

# 《統計實務》

一、甲市衛生單位欲對轄區內的水塔衛生進行調查，下表是該市的建築物數量分布：

甲市行政區	建築物分類			總計
	公寓	大廈	獨立透天厝	
A 區	400	100	500	1,000
B 區	300	300	200	800
C 區	100	300	300	700
總計	800	700	1,000	2,500

假設每棟建築（無論是公寓、大廈、獨立透天厝）都恰有一個水塔，且欲以4%的抽樣比例取出100個水塔進行衛生調查。

(一)設計一個分層隨機抽樣法（stratified random sampling），使得所取出之100個水塔，比例與上表一致。（5分）

(二)設計一個群集抽樣法（cluster sampling），使得所取出之100個水塔，比例與上表一致。（15分）

<b>試題評析</b>	本題是抽樣方法之內容，考分層抽樣法與群集抽樣法之設計。
<b>考點命中</b>	1.《高點·高上抽樣方法講義》，趙治勳編撰，第五章。 2.《高點·高上抽樣方法講義》，趙治勳編撰，第七章。

**答：**

(一)分層兩階段抽樣法且樣本配置均使用比例配置，先以甲市行政區作為第一階段之分層標準，再以建築物分類作為第二階段之分層標準。

規劃A區應隨機抽出  $100 \times \frac{1000}{2500} = 40$  戶，其中A區公寓有  $40 \times \frac{400}{1000} = 16$  戶，大廈有  $40 \times \frac{100}{1000} = 4$  戶，獨立透天厝有  $40 \times \frac{500}{1000} = 20$  戶。

規劃B區應隨機抽出  $100 \times \frac{800}{2500} = 32$  戶，其中B區公寓有  $32 \times \frac{300}{800} = 12$  戶，大廈有  $32 \times \frac{300}{800} = 12$  戶，獨立透天厝有  $32 \times \frac{200}{800} = 8$  戶。

規劃C區應隨機抽出  $100 \times \frac{700}{2500} = 28$  戶，其中C區公寓有  $28 \times \frac{100}{700} = 4$  戶，大廈有  $28 \times \frac{300}{700} = 12$  戶，獨立透天厝有  $28 \times \frac{300}{700} = 12$  戶。

天厝有  $28 \times \frac{300}{700} = 12$  戶。

(二)不等機率兩階段抽樣法且樣本配置均使用比例配置，先以甲市行政區作為第一階段之分群標準，再以建築物分類作為第二階段之分層標準。

於第一階段時，從A，B，C三個行政區中抽取一個行政區，每個行政區被選中之機率如下表：

行政區	被選中之機率
A	$\frac{1000}{2500} = 0.4$
B	$\frac{800}{2500} = 0.32$
C	$\frac{700}{2500} = 0.28$

第二階段就是由第一階段中被選出之行政區裡面按照建築物分類採用比例配置法抽出100戶，各行政區之樣本配置如下：

當第一階段抽出A區，A區公寓有  $100 \times \frac{400}{1000} = 40$  戶，大廈有  $100 \times \frac{100}{1000} = 10$  戶，獨立透天厝有  $100 \times \frac{500}{1000} = 50$  戶。

當第一階段抽出B區，B區公寓有  $100 \times \frac{300}{800} = 37.5 \approx 38$  戶，大廈有  $100 \times \frac{300}{800} = 37.5 \approx 38$  戶，獨立透天厝有  $100 - 38 - 38 = 24$  戶。

當第一階段抽出C區，C區公寓有  $100 \times \frac{100}{700} = 14.29 \approx 15$  戶，大廈有  $100 \times \frac{300}{700} = 42.86 \approx 43$  戶，獨立透天厝有  $100 - 15 - 43 = 42$  戶。

二、去年某一個國家（稱之為G族群）其3個不同年齡層的死亡人數與年中人口數如下表。表中另外也給出G群3個年齡層的人口分布百分比，以及另一個W標準人口分布百分比。

	0 到 14 歲	15 到 64 歲	65 歲以上	合計
G 族群各年齡層死亡數	600	5,400	1,200	7,200
G 族群各年齡層年中人口數	300,000	900,000	150,000	1,350,000
G 族群人口分布（百分比）	22.2%	66.7%	11.1%	100%
W 標準人口分布（百分比）	27.0%	65.0%	8.0%	100%

(一)計算G族群當年粗死亡率。(10分)

(二)就G族群的年齡別死亡率，依W標準人口分布重新計算死亡率。這與(一)的結果有多少差異？造成此差異的原因是什麼。(10分)

<b>試題評析</b>	本題是粗死亡率與標準化粗死亡率之計算題型，99年普考與103年關務特考也有考過，這些考古題都有收編於上課講義之題庫內。
<b>考點命中</b>	《高點·高上統計實務講義考題補充》，趙治勳編撰，第六章例4.43。

**答：**

$$(一) \frac{7200}{1350000} \times 1000 = 5.333 \quad (0/00)$$

$$(二) \frac{600}{300000} \times 0.27 + \frac{5400}{900000} \times 0.65 + \frac{1200}{150000} \times 0.08 = 5.08 \quad (0/00)$$

可得(一)與(二)兩個死亡率相差0.253(0/00)，造成此差異之原因是兩族群之人口結構差異大

三、我國過去某連續之6個年度家庭可支配所得最低的20%的平均值以L20表示，最高的20%的平均值以H20表示，如下表（單位：新臺幣萬元）：

年度	1	2	3	4	5	6
L20	29.6	30.1	30.9	31.7	32.0	32.9
H20	182.7	184.6	188.3	192.0	194.0	200.4

試根據此表說明該如何分析家庭所得最低20%與最高20%間的關係？(20分)

<b>試題評析</b>	本題是考家庭收支調查中，第五等分位組對第一等分位組之可支配所得倍數，於99年關務特考也有考過，獲得高分不難。
<b>考點命中</b>	《高點·高上統計實務講義考題補充》，趙治勳編撰，第三章例7。

**答：**

年度	1	2	3	4	5	6
L20	29.6	30.1	30.9	31.7	32.0	32.9
H20	182.7	184.6	188.3	192.0	194.0	200.4
最高20%對最低20%之可支配所得倍數	6.17	6.13	6.09	6.06	6.06	6.09

這6個年度可支配所得倍數都維持在六倍以上，前4個年度呈現遞減趨勢，到了第5個年度開始鈍化後又開始上升。

四、家庭收支調查由行政院主計總處主辦，每年以抽樣調查的方式舉辦。

(一)請以109年家庭收支調查為例，說明其抽樣方法。(12分)

(二)請以適當符號，寫出母體家戶平均收入的不偏估計。(8分)

<b>試題評析</b>	本題是考家庭收支調查之抽樣方法與參數估計，於99年、107年與108年都有考過，考生應該都有準備，獲得高分不難。
<b>考點命中</b>	《高點·高上統計實務講義考題補充》，趙治勳編撰，第三章例1,34,37。

**答：**

(一)採分層兩階段隨機抽樣法，以村里為第一階段，再以村里內之戶為第二階段(比例配置)，以縣市為副母體，再依各村里之就業人口產業結構及教育程度為分層變數，對各縣市每一村里進行分層，最多分為6層

(二)母體家戶平均收入之不偏估計量為  $\hat{Y} = \sum_{h=1}^L w_h \hat{Y}_h$

其中  $w_h = \frac{N_h}{N}$ ：第h層之母體比例

$\hat{Y}_h$ ：第h層之樣本家戶平均收入

五、為收集企業按季、年的經營狀況，政府也針對企業單位進行抽樣調查，以98年之工商企業經營概況調查為例，在製造業方面，以工商普查母體檔為抽樣母體，依照產業差異及重點產業分成38個副母體，各副母體再以分層隨機系統抽樣法選取樣本，樣本配置則以紐曼配置(Neyman Allocation)，配置副母體及其內各層的樣本數。

(一)請以適當符號，寫出各副母體及其內各層的樣本數配置公式。(12分)

(二)令X為調查項目中的產值，請以適當符號，寫出母體產值的不偏估計。(8分)

<b>試題評析</b>	本題是抽樣方法之內容，考分層抽樣法中總樣本數與各層樣本配置之計算公式。
<b>考點命中</b>	《高點·高上抽樣方法講義》，趙治勳編撰，第五章。

**答：**

(一)分層隨機抽樣法

總樣本數為  $n$ ，紐門配置之公式如下：

$$n = \frac{(\sum_{h=1}^L N_h S_h)^2}{VN^2 + \sum_{h=1}^L N_h S_h^2}, \quad L = 38 \quad \text{其中 } V = \left(\frac{B}{z_{\frac{\alpha}{2}}}\right)^2$$

各副母體之樣本配置  $n_h$ ，紐門配置之公式如下：

$$n_h = \frac{N_h S_h}{\sum_{h=1}^L N_h S_h} \times n$$

其中  $N_h$ ：第  $h$  個副母體之抽樣單位數 $N = \sum_{h=1}^L N_h$ ：母體之總抽樣單位數 $S_h^2$ ：第  $h$  個副母體之產值變異數 $B$ ：誤差界限 $1 - \alpha$ ：信賴水準(二)母體總產值  $Y$  之不偏估計量為

$$\hat{Y}_{st} = N\bar{y}_{st} = N \frac{\sum_{h=1}^L N_h \bar{y}_h}{N} = \sum_{h=1}^L N_h \bar{y}_h$$

其中  $\bar{y}_h$ ：第  $h$  個副母體中之樣本平均產值

【版權所有，重製必究！】