

《統計學概要》

試題評析

- 第一題：有關Poisson分配平均數與變異數及柴比雪夫不等式之應用，本題關鍵是要考生根據題目寫出每次校對費用與Poisson隨機變數之函數式。
- 第二題：有關共變數之計算，考古題也出現過相同的題型，本題難度不高，只要小心計算，要拿滿分不難。
- 第三題：本題屬卡方獨立性檢定，普考中常有的題型，只要小心計算，滿分不難。
- 第四題：有關兩相依常態母體平均數差之檢定，本題難度不高，但要注意題目缺少隨機樣本的假設，故沒有寫下這個必要假設會被扣分。
- 第五題：本題屬迴歸分析基本計算題。
- 第六題：有關ANOVA之計算題，只要小心計算，要拿滿分不難。
- 第七題：無母數統計方法之Wilcoxon signed rank test，檢定統計量之公式有點長，如果忘記推導過程，有可能背錯而失去分數。

- 一、某出版社出版之英語書籍初稿中每頁平均錯字個數 Y ，服從平均數為9之卜瓦松 (Poisson) 分布。今郵寄請工讀生校對，每次校對10頁，修正每一錯字之費用是3元，而支付郵費一次80元。
- (一) 試求每次校對費用之期望值與標準差。(10分)
- (二) 試求機率至少為0.75時，每次校對費用之上下界。(10分)

答：

令 $N(t)$ 表在 t 倍一頁內錯字之個數， $N(t) \sim \text{Poisson}(9t)$

(一)

$$E[3N(t=10) + 80] = 3(9)(10) + 80 = 350 \text{ (元)}$$

$$\sqrt{V[3N(t=10) + 80]} = \sqrt{9V[N(t=10)]} = \sqrt{9(9)(10)} = 28.46 \text{ (元)}$$

(二) 柴比雪夫不等式

令 $Y = 3N(t=10) + 80$ 表每次校對費用(元)， $Y \sim (350, (28.46)^2)$

$$\text{令 } 1 - \frac{1}{k^2} = 0.75 \Rightarrow k = 2$$

由柴比雪夫不等式，

$$P(|Y - 350| \leq (2)(28.46)) \geq 0.75 \Rightarrow P(293.08 \leq Y \leq 406.92) \geq 0.75$$

故在機率至少0.75下，每次校對費用之上下界為(293.08, 406.92) (元)

【高分閱讀】趙治勳老師上課講義5.5與4.5

二、若隨機變數 $Y_1, Y_2 \in \{0, 1, 2\}$ 且其聯合機率分布 (joint probability distribution) 如下：

		Y_1		
		0	1	2
Y_2	0	1/9	2/9	1/9
	1	2/9	2/9	0
	2	1/9	0	0

(一) 試求 Y_1 和 Y_2 之共變異數 (covariance)。(5分)

(二) 試求 $\bar{Y} = (Y_1 + Y_2)/2$ 之抽樣分布。(10分)

答：

(一)

$$E(Y_1) = (0)(4/9) + (1)(4/9) + (2)(1/9) = \frac{2}{3} = E(Y_2),$$

$$E(Y_1 Y_2) = \sum_{y_1=0}^2 \sum_{y_2=0}^2 f(y_1, y_2) = (1)(1)(2/9) = \frac{2}{9}$$

$$\text{Cov}(Y_1, Y_2) = E(Y_1 Y_2) - E(Y_1)E(Y_2) = -\frac{2}{9}$$

(二)

$\bar{Y} = \bar{y}$	0	0.5	1
$f_{\bar{Y}}(\bar{y})$	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

【高分閱讀】趙治勳老師上課講義4.3六(五)。

三、某娛樂公司欲了解電視節目型態與觀眾性別是否相關做了一項調查，抽訪了1200名成年觀眾，並詢問他們對政論節目的喜好，結果如下：

	喜歡	不喜歡	無意見
女性	361	228	17
男性	433	141	20

試問觀賞政論節目之偏好是否與性別有顯著關聯性 (association)？寫出檢定統計量之分布。
($\alpha = 5\%$, critical value = 5.99) (10分)

答：

H_0 : 偏好與性別獨立 vs H_1 : 偏好與性別相依

$$\text{T.S.: } \chi^2 = \sum_{i=1}^2 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{(2)}^2$$

R.R.: Reject H_0 at $\alpha = 0.05$ if $\chi^{2*} > \chi_{0.05(2)}^2 = 5.99$

$$\because \chi^{2*} = 27.167 \quad \therefore \text{reject } H_0$$

我們有足夠證據去推論偏好與性別具有關聯性。

【高分閱讀】趙治勳老師上課講義11.6一。

四、某A、B兩種飲料，經10位品評員評分，結果如下表：

品評員	1	2	3	4	5	6	7	8	9	10	合計	平均
飲料 A	4	3	7	2	1	3	3	1	4	2	30	3
B	5	3	6	2	4	2	6	5	7	5	45	4.5

若評分符合常態分布，試問兩種飲料之評分是否有顯著差異？寫出檢定統計量之分布。

($\alpha = 5\%$, critical value = 2.262) (10分)

答：

令 X, Y 分別表飲料A與飲料B之評分，且其母體平均數分別以 μ_1, μ_2 表示

設 $D_i = X_i - Y_i, i = 1, 2, \dots, 10$ 假設隨機樣本

母體: $D \sim N(\mu_D, \sigma_D^2)$ 其中 $\mu_D = \mu_1 - \mu_2$

樣本: $D_1, D_2, \dots, D_{10} \sim N(\mu_D, \sigma_D^2)$

點估計: $\bar{D} \sim N(\mu_D, \frac{\sigma_D^2}{10})$

$H_0: \mu_D = 0$ vs $H_1: \mu_D \neq 0$

T.S.: $T = \frac{\bar{D} - 0}{\frac{S_D}{\sqrt{10}}} \sim t_{(9)}$

R.R.: Reject H_0 at $\alpha = 0.05$ if $|T^*| > t_{0.025(9)} = 2.262$

$\therefore |T^*| = \left| \frac{-1.5 - 0}{\frac{1.9}{\sqrt{10}}} \right| = 2.497 \therefore \text{reject } H_0$

我們有足夠證據去推論兩種飲料之評分有差異。

【高分閱讀】趙治勳老師上課講義9.3二(六)。

五、自某大學統計系學生中隨機抽取10位，令 x_i 為第 i 位學生之微積分成績， y_i 為第 i 位學生之統計學成績，且 (x_i, y_i) 具二維常態分布， $i = 1, \dots, 10$ 。若 $\sum_{i=1}^{10} x_i = 460$ ， $\sum_{i=1}^{10} x_i^2 = 23634$ ， $\sum_{i=1}^{10} y_i = 760$ ，

$\sum_{i=1}^{10} y_i^2 = 59816$ ，及 $\sum_{i=1}^{10} x_i y_i = 36854$ 。

(一)試檢定微積分與統計學成績是否具線性關係？寫出檢定統計量之自由度。

($\alpha = 5\%$, $F = 5.32$) (10分)

(二)若某生微積分成績50分，試求其統計學成績之95%的預測區間。(10分)

答：

$$SS_{XY} = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{10} = 1894, \quad SS_X = \sum X_i^2 - \frac{(\sum X_i)^2}{10} = 2474,$$

$$SS_Y = \sum Y_i^2 - \frac{(\sum Y_i)^2}{10} = 2056$$

$$\hat{\beta}_1 = \frac{SS_{XY}}{SS_X} = 0.7656, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 40.7824, \quad \therefore \hat{y} = 40.7824 + 0.7656x$$

$$MSE = \frac{SSE}{10-2} = \frac{SS_Y - \hat{\beta}_1^2 SS_X}{8} = 75.735$$

$$S(\hat{\beta}_1) = \sqrt{\frac{MSE}{SS_X}} = 0.175$$

(一)

$$H_0 : \rho_{XY} = 0 \text{ vs } H_1 : \rho_{XY} \neq 0$$

$$\text{T.S.: } T = \frac{\hat{\beta}_1 - 0}{S(\hat{\beta}_1)} \sim t_{(8)}$$

$$\text{R.R.: Reject } H_0 \text{ at } \alpha = 0.05 \text{ if } |T^*| > t_{(8)0.025} = \sqrt{F_{0.05(1,8)}} = 2.3065$$

$$\therefore |T^*| = \left| \frac{0.7656 - 0}{0.175} \right| = 4.375 \quad \therefore \text{reject } H_0$$

我們有足夠證據去推論微積分與統計學成績具有線性關係。

(二)

當 $x_0 = 50$ 下，統計學成績之95%預測區間

$$(40.7824 + 0.7656(50) \mp (2.3065) \sqrt{75.735(1 + \frac{1}{10} + \frac{(50-46)^2}{2474})})$$

$$= (57.9484, 100.1764)$$

【高分閱讀】趙治勳老師，《迴歸分析熱門題庫》2-22~24頁第五、六節。

六、欲研究三種教學方法對學習成果的影響所得數據 y_{ij} 為第 i 種方法 ($i=1, 2, 3$) 下第 j 位 ($j=1, \dots, n_i$) 學生之測驗成績，結果如下：

	n_i	$\sum_{j=1}^{n_i} y_{ij}$	$\sum_{j=1}^{n_i} y_{ij}^2$					
y_{1j}	73	83	76	68	80	5	380	29018
y_{2j}	54	74	71		3	199	13433	
y_{3j}	79	95	87		3	261	22835	

試寫出ANOVA表並判斷三種教學方法之測驗成果是否有顯著差異。寫出檢定統計量之自由度。
($\alpha=5\%$, $F=4.46$) (15分)

答：

可得 $\bar{Y}_1 = 76, S_1^2 = 34.5$ ， $\bar{Y}_2 = 66.333, S_2^2 = 116.333$ ， $\bar{Y}_3 = 87, S_3^2 = 64$

$$SSE = \sum_{i=1}^3 (n_i - 1)S_i^2 = (4)(34.5) + (2)(116.333) + (2)(64) = 498.666$$

$$SSR = \sum_{i=1}^3 n_i (\bar{Y}_{i\cdot} - \bar{Y})^2 = (5)(76 - 76.364)^2 + (3)(66.333 - 76.364)^2 + (3)(87 - 76.364)^2 = 761.131$$

$$\text{其中 } \bar{Y} = \frac{(5)(76) + (3)(66.333) + (3)(87)}{5 + 3 + 3} = 76.364$$

ANOVA TABLE				
source	SS	d.f.	MS	F
教學方法	761.131	2	380.5655	$F^* = 6.105$
Error	498.666	8	62.33325	
Total	1259.797	10		

$H_0: \mu_1 = \mu_2 = \mu_3$ vs $H_1: \text{至少一個 } \mu_i \neq \mu_j, i \neq j = 1, 2, 3$

$$\text{T.S.: } F = \frac{MSR}{MSE} \sim F_{(2,8)}$$

R.R.: Reject H_0 at $\alpha = 0.05$ if $F^* > F_{0.05(2,8)} = 4.46$

$\because F^* = 6.105 \quad \therefore \text{reject } H_0$

我們有足夠的統計證據去推論三種教學方法之測驗成果有差異。

【高分閱讀】趙治勳老師上課講義10.3一。

七、資料 83 76 68 80 54 74 71 79 95 87 為來自同一母體之隨機樣本，試檢定母體中位數 (median) 是否為 85，求其 p-value。 ($\alpha=5\%$) (10分)

答：

Wilcoxon signed rank test

$H_0: \eta = 85$ vs $H_1: \eta \neq 85$ 其中 η 表母體中位數

Data: 83,76,68,80,54,74,71,79,95,87

等級: 8 5 2 7 1 4 3 6 10 9

因此 $T_+ = 19$, $n = 10$

$$\text{T.S.: } Z = \frac{|T_+ - \frac{1}{4}n(n+1)| - \frac{1}{2}}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}} \underset{\text{by C.L.T.}}{\sim} N(0,1)$$

R.R.: Reject H_0 at $\alpha = 0.05$ if $|Z^*| > z_{0.025} = 1.96$

$\because Z^* = 0.8154 \quad \therefore$ don't reject H_0

我們沒有足夠證據推論母體中位數不為 85。

$p\text{-value} = 2P(Z > 0.8154) = 0.412$

【高分閱讀】趙治勳老師上課講義 11.2 二。