$$

(3) 假如母體的標準差未知，請估計母體平均數的 **95%** 信賴區間。

母體標準差未知，應採用 **T** 分佈估計，**95%** 信賴區間為 $\bar{X}-2.093\frac{S}{\sqrt{n}} \le \mu \le \bar{X}+2.093\frac{S}{\sqrt{n}}$ ，在本題為

(7) 請問何種分布可以用來檢定或估計平均值的信賴區間？**(**假設母體的標準差已知**)**
常態分布

(8) 請問何種分布可以用來檢定或估計平均值的信賴區間？**(**假設母體的標準差未知**)**
T 分布

(9) 對於 **(5, 15)** 之間的均等分布，請寫出一個 **90%** 的信賴區間。
(5, 14)

(10) 對於 **(0, 2)** 之間的均等分布，請寫出一個 **95%** 的信賴區間。
(0.05, 1.95)

(11) 請說明檢定中的 **H0** 意義為何**?**
H0 為虛無假設，也就是可能被否定的假設，是對立假設的相反。
(12) 請說明檢定中的 **H1** 意義為何**?**
H1 為對立假設，也就是用來推翻虛無假設的另一方，當 **H1** 成立時就代表 **H0** 被推翻了。

$$3.35-2.093\frac{1.268}{\sqrt{20}}\leq\mu\leq 3.35+2.093\frac{1.268}{\sqrt{20}}$$

(4) 假如母體的標準差未知，請估計母體平均數的 **98%** 信賴區間。

母體標準差未知，應採用 **T** 分佈估計，**98%** 信賴區間為

$$\overline{X}-2.5395\frac{S}{\sqrt{n}}\leq\mu\leq\overline{X}+2.5395\frac{S}{\sqrt{n}}$$

，在本題為

$$3.35-2.5395\frac{1.268}{\sqrt{20}}\leq\mu\leq 3.35+2.5395\frac{1.268}{\sqrt{20}}$$

考生姓名： 學號：

檢定: 以下我們用 R 模擬銅板投擲 50 次，並用這些資料來檢定該銅板是否為公正銅板 (其中的 p 我們事先設定好了)，(a) 請說明下列 R 操作的中的 H0, H1 各為何?(b) 說明在 95% 信賴區間下，您會承認 H0 或者否認 H0，為甚麼? (15%)
 	
(1)
> x=sample(0:1, 50, replace=T, prob=c(p, 1-p))
> t.test(x, mu=0.5)
 	
One Sample t-test
 	
data: x
t = 2.3643, df = 49, p-value = 0.02207
alternative hypothesis: true mean is not equal to 0.5 (H1: 母體 mu≠ 0.5, H0: 母體 mu=0.5)
95 percent confidence interval:
0.5240067 0.7959933 (由於 0.5 不在區間內，所以我們否認 H0，承認 H1)
sample estimates:
mean of x
0.66
 	
(2)
> y=sample(0:1, 50, replace=T, prob=c(p, 1-p))
> t.test(y, mu=0.5)
 	
One Sample t-test
 	
data: y
t = -0.8461, df = 49, p-value = 0.4016
alternative hypothesis: true mean is not equal to 0.5 (H1: 母體 mu≠ 0.5, H0: 母體 mu=0.5)
95 percent confidence interval:
0.2974962 0.5825038 (由於 0.5 在區間內，所以我們無法否認 H0，也不能承認 H1)
sample estimates:
mean of x
0.44
	請說明下列 R 操作的意義。 (15%)
	(1)
	> x = sample(1:10, 25, replace=T) // 取樣 1:10 之間的整數 25 個 (取後放回)。
	> y=1+3*x // 按照 y=1+3x 做出 25 個 y 樣本
	> xy = data.frame(x, y) // 將 x,y 結合成框架變數
	> model = lm(y~x, data=xy) // 進行線性迴歸分析
	> model
	Call:
	lm(formula = y ~ x, data = xy)
	
	Coefficients:
	(Intercept) x
	1 3 // 回歸分析結果找出截距為 1，x 係數為 3 的算式，也就是 y=1+3x 的正確答案。
	
	(2)
	> x=runif(50, min=0, max=100) // 用均等分布取樣 (0,100) 之間的實數整數 50 個
	> y=3*x+4+rnorm(50, mean=0, sd=0.5)// 按照 y=3x+4+常態誤差 做出 50 個 y 樣本
	> xy=data.frame(x,y) // 將 x,y 結合成框架變數
	> model=lm(y~x, data=xy) // 進行線性迴歸分析
	> model
	
	Call:
	lm(formula = y ~ x, data = xy)
	
	Coefficients:
	(Intercept) x
	1. 3.899 3.000 // 回歸分析結果找出截距為 3.899，x 係數為 3 的算式，也就是 y=3x+3.899, 非常接近原本的 y=3x+4+常態誤差。

請說明下列 R 操作的意義，並加上註解。(10%)
(1)
> cor(x, x) // 計算 x 與 x 之間的相關係數
[1] 1 // 結果為 1
> cor(x, x+1)// 計算 x 與 x+1 之間的相關係數
[1] 1 // 結果為 1
(2)
> cor(x, -x)// 計算 x 與 -x 之間的相關係數
[1] -1// 結果為 -1
> cor(x, 0.5*x)// 計算 x 與 0.5x 之間的相關係數
[1] 1// 結果為 1
	請寫出您對這學期機率統計課程的建議 (10%)
	
	本題怎麼寫都得 10 分，但一定要有寫字，沒有建議就寫「沒有建議」。