

統計學 - 期末報告 06/20/2021 繳交

姓名：陳宣惠
座號：19
學號：409530093
系級：財金一 A

題目與答案：以下報告包含三個部分：ANOVA、Regression、Independence Test (並使用 Excel 輸出報表)，本報告共 6 頁。

1

ANOVA

以下表格為美國分別在北方、中部、南方和西方區域在一年內的自然排放廢氣的支出，我們想知道分別在這四個區域在一年內的自然排放廢氣的支出是否相同，所以使用 ANOVA Test - F 檢定驗證此宣稱是否為正確。

北方	中部	南方	西方
1608	449	509	591
779	1036	394	504
964	665	769	1011
1303	1213	753	463
1143	921	931	271
1695	1393	574	324
785	926	526	515
778	866	1096	599

Chapter9-10
Cumulative Review,
p.602, No.2

Excel 操作步驟順序

1. 點選上方的標籤「資料」
2. 點選「資料分析」
3. 點選分析工具中的「單因子變異數分析」
4. 選取原先的輸入的「資料範圍」
5. 勾選「類別軸標記是在第一列上」
6. 輸入 $\alpha = 0.1$
7. 輸出選項：點選「新工作表」，並命名為 ANOVA 報表

ANOVA 報表

摘要	組
	上方表格中的四大類別：北方、中部、南方及西方

個數	每一類別分別有 8 筆資料
總和	每一類別資料的加總
平均	每一類別的總和除以 8
變異數	每一類別的資料分散程度
變源	組間 SSB 組內 SSW
SS	加總為 Total Sum of Square - 報表上的「總和」
自由度	組數-1 資料數-組數 (n-1) (N-k)
	總和：資料數-1 (N-1)
MS	SS 除以個別自由度
F	MSB/MSW=F (下述本題之 F)
P-值	P-value 值
臨界值	P-值查表得出

單因子變異數分析

摘要

組	個數	總和	平均	變異數
北方	8	9055	1131.875	138780.7
中部	8	7469	933.625	87338.268
南方	8	5552	694	56246.857
西方	8	4278	534.75	50652.786

ANOVA

變源	SS	自由度	MS	F	P-值	臨界值
組間	1658955.63	3	552985.21	6.6420938	0.0016	2.2906
組內	2331130.25	28	83254.652			
總和	3990085.88	31				

檢定

檢定	$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ versus $H_1 : \text{至少有一個 } \mu_i \text{ 與其他母體平均數不相等}$
檢定統計量	$F = \frac{MSB}{MSW} \sim F_{\alpha(n-1, N-k)}$
檢定規則	在給定 $\alpha = 0.1$ 的信心水準下，當 $F > F_{0.1(3,28)} = 2.29$ ，拒絕 H_0
本題	$F = 6.6420938 > 2.29$ ，拒絕 H_0
結論	在給定 $\alpha = 0.1$ 的信心水準下，有足夠證據拒絕 H_0

Regression

以下表格為不同煮飯的區域大小間的瓦斯使用量的價格，我們想知道是否在這兩個類別資料間，有一個線性的迴歸關係，並使用相關係數檢定模型顯不顯著：

區域	價格
650	149
669	699
529	499
725	374
844	1599
445	187
669	1299
844	899
740	374
529	599
450	399
644	499
600	269
575	299
998	1999
529	519
265	99
530	109

Excel 操作步驟順序

1. 點選上方的標籤「資料」
2. 點選「資料分析」
3. 點選分析工具中的「迴歸」
4. 選取 Y 的範圍為 - 「價格」的資料；選取 X 的範圍為 - 「區域」的資料
5. 勾選「標記、信賴度、殘差 (殘差輸出)、殘差圖、樣本迴歸線圖」
6. 輸入信賴度為：95%
7. 輸出選項：點選「新工作表」，並命名為 Regression 報表

Review Exercise
Chapter 9, Section 9.3,
p.540, No.18

Regression 報表 - 摘要輸出 (標題)

<u>迴歸統計</u>	R 的倍數	兩筆資料的相關係數 r
	R 的平方	兩筆資料的判定係數 r^2
	調整的	把變數的數量也納入考量，形成

R 平方	r^2 的不偏估計量
標準誤	
觀察值個數	總共為 18

項目	迴歸變異 (SSR)	殘差或無法解釋的變異 (SSE)
自由度	1	資料數-2 (n-2)
	總和：資料數-1 (n-1)	
SS	迴歸與殘差加總為 Total SS - 報表上的「總和」	
ANOVA MS	SS 除以個別自由度	
F	MSR/MSE=F (下述本題之 F)	
顯著值	P-value 值	
t 統計量	區域 (下述本題之 T)	
	估計值	
係數	截距	$\widehat{\beta}_0$
	區域	$\widehat{\beta}_1$
方程式	$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$	
	$-853.28 = \bar{y} - 2.3346\bar{x}$	

摘要輸出									
迴歸統計									
R 的倍數	0.7518								
R 平方	0.5652								
調整的 R 平方	0.538								
標準誤	360.35								
觀察值個數	18								
ANOVA									
	自由度	SS	MS	F	顯著值				
迴歸	1	2700848.5	2700848.5	20.8	0.0003207				
殘差	16	2077615.3	129850.96						
總和	17	4778463.8							
	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
截距	-853.28	330.60451	2.5800636	0.02	-1554.1284	-152.4279	-1554.128	-152.42795	
區域	2.3346	0.5118954	4.5606583	3E-04	1.2494103	3.4197499	1.2494103	3.4197499	

檢定

檢定 $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$

$\beta_1 = 0$ 檢定統計量

$$T = \frac{\widehat{\beta}_1}{S_e \cdot \sqrt{\frac{1}{S_{xx}}}}$$

檢定規則 在給定 $\alpha = 0.05$ 的信心水準下，當 $|T| \geq t_{0.05(16)} = 2.12$ 時，拒絕 H_0

	本題	$T=4.5606583 > 2.12$ · 拒絕 H_0
	結論	在給定 $\alpha = 0.05$ 的信心水準下 · 有足夠證據拒絕 H_0 · 模型顯著
	檢定	$H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$
	檢定統計量	$F = \frac{MSR_{H_0}}{MSE} \rightarrow F_{\alpha(1, n-2)}$
$\beta_1 = 0$	檢定規則	在給定 $\alpha = 0.05$ 的信心水準下 · 當 $F > F_{0.05(1,16)} = 4.49$ · 拒絕 H_0
	本題	$F=20.8 > 4.49$ · 拒絕 H_0
	結論	在給定 $\alpha = 0.05$ 的信心水準下 · 有足夠證據拒絕 H_0 · 模型顯著
	檢定	$H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$
	檢定統計量	$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \xrightarrow{H_0} t(n-2)$
$\rho = 0$	檢定規則	在給定 $\alpha = 0.05$ 的信心水準下 · 當 $ T \geq t_{0.05(16)} = 2.12$ 時 · 拒絕 H_0
	本題	$T=4.5606583 > 2.12$ · 拒絕 H_0
	結論	在給定 $\alpha = 0.05$ 的信心水準下 · 有足夠證據拒絕 H_0 · 模型顯著

Independence Test

以下表格為某財務人員研究之資訊，主要研究學貸的貸款者與家庭所得之關係。我們想利用卡方之獨立性檢測這兩筆類別型資料是否有相關。

觀察值		貸款者				合計
		只有學生	只有家長	兩個都有	兩個均無	
家庭所得	少於35,000	149	34	10	311	504
	35,000~100,000	181	68	58	421	728
	大於100000	69	40	14	238	361
合計		399	142	82	970	1593

使用 Excel 計算出理論值：

(以第二行第二列為例，輸入公式為：=\$G3*C\$6/\$G\$6，再將右下角的十字拉至整個資料範圍，並個別加總，就可得出此表格)

理論值		貸款者				合計
		只有學生	只有家長	兩個都有	兩個均無	
家庭所得	少於35,000	126.237	44.9266	25.9435	306.893	504
	35,000~100,000	182.343	64.8939	37.4739	443.289	728
	大於100000	90.42	32.1795	18.5825	219.818	361
合計		399	142	82	970	1593

Chapter10.2,
p.567, No.24

再使用 Excel 計算出本題之檢定統計量：

項目	檢定統計量
	4.10450081
只有學生	0.009887842
	5.074264295
只有家長	2.657438984
	0.148670185
	1.900576395
兩個都有	9.798032442
	11.24297828
	1.130079223
	0.054971273
兩個均無	1.120750835
	1.503911796
加總為	38.74606235

檢定

檢定 H_0 ：貸款對象與家庭所得獨立 versus H_1 ：貸款對象與家庭所得相關

檢定統計量
$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi^2_{\alpha, ((r-1) \times (c-1))}$$

檢定規則 在給定 $\alpha = 0.01$ 的信心水準下，當 $\chi^2 > \chi^2_{0.01, (6)} = 16.812$ 時，拒絕 H_0

本題 $\chi^2 = 38.74606235 > 16.812$ ，拒絕 H_0

結論 在給定 $\alpha = 0.05$ 的信心水準下，有足夠證據拒絕 H_0