這份考卷共有四大題，每一題的題目雖然看起來很長，但答案都很單純；題目很長是因爲其中有許多說明是用來指引您如何回答問題的，只要您靜下心來慢慢思考，就不會覺得困難。

一、 表 1.1 是某校研究所入學申請人數與錄取人數。請問：

**表 1.1　研究所入學申請人數與錄取人數**

| | 入學申請人數 | 錄取人數 |
|---|---|---|
| 男性 | 8000 | 3500 |
| 女性 | 5000 | 1500 |

1. 男女生的入學申請許可率各為何？入學申請許可率有無性別差異？**(5%)**
2. 為了進一步瞭解差異的來源，我們將表 1.1 按學院別再細分為表 1.2。請問文學院的入學申請許可是否有性別差異？科學院呢？**(5%)**

**表 1.2　研究所入學申請人數與錄取人數**

| 性別 | 入學申請人數 | | 錄取人數 | |
|---|---|---|---|---|
| | 科學院 | 文學院 | 科學院 | 文學院 |
| 男性 | 6000 | 2000 | 3000 | 500 |
| 女性 | 1000 | 4000 | 500 | 1000 |

3. 前面兩小題的答案有無衝突？請解釋為什麼。**(10%)**

二、 要確定因果關係，最好是隨機分派被觀察者接受不同的處置（treatments），然後觀察處置後的結果。如此原因在前、結果在後，順序清楚，不容易發生倒果為因的情形。但許多時候隨機分派是不可能的，譬如說，我們想要瞭解性別對大學教授薪資的影響，我們不能隨機分派每個教授的性別（請每個老師來擲銅板，出現正面的就當女教授，出現反面的就當男教授）。大部份的情形是研究對象現有的狀況是什麼我們就照單全收，因此必須依賴理論來建立假設；但理論並非俯拾即是，所以有時候，我們會先利用常識來猜測可能的關係，建立初步的模型，然後再利用實際的資料驗證。只是根據常識來猜測，就容易出現因果倒置的問題。舉個例子來說，

> 在美國的 Arizona，人們死於呼吸系統疾病的比率很高。一個外國人猜測其原因可能是 Arizona 空氣不好的關係。但是美國人都知道，Arizona 的空氣乾燥又清淨，因此吸引了許多罹患呼吸系統疾病的人來居住 所以這個現象的產生可能是太多罹患呼吸系統疾病的人擁進 Arizona，抵銷了 Arizona 好空氣對呼吸系統疾病的克制能力，才使得居民死於呼吸系統疾病的比率很高。

在上述的例子裡，「Arizona 的空氣乾燥又清淨，並吸引了許多罹患呼吸系統疾病的人來居住」就是資料驗證，證明了這一點，我們就會傾向於接受第二種解釋。現在有一個新問題：「在美國，與子女同住的老人的健康狀況較沒有與子女同住的老人的健康狀況差。」請您參考 Arizona 的例子，提出兩種不同因果方向的解釋，並說明您將如何驗證這兩個假設。 **(20%)**

接 次 頁

三、 下面有 7 小題，每一題中有數個【】，各有 2 個答案選項，請圈選正確的一個。答案請按題目出現順序排列，例如，您認為第 1 小題的答案分別為❷❷❷❷，則寫「1. ❷❷❷❷」。(40%)

1. 當樣本是【❶ 隨機，❷ 樣本數非常大】時，【❶ 樣本平均數$\overline{X}$，❷ 母體平均數$\mu$】 是【❶ 樣本平均數$\overline{X}$，❷ 母體平均數$\mu$】的不偏估計式。也就是說，長期而言，前者（即第 2 個選項中的答案）的【❶ 期望值，❷ 標準差】等於後者（即第 3 個選項中的答案）的數值。
2. 從一個標準差為σ的母體多次抽出樣本數量為 n 的樣本，每一個樣本的平均數為$\overline{X}_i$，這些平均數($\overline{X}_i$)的機率分配是一個【❶ 均等分配，❷ 常態分配】，而此一分配的標準差為【❶ $\sigma/n$，❷ $\sigma/\sqrt{n}$】，稱為【❶ 標準誤，❷ 母體標準差】。這個現象稱之為【❶ 中央極限定理，❷ 貝氏定理】。
3. 將每一個樣本平均數除以【❶ 標準誤，❷ 母體標準差】，也就是將平均數標準化，可得到 z 分配。
4. 在$\overline{X}$左右加上約 【❶ $\sqrt{n}$，❷ 2】個標準誤，我們可得到一個數值範圍，這個數值範圍有 95% 的機會包含到母體平均數，我們稱其為【❶ $\overline{X}$，❷ $\mu$】的 95%信賴區間。
5. 一位統計學家建構一信賴水準為 95%的信賴區間，表示如果他終其一生做了 1000 次抽樣，會有【❶ 50 次，❷ 950 次】的信賴區間沒有包含到母體的參數。而且他【❶ 知道，❷不知道】哪一次的信賴區間是沒有包含到母體參數。
6. 大部份的時後，母體標準差 σ 都是未知的，只好利用【❶ 樣本平均數$\overline{X}$，❷ 樣本標準差 s】來估計。為了容納這個新增的不確定性，我們乃使用【❶ t 分配，❷ F 分配】來代替 z 分配。使用這個分配所得到的信賴區間長度會較 z 分配所得到的【❶ 窄，❷ 寬】，但當樣本數量愈【❶ 大，❷ 小】時，這個分配會愈接近 z 分配。
7. 在其他條件不變的情況下，為了得到較高的信賴水準，如從 95%上升至 99%，則信賴區間的長度必須變【❶ 窄，❷ 寬】，而這樣的變化使得估計更【❶ 精確，❷ 不精確】。另一個不改變信賴區間的長度但可以提高信賴水準的做法是【❶ 增加，❷ 減少】樣本數量。

四、 表 4.1 是三個機器上一週的產量資料，請您利用變異數分析（Analysis of Variance, ANOVA）來檢定這三個機器的產量是否有統計上的顯著差異。回答時必須列出 ANOVA Table，包含變異來源、平方和（SS）、自由度（d. f.）、平均平方和（MS）、F 值，並說明您的虛無假設與檢定結果（當自由度各為 2,12 時，$F_{.01}=6.93$）。 (20%)

**表 4.1　三個機器上一週的產量分配**

| | 機器 1 | 機器 2 | 機器 3 | |
|---|---|---|---|---|
| | 47 | 55 | 54 | |
| | 53 | 54 | 50 | |
| | 49 | 58 | 51 | |
| | 50 | 61 | 51 | |
| | 46 | 52 | 49 | |
| $\overline{X}_i$ | $\overline{X}_1=49$ | $\overline{X}_2=56$ | $\overline{X}_3=51$ | $\overline{\overline{X}}=52$ |
| $(\overline{X}_i-\overline{\overline{X}})$ | -3 | 4 | -1 | $\Sigma(\overline{X}_i-\overline{\overline{X}})=0$ |
| $(\overline{X}_i-\overline{\overline{X}})^2$ | 9 | 16 | 1 | $\Sigma(\overline{X}_i-\overline{\overline{X}})^2=26$ |
| $\sum_{t=1}^{5}(X_{it}-\overline{X}_i)^2$ | 30 | 50 | 14 | $\sum_{i=1}^{3}\sum_{t=1}^{5}(X_{it}-\overline{X}_i)=94$ |