

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。

一、假定有一母體中共有 $N=3$ 個單元，各單元編號 i 及其母體變數值 (value of population variable of interest) y_i 如下：

i	1	2	3
y_i	5	7	15

今以一取出放回之設計 D ，每次以下表之選擇機率 (draw-by-draw selection probability) 選擇一個單元觀察後放回，再以同樣的選擇機率選擇下一個單元，共選擇 $n=2$ 個樣本單元，選擇機率如下：

i	1	2	3
$P(i)$	$1/6$	$1/3$	$1/2$

(一)試求設計 D 下第一個單元($i=1$)之包含機率 (inclusion probability)，亦即以本設計選擇之一組樣本，其中包含第一個單元之機率。(5 分)

(二)若以觀察值樣本平均，記為 $\bar{y}_n$ ，推估本母體平均 μ ，請問 $\bar{y}_n$ 在設計 D 下是否為不偏估計？(5 分)

(三)請提出一個在設計 D 下， μ 的不偏估計量，請說明其何以不偏，並計算您提出的不偏估計在樣本 $s=(1,3)$ 時 μ 之估計值。(10 分)

二、在某鎮所進行的年度家庭醫療支出調查中，調查戶為以簡單隨機抽樣取出不放回 (simple random sampling without replacement) 在全鎮 10,000 戶中所選取之 1,000 戶樣本戶。調查結束後所得之家戶醫療支出樣本平均為 150 仟元。今欲進一步探討家戶收入在 50 仟元/月以下之低收入戶醫療支出概況，因無財稅資料可供查考，故僅能由樣本資料判定受訪戶是否為符合前述定義之低收入戶，但無法得知本鎮符合定義之低收入戶總戶數。

檢視樣本資料後，符合此一定義之受訪戶共有 100 戶，而該 100 戶之年醫療支出總和為 8,000 仟元，另該 100 戶之年醫療支出平方和為 2×10^{12} 。(亦即若令 y_i 為第 i 戶之年醫療支出， s_d 為樣本戶中之低收入戶集合，則 $\sum_{i \in s_d} y_i = 8 \times 10^6$ 以及 $\sum_{i \in s_d} y_i^2 = 2 \times 10^{12}$ 。)

(一)若以 80 仟元作為該鎮低收入戶之平均年醫療支出之估計量，請問該估計量是否為一不偏估計？並請說明理由或證明。(8 分)

(二)請估計(一)中估計量之 95%信賴區間，請以仟元為單位，並請四捨五入至小數點下第二位。(8 分)

(三)請問該鎮所有低收入戶之醫療總支出之不偏估計量為何？請以仟元為單位，並說明或證明該估計量之不偏性。(10 分)

(四)請估計(三)中估計量之 95%信賴區間，請以仟元為單位，並請四捨五入至整數位。(10 分)

(請接背面)

類 科：統計
科 目：抽樣方法

三、在某鎮所進行的家庭月支出調查中其主抽樣設計如下：依家戶所得將全鎮家戶分為高收入戶（家戶月收入 300 仟元以上之 100 戶），一般收入戶（家戶月收入 50 仟元至 299,999 元之 5,400 戶）及低收入戶（家戶月收入 49,999 元以下共 500 戶），然後在各類收入家戶中選擇欲觀察之樣本。

在各類收入戶中再視需要以不同之抽樣設計選擇該類收入戶樣本戶，在各類家戶中之次抽樣設計及調查方法如下：

1. 高收入戶採全查面訪；月支出平均值為 200 仟元，月支出標準差為 30 仟元。
2. 一般收入戶中，因其戶數眾多，故先在全鎮 40 個里中隨機選擇 5 個里，而在各選擇的里中，再以簡單隨機抽樣取出不放回選擇若干里內樣本戶。
3. 低收入戶中以簡單隨機抽樣取出不放回選擇 200 戶面訪調查；另因低收入戶之支出狀況為調查重點之一，為求慎重起見，再由該 200 樣本戶中以簡單隨機抽樣取出不放回選擇其中 50 戶，而該 50 戶另加以記帳調查蒐集其當月支出資料。

(一)請問本調查中之主抽樣設計為何？同時請說明本調查採本設計之可能原因及優點。(8 分)

(二)在一般收入戶中調查資料如下：

里編號	里內總戶數	里內樣本戶數	里內月支出樣本平均(仟)
2	100	25	80
18	400	50	75
19	120	30	60
30	200	40	90
38	800	100	65

請以比例估計 (ratio estimation) 推估一般收入戶之平均月支出，並請說明在此採比例估計之可能優點及其原因。(8 分)

(三)在低收入戶中若面訪戶之樣本為 s ，其中記帳戶樣本為 s' (s' 為 s 之子集合)，以下為樣本統計量（均以仟元為觀察單位）：

統計量	s	s'
面訪樣本平均	28	25
記帳樣本平均	NA	30
面訪樣本變異數	25	4
記帳樣本變異數	NA	9

另在 s' 中記帳與面訪所得之月支出相關係數為 0.68，請問本類家戶中所使用的抽樣設計為何？並請以適當的方式推估低收入戶之家庭月支出平均。(8 分)

(四)請推估本鎮家戶月支出平均，並說明（不須計算）其估計量之變異數估計程序。(10 分)

(五)明年度的調查仍欲採類似方式進行，然而因經費所限，故高收入戶將改採簡單隨機抽樣取出不放回之抽樣調查以取代全查，但仍欲將該類母體平均之最大推估誤差在 95% 之信心水準下，控制為 10 仟元以下，請問欲達此一精確度要求之所需最小樣本數。(10 分)