

類 科：統計
科 目：抽樣方法
考試時間：2 小時

座號：_____

※注意：(一)可以使用電子計算器，須詳列解答過程。

(二)不必抄題，作答時請將試題題號及答案依照順序寫在試卷上，於本試題上作答者，不予計分。

(三)本科目除專門名詞或數理公式外，應使用本國文字作答。

註：本試題可能使用之標準常態值如下： $Z_{0.05} = 1.645$ ， $Z_{0.025} = 1.96$

一、臺灣地區有 525 家 IC 設計公司，其中有 5 家公司規模特大。為了估計臺灣地區 IC 設計產業的總員工人數 Y ，除了將 5 家特大 IC 設計公司調查其員工人數得 y_1, y_2, \dots, y_5 以外，另以簡單隨機抽樣法（不歸還）從剩餘的 520 家 IC 設計公司中抽出 20 家公司為樣本，調查其員工人數得 y_6, \dots, y_{25} 。現欲採用下列兩種估計式以估計臺灣地區 IC 設計公司之總員工人數 Y ：

$$\hat{Y}_1 = \frac{525}{25} \sum_{i=1}^{25} y_i, \quad \hat{Y}_2 = \sum_{i=1}^5 y_i + \frac{520}{20} \sum_{i=6}^{25} y_i$$

(一)試證 \hat{Y}_1 及 \hat{Y}_2 是否皆為 Y 之不偏估計式。(6 分)

(二)請分別寫出 \hat{Y}_1 及 \hat{Y}_2 之均方誤差 (Mean Square Error; MSE) 的定義。(4 分)

(三)又已知 5 家特大 IC 設計公司員工總人數為 5,062，而其他 520 家 IC 設計公司員工總人數為 60,320；其他 520 家 IC 設計公司之員工人數的平方和為 11,842,980。試利用上列資料求估計式 \hat{Y}_1 及 \hat{Y}_2 的均方誤差之值，並請說明估計 Y 時，那一個估計式較好？(4 分)

二、自來水公司想調查臺中市某區住戶在 108 年 1~3 月的用水情況，該區共有 $N=20,000$ 戶，今以簡單隨機抽樣法調查了 $n=800$ 戶，得到下列統計資料： $\bar{y} = 22.5$ 噸， $s = 36$ 噸，有 320 戶用水總量超過了規定的用量標準。

(一)請估計該區住戶在 108 年 1~3 月的總用水量 Y 及其該估計之 95% 的信賴區間。(6 分)

(二)請估計該區在 108 年 1~3 月的總用水量超過了規定的用量標準的住戶數 T 。(6 分)

(三)又自來水公司想在 6 月份再做一次該區住戶用水總量的抽樣調查，若要求在 95% 水準下估計的相對誤差不超過 10%，則應該抽多少樣本 (戶) 才能達到上述之要求？(6 分)

三、若一母體共分為三層，其中各層母體大小分別為 $N_1=300$ ， $N_2=600$ 與 $N_3=100$ 。根據過去的調查結果知道這三層的母體標準差分別為 $\sigma_1=150$ ， $\sigma_2=75$ 與 $\sigma_3=100$ 。

(一)若調查總預算 (C) 只有 20,000 元，又固定成本 (c_0) 為 2,000 元，每一層調查單位成本均相同為 50 元，即 $c_1=c_2=c_3=50$ 元，我們想以分層隨機抽樣法估計式 \bar{y}_{st} 估計母體平均數 μ ，試依紐曼配置 (Neyman Allocation) 法，求滿足要求之總樣本大小 n 及各層所需之樣本大小 n_i ($i=1,2,3$) 分別為若干？(10 分)

(二)若我們想以分層隨機抽樣法估計式 \bar{y}_{st} 估計母體平均數 μ 的 95% 的誤差界限不超過 $B=20$ ，試依紐曼配置法，求滿足要求之總樣本大小 n 及各層所需之樣本大小 n_i ($i=1,2,3$) 分別為若干？(10 分)

四、某市轄區有 $N=250$ 個里，共有 $M=16,800$ 戶。今電力公司營業處想估計該市每戶裝置冷氣機之平均台數 μ_Y ，以簡單集體抽樣法自該市隨機抽出 10 個里進行調查，得如下調查資料：

里 (集體)	戶數 (M_i)	該里住戶裝 置冷氣機之 总台數 (y_i)	里 (集體)	戶數 (M_i)	該里住戶裝 置冷氣機之 总台數 (y_i)
1	150	480	6	120	216
2	72	252	7	50	240
3	80	440	8	100	230
4	90	162	9	71	213
5	125	200	10	42	252

註：上述調查資料經計算得 $\sum_{i=1}^{10} M_i = 900$ ， $\sum_{i=1}^{10} M_i^2 = 91,514$ ， $\sum_{i=1}^{10} y_i = 2,685$ ， $\sum_{i=1}^{10} y_i^2 = 819,777$ ， $\sum_{i=1}^{10} M_i y_i = 251,551$ ，在簡單集體抽樣法下，電力公司營

業處利用 $\bar{y}_{c1} = \frac{\sum_{i=1}^{10} y_i}{\sum_{i=1}^{10} M_i}$ ， $\bar{y}_{c2} = \frac{N}{M} \frac{1}{n} \sum_{i=1}^{10} y_i$ ， $\bar{y}_{c3} = \frac{1}{n} \sum_{i=1}^{10} \bar{y}_i$ 等 3 個估計式

(其中 $\bar{y}_i = \frac{y_i}{M_i}$) 來估計該市每戶裝置冷氣機之平均台數 μ_Y 。則：

(一)試以 \bar{y}_{c1} 、 \bar{y}_{c2} 、 \bar{y}_{c3} 來估計該市每戶裝置冷氣機之平均台數 μ_Y ，則 μ_Y 估計值分別為何？(9 分)

(二)若以估計式 \bar{y}_{c1} 估計 μ_Y ，則該估計在 95% 的信賴度下，其最大可能估計誤差界限 B 為何？(9 分)

(三)請根據上述調查資料，試分別求 \bar{y}_{c1} 、 \bar{y}_{c2} 、 \bar{y}_{c3} 等 3 個估計式的估計變異數 $\hat{V}(\bar{y}_{ci})$ ， $i=1,2,3$ 之值為何？並請說明哪一個估計式在估計 μ_Y 時有較好的估計效率？(6 分)

五、某市政府之行政區按區域別分為南區與北區，該市人事長想了解該市職工去年上半年（1~6 月份）與今年上半年（1~6 月份）因生病請假時數之變動比率，分別自南、北兩區各以簡單隨機抽樣抽出 100 名職工進行調查，得相關統計資料如下：

區域別	職工總數 (N_i)(人)	樣本數 (n_i)(人)	去年上半年樣本平均數(\bar{X}_i)	今年上半年樣本平均數(\bar{Y}_i)	去年上半年樣本標準差(S_{X_i})	今年上半年樣本標準差(S_{Y_i})	去年與今年上半年樣本相關係數($\hat{\rho}_{XY_i}$)
1(南區)	1,000	100	18.0	18.9	10.0	10.36	0.9841
2(北區)	1,500	100	7.5	4.8	5.45	3.50	0.5576

其中：

設 X_i ：去年上半年第 i 位職工因生病請假的時數；

Y_i ：今年上半年第 i 位職工因生病請假的時數；

\bar{X}_i ：去年上半年因生病請假的樣本平均時數；

\bar{Y}_i ：今年上半年因生病請假的樣本平均時數；

S_{X_i} ：去年上半年因生病請假時數的樣本標準差；

S_{Y_i} ：今年上半年因生病請假時數的樣本標準差；

$\hat{\rho}_{XY_i}$ ：去年與今年上半年因生病請假的時數的樣本相關係數。

又已知去年上半年(1~6 月份)南區職工因生病請假的總時數 $\tau_{X_1} = 16,300$ 小時，北區職工因生病請假的總時數 $\tau_{X_2} = 13,800$ 小時。根據上述資料：

- (一)試估計該市南區職工去年上半年（1~6 月份）至今年上半年（1~6 月份）因生病請假的時數之變動比率 R 為何？又在 95% 的信賴度下，其最大可能估計誤差界限 B 為何？（6 分）
- (二)對於(一)之估計問題，若要求在 95% 的信賴度下，其最大可能估計誤差界限 B 不超過 0.02，試問應抽出多少樣本才足夠？（6 分）
- (三)試以比率估計（ratio estimation）估計該市南區職工今年上半年（1~6 月份）因生病請假的平均時數 μ_{Y_1} 及該市北區職工今年上半年（1~6 月份）因生病請假的平均時數 μ_{Y_2} 。（6 分）
- (四)若視上述調查資料為分兩層（南區及北區），利用分層隨機抽樣法調查所得，請你幫忙該市人事長用層別比率估計式（separate ratio estimator）估計該市 2,500 位職工今年上半年（1~6 月份）因生病請假的平均時數 μ_Y 。又其在 95% 的信賴度下，其最大可能估計誤差界限 B 為何？（6 分）