# 【統計】

# 《抽樣方法》

**試題評析**

今年抽樣方法頗有深度，第一題含有簡單隨機抽樣法與分層抽樣法，其中第三小題計算非常繁複；第二題屬於比率估計法的基本重點題型：第三題爲群集抽樣法的基本重點題型，卻很容易被誤解爲簡單隨機抽樣法；第四題則屬一般統計學的基本重要概念，卻不易作答。雖然課堂上與講義都有詳細解說，如果不熟練，也不易拿到高分。預估一考生只能拿到40分左右；程度較高者，約可拿60分以上。

一、為了解台灣地區家庭上網的普及率，擬以電話訪問的方式進行資料蒐集及統計，其結果如下表所示。若希望推估整體家庭上網率的信賴區間為95%，且抽樣誤差不大於2%。請問在下列各狀況下，樣本數應抽取多少？

(一)在台灣地區整體家庭上網率未知的狀況下，採用簡單隨機抽樣。(10分)

(二)若依據往年調查的結果，已得知台灣地區的家庭上網率為61.01%，採用簡單隨機抽樣。(10分)

(三)若依區域別分層（分為台北市、高雄市、北部地區、中部地區、南部地區、東部地區等六層），各層的家庭上網率如下表所示。則請問採用紐曼配置下，總樣本數及各層樣本為何？(15分)

| 地區別 | 台北市 | 高雄市 | 北部地區 | 中部地區 | 南部地區 | 東部地區 |
|---|---|---|---|---|---|---|
| 母體數（單位：萬戶） | 93 | 54 | 239 | 164 | 154 | 19 |
| 上網率 | 0.7836 | 0.8073 | 0.7441 | 0.7219 | 0.6194 | 0.4811 |

答：

(一) $n=\dfrac{(z_{0.025})^2\cdot}{4\cdot d^2}=\dfrac{(1.96)^2}{4\cdot(0.02)^2}\cong 2401$

(二) $n_0=\dfrac{(z_{0.025})^2\cdot\left(\dfrac{N}{N-1}\times P\times Q\right)}{d^2}=\dfrac{(1.96)^2\cdot\left(\dfrac{723}{722}\times 0.6101\times 0.3899\right)}{(0.02)^2}\cong 2287.744$

$\Rightarrow\ n=\dfrac{N\times n_0}{N+n_0}=\dfrac{723\times 2287.744}{723+2287.744}\cong 550$

(三) $\sum_{h=1}^{5}N_h\cdot S_h=\sum_{h=1}^{5}N_h\cdot\sqrt{\dfrac{N_hP_hQ_h}{N_h-1}}$

$=93\times\sqrt{\dfrac{93\times 0.7836\times 0.2164}{93-1}}+54\times\sqrt{\dfrac{54\times 0.8073\times 0.1927}{54-1}}$

$$+239\times\sqrt{\frac{239\times0.7441\times0.2559}{239-1}}+164\times\sqrt{\frac{164\times0.7219\times0.2781}{164-1}}$$
$$+154\times\sqrt{\frac{154\times0.6194\times0.3806}{154-1}}+19\times\sqrt{\frac{19\times0.4811\times0.5189}{19-1}}$$

$$\cong 322.9899848$$

$$\sum_{h=1}^{5} N_h\cdot S_h^2=\sum_{h=1}^{5} N_h\cdot\frac{N_hP_hQ_h}{N_h-1}$$

$$=93\times\frac{93\times0.7836\times0.2164}{93-1}+54\times\frac{54\times0.8073\times0.1927}{54-1}$$
$$+239\times\frac{239\times0.7441\times0.2559}{239-1}+164\times\frac{164\times0.7219\times0.2781}{164-1}$$
$$+154\times\frac{154\times0.6194\times0.3806}{154-1}+19\times\frac{19\times0.4811\times0.5189}{19-1}$$

$$\cong 144.8762967$$

$$\Rightarrow\ n=\frac{\cdot\left(\sum_{h=1}^{6} N_h\cdot S_h\right)^2}{\frac{N^2\cdot d^2}{(z_{0.025})^2}+\sum_{h=1}^{6} N_h\cdot S_h^2}=\frac{(322.9899848)^2}{\left(\frac{723\times0.02}{1.96}\right)^2+144.8762967}\cong 524$$

| h | $N_h$ | $P_h$ | $S_h=\sqrt{\frac{N_hP_hQ_h}{N_h-1}}$ | $N_h\times S_h$ | $n_h=n\times\frac{N_h\times S_h}{\sum_{i=1}^{6}N_i\times S_i}$ |
|---|---|---|---|---|---|
| 1 | 93 | 0.7836 | 0.414021984 | 38.50404451 | 62 |
| 2 | 54 | 0.8073 | 0.39812301 | 21.49864254 | 35 |
| 3 | 239 | 0.7441 | 0.437281664 | 104.5103177 | 170 |
| 4 | 164 | 0.7219 | 0.449435255 | 73.70738182 | 120 |
| 5 | 154 | 0.6194 | 0.487118515 | 75.01625131 | 122 |
| 6 | 19 | 0.4811 | 0.513334037 | 9.753346703 | 16 |
| Sum | 723 | | | 322.9899846 | 524 |

【講義命中事實：(一)第一回p. 2-5及p. 3-9】

二、某家電製造商93年年終最後3個月之彩色電視機的銷售業績為2,300萬元。94年年底想在全公司業績結算前，先行估計彩色電視機的銷售業績，所以由全省100個銷售點中以簡單隨機抽樣法，抽取了5個銷售點，先行統計其銷售業績，這5個銷售點93年及94年彩色電視機的銷售業績如下表所示：

單位：萬元

| 銷售點 | 1 | 2 | 3 | 4 | 5 | 合計 |
|---|---|---|---|---|---|---|
| 93年業績(x) | 15 | 28 | 35 | 10 | 25 | 113 |
| 94年業績(y) | 20 | 35 | 32 | 18 | 23 | 128 |

請依93年業績為基礎，以比率估計法計算：
(一)94年年終全公司彩色電視機總業績之點估計值。（10分）
(二)94年年終全公司彩色電視機總業績之95%信賴區間估計值。（15分）

答：

| $i$ | $x_i$ | $y_i$ | $x_i y_i$ | $x_i^2$ | $y_i^2$ |
|---|---|---|---|---|---|
| 1 | 15 | 20 | 300 | 225 | 400 |
| 2 | 28 | 35 | 980 | 784 | 1225 |
| 3 | 35 | 32 | 1120 | 1225 | 1024 |
| 4 | 10 | 18 | 180 | 100 | 324 |
| 5 | 25 | 23 | 575 | 625 | 529 |
| Sum | 113 | 128 | 3155 | 2959 | 3502 |

(一) $\hat{Y}_R = \hat{R}\cdot X = \dfrac{y}{x}\cdot X = \dfrac{128}{113}\cdot(2300\times 4) = 10421$(萬元)。

(二) $s_x^2 = \dfrac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right] = \dfrac{1}{4}\left[2959 - \dfrac{(113)^2}{5}\right] = 101.3$

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right] = \frac{1}{4}\left[3502 - \frac{(128)^2}{5}\right] = 56.3$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right] = \frac{1}{4}\left[3155 - \frac{113\times 128}{5}\right] = 65.55$$

$$\hat{R} = \frac{128}{113} \approx 1.1327$$

$$\sqrt{\hat{V}(\hat{Y}_R)} = N\cdot\sqrt{(1-f)\cdot\frac{1}{n}[s_y^2 - 2(\hat{R})s_{yx} + (\hat{R})^2\cdot s_x^2]}$$

$$=100\sqrt{(1-\frac{5}{100})\cdot\frac{1}{5}\left[56.3-2\times1.1327\times65.55+(1.1327)^2\times101.3\right]}\cong 267.89$$

94年年終全公司彩色電視機總業績之95%信賴區間估計值爲

$$\hat{Y}_R \pm z_{0.025}\cdot\sqrt{\hat{V}(\hat{Y}_R)}=10421\pm1.96\times267.89=[9896(\text{萬元}),10946(\text{萬元})]$$

【講義命中事實：第二回p. 7-2及p. 7-8】

三、已知科學園區內有250家公司，從事研發工作的員工總人數為3,200人，為估計該科學園區內研發人員的平均薪資水準，以簡單隨機抽樣方式抽取園區內的30家公司進行調查，得知這30家公司的研發人員數總共有384人，這384人的薪資總計為1,986萬元，這30家公司薪資總計的標準差為45萬元。

(一)試求取該園區內平均每位研發人員薪資之估計值。（10分）

(二)試求取每位研發人員薪資估計值之變異數。（15分）

答：

(一) $\hat{\bar{\bar{Y}}}_{cl}\equiv\bar{\bar{y}}_{cl}=\frac{1}{M_0}\hat{Y}_{cl}=\frac{1}{M_0}\cdot N\cdot\bar{y}=\frac{1}{3200}\times250\times\frac{1986}{30}\cong 5.17(\text{萬元}/\text{人})$

(二) $\hat{V}(\bar{\bar{y}}_{cl})=(\frac{N}{M_0})^2\cdot\frac{s_e^2}{n}\cdot(1-f)=(\frac{250}{3200})^2\cdot\frac{2025}{30}\cdot(1-\frac{30}{250})\cong 0.3625(\text{萬元}/\text{人})^2$

【講義命中事實：第二回p .5-5及p. 5-6】

四、請說明95%信賴區間之意義？並探討信賴區間與抽樣誤差之關係。（15分）

答：

(一)若母體參數$\theta$之95%信賴區間爲$[L,U]$，其真正意義是：

如果我們可以不斷的隨機抽樣許多次，將每次取得的樣本代入隨機區間$[L,U]$計算後，得到的許多固定區間中，大約有95%的這些固定區間會包含$\theta$的真正值。

(二)假設母體參數$\theta$之點估計量爲$\hat{\theta}$，抽樣誤差爲$b$，則$\theta$之$100(1-\alpha)\%$信賴區間爲$[\hat{\theta}-b,\hat{\theta}+b]$，其中

$b=z_{\frac{\alpha}{2}}\cdot\sqrt{\hat{V}(\hat{\theta})}$。

【講義命中事實：第一回p. 2-4】