

## 111 年特種考試地方政府公務人員考試試題

等 別：三等考試

類 科：統計

科 目：迴歸分析

王瑋老師

甲、申論題部分：

一、一位統計分析師想瞭解身高 ( $Y_i$ ，以英寸為單位) 是否可以用手掌張開長度 ( $X_1$ ，以公分為單位) 和性別 ( $X_2$ ，男性是 1，女性是 0) 來預測？他收集 66 名大學生為樣本。所配適的線性迴歸模型如下：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, \quad i = 1, \dots, n$$

請依據表 1 回答下列問題。

表 1: ANOVA

Source	Sum of Squares	DF	Mean square	F test
Regression	840.8436	2		
Error	(1)	(3)	(5)	
(Lack of fit)	(2)	(4)		
(Pure error)	283.8476	45		
Total	1220.4394	65		

(一)請計算表 1 中(1)–(5)所列的線性迴歸的 ANOVA 相關訊息。(10 分)

(二)在顯著水準 5% 下，請檢定身高是否與手掌張開長度 ( $X_1$ ) 和性別 ( $X_2$ ) 有線性關係存在。請列出虛無假設/對立假設、檢定統計量及決策法則。在無需查表之下，你的建議結論為何？(5 分)

(三)在顯著水準 5% 下，請檢定線性迴歸模型是否有顯著的缺適 (lack of fit)？以了解線性迴歸模型是否足以描述身高與手掌張開長度 ( $X_1$ ) 和性別 ( $X_2$ ) 之間的關係。請列出虛無假設/對立假設、檢定統計量及決策法則。在無需查表之下，你的建議結論為何？請說明缺適檢定所需要之假設。(10 分)

## 【解題關鍵】

- 《考題難易》★☆☆☆☆
- 《破題關鍵》缺適度檢定屬課內基本題，雖不常考，但 ANOVA 表填表的難度不高，比較需要留意的是題(二)要求檢定是否有線性關係，此處應進行多元迴歸的整體檢定即可，而題(三)才是缺適度檢定的部分。
- 《命中特區》王瑋，迴歸分析，志光出版，頁 4-58 題 26 與頁 4-65 題 30。

## 【擬答】

(一)(1)  $1220.4394 - 840.8436 = 379.5958$ (2)  $379.5958 - 283.8476 = 95.7482$ (3)  $65 - 2 = 63$ (4)  $63 - 45 = 18$ (5)  $379.5958 \div 63 = 6.0253$ 

(二)將變異數分析表整理如下：

ANOVA表

Source	Sum of Squares	DF	Mean square	F test
Regression	840.8436	2	420.4218	69.7757
Error	379.5958	63	6.0253	
(Lack of fit)	95.7482	18	5.3193	0.8433
(Pure error)	283.8476	45	6.3077	
Total	1220.4394	65		

$H_0 : \beta_1 = \beta_2 = 0$   $H_1 : \beta_i$  不全為 0

$\alpha = 0.05$

$F^* = 69.7757 \in C$

$C : \{F^* > F_{0.05}(2, 63)\}$

沒有 F 表的情況下，觀察此處 F 檢定統計量已足夠大，故拒絕  $H_0$ ，故有顯著證據說身高與手掌張開長度和性別有線性關係存在。

(三)  $H_0 : \text{線性迴歸模型成立}$   $H_1 : \text{線性迴歸模型不成立}$

$\alpha = 0.05$

$F^* = 0.8433 \notin C$

$C : \{F^* > F_{0.05}(18, 45)\}$

沒有 F 表的情況下，觀察此處 F 檢定統計量非常小，故無法拒絕  $H_0$ ，故沒有顯著證據說線性迴歸模型不成立，即迴歸模型沒有顯著的缺適。

二、一位統計分析師分析奧林匹克男子田徑短跑 200 公尺數據，包含 1900 年至 2020 年間舉行的 28 次男子 200 公尺奧林匹克短跑比賽獲金牌的秒數，其中第一次和第二次世界大戰期間沒有舉辦奧運會，而 2020 年奧林匹克運動會因為 COVID-19 疫情實際是 2021 年在日本東京舉行。因此資料包含 year (以年為單位) 和 Y (以秒為單位)，其散布圖在圖 1。

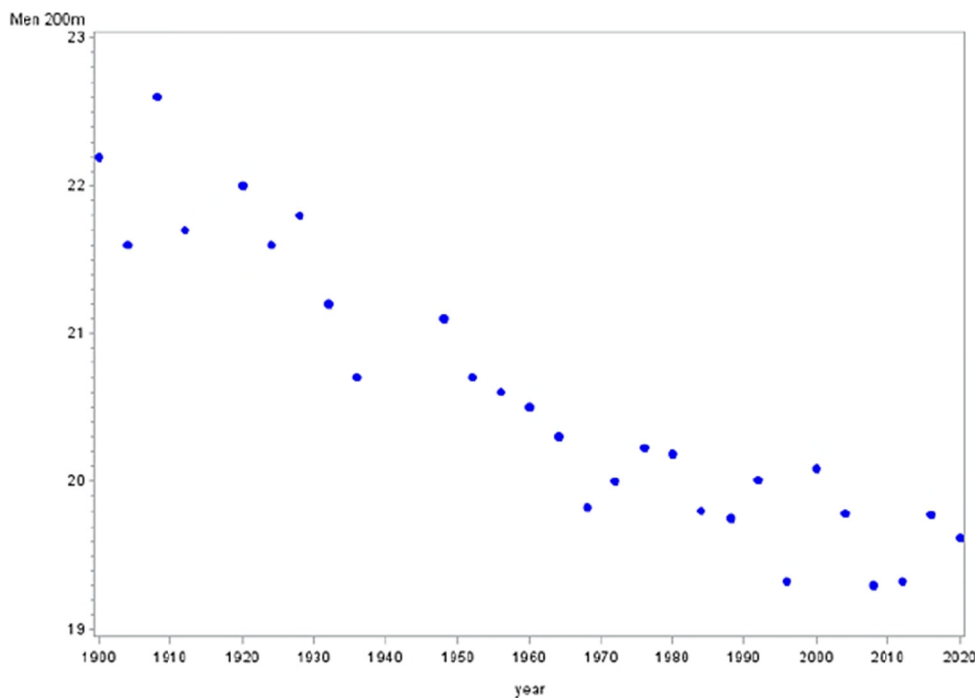


圖 1 奧林匹克年份和男子田徑短跑 200 公尺秒數散布圖

這位統計分析師重新定義變數，他把“西元年 (year)”平減 1963，並定義新的解釋變數  $X$ ，也就是  $X = \text{year} - 1963$ 。樣本相關資訊如下，其中  $n$  為樣本數，請依據這些資訊回答問題。

$$\bar{X} = -0.1429, \bar{Y} = 20.5582, S_{XY} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = -888.2171$$

$$S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2 = 36859.4286, S_{YY} = \sum_{i=1}^n (Y_i - \bar{Y})^2 = 24.3354$$

(一)請計算 $(X, Y)$ 的皮爾森相關係數。(5分)

(二)該統計分析師配適模型 $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ，此處 $\varepsilon_i$ 是誤差項。請寫出以最小平方估計法所得到的估計迴歸線，並推導共變異數 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ ，也就是 $Cov(\hat{\beta}_0, \hat{\beta}_1)$ 。(10分)

(三)在顯著水準 0.05 之下，請檢定 $H_0: \beta_1 = 0$ 是否顯著？請詳述檢定統計量之值、決策法則和結論。請問年份和獲金牌的秒數之間是否存在線性關係？以此資料是否可以推論人類在田徑短跑越跑越快？ $t$ 分配臨界值， $t_{0.025}(26) = -2.0555$ ， $t_{0.025}(27) = -2.0518$ 。(10分)

**【解題關鍵】**

1. 《考題難易》★☆☆☆☆
2. 《破題關鍵》簡單線性迴歸的基本題，因為已經提供 X 與 Y 的三個平方和，所以計算量不會很大，屬於必拿的分數。
3. 《命中特區》王瑋，迴歸分析，志光出版，頁 2-39 題 29~頁 2-47 題 33。

**【擬答】**

$$(一) r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = \frac{-888.2171}{\sqrt{36859.4286 \times 24.3354}} = -0.9378$$

$$(二) 1. \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{-888.2171}{36859.4286} = -0.0241$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 20.5582 - (-0.0241) \times (-0.1429) = 20.5548$$

所以最小平方方法之迴歸線 $\hat{Y} = 20.5548 - 0.0241X$

$$2. Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{Y} - \hat{\beta}_1 \bar{X}, \hat{\beta}_1) = Cov(\bar{Y}, \hat{\beta}_1) - \bar{X}Cov(\hat{\beta}_1, \hat{\beta}_1)$$

$$= 0 - \bar{X}V(\hat{\beta}_1) = -\frac{\bar{X}}{S_{XX}}\sigma^2$$

$\sigma^2$  可以 MSE 估計之。

$$(三) MSE = \frac{SSTO - SSR}{n-2} = \frac{S_{YY} - \hat{\beta}_1^2 \cdot S_{XX}}{10-2}$$

$$= \frac{24.3354 - (-0.0241)^2 \cdot 36859.4286}{28-2} = 0.1126$$

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

$$\alpha = 0.05$$


$$T^* = \frac{\hat{\beta}_1}{\sqrt{\frac{MSE}{S_{XX}}}} = \frac{-0.0241}{\sqrt{\frac{0.1126}{36859.4286}}} = 13.79 \in C$$

$$C: \{|T^*| > t_{0.975}(26) = 2.0555\}$$

拒絕 $H_0$ ，有顯著證據說年份和獲金牌的秒數之間存在線性關係。


若要推論人類在田徑短跑是否越跑越快，可進行 $H_1: \beta_1 < 0$ 之左尾檢定，以驗證年份和獲金牌的秒數是否為負相關。

## 五大學習方式 上課超便利




**現場面授**

名師現場面對面  
即時互動解答疑惑




**直播教學**

即時登入直播跟課  
掌握進度免等待




**視訊課程**

手機APP預約上課  
輔導期間 無限重覆看課



**WIFI看課**

專屬WIFI教室  
讓你學習時間更彈性



**在家學習**

使用在家補課點數  
即可在家複習上課  
(以老師授權科目為主)

持地方特考准考證享專案優惠(詳細請洽全國各班門市)

三、一位統計分析師受託分析 20 名年齡 40~60 歲高血壓患者的血壓相關數據，以評估可能影響血壓的重要因素，資料描述如下：

血壓 ( $Y$ ，反應變數，以 mmHg 為單位)，年齡 ( $X_1$ ，以年為單位)，重量 ( $X_2$ ，公斤)，體表面積 ( $X_3$ ，平方公尺)，高血壓病史 ( $X_4$ ，以年為單位)，基礎脈搏 ( $X_5$ ，以每分鐘為單位)，壓力指數 ( $X_6$ ，0-100 為範圍)。部分統計套裝軟體輸出結果在表2和表3。

表2

反應變數	5個解釋變數	判定係數 $R_j^2$
$X_1$	$X_2 - X_6$	0.451
$X_2$	$X_1, X_3 - X_6$	0.925
$X_3$	$X_1 - X_2, X_4 - X_6$	0.905
$X_4$	$X_1 - X_3, X_5 - X_6$	0.196
$X_5$	$X_1 - X_4, X_6$	0.754
$X_6$	$X_1 - X_5$	0.416

表3

解釋變數	Type I SS	偏判定係數
$X_1$	$SSR(X_1)$ 243.266	$R_{Y, X_1}^2$ 0.4344
$X_2$	$SSR(X_2   X_1)$ 306.886	$R_{Y, X_2   X_1}^2$ 0.96891
$X_3$	$SSR(X_3   X_1, X_2)$ 0.765	$R_{Y, X_3   X_1, X_2}^2$ 0.07763
$X_4$	$SSR(X_4   X_1, X_2, X_3)$ 0.250	$R_{Y, X_4   X_1, X_2, X_3}^2$ 0.02755
$X_5$	$SSR(X_5   X_1, X_2, X_3, X_4)$ 0.965	$R_{Y, X_5   X_1, X_2, X_3, X_4}^2$ 0.1092
$X_6$	$SSR(X_6   X_1, X_2, X_3, X_4, X_5)$ 1.023E-04	$R_{Y, X_6   X_1, X_2, X_3, X_4, X_5}^2$ 1.3E-05

(-)這位分析師一開始採用(1)式中模型 1 的複迴歸分析，他擔心有多重共線性 (Multicollinearity) 問題。

模型 1：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \varepsilon_i, \quad i = 1, \dots, n$$

請協助這位分析師利用表 2 判斷是否有嚴重的多重共線性，並說明模型 1 是否合適？如果

公職王歷屆試題 (111 地方特考)

不合適，請詳細說明原因和判斷方法。(5分)

(二)表 3 第二欄的定義，若  $SSR(X_i | X_j)$  代表給定  $X_j$  已在模型中， $X_i$  加入模型中的額外平方和 (extra sum of squares)。請計算  $SSR(X_1, X_2, X_3, X_4, X_5, X_6)$ 。最後一欄符號代表偏判定係數 (coefficient of partial determination)。請說明偏判定係數  $R^2_{Y, X_3 | X_1, X_2}$  的計算式及其意義。

請利用表 3 結果，建議分析師採用那些變數，詳細說明理由和判斷方法。(10分)

(三)請利用表 3 結果及  $SST = 560$ ， $SSR(X_1, X_2, X_5) = 551.568$ ，計算  $SSR(X_5 | X_1, X_2)$  和偏判定係數  $R^2_{Y, X_5 | X_1, X_2}$ 。(10分)

【解題關鍵】

1. 《考題難易》★★★★☆☆
2. 《破題關鍵》多元共線性判斷透過計算 VIF 值，在 109 年地特與 101 年高考迴歸有類似考題；序列平方和與偏判定係數是近年迴歸分析的重要考題，本題雖沒有要求進行偏 F 檢定，但可就額外平方和或偏判定係數是否夠大來決定是否需要加入新的變項，106 年高考迴歸有相似命題。
3. 《命中特區》王瑋，迴歸分析，志光出版，頁 7-89 題 34、頁 7-110 題 43~頁 7-111 題 44 頁。

【擬答】

(一)可透過變異數膨脹因子(VIF)來判斷是否存在共線性問題

$$VIF_{X_i} = \frac{1}{1 - R^2(X_i \text{ 對其他自變數作複迴歸})}$$

$$VIF_{X_1} = \frac{1}{1 - 0.451} = 1.82 < 10$$

$$VIF_{X_2} = \frac{1}{1 - 0.925} = 13.3 > 10$$

$$VIF_{X_3} = \frac{1}{1 - 0.905} = 10.53 > 10$$

$$VIF_{X_4} = \frac{1}{1 - 0.196} = 1.24 < 10$$

$$VIF_{X_5} = \frac{1}{1 - 0.754} = 4.07 < 10$$

$$VIF_{X_6} = \frac{1}{1 - 0.416} = 1.71 < 10$$

VIF 值大於 10，則此  $X_i$  可被其他自變數解釋或取代。此例變數  $X_2$  與  $X_3$  之 VIF 值大於 10，則這兩個解釋變數存在嚴重共線性問題。

解決的辦法可去除此兩者相依的自變數，或採用脊迴歸方法來消除共線性問題。

(二) 1.  $SSR(X_1, X_2, X_3, X_4, X_5, X_6)$

$$= SSR(X_1) + SSR(X_2 | X_1) + SSR(X_3 | X_1, X_2)$$

$$+ SSR(X_4 | X_1, X_2, X_3) + SSR(X_5 | X_1, X_2, X_3, X_4)$$

$$+ SSR(X_6 | X_1, X_2, X_3, X_4, X_5)$$

$$= 243.266 + 306.886 + 0.765 + 0.250 + 0.965 + 0.0001$$

$$= 552.1321$$

$$2. R^2_{Y, X_3 | X_1, X_2} = \frac{SSR(X_3 | X_1, X_2)}{SSE(X_1, X_2)}$$

表示模型中有  $X_1, X_2$  時，新加入  $X_3$  後所降低的未解釋變異比例。

3. 由表 3 可以看出當新增  $X_1$  與  $X_2$  變項時，額外平方和與偏判定係數都很大，代表需要納入這兩個變項，然而當加入  $X_3$  時，額外平方和僅 0.76 與偏判定係數 0.0776，代表已有



公職王歷屆試題 (111 地方特考)

$X_1, X_2$  下，加入  $X_3$  無法提供足夠的解釋力，所以最佳模型可考慮僅放入  $X_1, X_2$  變項。

$$\text{(三)} SSR(X_1, X_2) = SSR(X_1) + SSR(X_2 | X_1) = 243.266 + 306.886 = 550.152$$

$$SSR(X_5 | X_1, X_2) = SSR(X_1, X_2, X_5) - SSR(X_1, X_2)$$

$$= 551.568 - 550.152 = 1.416$$

$$SSE(X_1, X_2) = SST - SSR(X_1, X_2)$$

$$= 560 - 550.152 = 9.848$$

$$R_{Y, X_5 | X_1, X_2}^2 = \frac{SSR(X_5 | X_1, X_2)}{SSE(X_1, X_2)} = \frac{1.416}{9.848} = 0.144$$

四、一位教師擬瞭解學生的測試表現是否受智商和教學方法所影響，以 60 名學生為實驗對象，在採用三種教學方法之下，獲得測試成績  $Y$ ，智商  $X$ 。前兩種教學方法  $M_1, M_2$  變數定義如下。

$$M_1 = \begin{cases} 1 & \text{教學法1} \\ 0 & \text{其他} \end{cases} \quad M_2 = \begin{cases} 1 & \text{教學法2} \\ 0 & \text{其他} \end{cases}$$

這位教師分別考慮的模型如下：

模型 1： $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

模型 2： $Y_i = \beta_0 + \beta_2 M_{1i} + \beta_3 M_{2i} + \varepsilon_i$

模型 3： $Y_i = \beta_0 + \beta_1 X_i + \beta_2 M_{1i} + \beta_3 M_{2i} + \varepsilon_i$

請使用表 4 部分電腦輸出 3 個模型的變異數分析 (ANOVA, Analysis of Variance) 報表來回答下列問題。

(一)在考慮模型 3 之下，請檢定智商  $X$  該解釋變數對於解釋測試成績是否有顯著的解釋能力。

請用顯著水準  $\alpha = 0.05$  檢定並詳述檢定統計量之值、決策法則、結論和所需之假設。t 分配臨界值， $t_{0.975}(56) = 2.0032$ 。(10 分)

(二)在考慮模型 3 之下，請檢定教學方法  $M_1$  和  $M_2$  這兩個虛擬變數是否在模型 3 對預測學生測試成績有效應。請在顯著水準  $\alpha = 0.05$ ，檢定  $H_0: \beta_2 = \beta_3 = 0$ ，請詳述檢定統計量之值、

決策法則、結論和所需之假設。F 分配左尾臨界值， $F_{0.95}(1, 56) = 4.0130$ ，

$F_{0.95}(2, 56) = 3.1619$ 。(10 分)

(三)請使用表 4 說明那一種教學方法最能提升測試成績，須說明論述。(5 分)

表4

模型1 ANOVA表

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F value	P-value
Regression	1	816.928	816.928	14.72	0.0003
Error	58	3219.255	55.504		
Total	59	4036.183			

模型2 ANOVA表

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F value	P-value
Regression	2	2880.033	1440.017	71	P-value
Error	57	1156.150	20.283		
Total	59	4036.183			

模型3 ANOVA表和參數估計

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F value	P-value
Regression	3	3512.745	1170.915	125.27	<.0001
Error	56	523.438	9.347		
Total	59	4036.183			

模型3參數估計					
Variable	DF	Estimate	Standard Error	t value	P-value
Intercept	1	56.024	4.306	13.01	<.0001
$X$	1	0.350	0.043	8.14	<.0001
$M_1$	1	-15.770	0.967	-16.3	<.0001
$M_2$	1	-11.943	0.972	-12.28	<.0001

## 【解題關鍵】

- 《考題難易》★★☆☆☆
- 《破題關鍵》同時考慮屬量與虛擬變項的迴歸模型常見於考古題，尤其是迴歸係數的含意與解釋特別重要；試卷提供不同模型 ANOVA 表，間接暗示採用偏 F 檢定來進行選模，亦是近年考試的熱門題型。
- 《命中特區》王瑋，抽樣方法，志光出版，頁 7-43 題 17~頁 7-55 題 20。

## 【擬答】

$$(\rightarrow) H_0: \beta_1 = 0 \quad H_1: \beta_1 \neq 0$$

$$\alpha = 0.05$$

$$T^* = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} = 8.14 \in C$$

$$C: \{|T^*| > t_{0.975}(56) = 2.0032\}$$

並且表 4 中對應之  $p < 0.0001$

拒絕  $H_0$ ，有顯著證據說智商  $X$  對於解釋測試成績有顯著的解釋能力

$$(\rightarrow) H_0: \beta_2 = \beta_3 = 0$$

$$\alpha = 0.05$$

$$F^* = \frac{SSR(X_2, X_3 | X_1) / 2}{MSE(X_1, X_2, X_3)} = \frac{[SSR(X_1, X_2, X_3) - SSR(X_1)] / 2}{MSE(X_1, X_2, X_3)}$$

$$= \frac{(3512.745 - 816.928) / 2}{9.347} = 144.208 \in C$$

$$C: \{F^* > F_{0.95}(2, 56) = 3.1619\}$$

拒絕  $H_0$ ，有顯著證據說教學方法對預測學生測試成績有顯著效應

$$(\rightarrow) \text{教學法 1 } (M_1 = 1, M_2 = 0) \Rightarrow \hat{Y} = 56.024 + 0.35X - 15.77$$

$$\text{教學法 2 } (M_1 = 0, M_2 = 1) \Rightarrow \hat{Y} = 56.024 + 0.35X - 11.943$$

$$\text{教學法 3 } (M_1 = 0, M_2 = 0) \Rightarrow \hat{Y} = 56.024 + 0.35X$$

由此可知教學法 3 的測試成績高於教學法 1 與教學法 2，所以教學法 3 最能提升測試成績。

志光×保成×學儒  
做你的神兵利器

# 穩佔高普 穩穩上榜

## 3大課程圈 穩固你的考取實力

不怕非本科 不怕沒基礎

### ●打基礎

✍️ 基礎班&架構班

正規課前專屬導讀課，掌握專業科目基本觀念架構，快速釐清各科完整脈絡。

### ●增實力

💡 正規班

· 重點科目採多元師資教學，可汲取多位精華  
· 同類科開立多循環課程，可配合旁聽加強弱科

### ●抓考點

📊 總複習

考前重要章節總整理，補充最新修法時事，關鍵時刻帶你衝對方向不失分。

依各區規劃為主，請洽全國門市

志光×保成×學儒 為你絕佳助攻

# 四大衝刺課程 帶你直攻112高普考

## 題庫班

讀書精熟+答題精準=快速上榜

題庫演練

精準教學

解題技巧

## 測驗易點通

埋頭苦練 不如讓老師點通你的學習之路

常考題型知識強化

易錯題型觀念釐清

## 總複習

考點update! 時事修法update!

關鍵  
考點

最新  
考情

考前  
複習

短期  
密集

## 作文實戰班

作文學得好，同時提升寫作能力與論述邏輯

高分  
寫作指引

架構  
分層演練

強化  
論述深度

新式  
作文教戰

依各區規劃為主，請洽全國門市