

《抽樣方法》

本試題可能使用之標準常態值如下：

$$z_{0.025}=1.96, z_{0.05}=1.645$$

一、美麗市共有20個里，75000位居民。為了解美麗市居民的健康情形，衛生單位分別自各里中隨機抽出30位居民進行檢查。試問：（每小題5分，共25分）

- (一)這是何種抽樣方法？
- (二)抽樣母體是什麼？
- (三)母體大小為多少？
- (四)抽樣單位是什麼？
- (五)樣本大小為多少？

試題評析	本題是在考專有名詞之意涵。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁4。

答：

- (一)分層隨機抽樣法
- (二)美麗市
- (三) $N=75000$
- (四)20個里中居民
- (五) $n=20*30=600$

二、聯合便利超商共有40家門市（編號1~40）。其中編號1, 2, 8, 9, 12, 13, 19, 20, 21, 22, 23, 24, 29, 30, 31, 37的門市設有生鮮部。現欲利用系統抽樣來估計設有生鮮部的門市占所有門市的比例 p 。

- (一)請列出所有可能的「從5取1」系統抽樣樣本。（10分）
- (二)請根據題(一)計算樣本比例 \hat{p} 的期望值及真實變異數，並且證明 \hat{p} 為 p 之不偏估計式。（10分）
- (三)若「從5取1」系統抽樣的起始值為3，試估計設有生鮮部的門市占所有門市的比例 p 之95%信賴區間。（5分）

試題評析	本題是考系統抽樣法，題型在考古題中也曾經出現過，要拿到分數並不難。
考點命中	《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁17。

答：

$$(一) \text{令 } Y_i = \begin{cases} 1, \text{設有生鮮部} \\ 0, \text{其他} \end{cases}$$

【版權所有，重製必究！】

從1至5隨機抽一數	系統樣本編號	系統樣本觀察值
1	1,6,11,16,21,26,31,36	1,0,0,0,1,0,1,0
2	2,7,12,17,22,27,32,37	1,0,1,0,1,0,0,1
3	3,8,13,18,23,28,33,38	0,1,1,0,1,0,0,0
4	4,9,14,19,24,29,34,39	0,1,0,1,1,1,0,0
5	5,10,15,20,25,30,35,40	0,0,0,1,0,1,0,0

$$(二) E(\hat{p}) = \frac{3}{8} \times \frac{1}{5} + \frac{4}{8} \times \frac{1}{5} + \frac{3}{8} \times \frac{1}{5} + \frac{4}{8} \times \frac{1}{5} + \frac{2}{8} \times \frac{1}{5} = \frac{2}{5}$$

$$p = \frac{16}{40} = \frac{2}{5}$$

由於 $E(\hat{p}) = \frac{2}{5} = p$ ，故 \hat{p} 為 p 之不偏估計量

$$V(\hat{p}) = \sum_{i=1}^5 (P_i - P)^2 \frac{1}{5} = \frac{1}{5} \left[\left(\frac{3}{5} - \frac{2}{5}\right)^2 + \left(\frac{4}{5} - \frac{2}{5}\right)^2 + \left(\frac{3}{5} - \frac{2}{5}\right)^2 + \left(\frac{4}{5} - \frac{2}{5}\right)^2 + \left(\frac{2}{5} - \frac{2}{5}\right)^2 \right] = \frac{2}{25}$$

$$(三) \hat{p} = \frac{3}{8}, s_{\hat{p}} = \sqrt{(1-f) \frac{\frac{n}{n-1} pq}{n}} = \sqrt{\left(1 - \frac{8}{40}\right) \frac{\left(\frac{3}{8}\right)\left(\frac{5}{8}\right)}{40-1}} = 0.06934$$

p 之 95% 信賴區間為 $(\hat{p} \pm 1.96s_{\hat{p}}) = (0.2391, 0.5109)$

三、國家公園內有塊占地1000公頃的林地，因遭蟲害導致部分樹木枯死。現公園管理處欲利用空中照相技術及部分實地勘察來估計遭蟲害而死亡之樹木的總數。管理處利用空拍圖將林地以每單位5公頃劃分，設 x 表示利用空拍圖計算所得之每單位區域死亡樹木的數量， y 表示實地勘察計算所得之每單位區域死亡樹木的數量。從空拍圖計算所得之死亡樹木的總量 $\tau_x = 4200$ ，實地勘察10個單位區域，計算得下列的資料：

$$\sum_{i=1}^{10} x_i = 234, \sum_{i=1}^{10} y_i = 306, \sum_{i=1}^{10} x_i y_i = 8652, \sum_{i=1}^{10} x_i^2 = 6660, \sum_{i=1}^{10} y_i^2 = 11348。$$

- (一)請利用迴歸估計式 (regression estimator) 估計這1000公頃林地因蟲害死亡的樹木總數，並求算其變異數。(10分)
- (二)請利用差量估計式 (difference estimator) 估計這1000公頃林地因蟲害死亡的樹木總數，並求算其變異數。(10分)
- (三)請計算迴歸估計式對差量估計式的相對效率 (relative efficiency) 且評估那種估計式較適合本個案。(5分)

試題評析 本題是考迴歸簡單估計法，屬於常考之題型，要拿到分數並不難。

考點命中 《高點·高上抽樣方法講義》第一回，趙治勳編撰，頁86。

答：注意：1000公頃的林地，每5公頃劃分，共分成200個單位，故母體數為 $N=200$ ，樣本數 $n=10$

$$SS_x = 6660 - \frac{(234)^2}{10} = 1184.4, SS_y = 11348 - \frac{(306)^2}{10} = 1984.4,$$

$$SS_{xy} = 8652 - \frac{(234)(306)}{10} = 1491.6$$

$$b_1 = \frac{SS_{xy}}{SS_x} = 1.2594, s_e^2 = MSE = \frac{SS_y - b_1^2 SS_x}{n-2} = 13.2296, \bar{X} = \frac{4200}{200} = 21,$$

$$\bar{x} = \frac{234}{10} = 23.4, \bar{y} = \frac{306}{10} = 30.6$$

$$(一) \hat{Y}_h = N[\bar{y} + b_1(\bar{X} - \bar{x})] = 200[30.6 + 1.2594(21 - 23.4)] = 5515.488(\text{棵})$$

$$v(\hat{Y}_{lr}) = N^2(1-f) \frac{s_e^2}{n} = 200^2(1-\frac{10}{200}) \frac{13.2296}{10} = 50272.48$$

$$(二) \hat{Y}_{lr} = N[\bar{X} + (\bar{y} - \bar{x})] = 200[21 + (30.6 - 23.4)] = 5640 \text{ (棵)}$$

$$v(\hat{Y}_{lr}) = N^2(1-f) \frac{s_D^2}{n} = 200^2(1-\frac{10}{200}) \frac{20.6222}{10} = 78364.36$$

$$\begin{aligned} \text{其中 } s_D^2 &= s_Y^2 + s_X^2 - 2rs_Xs_Y = \frac{SS_Y}{n-1} + \frac{SS_X}{n-1} - 2 \frac{SS_{XY}}{\sqrt{SS_X}\sqrt{SS_Y}} \sqrt{\frac{SS_X}{n-1}} \sqrt{\frac{SS_Y}{n-1}} \\ &= \frac{SS_Y}{n-1} + \frac{SS_X}{n-1} - 2 \frac{SS_{XY}}{\sqrt{n-1}\sqrt{n-1}} \\ &= \frac{1984.4}{10-1} + \frac{1184.4}{10-1} - 2 \frac{1491.6}{\sqrt{10-1}\sqrt{10-1}} = 20.6222 \end{aligned}$$

(三)迴歸估計式對差量估計式之相對效率為 $\frac{50272.48}{78364.36} = 64.15\%$ ，故迴歸估計式較適合本案。

四、文化部想了解幸福社區居民每年在娛樂方面的消費支出情形。今將幸福社區依地理位置劃分為200個區集，現自200個區集中隨機抽出8個區集，調查每區集的居民人數及每年在娛樂方面的消費金額。調查資料如下：

區集	居民人數	消費金額(單位：萬元)
1	12	35
2	14	34
3	3	7
4	20	47
5	12	32
6	8	28
7	10	36
8	6	9

(一)請估計每位居民每年的平均娛樂消費金額。(5分)

(二)請估計全社區每年的總娛樂消費金額及其95%信賴區間。(10分)

(三)若經調查得知全社區共有2500位居民，請利用此資料估計全社區每年的總娛樂消費金額。(5分)

(四)請依據估計的準確度，比較題(二)及題(三)這兩種估計方法，並說明何種估計方法較適合本個案。(5分)

試題評析	本題是考群集隨機抽樣法及比率群集估計法，其中群集隨機抽樣法在歷屆考古題中都常出現，而比率群集估計法考古題不多但仍有出現過，且講義都有其相關的內容介紹與例題，要拿到滿分並不難。
考點命中	《高點抽樣方法講義》第一回，趙治勳編撰，頁45，頁74。

答：

(一)本小題中M未知，利用比率群集估計法

$$\bar{y}_{rel} = \frac{\bar{y}_t}{M_t} = \frac{(35+34+7+47+32+28+36+9)/8}{(12+14+3+20+12+8+10+6)/8} = \frac{228}{85} = 2.6824 \text{ (萬元)}$$

(二) 本小題中M未知，利用比率群集估計法

$$\hat{Y}_{rcl} = M\bar{y}_{rcl} \stackrel{M=N\bar{M}_t}{=} 200 \times \frac{85}{8} \times 2.6824 = 5700.1 \text{ (萬元)}$$

$$s_{\hat{Y}_{rcl}} = N\sqrt{(1-f)\frac{s_c^2}{n}} = 200\sqrt{(1-\frac{8}{200})\frac{34.651}{8}} = 407.8294$$

$$\begin{aligned} \text{其中 } s_c^2 &= s_Y^2 + \bar{y}_{rcl}^2 s_M^2 - 2\bar{y}_{rcl} r' s_M s_Y \\ &= 189.4286 + (2.6824)^2 27.125 - 2(2.6824)(0.91)(5.2082)(13.7633) \\ &= 34.651 \end{aligned}$$

$$Y \text{ 之 } 95\% \text{ 信賴區間為 } (\hat{Y}_{rcl} \pm 1.96s_{\hat{Y}_{rcl}}) = (4900.7544, 6499.4456)$$

(三) 本小題中M已知，利用群集隨機抽樣法估計

$$\hat{Y}_{cl} = M\bar{y}_{cl} = N\bar{y}_t = 200 \times \frac{228}{8} = 5700 \text{ (萬元)}$$

$$s_{\hat{Y}_{cl}} = Ms_{\bar{y}_{cl}} = N\sqrt{(1-f)\frac{s_t^2}{n}} = 200\sqrt{(1-\frac{8}{200})\frac{189.4286}{8}} = 953.5498$$

(四) 由於 $s_{\hat{Y}_{cl}} > s_{\hat{Y}_{rcl}}$ ，故比率群集估計法之精確度較高，因此較適合本案。

【版權所有，重製必究！】