

113 年特種考試地方政府公務人員及 離島地區公務人員考試試題

等 別：三等考試

王瑋老師解題

類 科：統計

科 目：抽樣方法與迴歸分析

一、有母體如下：

i	1	2	3	4
y_i	10	30	10	50

考慮以下設計於該母體中選擇兩個樣本：

1. 將母體分為 (1, 2) 及 (3, 4) 兩組。
2. 以簡單隨機抽樣選擇一個母體單元 i ，觀察得 y_i 。
3. 若 $y_i > 20$ ，則以 i 所在該組之另一單元為第二個樣本單元，反之若 $y_i \leq 20$ ，則由另一組隨機選擇第二個樣本單元。例如若第一個被選到的單元是第 1 個單元，則下一個單元則從第 3 跟第 4 個單元中隨機選擇，若第一個被選擇到的是第 4 個單元，則第二個被選單元必須是第 3 個單元。

請回答下列問題：（每小題 5 分，共 20 分）

- (一) 請問單元 2 之包含機率 (inclusion probability)，亦即在本設計下選擇之樣本組合中包含單元 2 之機率。
- (二) 若以觀察值樣本平均 \bar{y} 為母體平均 μ 之估計量，請問該估計量在本設計及本母體下是否不偏？
- (三) 請計算 \bar{y} 之均方誤 (mean squared error)。
- (四) 請計算 \bar{y} 之變異數。

《考題難易》★★★★☆

《解題關鍵》此題為適應性抽樣 (adaptive sampling)，本屬課外內容，但同學未必需要理解何謂適應性抽樣，僅需針對題意進行窮舉所有的可能樣本，仍可得到抽樣分配，進而推算偏誤與均方誤。可參考 110 高考幾乎相同考題。

《命中特區》王瑋，抽樣方法總複習，頁 79 題 43。

【擬答】

(一) 第 1 個單元抽到 #1，第 2 單元可抽到 #3, #4：

地區	(1,3)	(1,4)
\bar{y}	10	30
$f(\bar{y})$	$\frac{1}{8}$	$\frac{1}{8}$

第 1 個單元抽到 #2，第 2 單元可抽到 #1：

地區	(2,1)
\bar{y}	20
$f(\bar{y})$	$\frac{1}{4}$

公職王歷屆試題 (113 地方政府特考)

第 1 個單元抽到#3，第 2 單元可抽到#4：

地區	(3,4)
\bar{y}	30
$f(\bar{y})$	$\frac{1}{4}$

第 1 個單元抽到#4，第 2 單元可抽到#1,#2：

地區	(4,1)	(4,2)
\bar{y}	30	40
$f(\bar{y})$	$\frac{1}{8}$	$\frac{1}{8}$

根據上述整理，包含#2 的機率為 $\pi_2 = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$

註：包含#1 的機率為 $\pi_1 = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{5}{8}$

包含#3 的機率為 $\pi_3 = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$

包含#4 的機率為 $\pi_4 = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \frac{5}{8}$

(二) \bar{y} 抽樣分配整理如下：

\bar{y}	10	20	30	40
$f(\bar{y})$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$

$E(\bar{y}) = 10 \times \frac{1}{8} + 20 \times \frac{1}{4} + 30 \times \frac{1}{2} + 40 \times \frac{1}{8} = 26.25 \neq \mu$ 為有偏估計式

其中 $\mu = \frac{10 + 30 + 10 + 50}{4} = 25$

$bias(\bar{y}) = 9 - 10 = -1$

(三) $E(\bar{y}^2) = 10^2 \times \frac{1}{8} + 20^2 \times \frac{1}{4} + 30^2 \times \frac{1}{2} + 40^2 \times \frac{1}{8} = 762.5$

$Var(\bar{y}) = E(\bar{y}^2) - (E(\bar{y}))^2 = 762.5 - 26.25^2 = 73.4375$

$MSE(\bar{y}) = bias^2 + Var(\bar{y}) = (26.25 - 25)^2 + 73.4375 = 75$

(四) 根據題(三) $Var(\bar{y}) = 73.4375$

公職王歷屆試題 (113 地方政府特考)

二、人行道垃圾筒的設置可以方便行人處理隨身垃圾，但也因為附近住家丟棄居家垃圾而造成髒亂，某里長為了解里民是否贊成增設人行道垃圾筒，對里民進行一項問卷調查。該里家戶共 950 戶，居住狀況可以分為獨棟別墅 10 戶，每棟 4~5 戶之舊式公寓 50 棟共 240 戶，以及集合式大樓 4 座共 700 戶三類。抽樣設計及觀察之樣本資料彙整如下：

1. 獨棟別墅之 10 戶全查。贊成比例 0.2。

2. 以簡單隨機抽樣取出不放回選擇 5 棟舊式公寓，再調查所選公寓中之全體住戶。各樣本公寓戶數及贊成之比例如下：

公寓編號	公寓戶數	贊成比例
15	4	1
27	4	0.5
35	5	1
39	5	0.6
47	5	0

3. 每座集合式大樓以簡單隨機抽樣取出不放回各選擇 20 戶。各大樓戶數及贊成之樣本比例如下：

大樓編號	戶數	贊成比例
1	75	0.2
2	400	0.3
3	100	0.2
4	125	0.4

請回答下列問題：

(一) 請問舊式公寓住戶之抽樣設計為何？並請以不偏估計推估舊式公寓住家贊成之比例，並推估該不偏估計量之變異數估計。(12 分)

(二) 請問集合式大樓住戶之抽樣設計為何？請以不偏估計推估集合式大樓住家贊成之比例，並推估該不偏估計量之變異數估計。(12 分)

(三) 請估計本里住家贊成之比例以及該估計量在 95% 信心水準下之最大誤差 ($z_{0.025}=1.96$)。(6 分)

《考題難易》★★★★★

《解題關鍵》本題二階段抽樣法雖不屬課外，但要聯想到以二階段方法來結合不同的抽樣設計並不容易。前兩題是基本的群集隨機抽樣與分層隨機抽樣，可以仿照 110 高考與 108 年高考相同的二階段抽樣沿伸題，可參考黃文隆抽樣方法 P.6-15 頁有相同範例。

《命中特區》王瑋，抽樣方法題庫班，頁 53-59 題九~題十一完全命中。

【擬答】

(一) 自 50 棟舊式公寓抽出 5 棟，並且調查抽到公寓的每一戶，詢問其贊成比例，此為群集隨機抽樣。

$$p_{cl} = \frac{N}{M} \bar{A}_t = \frac{50}{240} \cdot \frac{4 \times 1 + 4 \times 0.5 + \dots + 5 \times 0}{5} = \frac{50}{240} \times 2.8 = 0.583333$$

$$\text{var}(p_{cl}) = \frac{N^2}{M^2} (1-f) \frac{s_t^2}{n} = \frac{50^2}{240^2} \left(1 - \frac{5}{50}\right) \frac{3.7}{5} = 0.028906$$

(二) 4 座集合大樓分別隨機抽出 20 戶，為分層隨機抽樣。

$$p_{st} = \sum_{h=1}^4 W_h p_h = \frac{75}{700} \times 0.2 + \frac{400}{700} \times 0.3 + \frac{100}{700} \times 0.2 + \frac{125}{700} \times 0.4 = 0.292857$$

$$\begin{aligned} \text{var}(p_{st}) &= \frac{1}{N^2} \sum_{h=1}^L N_h (N_h - n_h) \frac{\frac{n_h}{n_h - 1} p_h q_h}{n_h} \\ &= \frac{1}{700^2} \left[75(75 - 20) \frac{0.2 \times 0.8}{19} + \dots + 125(125 - 20) \frac{0.4 \times 0.6}{19} \right] = 0.003975 \end{aligned}$$

(三)採用兩階段群集抽樣估計本里住家贊成之比例，

第一階段抽樣單位是居住的類型，第二階段抽樣單位則是戶。

$$\begin{aligned} \hat{A}_2 &= \sum_{i=1}^{10} A_i + \hat{A}_{cl} + \hat{A}_{st} \\ &= (10 \times 0.2) + 50 \times 2.8 + 700 \times 0.292857 = 347 \end{aligned}$$

$$p_2 = \frac{1}{950} \hat{A}_2 = 0.3653$$

$$\begin{aligned} \text{var}(\hat{A}_2) &= N^2 (1 - f_1) \frac{s_{1b}^2}{n} + \frac{N}{n} \sum M_i^2 (1 - f_{2i}) \frac{s_{2i}^2}{m_i} \\ &= 3^2 (1 - \frac{3}{3}) \frac{s_{b1}^2}{3} + \frac{3}{3} \left(10^2 (1 - \frac{10}{10}) \frac{s_{21}^2}{10} + 50^2 (1 - \frac{5}{50}) \frac{s_{22}^2}{5} + 700^2 (1 - \frac{80}{700}) \frac{s_{23}^2}{80} \right) \\ &= 50^2 (1 - \frac{5}{50}) \frac{0.028906}{5} + 700^2 (1 - \frac{80}{700}) \frac{0.003975}{80} = 34.572075 \end{aligned}$$

$$s_{p_2} = \frac{1}{950} \sqrt{34.572075} = 0.006189$$

本里住家贊成之比例 95% 最大誤差為

$$B = 1.96 \times 0.006189 = 0.0121$$

志光 × 學儒 × 保成

上榜生唯一指定

效率考取關鍵

全國狀元 雙料金榜



113 高考統計狀元
113 普考統計狀元 林○凱

統計學老師上課嚴謹並有條理的介紹統計觀念和公式，將抽象內容用較生活化或白話的解釋讓我們理解其中含義，以及大量題目的練習，將內容量很龐大且是我們本組最重要的統計學，變成我們能在考場上感到得心應手的專業科目。我本身不太會寫作文，經由老師批改後，老師會建議我可以舉哪些例子來增加文章的說服力和深度廣度。

全國狀元 雙料金榜



113 高考經建行政狀元
113 普考經建行政榜眼 陳○霖

財政學老師的課程生動有趣，總能將艱深的理論與實際生活中的例子結合起來，讓我在學習中感受到樂趣。統計學老師教學風格極啟發性，能夠引導我們從不同的角度思考問題，在他的指導下，我的解題能力有顯著提升，這對於申論題的準備尤其重要。當我在解題中遇到瓶頸時，老師的上課內容總能給我提供寶貴的建議，讓我找到突破的方向。

公職王歷屆試題 (113 地方政府特考)

三、教師收集 30 位學生考試成績 (Y) 與讀書時間 (X) 資料，用以了解兩者之間的關係，得到

以下的數據：

$$\sum_{i=1}^{30} x_i = 396, \quad \sum_{i=1}^{30} y_i = 1644.302, \quad \sum_{i=1}^{30} x_i^2 = 5852, \quad \sum_{i=1}^{30} y_i^2 = 94202.131,$$

$$\sum_{i=1}^{30} x_i * y_i = 23255.832$$

(一)以 Y 做為應變數 (dependent variable)，X 做為自變數 (independent variable)，假設簡單線性迴歸模型為 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ，其中 ε_i 為相互獨立且具常態分配 $N(0, \sigma^2)$ 的隨機誤差，試以最小平方法 (least squares method) 求出 β_0 及 β_1 的估計值。(8 分)

(二)若以 $x_i^* = \frac{x_i - \bar{x}}{s_x}$ ， $y_i^* = \frac{y_i - \bar{y}}{s_y}$ 分別做為新的自變數及新的應變數 (其中 \bar{x}, \bar{y} 為原本資料的樣本平均數， s_x, s_y 為原本資料的樣本標準差)，建立新的迴歸模型 $Y_i^* = \beta_0^* + \beta_1^* X_i^* + \varepsilon_i^*$ ，試以最小平方法求出 β_0^* 及 β_1^* 的估計值。(8 分)

(三)若學生的居住地區分為北、中、南三個地區，今定義三個虛擬變數 (dummy variable) D_1, D_2, D_3 ，其中 $D_1=1$ 代表居住北部， $D_1=0$ 代表其他； $D_2=1$ 代表居住中部， $D_2=0$ 代表其他； $D_3=1$ 代表居住南部， $D_3=0$ 代表其他。如果以 Y 做為應變數，X, D_1, D_2, D_3 做為自變數建立複迴歸模型，請問有何問題？(4 分)

(四)如果以 Y 做為應變數，X, D_1, D_2 做為自變數建立的複迴歸方程式為

$$\hat{Y} = 20.03 + 2.02X + 1.03D_1 + 3.12D_2, \quad \text{試求以 Y 做為應變數，X, } D_2, D_3 \text{ 做為自變數建立的複迴歸方程式為何？(5 分)}$$

《考題難易》★★☆☆☆

《解題關鍵》本題的前兩個小題包含簡單線性迴歸與變數標準化，屬於課內基本題，最近的 112 地特也考過。第 3 小題與第 4 小題是虛擬變數的應用題，可參考 103 普考統計學以及 112 地特迴歸類似的觀念題。

《命中特區》王瑋，迴歸分析，志光出版，頁 3-29 題 15，頁 7-164 題 67，頁 7-169 題 71；迴歸分析題庫班，頁 70 自編試題三完全命中。

【擬答】

$$\text{(一)} SS_x = 5852 - \frac{396^2}{30} = 624.8$$

$$SS_y = 94202.131 - \frac{1644.302^2}{30} = 4077.82876$$

$$SS_{xy} = 23255.832 - \frac{396 \times 1644.302}{30} = 1551.0456$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{1551.0456}{624.8} = 2.482$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \frac{1644.302}{30} - 2.482 \times \frac{396}{30} = 22.041$$

$$\text{(二)} \bar{x}^* = \sum \frac{(x_i - \bar{x})}{s_x} = 0, \quad \bar{y}^* = \sum \frac{(y_i - \bar{y})}{s_y} = 0$$

$$\hat{\beta}_1^* = \frac{\sum x_i^* y_i^* - n\bar{x}^* \bar{y}^*}{\sum x_i^{*2} - n\bar{x}^{*2}} = \frac{\sum \frac{(x_i - \bar{x})}{s_x} \cdot \frac{(y_i - \bar{y})}{s_y}}{\sum \frac{(x_i - \bar{x})^2}{s_x^2}} = \frac{s_x}{s_y} \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \sqrt{\frac{SS_x}{SS_y}} \frac{SS_{xy}}{SS_x} = \frac{SS_{xy}}{\sqrt{SS_x SS_y}} = \frac{1551.0456}{\sqrt{624.8 \times 4077.82876}} = 0.972$$

$$\hat{\beta}_0^* = \bar{y}^* - \hat{\beta}_1^* \bar{x}^* = 0 - \hat{\beta}_1^* \times 0 = 0$$

(三)若有 k 個分類，需設定 $k-1$ 個虛擬變數。但若三分類，亦設定了三個虛擬變數，造成線性相依的情況，會使得方程式無法求得估計式(即三個方程式卻要求解四個未知數)。

(四)根據 X, D_1, D_2 做為自變數的複迴歸方程式

$$\text{北區 } (D_1 = 1, D_2 = 0) \Rightarrow \hat{Y} = 20.03 + 2.02X + 1.03 = 21.06 + 2.02X$$

$$\text{中區 } (D_1 = 0, D_2 = 1) \Rightarrow \hat{Y} = 20.03 + 2.02X + 3.12 = 23.15 + 2.02X$$

$$\text{南區 } (D_1 = 0, D_2 = 0) \Rightarrow \hat{Y} = 20.03 + 2.02X$$

若採用 X, D_2, D_3 做為自變數的複迴歸方程式

$$Y = \beta_0 + \beta_1 X + \beta_2 D_2 + \beta_3 D_3$$

$$\text{北區 } (D_2 = 0, D_3 = 0) \Rightarrow \hat{Y} = \beta_0 + \beta_1 X = 21.06 + 2.02X$$

$$\text{中區 } (D_2 = 1, D_3 = 0) \Rightarrow \hat{Y} = (\beta_0 + \beta_2) + \beta_1 X = 2.15 + 2.02X$$

$$\text{南區 } (D_2 = 0, D_3 = 1) \Rightarrow \hat{Y} = (\beta_0 + \beta_3) + \beta_1 X = 20.03 + 2.02X$$

所以比對上述係數的關係，可知採用 X, D_2, D_3 的複迴歸方程式為

$$\hat{Y} = 21.06 + 2.02X + 2.09D_2 - 1.03D_3$$

感謝
志光 學儒 保成

讓我們 高分考取

經建行政	經建行政	商業行政	統計
<p>高考前三佔二</p> <p>狀元陳○霖/榜眼林○穎</p>	<p>普考前三佔二</p> <p>狀元林○穎/榜眼陳○霖</p>	<p>普考前三佔二</p> <p>榜眼邱○恩/探花彭○儀</p>	<p>普考前三佔二</p> <p>狀元林○凱/探花林○甫</p>
<p>👑 四個月考取</p> <p>雙料金榜 梁○傑</p> <p>113高普考經建行政</p>	<p>👑 五個月應屆考取</p> <p>雙料金榜 周○森</p> <p>113高普考經建行政</p>	<p>👑 五個月考取</p> <p>雙料金榜 朱○萱</p> <p>113高普考經建行政</p>	<p>👑 六個月考取</p> <p>林○慶</p> <p>113普考會計</p>
<p>👑 六個月考取</p> <p>吳○峰 113高考會計</p>	<p>👑 九個月應屆考取</p> <p>王○期 113普考財稅行政</p>	<p>👑 十個月應屆考取</p> <p>許○惠 113普考會計</p>	<p>👑 應屆考取</p> <p>高○真 113普考財稅行政</p>
<p>👑 一年考取</p> <p>雙料金榜 陳○霖</p> <p>113高普考經建行政</p>	<p>👑 一年應屆考取</p> <p>雙料金榜 游○欣</p> <p>113高普考經建行政</p>	<p>👑 一年考取</p> <p>雙料金榜 彭○禎</p> <p>113高普考會計</p>	<p>👑 一年應屆考取</p> <p>連過三榜 許○谷</p> <p>113高普考/113初等考狀元財稅行政</p>
<p>👑 一年考取</p> <p>鄭○仁 113普考財稅行政</p>	<p>👑 一年考取</p> <p>謝○榮 113高考統計</p>	<p>👑 一年考取</p> <p>王○鑫 113高考統計</p>	<p>🏆 KEEP FOR YOU</p>

公職王歷屆試題 (113 地方政府特考)

四、利用 20 個樣本，計算複迴歸模型 $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$ 參數估計如下：

	估計值 (estimate)	標準誤 (standard error)
β_0	20.03	10.15
β_1	2.02	1.05
β_2	-1.03	2.01
β_3	3.12	1.56

並計算出複判定係數 (coefficient of multiple determination) 為 $R^2 = 0.8$ 。

(一) 計算並解釋調整複判定係數 (adjusted coefficient of multiple determination)。其與複判定係數的差異為何？(8 分)

(二) 在顯著水準 $\alpha = 0.05$ 下，試檢定 $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ vs. $H_1: \beta_i$ 不全為 0, $i = 1, 2, 3$ 。(10 分)

(三) 在顯著水準 $\alpha = 0.05$ 下，試檢定 $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ 。(7 分)

參考之查表值：

$$F_{0.05}(3, 16) = 3.239, \quad F_{0.05}(3, 18) = 3.160, \quad F_{0.05}(3, 20) = 3.098,$$

$$t_{0.05}(16) = 1.746, \quad t_{0.05}(18) = 1.734, \quad t_{0.05}(20) = 1.725,$$

$$t_{0.025}(16) = 2.120, \quad t_{0.025}(18) = 2.101, \quad t_{0.025}(20) = 2.086$$

《考題難易》★☆☆☆☆

《解題關鍵》多元線性迴歸基本題，包括常考的調整後判定係數的意義以及整體檢定與個別檢定，是不能掉分的考題。

《命中特區》王瑋，抽樣方法，志光出版，頁 7-24 題 13~題 16、頁 7-33 題 19~題 21。

【擬答】

$$(一) R^2 = 1 - \frac{SSE}{SST_0} = 0.8 \Rightarrow \frac{SSE}{SST_0} = 0.2$$

$$\text{調整複判定係數 } R_a^2 = 1 - \frac{SSE}{SST_0} \frac{n-1}{n-4} = 1 - 0.2 \times \frac{20-1}{20-4} = 0.7625$$

當模式引入一新的解釋變數時，只要此一解釋變數稍具解釋力，模式之複判定係數必然提高。然而，解釋變數每增加一個，代表誤差自由度隨之減少一個，代表模式之解釋能力可能降低，複判定係數未能顯示此情形，故以調整後判定係數替代之：

$$R_a^2 = 1 - \frac{SSE}{SST_0} \left(\frac{n-1}{n-p} \right) \times 100\% , \text{ 因此調整判定係數可用於模型選擇。}$$

(二) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$ $H_1: \beta_i$ 不全為 0

$$\alpha = 0.05$$

$$F^* = \frac{n-p}{p-1} \frac{R^2}{1-R^2} = \frac{20-4}{4-1} \times \frac{0.8}{1-0.8} = 21.333 \in C$$

$$C: \{F^* > F_{0.05}(3, 16) = 3.239\}$$

拒絕 H_0 ，有顯著證據 β_i 不全為 0，代表複迴歸模型顯著。

(三) $H_0: \beta_3 = 0$ $H_1: \beta_3 \neq 0$

$$\alpha = 0.05$$

$$T^* = \frac{\hat{\beta}_3}{S(\hat{\beta}_3)} = \frac{3.12}{1.56} = 2 \notin C$$

$$C: \{|T^*| > t_{0.025}(16) = 2.120\}$$

不拒絕 H_0 ，沒有顯著證據 β_3 不全為 0，代表 X_3 變項不顯著。