

111 年公務人員高等考試三級考試試題

等 別：三等考試

類 科：統計

科 目：抽樣方法

王瑋老師

甲、申論題部分：

一、南興社區疑似爆發某流行病之群聚感染，為規劃後續防疫措施，亟需先行瞭解該社區染病人數。現有兩種檢測方法可用於檢測是否染病，檢測方法 A 可迅速得到結果，但其精確程度較不理想，而檢測方法 B 雖可提供相當精確之結果，但其所需檢驗程序繁複，若要爭取時效，僅能施測 30 位。今先以方法 A 對該社區所有之 500 位居民進行檢測，驗得其陽性比例為 0.6，再以簡單隨機抽樣取出不放回之方式，由全社區之居民中選擇 30 位以方法 B 進行施測，驗得其陽性為 15 人，而這 30 位受測者以方法 A 施測之陽性人數為 13 人，且 30 人中有 26 位以方法 A 及方法 B 檢驗之結果相同。請回答下列問題：

- (1) 若僅以方法 B 施測所得之結果推估本社區染病人數，請問其估計值及 95% 信賴區間為何？（5 分）
- (2) 是否可以整合兩種方法推估本社區染病人數？若您認為可以，請敘述您採用的推估方式、本社區染病人數推估值及其 95% 信賴區間，同時請問在上述兩種推估方式中，您會採取那一種及您的理由。（15 分）

《考題難易》★★★★☆

《解題關鍵》相當經典且特別的考題，考生臨場要能聯想到事後分層並不容易，主要核心觀念在於方法 B 是抽樣的主要程序，可把方法 A 視為母體資訊，這樣的想法才能正確得到解題的靈感。102 年關務與 108 年薦任皆有事後分層的考題，可參考王瑋，抽樣方法 P.3-48 與 P3-51 類似試題演練。

【擬答】

- (1) 以方法 B 施測陽性比例之不偏估計值為 $p = \frac{15}{30} = 0.5$

故以此推估本社區染病人數估計值為

$$\hat{A} = Np = 500 \times 0.5 = 250 \text{ (人)}$$

$$\begin{aligned} s_{\hat{A}} &= N \sqrt{(1-f) \frac{s^2}{n}} = N \sqrt{(1-f) \frac{\frac{n}{n-1} pq}{n}} \\ &= 500 \sqrt{\left(1 - \frac{30}{500}\right) \frac{0.5 \times 0.5}{30-1}} = 45.0096 \end{aligned}$$

本社區染病人數之 95% 信賴區間為

$$\hat{A} \pm Z \cdot s_{\hat{A}}$$

$$\Rightarrow (250 \pm 1.96 \times 45.0096) \Rightarrow (161.7812, 338.2188)$$

- (2) 流行病社區感染並無完整的分層母體資訊，我們只能在樣本被選取之後，才可能設定抽樣

單位到正確的層，即事後分層。在此可將 A 方法視為分層，但我們沒有確切的分層資訊，僅知道陽性約略佔 0.6，即 $A_1 = 0.6$ 與 $A_2 = 0.4$ ，從中抽取 $n = 30$ 樣本，根據題意結果可整理成下表：

B 方法	A 方法		合計
	陽性	陰性	
陽性	x	y	15
陰性	w	z	15
合計	13	17	30

其中 $x + z = 26$ ， $x + y = 15$ ， $x + w = 13$

可聯立解得 $x = 12, y = 3, w = 1, z = 14$

$$p_{pst} = \sum_{h=1}^L A_h p_h = 0.6 \times \frac{12}{13} + 0.4 \times \frac{3}{17} = 0.6244$$

$$\hat{A}_{pst} = N p_{pst} = 500 * 0.6244 = 312.2$$

$$\begin{aligned} \text{var}(p_{pst}) &\approx \left(\frac{N-n}{Nn} \right) \sum_{i=1}^L A_h \cdot s_h^2 + \frac{1}{n^2} \sum_{i=1}^L (1-A_h) \cdot s_h^2 \\ &= \frac{500-30}{500 \times 30} \left(0.6 \times \frac{13 \times \frac{12}{13} \times \frac{1}{13}}{12} + 0.4 \times \frac{17 \times \frac{3}{17} \times \frac{14}{17}}{16} \right) \\ &\quad + \frac{1}{30^2} \left(0.4 \times \frac{13 \times \frac{12}{13} \times \frac{1}{13}}{12} + 0.6 \times \frac{17 \times \frac{3}{17} \times \frac{14}{17}}{16} \right) \end{aligned}$$

$$= 0.00338145 + 0.00013713 = 0.00351858$$

$$s_{\hat{A}_{pst}} = 500 \times \sqrt{0.00351858} = 29.6588$$

染病人數之 95% 信賴區間為

$$\hat{A}_{pst} \pm Z \cdot s_{\hat{A}_{pst}}$$

$$\Rightarrow (312.2 \pm 1.96 \times 29.6588) \Rightarrow (254.0687, 370.3313)$$

因為 $s_{\hat{A}} > s_{\hat{A}_{pst}}$ ，所以採整合兩種方法的事後分層抽樣比僅採用方法 B 的簡單隨機抽樣精確。

志光 × 保成 × 學儒

測驗易點通

全國首創

O!ops 你又踩雷了嗎?

答題測驗就像玩踩地雷，總是在賭一把運氣？
已經錯過的題目，總是一錯再錯？

埋頭苦練，不如讓老師
點通你的學習之路

一點就通！

常考題型

知識強化

同樣的出題範圍一考再考，卻還是選不出答案，測驗題不能硬背，唯有讓老師帶你一觀出題知識的原貌，弄清題目在考什麼，才是唯一正解。

易錯題型

觀念釐清

彙整全國最大公職王線上網站測驗中，考生最高頻率答錯的試題，針對試題透徹分析出最易混淆的考點，加強授課、觀念釐清。

— 完整課程資訊詳洽全國志光・保成・學儒門市 —



二、在某大學所進行的某滿意度調查中，樣本為以簡單隨機抽樣取出不放回所選擇出的 $n = 400$ 位大學部學生，該校大學部學生共 5000 名，400 位學生依其年級分類及各年級滿意樣本比例整理如下表：

	大一	大二	大三	大四
樣本數(x_i)	200	100	50	50
滿意比例(p_i)	0.4	0.8	0.8	0.8

則全體學生滿意比例之不偏估計為 $\frac{200 \cdot 0.4 + 100 \cdot 0.8 + 50 \cdot 0.8 + 50 \cdot 0.8}{400} = 0.6$

(1) 根據註冊組提供之資料，該校大學部學生之各年級比例如下：

	大一	大二	大三	大四
母體比例(Q_i)	0.3	0.3	0.2	0.2

請檢定在 $\alpha = 0.05$ 之水準下樣本比例與母體比例是否相符（請參考下表之卡方分配 95% 百分位數）。（5 分）

自由度	1	5	10	20
$\chi^2_{0.95}$	3.84	11.07	18.31	31.41

(2) 根據子題(1)之結論，請問滿意度 0.6 的結果是否應加以修正？請敘明理由，若您認為需要修正，請提供修正後之母體比例。（10 分）

《考題難易》★☆☆☆☆

《解題關鍵》抽樣方法考題中少見出現的適合度檢定，應可輕鬆拿分。第 2 小題仍是事後分層的觀念，但相對較為簡單。可參考王瑋，抽樣方法 P.3-48 與 P.3-51 類似試題演練。

【擬答】

(1)應採用卡方適合度檢定

	大一	大二	大三	大四	總計
O_i	200	100	50	50	400
P_i	0.3	0.3	0.2	0.2	1
E_i	120	120	80	80	400

期望值 $E_i = n \times P_i$ H_0 ：樣本比例與母體相符 H_1 ：樣本比例與母體不相符 $\alpha = 0.05$ ， $df = (3 - 1) = 2$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(200 - 120)^2}{120} + \dots + \frac{(50 - 80)^2}{80}$$

$$= 79.17 \in C$$

$$C: \{\chi^2 > \chi_{0.95}^2(3) = 7.82\}$$

拒絕 H_0 ，有顯著證據說樣本比例與母體不相符

(2)因為樣本比例與母體比例不相符，所以應修正滿意度估計值

採事後分層進行比例估計

$$p_{pst} = \sum_{h=1}^L A_h p_h = 0.3 \times 0.4 + 0.3 \times 0.8 + 0.2 \times 0.8 + 0.2 \times 0.8 = 0.68$$

志光
保成
學儒

112年 虛實整合

多元學習新型態

重聽OK
旁聽OK

突破傳統上課形式 5大方式彈性又便利

| 面授學習 | 直播學習 | 在家學習 | 視訊學習 | Wifi學習 |

◆學習◆
零時差同類科各班別
皆可同步直播上課◆服務◆
零死角服務緊貼需求
隨時掌握學習狀況線上
課業諮詢老師
申論批閱雙師資
雙循環多元
補課方式上榜生
經驗親授時事
專題講座歷屆試題
練習班導師
制度

各班服務略有不同，詳情請洽全國志光、保成、學儒門市

三、金芒鄉是主要的芒果產區，鄉內農戶有專門種植芒果之專業果農，也有混種其他作物而僅部分種植芒果之農戶，專業芒果之農戶名冊可由農林漁牧業普查資料中獲得，但兼種芒果之農戶時有變動，無法經由普查資料中確認，惟為求精準推估金芒鄉本年度芒果產值，亦須將兼種芒果之農戶之芒果產值列入。在本鄉一共 500 戶農戶中，有 100 戶為專業種植芒果的農戶，而此 100 戶中，有 5 戶為大型農戶，95 戶為中小型農戶。本調查以下列方式選擇樣本農戶進行金芒鄉本年度之調查：5 戶大型專業芒果農戶全數加以調查。接下來在全金芒鄉之 11 個村中，以簡單隨機抽樣取出不放回之方式選擇其中 6 個村，每個樣本村中再以簡單隨機抽樣取出不放回之方式選擇 5 戶中小型專業芒果農戶，若該村中小型專業芒果農戶數不足或等於 5 戶，則該村中所有中小型專業芒果農戶全數加以調查。另外以簡單隨機抽樣取出不放回之方式在其餘農戶中選擇其中 50 戶作為樣本戶。

本調查各類農戶所得資料綜整如下：

(a) 大型專業芒果農戶：平均值為 125 萬元，標準差為 20 萬元。

(b) 中小型專業芒果農戶：6 個樣本村之資料如下：

村別	文新村	文景村	文佑村	金河村	同億村	同欣村
總戶數	16	4	3	12	2	13
平均值(萬元)	30	25	30	25	20	30
標準差(萬元)	5	10	6	5	8	8

(c) 其餘農戶：50 戶樣本戶中，有 10 戶種植芒果，而這 10 戶之平均芒果產值為 12 萬元，標準差為 8 萬元。

請回答下列問題（請以萬元為單位作答）：

- (1) 請提出兩種不同估計方式，推估金芒鄉中小型專業芒果農戶之年度芒果總產值，以及其相對應之變異數估計，並比較您提出之方法於本調查之適用性。（20 分）
- (2) 若以其他農戶之 50 戶樣本戶中，10 戶有兼種芒果之農戶的平均芒果產值 12 萬元，做為全體兼種芒果農戶之平均芒果產值推估值，請問其變異數估計值為何？（5 分）
- (3) 請推估金芒鄉兼種芒果農戶之年度芒果總產值，並提出其變異數估計值。（10 分）

《考題難易》★★★☆☆

《解題關鍵》乍看之下與 110 年高考如出一轍的考題，而且抽樣設計更為複雜，但仔細逐題閱讀，發現沒有想像中困難。第 1 小題僅是兩階段簡單群集抽樣法與二階段比率機率的變化題型，第 2 小題與第 3 小題則是次母體的問題，次母體的考題，很類似 103 年高考的考法，109 地特亦有相似題，可參考王瑋，抽樣方法 P.3-57 至 P3-61 完全相同試題。

【擬答】

(1)

1. 採用兩階段簡單群集抽樣法來估計總產值

$$\hat{Y}_{cl2} = N \frac{\sum M_i \bar{y}_i}{n} = 11 \cdot \frac{16 \times 30 + \dots + 13 \times 30}{6} = 2566.6667$$

$$s_{1b}^2 = \frac{\sum (\hat{Y}_i - \bar{\hat{Y}})^2}{n-1} = 33106.6667$$

$$\sum_{i=1}^n M_i^2 (1 - f_{2i}) \frac{s_{2i}^2}{m_i}$$

$$= 16^2 \left(1 - \frac{5}{16}\right) \frac{5^2}{5} + 4^2 \left(1 - \frac{4}{4}\right) \frac{10^2}{4} + \dots + 13^2 \left(1 - \frac{5}{13}\right) \frac{8^2}{5} = 2631.2$$

$$\begin{aligned} \text{var}(\hat{Y}_{cl2}) &= N^2(1-f_1) \frac{s_{1b}^2}{n} + \frac{N}{n} \sum_{i=1}^n M_i^2 (1-f_{2i}) \frac{s_{2i}^2}{m_i} \\ &= 11^2 \left(1 - \frac{6}{11}\right) \frac{33106.6667}{6} + \frac{11}{6} \times 2631.2 = 308301.6444 \end{aligned}$$

2. 採用兩階段比率機率抽樣法來估計總產值

$$\hat{Y}_{pps2} = \frac{M}{n} \sum_{i=1}^n \bar{y}_i = 95 \cdot \frac{30+25+30+25+20+30}{6} = 2533.3333$$

$$\text{var}(\hat{Y}_{pps}) = \frac{M^2}{n} \cdot \frac{\sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2}{n-1} = \frac{95^2}{6} \times 16.6667 = 25069.4444$$

所以本調查 $\text{var}(\hat{Y}_{pps2}) < \text{var}(\hat{Y}_{cl2})$ ，因此兩階段比率機率抽樣法來估計總產值較適合。

$$(2) \bar{y}_1 = 12, s_1^2 = 8^2 = 64$$

而 N_1 為有兼種芒果之農戶個數，為次母體但未知

所以抽出率 $f = \frac{n_1}{N_1}$ 以 $f' = \frac{n}{N}$ 估計之

$$\text{var}(\bar{y}_1) = (1-f') \frac{s_1^2}{n_1} = \left(1 - \frac{50}{400}\right) \frac{64}{10} = 5.6$$

$$(3) \hat{Y}_1 = N_1 \bar{y}_1, \text{ 但 } N_1 \text{ 未知}$$

$$\text{故以 } \hat{Y}_1 = \frac{N}{n} \sum_{i=1}^n y_i = \frac{400}{50} \times 120 = 960 \xrightarrow{\text{估計}} Y_1$$

$$\text{當 } N_1 \text{ 未知時, } s_{\hat{Y}_1} = N \sqrt{(1-f') \frac{s_n^2}{n}}$$

其中 s_n^2 為調整後的樣本 n 的變異數。其方式為將不屬於次母體的樣本數值改為 0，然後在計算調整後樣本變異數。

$$s_n^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{(9 \times 8^2 + 10 \times 12^2) - \frac{120^2}{50}}{50-1} = 35.2653$$

$$\begin{aligned} \text{var}(\hat{Y}_1) &= N^2(1-f') \frac{s_n^2}{n} \\ &= 400^2 \left(1 - \frac{50}{400}\right) \frac{35.2653}{50} = 98742.8571 \end{aligned}$$

志光 × 保成 × 學儒

跟著學長姐 你也可以高分上榜

普考狀元 高榜眼 黃○慧 110高普考·經建行政

老師的許多重要考點以及解題架構都講解十分清楚，如果沒有法律基礎，只要跟著老師的步調，也能拿到平均分數。成功從來都沒有捷徑，想要怎樣的人生就得自己爭取，只要堅持到最後一刻，相信榜單上一定有你。

雙料金榜 應屆考取 傅○ 110高普考·財稅行政

這裡的師資很齊全，後備的服務資源也很完善。我是在校期間報名校園專案課程，這個課程有多元的上課方式，讓我可以一邊完成學校的課業，也不會落下補習班的進度，也因此我才能在畢業就考取公職。

雙料金榜 莊○家 110高普考·統計

考試與學習是長期的，不應該是考試前要熬夜苦讀，考完試就可以不用繼續念書，應該視準備考試為生活的一部分，想放鬆就放鬆，但還是要持續的念，數學是累積學習的科目，慢慢理解每一個細節，就有機會上榜。

一年考取 陳○宇 110高普考·財稅行政

正規班幾乎整天都會上課，在當天課程結束後，我至少會花兩個小時以上的時間將當日的課程做複習，加深自己的印象。除了補習班提供的教材資源外，並沒有額外購買其他參考書。補習班提供課本內容非常詳盡，準備國考綽綽有餘。

四、今有一母體，其中有三個單元，各單元編號 i 及其主要母體變量值 y_i 如下表：

i	1	2	3
y_i	7	5	3

若要從該母體中選擇 2 個單元作為樣本，在不考慮抽樣順序及可重複選擇之情形下，共有 6 種可能之樣本組合，考慮一抽樣設計為每組樣本組合 $s = (i, j)$ ，被選擇到之選擇機率為 $p(s) = (i + j) / 24$ ，例如樣本組合 (2,3) 之選擇機率為 $5/24$ ，(1,1) 被選擇之機率為 $2/24$等。

請回答下列問題（建議以分數計算及作答）：

(1) 請問在本抽樣設計下，第 i 個單元被選到之包含機率 (Inclusion Probability) 若為 π_i ，則 π_2 為何？(5 分)

(2) 請問若以樣本平均 $\bar{y} = \frac{\sum_{i \in s} y_i}{2}$ 作為母體平均 $\mu = \frac{\sum_{i=1}^3 y_i}{3}$ 之估計量，則 $\bar{y}$ 在本設計下推估 μ

之偏誤 (Bias) 為何？(5 分)

(3) $\bar{y}$ 在本設計下之均方誤差 (Mean square error; MSE) 為何？(5 分)

(4) 請就本設計提出一不偏估計量，請問您提出之估計量在樣本組合分別為 (2,3) 及 (3,3) 的情況下之估計值為何？並證明或以本母體為例驗證該估計量為不偏。(10 分)

(5) 請試述您要如何評估您所提出之不偏估計與 $\bar{y}$ 在本設計下孰優孰劣？(不須計算實際數值)(5 分)

《考題難易》★★★★★

《解題關鍵》此題與 110 年高考的出法頗為類似，同樣進行一個不常見的抽樣設計，依序要求描述抽樣過程與採用簡單隨抽樣會產生有偏的結果，並且也要求偏誤與均方誤差，可參考王瑋，111 年抽樣方法總複習教材 P.33 類似試題演練。本題較困難處是要求提出一個不偏估計量，此題臨場不易想到，而且還需要驗證，所以並不容易拿分。事

實上題目有提供暗示，要求算第 i 個單元被選到之包含機率，所以要聯想到 Horvitz-Thompson 估計量，可參考 103 年高考與 100 年關務，王瑋，抽樣方法 P.5-28 與 P.5-33 類似試題。

【擬答】

(1) 在此列出全部的樣本組合

$s = (i, j)$	(y_i, y_j)	$p(s) = (i + j) / 24$	$\bar{y}$
(1, 1)	(7, 7)	2/24	7
(1, 2)	(7, 5)	3/24	6
(1, 3)	(7, 3)	4/24	5
(2, 2)	(5, 5)	4/24	5
(2, 3)	(5, 3)	5/24	4
(3, 3)	(3, 3)	6/24	3

$$\pi_2 = \frac{3}{24} + \frac{4}{24} \times 2 + \frac{5}{24} = \frac{16}{24}$$

可知，包含#2 的機率

(2) 由(1)得的抽樣分布為

$\bar{y}$	3	4	5	6	7
$f(\bar{y})$	$\frac{6}{24}$	$\frac{5}{24}$	$\frac{8}{24}$	$\frac{3}{24}$	$\frac{2}{24}$

$$E(\bar{y}) = 3 \times \frac{6}{24} + 4 \times \frac{5}{24} + \dots + 7 \times \frac{2}{24} = \frac{110}{24}$$

$$\mu = \frac{7+5+3}{3} = 5$$

$$bias = \frac{110}{24} - 5 = -\frac{5}{12}$$

$$(3) Var(\bar{y}) = E(\bar{y}) - (E(\bar{y}))^2 = 22.5 - \left(\frac{110}{24}\right)^2 = \frac{215}{144}$$

$$MSE = bias^2 + Var(\bar{y}) = \left(-\frac{5}{12}\right)^2 + \frac{215}{144} = \frac{5}{3}$$

$$(4) \text{取不偏估計量 } \bar{y}_{HT} = \frac{1}{N} \sum_{i=1}^n \frac{Y_i}{\pi_i} \xrightarrow{\text{估計}} \mu$$

$$\pi_1 = \frac{2}{24} \times 2 + \frac{3}{24} + \frac{4}{24} = \frac{11}{24}$$

包含#1 的機率

$$\pi_3 = \frac{4}{24} + \frac{5}{24} + \frac{6}{24} \times 2 = \frac{21}{24}$$

包含#3 的機率

$$s = (2, 2), \quad \bar{y}_{HT} = \frac{1}{3} \left(\frac{5}{16/24} + \frac{5}{16/24} \right) = 5$$

$$s = (3, 3), \quad \bar{y}_{HT} = \frac{1}{3} \left(\frac{3}{21/24} + \frac{3}{21/24} \right) = \frac{16}{7}$$

其他抽樣單位所對應的估計量如下所示

$s = (i, j)$	(y_i, y_j)	$p(s) = (i + j) / 24$	$\bar{y}_{HT}$
--------------	--------------	-----------------------	----------------

(1, 1)	(7, 7)	2/24	112/11
(1, 2)	(7, 5)	3/24	167/22
(1, 3)	(7, 3)	4/24	480/77
(2, 2)	(5, 5)	4/24	5
(2, 3)	(5, 3)	5/24	51/14
(3, 3)	(3, 3)	6/24	16/7

$$E(\bar{y}_{HT}) = \frac{112}{11} \times \frac{6}{24} + \frac{176}{22} \times \frac{5}{24} + \dots + \frac{16}{7} \times \frac{2}{24} = 5 = \mu, \text{ 為不偏估計量}$$

(5) 相對於 $\bar{y}$ 為有偏估計量， $\bar{y}_{HT}$ 具有不偏性的優點。

除此之外， $\bar{y}_{HT}$ 為不等機率估計法，每個抽樣單位的權重與其抽到的機率成正比，除較合理外，理論上變異數亦較小。

志光 × 保成 × 學儒

適合**非上榜不可**的你

高普考取班**8**大保障 **一次繳費 輔導至考取**

學費省很大

考取班全年課程不間斷，一次繳清學費輔導至考取。

課程最完整

完整課程循環，基礎班→正規班→專題課→總複習…等，全部擁有。

上榜賺獎金

報名考取班第一年考取同職等考試，頒發高額獎學金。

學習最便利

輔導期間可依自己時間選擇面授或視訊學習，提高學習效率。

師資最多元

重點科目安排多元師資，雙循環教學，可旁聽加強弱科，強化上榜實力。

加選最超值

輔導期間可加選其他科目增加考試機會，加選另享專案優惠。

榜單最實在

年年榜單見證，錄取人數最多，錄取率最高，奪榜實力全國第一。

公約有保障

考取班簽訂公約，保障您的權利與義務至考取為止。

— 完整課程資訊詳洽全國志光·保成·學儒門市 —