十五、複迴歸分析

Chapter 15　Multiple Regression Analysis

追尋那道光　李明聰

目錄

學習目標

知識(認知)

1.可以清楚陳述複迴歸分析的意涵。

2.可以清楚陳述在複迴歸分析中複判定係數的意涵。

3.可以說明各種狀況下，斜率顯著性檢定的程序和標準。

4.評價各種情境下，複迴歸分析的使用價值。

技能

1.能夠計算各種情境下的迴歸係數與截距樣本統計值。

2.能夠利用斜率顯著性檢定，提出統計推論。

3.綜合所學，能夠於實務領域中，依據特定情境的需求進行複迴歸分析程序。

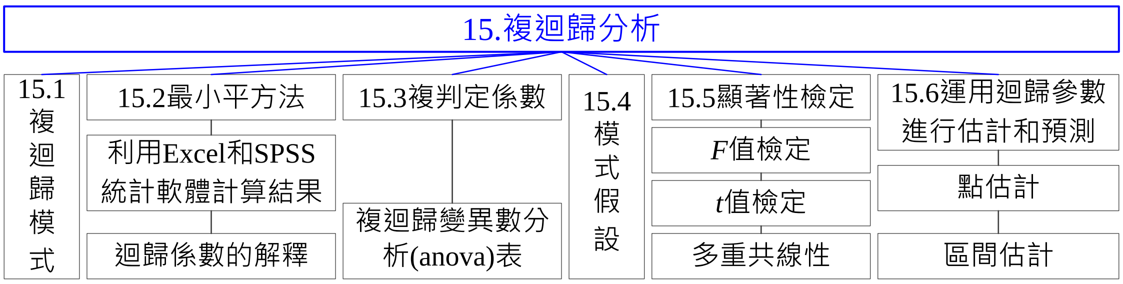
態度(情意)

1.意識到在日常生活或未來工作環境中，複迴歸分析的重要性。

2.在各種情境下，依循複迴歸分析的程序，接受統計推論所傳達的意涵。

若欲分析的變數只有一個自變數(自變項)和另一個依變數(依變項)時，兩者的關係趨近於比例關係(線性關係、直線關係)時，歸類為**簡單線性迴歸分析**(simple linear regression analysis)。在迴歸程序中若有超過兩個(含兩個)的自變數與一個依變數時，歸類為**複迴歸分析**、**多重迴歸分析**或**多元迴歸分析**(multiple regression analysis)。透過前一章簡單線性迴歸分析後，了解到可以利用一個自變數預估另一個依變數的數值，在管理實務領域中，經常需要使用多個自變數，期望能夠客觀與準確地預測另一個依變數的數值，故，凸顯出學習複迴歸分析的價值。

迴歸分析不適合運用於特例的預測，因為特例的相關狀態與一般群體完全不一樣，千萬不要認為統計學就可以解釋全部的情境。



章節結構圖

15.1 複迴歸模式

複迴歸分析中自變數有兩個(含)以上，依變數僅有一個 Y。同時探索兩個(含)以上(一般使用符號 k 或 p，代表自變數數量)自變數對一個依變數的關係，即屬於複迴歸分析、多重迴歸分析或多元迴歸分析。

確定性數學模式

敘述自變數(X_1、X_2、X_3、…、X_k)和依變數 Y 之間關係的方程式。

$$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki}\text{，其中 } i = 1, \ldots, n$$

複迴歸模式

敘述自變數(X_1、X_2、X_3、…、X_k)、依變數 Y 和誤差項 ε_i 之間關係的方程式稱為複迴歸模式、多元迴歸模式(multiple regression model)或一階複迴歸線性模式(first-order multiple linear regression model)。

$$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki} + \varepsilon_i\text{，其中 } i = 1, \ldots, n$$

其中 y_i = 依變數 Y 第 i 個觀測值的實際觀測值

x_{ki} = 第 k 個自變數 X 第 i 個觀測值

β_0 = 複迴歸模式的參數(parameter)，截距(intercept)。數值可能範圍-∞~+∞。

$\beta_1, \ldots, \beta_k$ = 複迴歸模式的參數，偏迴歸係數(partial regression coefficient)或迴歸係數(regression coefficient)。數值可能範圍-∞~+∞。

ε_i = 第 i 個觀測值的隨機變數，屬於隨機誤差(random error)，讀音 epsilon。此誤差項(error term)屬於在 X 和 Y 線性關係上無法解釋的依變數 Y 變異性、波動性、變動性。

n = 觀測值(組)數量

$k =$ 自變數數量(個數)，$k > 0$，正整數

複迴歸方程式

複迴歸模式中假設誤差項 ε_i 的平均值或期望值等於 0。因此，依變數 y_i 的平均值$\bar{y}_i$或期望值 E(y_i)即等於 $\beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \beta_3\times x_{3i} + \ldots + \beta_k\times x_{ki}$。敘述自變數($X_1$、$X_2$、$X_3$、…、$X_k$)與依變數 y_i 的平均值$\bar{y}_i$或期望值 $E(y_i)$之間關係的方程式稱為複迴歸方程式、多元迴歸方程式(multiple regression equation)或一般複迴歸方程式(general multiple regression equation)。

$$\bar{y}_i = E(y_i) = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \beta_3\times x_{3i} + \ldots + \beta_k\times x_{ki}\text{，其中 } i = 1, \ldots, n$$

其中$\bar{y}_i$ = 依變數第 i 個觀測點之觀測值的平均值

$E(y_i) =$ 依變數第 i 個觀測點之觀測值的期望值

$y_i =$ 依變數 Y 第 i 個觀測值的實際觀測值(變量)

$x_{ki} =$ 第 k 個自變數 X 第 i 個觀測值(變量)

$\beta_0 =$ 複迴歸模式的參數(parameter)，截距(intercept)。數值可能範圍$-\infty\sim+\infty$。

$\beta_1,\ldots, \beta_k =$ 複迴歸模式的參數，偏迴歸係數(partial regression coefficient)或迴歸係數(regression coefficient)。數值可能範圍$-\infty\sim+\infty$。

$n =$ 觀測值(組)數量

$k =$ 自變數數量(個數)，$k > 0$，正整數

估計複迴歸方程式或估計多元迴歸方程式(Estimated multiple regression equation)

大部分情況下，複迴歸模式中的迴歸參數($\beta_0, \beta_1,\ldots, \beta_k$)皆不知其實際數值。僅可以利用樣本數值估算，分別以樣本統計值 $b_0, b_1,\ldots, b_k$ 作為迴歸參數 $\beta_0, \beta_1,\ldots, \beta_k$ 的點估計值。

$$\hat{y}_i = b_0 + b_1\times x_{1i} + b_2\times x_{2i} + b_3\times x_{3i} + \ldots + b_k\times x_{ki}\text{，其中 } i = 1, \ldots, n$$

其中$\hat{y}_i$ = 依變數 Y 第 i 個觀測值的估計值

$x_{ki} =$ 第 k 個自變數 X 第 i 個觀測值

$b_0 =$ 複迴歸模式的統計值(statistic)，截距(intercept)。數值可能範圍$-\infty\sim+\infty$。

$b_1,\ldots, b_k =$ 複迴歸模式的統計值，偏迴歸係數(partial regression coefficient)或迴歸係數(regression coefficient)。數值可能範圍$-\infty\sim+\infty$。

$n =$ 觀測值(組)數量

$k =$ 自變數數量(個數)

名稱	模式或方程式
確定性數學模式(Deterministic model)	$y_i = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \beta_3\times x_{3i} + \ldots + \beta_k\times x_{ki}$
複迴歸模式(Multiple regression model)	$y_i = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \beta_3\times x_{3i} + \ldots + \beta_k\times x_{ki} + \varepsilon_i$
複迴歸方程式(Multiple regression equation)	$E(y_i) = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \beta_3\times x_{3i} + \ldots + \beta_k\times x_{ki}$
估計複迴歸方程式(Estimated multiple regression equation)	$\hat{y}_i = b_0 + b_1\times x_{1i} + b_2\times x_{2i} + b_3\times x_{3i} + \ldots + b_k\times x_{ki}$

15.2 最小平方法

最小平方法(Least squares method)是依據依變數觀測值 y_i 與預測值$\hat{y}_i$之差的平方和(sum of squares due to error, SSE)必須維持最小數值為基準。

最小平方法數學法則

$$\text{Min SSE} = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \sum_{i=1}^{n}(y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i}-b_3\times x_{3i}-\cdots-b_k\times x_{ki})^2$$

其中 y_i = 依變數 Y 第 i 個觀測值的實際觀測值

$\hat{y}_i$ = 在自變數為 x_i 時依變數 y_i 的估計值；依變數 Y 第 i 個觀測值的估計值或預測值

x_{ki} = 第 k 個自變數 X 第 i 個觀測值

b_0 = 複迴歸模式 $\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}+\beta_3\times x_{3i}+\ldots+\beta_k\times x_{ki}$ 中，參數(parameter) β_0 的估計值，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 複迴歸模式 $\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}+\beta_3\times x_{3i}+\ldots+\beta_k\times x_{ki}$ 中，參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)。數值可能範圍-∞~+∞。

複迴歸分析中，對於迴歸參數的統計值估算，皆是使用矩陣方式計算，較為繁雜，不易使用表格與計算機運算。故此兩個自變數以上的估計複迴歸方程式分析中使用統計軟體估算迴歸參數的統計值。

在兩個自變數的估計複迴歸方程式中($\hat{y}_i = b_0 + b_1\times x_{1i} + b_2\times x_{2i}$)，迴歸參數的估計值計算比較簡單，迴歸係數的估計值公式：

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}$$

$$b_0 = \bar{y}_i - b_1\times\bar{x}_1 - b_2\times\bar{x}_2$$

其中 y_i = 依變數第 i 個觀測值的實際觀測值

$\bar{y}$ = 依變數所有樣本觀測值的平均值

$\hat{y}_i$ = 在自變數為 x_i 時依變數 y_i 的估計值；依變數第 i 個觀測值的估計值

x_{ki} = 第 k 個自變數第 i 個觀測值

$\bar{x}_1$ = 第 1 個自變數所有樣本觀測值的平均值

$\bar{x}_2$ = 第 2 個自變數所有樣本觀測值的平均值

b_0 = 複迴歸模式 $\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}$ 中，參數(parameter) β_0 的估計值，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 複迴歸模式 $\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}$ 中，參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)。數值可能範圍-∞~+∞。

b_2 = 複迴歸模式 $\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}$ 中，參數(parameter) β_2 的估計值，迴歸係數(regression coefficient)。數值可能範圍-∞~+∞。

範例 15.1 阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、虱目魚套餐定價 x_{2i}(新台幣：元)和平均每日販售虱目魚套餐數 y_i(單位：套)列於下表。請計算出估計複迴歸方程式的統計值。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
10	275	96	280

題解：利用公式計算結果

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	150	156	-34.20	24.50	-41.20	1169.64	600.25	1409.04	-1009.40	-837.90
2	160	140	180	-24.20	14.50	-17.20	585.64	210.25	416.24	-249.40	-350.90
3	179	125	190	-5.20	-0.50	-7.20	27.04	0.25	37.44	3.60	2.60
4	160	137	170	-24.20	11.50	-27.20	585.64	132.25	658.24	-312.80	-278.30
5	190	120	198	5.80	-5.50	0.80	33.64	30.25	4.64	-4.40	-31.90
6	210	99	250	25.80	-26.50	52.80	665.64	702.25	1362.24	-1399.20	-683.70
7	178	119	189	-6.20	-6.50	-8.20	38.44	42.25	50.84	53.30	40.30
8	160	145	168	-24.20	19.50	-29.20	585.64	380.25	706.64	-569.40	-471.90
9	180	124	191	-4.20	-1.50	-6.20	17.64	2.25	26.04	9.30	6.30
10	275	96	280	90.80	-29.50	82.80	8244.64	870.25	7518.24	-2442.60	-2678.60
合計	1842	1255	1972	0.00	0.00	0.00	11953.60	2970.50	12189.60	-5921.00	-5284.00
平均值	184.2	125.5	197.2								

$$b_1=\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}=$$

$$\frac{2970.5\times12189.6-(-5284.0)\times(-5921.0)}{11953.6\times2970.5-[-5284.0]^2}=\frac{4922642.8}{7587512.8}=0.6488$$

$$b_2=\frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}=$$

$$\frac{11953.6\times(-5921.0)-(-5284.0)\times(12189.6)}{11953.6\times2970.5-[-5284.0]^2}=\frac{-6367419.2}{7587512.8}=-0.8392$$

$b_0=\bar{y}_i-b_1\times\bar{x}_1-b_2\times\bar{x}_2=197.2-0.6488\times184.2-(-0.8392)\times125.5=183.0136$

估計複迴歸方程式$\hat{y}_i=183.0136+0.6488\times x_{1i}-0.8392\times x_{2i}$

答案：估計複迴歸方程式統計值 $b_1=0.6488$；$b_2=-0.8392$；$b_0=183.0136$

利用 Excel 和 SPSS 統計軟體計算結果

Excel 2007 軟體中選擇檔案→選項→執行(G)...→勾選分析工具箱→確定。回到 excel 視窗→資料→資料分析→迴歸→確定。

單獨解析行銷費用 x_{1i} 對銷售套餐數量 y_i 的估計簡單線性迴歸方程式

$$\hat{y}_i=b_0+b_1\times x_{1i}=9.363+1.020\times x_{1i}$$

行銷費用 x_{1i} 可以解釋銷售套餐數量 y_i 之 92.85%變異量。

摘要輸出

迴歸統計	
R 的倍數	0.9636
R 平方	0.9285
調整的 R 平方	0.9196
標準誤	10.9393
觀察值個數	10

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	12430.26	12430.26	103.87	0.000
殘差	8	957.34	119.67		

	自由度	SS	MS	*F*	顯著值
總和	9	13387.60			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%
截距	9.363	18.752	0.499	0.631	-33.9	52.6
X 變數 1	1.020	0.100	10.192	0.000	0.8	1.3

殘差輸出

觀察值	預測 Y	殘差	標準化殘差
1	162.32	-6.3248	-0.6132
2	172.52	7.4778	0.7250
3	191.90	-1.8973	-0.1840
4	172.52	-2.5222	-0.2446
5	203.11	-5.1145	-0.4959
6	223.51	26.4906	2.5685
7	190.88	-1.8776	-0.1820
8	172.52	-4.5222	-0.4385
9	192.92	-1.9171	-0.1859
10	289.79	-9.7927	-0.9495

單獨解析虱目魚套餐定價 x_{2i} 對銷售套餐數量 y_i 的估計簡單線性迴歸方程式

$$\hat{y}_i = b_0 + b_1 \times x_{2i} = 447.355 - 1.993 \times x_{2i}$$

虱目魚套餐定價 x_{2i} 可以解釋銷售套餐數量 y_i 之 88.16%變異量。

摘要輸出

迴歸統計	
R 的倍數	0.9389
R 平方	0.8816
調整的 R 平方	0.8668
標準誤	14.0778
觀察值個數	10

ANOVA

	自由度	SS	MS	*F*	顯著值
迴歸	1	11802.13	11802.13	59.55	0.000
殘差	8	1585.47	198.18		
總和	9	13387.60			

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%
截距	447.355	32.720	13.672	0.0000	371.9	522.8
X 變數 1	-1.993	0.258	-7.717	0.0001	-2.6	-1.4

殘差輸出

觀察值	預測 Y	殘差	標準化殘差
1	148.4	7.6350	0.5752
2	168.3	11.7024	0.8817
3	198.2	-8.1966	-0.6176
4	174.3	-4.2774	-0.3223

觀察值	預測 Y	殘差	標準化殘差
5	208.2	-10.1630	-0.7657
6	250.0	-0.0216	-0.0016
7	210.2	-21.1562	-1.5940
8	158.3	9.6687	0.7285
9	200.2	-9.1899	-0.6924
10	256.0	23.9986	1.8081

行銷費用 x_{1i} 和虱目魚套餐定價 x_{2i} 對銷售套餐數量 y_i 的估計複迴歸方程式

$$\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} = 183.014 + 0.649 \times x_{1i} - 0.839 \times x_{2i}$$

行銷費用 x_{1i} 和虱目魚套餐定價 x_{2i} 可以解釋銷售套餐數量 y_i 之 95.1 %變異量。

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	X2, X1[a]	.	Enter

[a] All requested variables entered.
[b] Dependent Variable: Y

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.981[a]	.962	.951	8.5383

[a] Predictors: (Constant), X2, X1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12877.281	2	6438.641	88.318	.000[a]
	Residual	510.319	7	72.903		
	Total	13387.600	9			

[a] Predictors: (Constant), X2, X1
[b] Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	183.014	71.638		2.555	.038
	X1	.649	.169	.613	3.840	.006
	X2	-.839	.339	-.395	-2.476	.042

[a] Dependent Variable: Y

殘差輸出

觀察值	預測 Y	殘差	標準化殘差
1	154.45	1.55	0.21
2	169.33	10.67	1.42
3	194.25	-4.25	-0.56
4	171.85	-1.85	-0.25
5	205.58	-7.58	-1.01
6	236.18	13.82	1.84
7	198.63	-9.63	-1.28
8	165.14	2.86	0.38
9	195.73	-4.73	-0.63
10	280.87	-0.87	-0.11

迴歸係數的解釋

在簡單線性迴歸方程式中，從統計值 b_1 可以解釋自變數 x_i 變動一個單位時，對依變數估計值產生的改變量。因此，從簡單線性迴歸方程式中，可以解釋行銷費用 x_{1i} 對銷售套餐數量 y_i 的估計值影響，當行銷費用 x_{1i} 增加(減少)一個單位時，對銷售套餐數量 y_i 的估計值$\hat{y}_i$增加(減少)1.020 個單位數。虱目魚套餐定價 x_{2i} 增加(減少)一個單位時，對銷售套餐數量 y_i 的估計值$\hat{y}_i$減少(增加)1.993 個單位數。

$$\hat{y}_i = b_0 + b_1 \times x_{1i} = 9.363 + 1.020 \times x_{1i}$$
$$\hat{y}_i = b_0 + b_2 \times x_{2i} = 447.355 - 1.993 \times x_{2i}$$

在複迴歸方程式中，從統計值 b_1、b_2、...和 b_k 可以解釋自變數 x_i 變動一個單位時，在其他自變數不變的情況下，對依變數估計值產生的改變量。因此，從複迴歸方程式中，可以解釋行銷費用 x_{1i} 對銷售套餐數量 y_i 的估計值$\hat{y}_i$影響，當虱目魚套餐定價 x_{2i} 固定不變時，行銷費用 x_{1i} 增加(減少)一個單位時，對銷售套餐數量 y_i 的估計值$\hat{y}_i$增加(減少)0.649 個單位數。虱目魚套餐定價 x_{2i} 對銷售套餐數量 y_i 的估計值$\hat{y}_i$影響，當行銷費用 x_{1i} 固定不變時，虱目魚套餐定價 x_{2i} 增加(減少)一個單位時，對銷售套餐數量 y_i 的估計值$\hat{y}_i$減少(增加)0.839 個單位數。

$$\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} = 183.014 + 0.649 \times x_{1i} - 0.839 \times x_{2i}$$

答案：使用統計軟體運算獲得之估計複迴歸方程式統計值 $b_1 = 0.649$；$b_2 = -0.839$；$b_0 = 183.014$

練習 15.1 叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(單位：新台幣元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表。請計算出估計複迴歸方程式的統計值。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：利用公式計算結果

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0
平均值	173.0	4.6	174.0								

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{74.40\times4680-(-328.0)\times(-664.0)}{4610\times74.40-[-328.0]^2} = \frac{130400}{235400} = 0.5540$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{4610\times(-664.0)-(-328.0)\times(4680)}{4610\times74.40-[-328.0]^2} =$$

$\frac{-1526000}{235400}$ = -6.4826

$b_0 = \bar{y}_i - b_1 \times \bar{x}_1 - b_2 \times \bar{x}_2$ = 174.0 – 0.5540 × 173.0 – (-6.4826) × 4.6 = 107.9864

估計複迴歸方程式 $\hat{y}_i$ = 107.9864 + 0.5540 × x_{1i} – 6.4826 × x_{2i}

答案：估計複迴歸方程式統計值 b_1 = 0.5540；b_2 = –6.4826；b_0 = 107.9864

練習 15.2　吻別連鎖咖啡店有 10 營業點，每個營業點個別的前一日訓練費用 x_{1i}(單位：新台幣百元)、方圓兩公里內競爭咖啡店數量 x_{2i}(單位：家)和前一日營業金額 y_i(單位：新台幣百元)列於下表。請計算出估計複迴歸方程式的統計值。

營業點 i	訓練費用 x_{1i}	競爭咖啡店數量 x_{2i}	營業金額 y_i
1	50	5	160
2	60	3	180
3	82	2	190
4	65	4	170
5	90	1	190
6	110	1	240
7	75	9	130
8	60	8	150
9	80	7	160
10	90	6	170

							$(x_{1i} - \bar{x}_1)^2$	$(x_{2i} - \bar{x}_2)^2$	$(x_{1i} - \bar{x}_1) \times (y_i - \bar{y})$	$(x_{2i} - \bar{x}_2) \times (y_i - \bar{y})$	$(x_{1i} - \bar{x}_1) \times (x_{2i} - \bar{x}_2)$

15.3 複判定係數

在簡單線性迴歸分析中使用**判定係數**(coefficient of determination)，評估簡單線性迴歸方程式中的適合度(goodness of fit)。同樣的評估程序亦可使用於複迴歸方程式，在複迴歸方程式中利用複判定係數、複迴歸決定係數、複迴歸判定係數或多元決定係數(multiple coefficient of determination)。

總平方和或總變異(sum of squares total, SST)可以分解成迴歸造成的平方和、迴歸項平方和(sum of squares due to regression, SSR)或可解釋的變異和誤差造成的平方和、誤差項平方和(sum of squares due to error, SSE)、不可解釋的變異或隨機變異兩部分。

$$\text{SST} = \text{SSR} + \text{SSE}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 + \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

複判定係數、多元判定係數(multiple coefficient of determination)或判定係數(coefficient of determination)即是迴歸造成的平方和(sum of squares due to regression, SSR)佔總平方和(sum of squares total, SST)的比例，常使用 R^2 或 r^2 符號代表。複判定係數代表迴歸方程式可以解釋(說明)依變數 Y 變異量的比例。R 稱為多元相關係數或複相關係數(multiple correlation coefficient)。

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1-\frac{\text{SSE}}{\text{SST}} = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$R = \sqrt{R^2} = \sqrt{\frac{\text{SSR}}{\text{SST}}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \sqrt{1-\frac{\text{SSE}}{\text{SST}}} = \sqrt{1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

利用複判定係數 R^2 評估複迴歸方程式的適合度時，當樣本數量較少或自變數較多，會使自由度降低，導致對複判定係數 R^2 有高估，自變數對可解釋之變異量(SSR)的影響。若在複迴歸模式中加入一些與模式無關的自變數，複判定係數 R^2 會增加，因此，無法客觀的代表複迴歸模式的解釋能力。故，建議使用調整判定係數、調整複判定係數或調整後的決定係數(adjusted coefficient of multiple determination; adjusted multiple coefficient of determination; adjusted R square, R_a^2, $\bar{R}^2$, adjusted R^2 or adj - R^2)為評量複迴歸方程式的解釋能力。

$$R_a^2 = \bar{R}^2 = 1-(1-R^2)\times\frac{n-1}{n-k-1} = 1-\frac{\text{SSE}}{\text{SST}}\times\frac{n-1}{n-k-1} = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\times\frac{n-1}{n-k-1}$$

其中 $\hat{y}_i$ = 依變數第 i 個觀測值的估計值 n = 觀測值數量 k = 自變數數量(個數)

從上述調整判定係數公式中，發現 $\frac{n-1}{n-k-1} > 1$，因此複判定係數 R^2 > 調整判定係數 R_a^2。當樣本數量 n 和自變數數量 k 在特定的比值下，可能出現 $\frac{\text{SSE}}{\text{SST}} > \frac{n-k-1}{n-1}$，調整判定係數 R_a^2 會出現負值($R_a^2 < 0$)。當調整判定係數(adjusted R square)是負值時，一般統計軟體會出現調整判定係數(adjusted R square)等於 0 的結果。

複迴歸變異數分析(anova)表

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)	$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	k	$\text{MSR} = \frac{\text{SSR}}{k}$	$F = \frac{\text{MSR}}{\text{MSE}}$
誤差項(隨機項)(error)	$\text{SSE} = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$n-k-1$	$\text{MSE} = \frac{\text{SSE}}{n-k-1}$	
總變異(total)	$\text{SST} = \sum_{i=1}^{n}(y_i-\bar{y})^2$	$n-1$		

範例 15.2 阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。請計算出估計複迴歸方程式的複判定係數和調整判定係數數值。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
10	275	96	280

題解：估計複迴歸方程式$\hat{y}_i$ = 183.0136 + 0.6488 × x_{1i} – 0.8392 × x_{2i}

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	150	150	156	154.451	-42.749	1827.450	-41.200	1697.440
2	160	140	180	169.331	-27.869	776.675	-17.200	295.840
3	179	125	190	194.246	-2.954	8.727	-7.200	51.840
4	160	137	170	171.849	-25.351	642.688	-27.200	739.840
5	190	120	198	205.579	8.379	70.200	0.800	0.640
6	210	99	250	236.177	38.977	1519.230	52.800	2787.840
7	178	119	189	198.632	1.432	2.052	-8.200	67.240
8	160	145	168	165.135	-32.065	1028.156	-29.200	852.640
9	180	124	191	195.734	-1.466	2.149	-6.200	38.440
10	275	96	280	280.866	83.666	6999.955	82.800	6855.840
合計	1842	1255	1972	1972.000	0.000	12877.281	0.000	13387.600
平均值	184.2	125.5	197.2					

SSR = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ = 12877.281　　SST = $\sum_{i=1}^{n}(y_i - \bar{y})^2$ = 13387.600

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12877.281	2	6438.641	88.318	.000[a]
	Residual	510.319	7	72.903		
	Total	13387.600	9			

[a] Predictors: (Constant), X2, X1　[b] Dependent Variable: Y

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{12877.281}{13387.600} = 0.9619$$

$$R_a^2 = \bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1} = 1 - (1 - 0.9619) \times \frac{10-1}{10-2-1} = 1 - (0.0381) \times \frac{9}{7} = 0.9510$$

答案：複判定係數 R^2 = 0.9619 和調整判定係數R_a^2 = 0.9510

練習 15.3　叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表。請計算出估計複迴歸方程式的複判定係數和調整判定係數數值。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：利用公式計算迴歸方程式統計值結果

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i} - \bar{x}_1$	$x_{2i} - \bar{x}_2$	$y_i - \bar{y}$	$(x_{1i} - \bar{x}_1)^2$	$(x_{2i} - \bar{x}_2)^2$	$(x_{1i} - \bar{x}_1) \times (y_i - \bar{y})$	$(x_{2i} - \bar{x}_2) \times (y_i - \bar{y})$	$(x_{1i} - \bar{x}_1) \times (x_{2i} - \bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0
平均值	173.0	4.6	174.0								

$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{74.40\times4680-(-328.0)\times(-664.0)}{4610\times74.40-[-328.0]^2} = \frac{130400}{235400} = 0.5540$

$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{4610\times(-664.0)-(-328.0)\times(4680)}{4610\times74.40-[-328.0]^2} = \frac{-1526000}{235400} = -6.4826$

$b_0 = \bar{y}_i - b_1\times\bar{x}_1 - b_2\times\bar{x}_2 = 174.0 - 0.5540\times173.0 - (-6.4826)\times4.6 = 107.9864$

估計複迴歸方程式$\hat{y}_i = 107.9864 + 0.5540\times x_{1i} - 6.4826\times x_{2i}$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$\hat{y}_i-\bar{y}$	$(\hat{y}_i-\bar{y})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
1	150	5	160	158.666	-15.334	235.128	-14	196
2	160	3	180	177.171	3.171	10.054	6	36
3	180	2	190	194.732	20.732	429.831	16	256
4	160	4	170	170.688	-3.312	10.968	-4	16
5	190	1	190	206.754	32.754	1072.855	16	256
6	210	1	240	217.833	43.833	1921.374	66	4356
7	140	9	130	127.196	-46.804	2190.590	-44	1936
8	160	8	150	144.758	-29.242	855.103	-24	576
9	180	7	160	162.319	-11.681	136.435	-14	196
10	200	6	170	179.881	5.881	34.587	-4	16
合計	1730	46	1740	1740.000	0.000	6896.924	0	7840
平均值	173.0	4.6	174.0					

$SSR = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = 6896.924$　　　SST = $\sum_{i=1}^{n}(y_i-\bar{y})^2 = 7840.00$

$R^2 = \frac{SSR}{SST} = \frac{6896.924}{7840.00} = 0.8797$

$R_a^2 = \bar{R}^2 = 1-(1-R^2)\times\frac{n-1}{n-k-1} = 1-(1-0.8797)\times\frac{10-1}{10-2-1} = 1-(0.1203)\times\frac{9}{7} = 0.8453$

答案：複判定係數 $R^2 = 0.8797$ 和調整判定係數$R_a^2 = 0.8453$

練習 15.4　某房屋仲介公司欲了解中古房屋之售價 Y(單位：萬元)與地坪 X_1(單位：坪)、建坪 X_2(單位：坪)、屋齡 X_3(單位：年)及所在區域 X_4(1 代表都會區，0 代表非都會區)之關係，其迴歸分析變異數分析表如下所示，請計算複判定係數和調整判定係數。

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)	745528.244			
誤差項(隨機項)(error)		16	95.931	
總變異(total)				

題解：依據下列複迴歸變異數分析表，運算出上述題目中變異數分析表的缺值。

複迴歸變異數分析(anova)表

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)	$SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	k	$MSR=\frac{SSR}{k}$	$F=\frac{MSR}{MSE}$
誤差項(隨機項)(error)	$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$n-k-1$	$MSE=\frac{SSE}{n-k-1}$	
總變異(total)	$SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$	$n-1$		

本題目中一共有地坪 X_1(單位：坪)、建坪 X_2(單位：坪)、屋齡 X_3(單位：年)及所在區域 X_4(1 代表都會區，0 代表非都會區)四個變數，故 $k=4$(自變數數量)。$n-k-1=16\rightarrow n=21$。

已知誤差項均方和自由度數值，即可透過 $MSE=\frac{SSE}{n-k-1}$關係，獲得誤差項平方和 $SSE=MSE\times(n-k-1)=95.931\times16=1534.896$。

已知迴歸項和誤差項平方和數值，即可透過 $SST=SSR+SSE$，獲得總變異的平方和 $SST=SSR+SSE=745528.244+1534.896=747063.140$。

已知迴歸項平方和和自由度數值，即可透過 $MSR=\frac{SSR}{k}$，獲得迴歸項的均方 $MSR=\frac{SSR}{k}=\frac{745528.244}{4}=186382.061$。

$F=\frac{MSR}{MSE}=\frac{186382.061}{95.931}=1942.876$

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)	745528.244	4	186382.061	1942.876
誤差項(隨機項)(error)	1534.896	16	95.931	
總變異(total)	747063.140	20		

複判定係數 $R^2=\frac{SSR}{SST}=\frac{745528.244}{747063.140}=0.997945$

調整判定係數$R_a^2=\bar{R}^2=1-(1-R^2)\times\frac{n-1}{n-k-1}=1-(1-0.997945)\times\frac{21-1}{21-4-1}=1-(0.0020546)\times\frac{20}{16}=0.997432$

答案：複判定係數 $R^2=0.9979$ 和調整判定係數$R_a^2=0.9974$

練習 15.5 欲使用自變數 X_1、X_2、X_3 和 X_4 對依變數 Y 進行複迴歸分析，其模式為 $y_i=\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}+\beta_3\times x_{3i}+\beta_4\times x_{4i}$ 獲得下列部分變異數分析(ANOVA)表資料，請將缺的數值填入。

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)				
誤差項(隨機項)(error)	70			
總變異(total)	220	47		

題解：自變數數量 $k=4$

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F 值
迴歸項(regression)	220 - 70 = 150	$k=4$	$\frac{150}{4}=37.5$	$\frac{37.5}{1.6279}=23.04$
誤差項(隨機項)(error)	70	47 – 4 = 43	$\frac{70}{43}=1.6279$	
總變異(total)	220	47		

15.4 模型假設

複迴歸模式

敘述自變數(X_1、X_2、X_3、…、X_k)、依變數 Y 和誤差項 ε_i 之間關係的方程式稱為複迴歸模式、多元迴歸

模式(multiple regression model)或複迴歸線性模式(multiple regression linear model)。

$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki} + \varepsilon_i$，其中 $i = 1, \ldots, n$

在複迴歸模式或多元迴歸模式與簡單線性迴歸模式對於誤差項 ε_i 的假設皆相同。

複迴歸模式或多元迴歸模式之顯著性檢定必須依據下列誤差項或殘差項 ε_i 的假設條件。

A.全部觀測點之誤差項 ε_i 的平均值或期望值為 0。$\bar{\varepsilon}_i = E(\varepsilon_i) = 0$。

B.各觀測點之誤差項 ε_i 之間相互獨立。$Cov(\varepsilon_i, \varepsilon_j) = 0$，在 $i \neq j$，其中 $i, j = 1,\ldots, n$。任何兩個誤差項不相關。

C.各觀測點之誤差項 ε_i 之變異數 σ^2 皆相等[標準(偏)差 σ 皆相等]。$V(\varepsilon_i) = \sigma^2$。殘差項之變異數具有均一性(齊一性)。

D.全部觀測點之誤差項 ε_i 屬於常態分布，$\varepsilon_i \sim N$。

E.誤差項 ε_i 與 k 個自變數的數值沒有關係。$Cov(\varepsilon_i, x_{ji}) = 0$，其中 $i = 1,\ldots, n$，$j = 1,\ldots, k$。

15.5 斜率顯著性檢定

在複迴歸分析中，先利用 F 值檢定所有自變數 $X_1,\ldots, X_k$ 和依變數 Y 的關係是否達到顯著水準。此 F 值檢定稱為整體顯著性(overall significance)檢定或聯合檢定(global test)。

若在 F 值檢定中發現所有自變數 $X_1,\ldots, X_k$ 和依變數 Y 的關係達到顯著水準。進一步利用 t 值檢定分別各自變數 $X_1,\ldots, X_k$ 與依變數 Y 的關係。此 t 值檢定稱為個別顯著性(individual significance)檢定或個別迴歸係數檢定(Testing individual regression coefficients)。

F 值檢定

利用 F 機率分布檢定樣本資料迴歸方程式中斜率 β_1、β_2、β_3、…和 β_k 是否等於 0，使用於驗證迴歸關係的顯著性。

迴歸造成的均方(mean square due to regression)、迴歸均方或迴歸平均平方和(mean square regression, MSR)是迴歸項平方和(sum of squares due to regression, SSR)除以迴歸自由度(regression degrees of freedom)獲得。迴歸自由度(regression degrees of freedom)等於自變數之個數 k。

$$\text{MSR} = \frac{SSR}{df} = \frac{SSR}{k} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{k}$$

在自變數 X 與依變數 y_i 的複迴歸模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki} + \varepsilon_i$ 中顯示，誤差項 ε_i 的變異數 σ^2 亦即是依變數 y_i 在迴歸模式中的變異數。

誤差均方或誤差平均平方和(mean square error, MSE)為誤差項 ε_i 之變異數 σ^2 的估計值(可表示為 S^2)，可由殘差平方和(sum square error, SSE)除以其自由度(degree of freedom, df)獲得。在計算殘差平方和(sum square error, SSE)時，需先估算迴歸模式的參數(β_0、β_1、β_2、β_3、…、β_k)，因此殘差平方和的自由度為 $n - k - 1$。

$$\text{MSE} = S^2 = \frac{\text{SSE}}{df} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-k-1}$$

利用 F 值檢定的程序

A.設定顯著水準 α。

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0,\ldots$, or $\beta_k \neq 0$。

D.計算檢定統計值—F 值

$$檢定統計值\ F = \frac{\text{MSR}}{\text{MSE}}$$

E.若檢定統計值 $F \leq$ 臨界值 $F_{\alpha,k,n\text{-}k\text{-}1}$，接受虛無假設 H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$。

F.若檢定統計值 $F >$ 臨界值 $F_{\alpha,k,n\text{-}k\text{-}1}$，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$，接受對立假設 H_1: $\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 \neq 0, \ldots$, or $\beta_k \neq 0$。

其中 $F_{\alpha,k,n\text{-}k\text{-}1}$：為顯著水準 α，分子自由度 k，分母自由度 $n-k-1$ 的 F 分布數值(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

範例 15.3 阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。請利用 F 值法檢定複迴歸方程式中斜率 β_1 和 β_2 是否全部等於 0。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191
10	275	96	280

題解：

變異數分析(ANOVA)

變異來源(source of variation)	自由度(degrees of freedom)	平方和(sum of square)	均方(mean square)	F 值	顯著值
迴歸	2	12877.2812	6438.6406	88.3183	0.0000
殘差	7	510.3188	72.9027		
總和	9	13387.6000			

A.設定顯著水準 $\alpha = 0.05$，臨界值 $F_{\alpha,k,n\text{-}k\text{-}1} = F_{0.05,2,10\text{-}2\text{-}1} = F_{0.05,2,7} = 4.7374$(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

D.計算檢定統計值—F 值

$$檢定統計值\ F = \frac{\text{MSR}}{\text{MSE}} = \frac{6438.641}{72.903} = 88.3179$$

E.檢定統計值 $F = 88.3179 >$ 臨界值 $F_{\alpha,k,n\text{-}k\text{-}1} = 4.7374$，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

答案：複迴歸方程式中斜率 β_1 和 β_2 沒有全部等於 0

練習 15.6 奇遇連鎖餐廳有 14 個營業點，每個營業點個別的平均每月行銷費用(x_{1i}，單位：萬元)、訓練費用(x_{2i}，單位：萬元)和營業額(y_i，單位：十萬元)，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請利用顯著水準 $\alpha = 0.05$ 檢定複迴歸方程式中斜率 β_1 和 β_2 是否全部等於 0。迴歸與殘差平方和(SS)如下所示。假設此資料適合於有母數統計方法(parametric statistical method)分析。

	SS

迴歸	3560
殘差	440

題解：

變異來源(source of variation)	自由度(degrees of freedom)	平方和(sum of square)	均方(mean square)	F 值
迴歸	2	3560	1780	44.5
殘差	11	440	40	
總和	13	4000		

A.設定顯著水準 α = 0.05，臨界值 $F_{\alpha,k,n-k-1} = F_{0.05,2,11} = F_{0.05,2,14-2-1}$ = 3.9823(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

D.計算檢定統計值—F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{1780}{40} = 44.5$$

E.檢定統計值 F = 44.5 > 臨界值 $F_{\alpha,k,n-k-1}$ = 3.9823，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

答案：複迴歸方程式中斜率 β_1 和 β_2 沒有全部等於 0

練習 15.7 叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請利用 F 值法檢定複迴歸方程式中斜率 β_1 和 β_2 是否全部等於 0。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：利用公式計算結果

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
平均值	173.0	4.6	174.0								

$$b_1=\frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}=\frac{74.40\times4680-(-328.0)\times(-664.0)}{4610\times74.40-[-328.0]^2}$$

$$=\frac{130400}{235400}=0.5540$$

$$b_2=\frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}=\frac{4610\times(-664.0)-(-328.0)\times(4680)}{4610\times74.40-[-328.0]^2}$$

$$=\frac{-1526000}{235400}=-6.4826$$

$b_0=\bar{y}_i-b_1\times\bar{x}_1-b_2\times\bar{x}_2=174.0-0.5540\times173.0-(-6.4826)\times4.6=107.9864$

估計複迴歸方程式$\hat{y}_i=107.9864+0.5540\times x_{1i}-6.4826\times x_{2i}$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$\hat{y}_i-\bar{y}$	$(\hat{y}_i-\bar{y})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
1	150	5	160	158.666	-15.334	235.128	-14	196
2	160	3	180	177.171	3.171	10.054	6	36
3	180	2	190	194.732	20.732	429.831	16	256
4	160	4	170	170.688	-3.312	10.968	-4	16
5	190	1	190	206.754	32.754	1072.855	16	256
6	210	1	240	217.833	43.833	1921.374	66	4356
7	140	9	130	127.196	-46.804	2190.590	-44	1936
8	160	8	150	144.758	-29.242	855.103	-24	576
9	180	7	160	162.319	-11.681	136.435	-14	196
10	200	6	170	179.881	5.881	34.587	-4	16
合計	1730	46	1740	1740.000	0.000	6896.924	0	7840
平均值	173.0	4.6	174.0					

$SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=6896.924$ $SST=\sum_{i=1}^{n}(y_i-\bar{y})^2=7840.00$

變異數分析(ANOVA)

變異來源(source of variation)	自由度(degrees of freedom)	平方和(sum of square)	均方(mean square)	F 值	顯著值
迴歸	2	6896.924	3448.462	25.5963	0.0000
殘差	7	943.076	134.725		
總和	9	7840.00			

A.設定顯著水準 $\alpha=0.05$，臨界值 $F_{\alpha,k,n-k-1}=F_{0.05,2,10-2-1}=F_{0.05,2,7}=4.7374$(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1=\beta_2=0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1\neq0$ or $\beta_2\neq0$。

D.計算檢定統計值–F 值

$$\text{檢定統計值 } F=\frac{MSR}{MSE}=\frac{3448.462}{134.725}=25.5963$$

E.檢定統計值 $F=25.5963>$ 臨界值 $F_{\alpha,k,n-k-1}=4.7374$，拒絕虛無假設 H_0: $\beta_1=\beta_2=0$，接受對立假設 H_1: $\beta_1\neq0$ or $\beta_2\neq0$。

答案：複迴歸方程式中斜率 β_1 和 β_2 沒有全部等於 0

***t* 值檢定**

經過 F 值檢定確認所有自變數 $X_1,\ldots,X_k$ 和依變數 Y 的關係是否達到顯著水準，若有達到顯著性相關水準。後續利用 t 值檢定法檢定個別自變數 $X_1,\ldots,X_k$ 和依變數 Y 的關係是否達到顯著相關水準。利用樣本資

料檢定複迴歸方程式中個別自變數參數 β_i 是否等於 0，才可以進一步決定是否接受複迴歸分析的結果。若依變數母體變異數σ_y^2已知時，可以運用標準化 z 值進行檢定或區間估計。若依變數母體變異數σ_y^2未知時，必須使用其估計值—依變數樣本變異數S_y^2或$S_{y|x_1x_2}^2$取代

$$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2-b_1\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-b_2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{n-k-1}$$

運用依變數樣本變異數S_y^2估算依變數母體變異數σ_y^2時，相對應的 b_0、b_1 與 b_2 估算的變異數依序為：

$$S_{b_0}^2 = \left[\frac{\bar{x}_1^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+\bar{x}_2^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times\bar{x}_1\times\bar{x}_2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right]\times S_y^2$$

$$S_{b_1}^2 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}\times S_y^2$$

$$S_{b_2}^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}\times S_y^2$$

利用 t 值檢定的程序

A. 設定顯著水準 α。

B. 虛無假設(null hypothesis) H_0: $\beta_i = 0$。

C. 對立假設(alternative hypothesis) H_1: $\beta_i \neq 0$。

D. 計算檢定統計值—t 值

$$\text{檢定統計值 } t = \frac{b_i}{s_{b_i}}$$

E. 若左側臨界值$-t_{\frac{\alpha}{2},n-k-1}$ < 檢定統計值 t < 右側臨界值$t_{\frac{\alpha}{2},n-k-1}$，接受虛無假設 H_0: $\beta_i = 0$。

F. 若檢定統計值 t < 左側臨界值$-t_{\frac{\alpha}{2},n-k-1}$或檢定統計值 t > 右側臨界值$t_{\frac{\alpha}{2},n-k-1}$，拒絕虛無假設 H_0: β_i = 0，接受對立假設(alternative hypothesis) H_1: $\beta_i \neq 0$。

其中$t_{\frac{\alpha}{2},n-k-1}$：為顯著水準$\frac{\alpha}{2}$，自由度 $n-k-1$ 的 t 分布數值(使用 Excel 軟體 T.INV 函數查詢獲得)。在兩個自變數的情況下，可以簡化為$t_{\frac{\alpha}{2},n-3}$。

k = 自變數數量(個數)，$k > 0$，正整數。

經過統計驗證的結果顯示，若接受虛無假設 H_0: $\beta_i = 0$ 時，顯示依變數 $E(y_i)$和自變數 x_i 之間沒有足夠的證據證明兩者關係存在；若接受對立假設(alternative hypothesis) H_1: $\beta_i \neq 0$ 時，代表依變數 $E(y_i)$和自變數 x_i 之間有統計上的相關性存在。在進行統計驗證時，將依據迴歸方程式斜率 β_i 之估計值 b_i 之抽樣分布資料。

範例 15.4　阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。請利用 t 值法在顯著水準 α = 0.05 分別檢定複迴歸方程式中斜率 β_1 和 β_2 是否等於 0。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191
10	275	96	280

題解：估計複迴歸方程式$\hat{y}_i = 183.0136 + 0.6488\times x_{1i} - 0.8392\times x_{2i}$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	150	150	156	154.451	1.5487	2.3984
2	160	140	180	169.331	10.6689	113.8251
3	179	125	190	194.246	-4.2459	18.0279
4	160	137	170	171.849	-1.8487	3.4177
5	190	120	198	205.579	-7.5785	57.4340
6	210	99	250	236.177	13.8227	191.0669
7	178	119	189	198.632	-9.6323	92.7818
8	160	145	168	165.135	2.8649	8.2075
9	180	124	191	195.734	-4.7339	22.4099
10	275	96	280	280.866	-0.8657	0.7495
合計	1842	1255	1972	1972.000	0.0000	510.3188
平均值	184.2	125.5	197.2			

第一種算法$S_y^2 = S_{y|x_1x_2}^2 = MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{510.3188}{10-2-1} = 72.9027$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$
1	150	150	156	-34.20	24.50	-41.20	1697.44	1409.04	-1009.40
2	160	140	180	-24.20	14.50	-17.20	295.84	416.24	-249.40
3	179	125	190	-5.20	-0.50	-7.20	51.84	37.44	3.60
4	160	137	170	-24.20	11.50	-27.20	739.84	658.24	-312.80
5	190	120	198	5.80	-5.50	0.80	0.64	4.64	-4.40
6	210	99	250	25.80	-26.50	52.80	2787.84	1362.24	-1399.20
7	178	119	189	-6.20	-6.50	-8.20	67.24	50.84	53.30
8	160	145	168	-24.20	19.50	-29.20	852.64	706.64	-569.40
9	180	124	191	-4.20	-1.50	-6.20	38.44	26.04	9.30
10	275	96	280	90.80	-29.50	82.80	6855.84	7518.24	-2442.60
合計	1842	1255	1972	0.00	0.00	0.00	13387.60	12189.60	-5921.00
平均值	184.2	125.5	197.2						

第二種算法$S_y^2 = S_{y|x_1x_2}^2 = MSE = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2-b_1\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-b_2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{n-k-1} = \frac{13387.60-0.6488\times12189.60-(-0.8392)\times(-5921.00)}{10-2-1} = \frac{13387.60-7908.612-4968.903}{7} = \frac{510.3188}{7} = 72.9027$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	150	156	-34.20	24.50	-41.20	1697.44	1169.64	600.25	-837.90
2	160	140	180	-24.20	14.50	-17.20	295.84	585.64	210.25	-350.90
3	179	125	190	-5.20	-0.50	-7.20	51.84	27.04	0.25	2.60
4	160	137	170	-24.20	11.50	-27.20	739.84	585.64	132.25	-278.30
5	190	120	198	5.80	-5.50	0.80	0.64	33.64	30.25	-31.90
6	210	99	250	25.80	-26.50	52.80	2787.84	665.64	702.25	-683.70
7	178	119	189	-6.20	-6.50	-8.20	67.24	38.44	42.25	40.30
8	160	145	168	-24.20	19.50	-29.20	852.64	585.64	380.25	-471.90
9	180	124	191	-4.20	-1.50	-6.20	38.44	17.64	2.25	6.30
10	275	96	280	90.80	-29.50	82.80	6855.84	8244.64	870.25	-2678.60
合計	1842	1255	1972	0.00	0.00	0.00	13387.60	11953.60	2970.50	-5284.00
平均值	184.2	125.5	197.2							

$S_{b_1}^2 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)]^2}\times S_y^2 = \frac{2970.50}{11953.60\times2970.50-[-5284.00]^2}\times 72.9027 = \frac{2970.50}{7587512.8000}\times 72.9027 = 0.0285$

$S_{b_2}^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)]^2}\times S_y^2 = \frac{11953.60}{11953.60\times2970.50-[-5284.00]^2}\times 72.9027 =$

$\frac{11953.60}{7587512.8000} \times 72.9027 = 0.1149$

$S_{b_0}^2 = \left[\frac{\bar{x}_1^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+\bar{x}_2^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times\bar{x}_1\times\bar{x}_2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right] \times S_y^2 = \left[\frac{184.2^2\times2970.50+125.5^2\times11953.60-2\times184.2\times125.5\times(-5284.00)}{11953.60\times2970.50-[-5284.00]^2}+\frac{1}{10}\right] \times 72.9027 = \left[\frac{533361696.82}{7587512.8000}+\frac{1}{10}\right] \times 72.9027 = \left[70.2947+\frac{1}{10}\right] \times 72.9027 = 5131.9606$

$S_{b_1} = \sqrt{S_{b_1}^2} = \sqrt{0.0285} = 0.1689$ $S_{b_2} = \sqrt{S_{b_2}^2} = \sqrt{0.1149} = 0.3389$

$S_{b_0} = \sqrt{S_{b_0}^2} = \sqrt{5131.9606} = 71.6377$

對 β_1 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，左側臨界值 $-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$；右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_1}{S_{b_1}} = \frac{0.6488}{0.1689} = 3.8403$$

E.檢定統計值 $t = 3.8403$ > 右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = 2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$。

對 β_2 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，左側臨界值 $-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$；右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_2 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_2}{S_{b_2}} = \frac{-0.8392}{0.3389} = -2.4762$$

E.檢定統計值 $t = -2.4762$ < 左側臨界值 $-t_{\frac{\alpha}{2},n-k-1} = -2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_2 = 0$，接受對立假設 H_1: $\beta_2 \neq 0$。

對 β_0 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，左側臨界值 $-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$；右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_0 = 0$。代表當兩個自變數數值都等於 0 時，依變數數值等於 0。

C.對立假設(alternative hypothesis) H_1: $\beta_0 \neq 0$。代表當兩個自變數數值都等於 0 時，依變數數值不會等於 0。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_0}{S_{b_0}} = \frac{183.0136}{71.6377} = 2.5547$$

E.檢定統計值 $t = 2.5547$ > 右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = 2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_0 = 0$，接受對立假設 H_1: $\beta_0 \neq 0$。代表當兩個自變數數值都等於 0 時，依變數數值不會等於 0。

答案：迴歸方程式中參數 β_0、β_1 和 β_2 皆不等於 0

	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	183.0136	71.6377	2.5547	0.0378	13.6173	352.4098	13.6173	352.4098
X 變數 1	0.6488	0.1689	3.8403	0.0064	0.2493	1.0483	0.2493	1.0483
X 變數 2	-0.8392	0.3389	-2.4762	0.0424	-1.6406	-0.0378	-1.6406	-0.0378

練習 15.8　叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請利用 t 值法在顯著水準 $\alpha = 0.05$ 分別檢定複迴歸方程式中斜率 β_1 和 β_2 是否等於 0。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：利用表格協助公式計算出迴歸係數

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i} - \bar{x}_1$	$x_{2i} - \bar{x}_2$	$y_i - \bar{y}$	$(x_{1i} - \bar{x}_1)^2$	$(x_{2i} - \bar{x}_2)^2$	$(x_{1i} - \bar{x}_1) \times (y_i - \bar{y})$	$(x_{2i} - \bar{x}_2) \times (y_i - \bar{y})$	$(x_{1i} - \bar{x}_1) \times (x_{2i} - \bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0
平均值	173.0	4.6	174.0								

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 \times \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{74.40\times4680-(-328.0)\times(-664.0)}{4610\times74.40-[-328.0]^2}$$

$$= \frac{130400}{235400} = 0.5540$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{4610\times(-664.0)-(-328.0)\times(4680)}{4610\times74.40-[-328.0]^2}$$

$$= \frac{-1526000}{235400} = -6.4826$$

$b_0 = \bar{y}_i - b_1 \times \bar{x}_1 - b_2 \times \bar{x}_2 = 174.0 - 0.5540 \times 173.0 - (-6.4826) \times 4.6 = 107.9864$

估計複迴歸方程式 $\hat{y}_i = 107.9864 + 0.5540 \times x_{1i} - 6.4826 \times x_{2i}$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	150	5	160	158.666	-15.334	235.128	-14	196
2	160	3	180	177.171	3.171	10.054	6	36
3	180	2	190	194.732	20.732	429.831	16	256
4	160	4	170	170.688	-3.312	10.968	-4	16
5	190	1	190	206.754	32.754	1072.855	16	256

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
6	210	1	240	217.833	43.833	1921.374	66	4356
7	140	9	130	127.196	-46.804	2190.590	-44	1936
8	160	8	150	144.758	-29.242	855.103	-24	576
9	180	7	160	162.319	-11.681	136.435	-14	196
10	200	6	170	179.881	5.881	34.587	-4	16
合計	1730	46	1740	1740.000	0.000	6896.924	0	7840
平均值	173.0	4.6	174.0					

$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 6896.924$ $\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 7840.00$

變異數分析(ANOVA)

變異來源(source of variation)	自由度(degrees of freedom)	平方和(sum of square)	均方(mean square)	F 值	顯著值
迴歸	2	6896.924	3448.462	25.5963	0.0000
殘差	7	943.076	134.725		
總和	9	7840.00			

A.設定顯著水準 $\alpha = 0.05$，臨界值 $F_{\alpha,k,n-k-1} = F_{0.05,2,10-2-1} = F_{0.05,2,7} = 4.7374$(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

D.計算檢定統計值—F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{3448.462}{134.725} = 25.5963$$

E.檢定統計值 $F = 25.5963 >$ 臨界值 $F_{\alpha,k,n-k-1} = 4.7374$，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	150	5	160	158.6661	1.3339	1.7793
2	160	3	180	177.1708	2.8292	8.0045
3	180	2	190	194.7324	-4.7324	22.3953
4	160	4	170	170.6882	-0.6882	0.4736
5	190	1	190	206.7545	-16.7545	280.7119
6	210	1	240	217.8335	22.1665	491.3548
7	140	9	130	127.1963	2.8037	7.8609
8	160	8	150	144.7579	5.2421	27.4800
9	180	7	160	162.3195	-2.3195	5.3799
10	200	6	170	179.8811	-9.8811	97.6352
合計	1730	46	1740	1740.0000	0.0000	943.0756
平均值	173.0	4.6	174.0			

$$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{943.0756}{10-2-1} = 134.7251$$

$$S_{b_1}^2 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}\times S_y^2 = \frac{74.40}{4610\times 74.40-[-328.0]^2}\times 134.7251 = \frac{74.40}{235400}\times 134.7251 = 0.0426$$

$$S_{b_2}^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}\times S_y^2 = \frac{4610}{4610\times 74.40-[-328.0]^2}\times 134.7251 = \frac{4610}{235400}\times 134.7251 = 2.6384$$

$$S_{b_0}^2 = \left[\frac{\bar{x}_1^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+\bar{x}_2^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times\bar{x}_1\times\bar{x}_2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right]\times S_y^2 = \left[\frac{173.0^2\times 74.40+4.6^2\times 4610-2\times 173.0\times 4.6\times(-328.0)}{4610\times 74.40-[-328.0]^2}+\frac{1}{10}\right]\times 134.7251 = \left[\frac{2846310}{235400}+\frac{1}{10}\right]\times 134.7251 =$$

$\left[12.0914+\frac{1}{10}\right]\times 134.7251 = 1642.484$

$S_{b_1} = \sqrt{S_{b_1}^2} = \sqrt{0.0426} = 0.2064$ $S_{b_2} = \sqrt{S_{b_2}^2} = \sqrt{2.6384} = 1.6243$

$S_{b_0} = \sqrt{S_{b_0}^2} = \sqrt{1642.484} = 40.5276$

對 β_1 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，右側臨界值$t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$；左側臨界值$-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_1}{S_{b_1}} = \frac{0.5540}{0.2064} = 2.6845$$

E.檢定統計值 $t = 2.6845 > $ 右側臨界值$t_{\frac{\alpha}{2},n-k-1} = 2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$。

對 β_2 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，右側臨界值$t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$；左側臨界值$-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_2 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_2}{S_{b_2}} = \frac{-6.4826}{1.6243} = -3.9910$$

E.檢定統計值 $t = -3.9910 < $ 左側臨界值$-t_{\frac{\alpha}{2},n-k-1} = -2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_2 = 0$，接受對立假設 H_1: $\beta_2 \neq 0$。

對 β_0 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，右側臨界值$t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},10-2-1} = t_{0.025,7} = 2.3646$；左側臨界值$-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},10-2-1} = -t_{0.025,7} = -2.3646$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_0 = 0$。代表當兩個自變數數值都等於 0 時，依變數數值等於 0。

C.對立假設(alternative hypothesis) H_1: $\beta_0 \neq 0$。代表當兩個自變數數值都等於 0 時，依變數數值不會等於 0。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_0}{S_{b_0}} = \frac{107.9864}{40.5276} = 2.6645$$

E.檢定統計值 $t = 2.6645 > $ 右側臨界值$t_{\frac{\alpha}{2},n-k-1} = 2.3646$，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_0 = 0$，接受對立假設 H_1: $\beta_0 \neq 0$。代表當兩個自變數數值都等於 0 時，依變數數值不會等於 0。

答案：迴歸方程式中參數 β_0、β_1 和 β_2 皆不等於 0

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%
截距	107.9864	40.5276	2.6645	0.0323	12.1539	203.8189
X 變數 1	0.5540	0.2064	2.6845	0.0313	0.0660	1.0419

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%
X 變數 2	-6.4826	1.6243	-3.9910	0.0053	-10.3235	-2.6417

練習 15.9　某旅行社有 8 位業務員，上週工作時數 x_{1i}(單位：小時)、年資 x_{2i}(單位：年)和上週業績 y_i(單位：新台幣萬元)列於下表，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請利用 t 值法在顯著水準 $\alpha = 0.05$ 分別檢定複迴歸方程式中斜率 β_1 和 β_2 是否等於 0。

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i
1	44	5	130
2	50	3	120
3	47	2	95
4	40	4	119
5	46	1	90
6	52	1	92
7	49	9	130
8	60	8	150

題解：利用表格協助公式計算出迴歸係數

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	44	5	130	-4.5	0.875	14.25	20.25	0.7656	-64.125	12.4688	-3.9375
2	50	3	120	1.5	-1.125	4.25	2.25	1.2656	6.375	-4.7813	-1.6875
3	47	2	95	-1.5	-2.125	-20.75	2.25	4.5156	31.125	44.0938	3.1875
4	40	4	119	-8.5	-0.125	3.25	72.25	0.0156	-27.625	-0.4063	1.0625
5	46	1	90	-2.5	-3.125	-25.75	6.25	9.7656	64.375	80.4688	7.8125
6	52	1	92	3.5	-3.125	-23.75	12.25	9.7656	-83.125	74.2188	-10.9375
7	49	9	130	0.5	4.875	14.25	0.25	23.7656	7.125	69.4688	2.4375
8	60	8	150	11.5	3.875	34.25	132.25	15.0156	393.875	132.7188	44.5625
合計	388	33	926	0.0	0.000	0.00	248.00	64.8750	328.000	408.2500	42.5000
平均值	48.5	4.125	115.75								

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 \times \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{64.8750\times328-42.50\times408.25}{248\times64.875-[42.5]^2} = \frac{3928.38}{14282.8} = 0.27504$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{248\times408.25-42.50\times328}{248\times64.875-[42.5]^2} = \frac{87306}{14282.8} = 6.11269$$

$b_0 = \bar{y}_i - b_1 \times \bar{x}_1 - b_2 \times \bar{x}_2 = 115.75 - 0.27504 \times 48.5 - 6.11269 \times 4.125 = 77.19556$

估計複迴歸方程式 $\hat{y}_i = 77.1956 + 0.2750 \times x_{1i} - 6.1127 \times x_{2i}$

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$\hat{y}_i$	$\hat{y}_i-\bar{y}$	$(\hat{y}_i-\bar{y})^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$
1	44	5	130	119.8609	4.1109	16.8996	14.25	203.0625
2	50	3	120	109.2858	-6.4642	41.7860	4.25	18.0625
3	47	2	95	102.3480	-13.4020	179.6143	-20.75	430.5625
4	40	4	119	112.6480	-3.1020	9.6221	3.25	10.5625
5	46	1	90	95.9602	-19.7898	391.6346	-25.75	663.0625
6	52	1	92	97.6105	-18.1395	329.0414	-23.75	564.0625
7	49	9	130	145.6869	29.9369	896.2166	14.25	203.0625
8	60	8	150	142.5997	26.8497	720.9045	34.25	1173.0625
合計	388	33	926	926.0000	0.0000	2585.7192	0.00	3265.5000
平均值	48.5	4.125	115.75					

$SSR = \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = 2585.7192$　　$SST = \sum_{i=1}^{n}(y_i-\bar{y})^2 = 3265.50$

變異數分析(ANOVA)

變異來源(source of variation)	自由度(degrees of freedom)	平方和(sum of square)	均方(mean square)	F 值	顯著值
迴歸	2	2585.7192	1292.8596	9.5094	0.0198
殘差	5	679.7808	135.9562		
總和	7	3265.5000			

A.設定顯著水準 $\alpha = 0.05$，臨界值 $F_{\alpha,k,n-k-1} = F_{0.05,2,8-2-1} = F_{0.05,2,5} = 5.7861$(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

D.計算檢定統計值—F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{1292.8596}{135.9562} = 9.5094$$

E.檢定統計值 $F = 9.5094$ $>$ 臨界值 $F_{\alpha,k,n-k-1} = 5.7861$，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$。

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	44	5	130	119.8609	10.1391	102.8012
2	50	3	120	109.2858	10.7142	114.7943
3	47	2	95	102.3480	-7.3480	53.9927
4	40	4	119	112.6480	6.3520	40.3473
5	46	1	90	95.9602	-5.9602	35.5245
6	52	1	92	97.6105	-5.6105	31.4777
7	49	9	130	145.6869	-15.6869	246.0781
8	60	8	150	142.5997	7.4003	54.7649
合計	388	33	926	926.0000	0.0000	679.7808
平均值	48.5	4.125	115.75			

$$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{679.7808}{8-2-1} = 135.9562$$

$$S_{b_1}^2 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} \times S_y^2 = \frac{64.8750}{248\times64.8750-[42.5]^2} \times 135.9562 = \frac{64.8750}{14282.8} \times 135.9562 = 0.6175$$

$$S_{b_2}^2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} \times S_y^2 = \frac{248}{248\times64.8750-[42.5]^2} \times 135.9562 = \frac{248}{14282.8} \times 135.9562 = 2.3607$$

$$S_{b_0}^2 = \left[\frac{\bar{x}_1^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+\bar{x}_2^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times\bar{x}_1\times\bar{x}_2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right] \times S_y^2 = \left[\frac{48.5^2\times64.8750+4.125^2\times248-2\times48.5\times4.125\times(42.5)}{248\times64.875-[42.5]^2}+\frac{1}{8}\right] \times 135.9562 = \left[\frac{139816.8}{14282.8}+\frac{1}{8}\right] \times 135.9562 = \left[9.7892+\frac{1}{8}\right] \times 135.9562 = 1347.9$$

$$S_{b_1} = \sqrt{S_{b_1}^2} = \sqrt{0.6175} = 0.7858 \qquad S_{b_2} = \sqrt{S_{b_2}^2} = \sqrt{2.3607} = 1.5365$$

$$S_{b_0} = \sqrt{S_{b_0}^2} = \sqrt{1347.9} = 36.7137$$

對 β_1 是否等於 0 進行假設檢定

A.設定顯著水準 $\alpha = 0.05$，左側臨界值 $-t_{\frac{\alpha}{2},n-k-1} = -t_{\frac{0.05}{2},8-2-1} = -t_{0.025,5} = -2.5706$；右側臨界值 $t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},8-2-1} = t_{0.025,5} = 2.5706$(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_1 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_1}{S_{b_1}} = \frac{0.2750}{0.7858} = 0.3500$$

E.左側臨界值-$t_{\frac{\alpha}{2},n-k-1}$ = -2.5706 < 檢定統計值 t = 0.3500 < 右側臨界值$t_{\frac{\alpha}{2},n-k-1}$ = 2.5706，檢定統計值位於接受虛無假設區域，推論接受虛無假設 H_0: $\beta_1 = 0$。

對 β_2 是否等於 0 進行假設檢定

A.設定顯著水準 α = 0.05，左側臨界值-$t_{\frac{\alpha}{2},n-k-1}$ = -$t_{\frac{0.05}{2},8-2-1}$ = -$t_{0.025,5}$ = -2.5706；右側臨界值$t_{\frac{\alpha}{2},n-k-1}$ = $t_{\frac{0.05}{2},8-2-1}$ = $t_{0.025,5}$ = 2.5706(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_2 = 0$。

C.對立假設(alternative hypothesis) H_1: $\beta_2 \neq 0$。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_2}{S_{b_2}} = \frac{6.1127}{1.5365} = 3.9784$$

E.檢定統計值 t = 3.9784 > 右側臨界值$t_{\frac{\alpha}{2},n-k-1}$ = 2.5706，檢定統計值位於接受對立假設區域，推論拒絕虛無假設 H_0: $\beta_2 = 0$，接受對立假設 H_1: $\beta_2 \neq 0$。

對 β_0 是否等於 0 進行假設檢定

A.設定顯著水準 α = 0.05，左側臨界值-$t_{\frac{\alpha}{2},n-k-1}$ = -$t_{\frac{0.05}{2},8-2-1}$ = -$t_{0.025,5}$ = -2.5706；右側臨界值$t_{\frac{\alpha}{2},n-k-1}$ = $t_{\frac{0.05}{2},8-2-1}$ = $t_{0.025,5}$ = 2.5706(使用 Excel 軟體 T.INV 函數查詢獲得)。

B.虛無假設(null hypothesis) H_0: $\beta_0 = 0$。代表當兩個自變數數值都等於 0 時，依變數數值等於 0。

C.對立假設(alternative hypothesis) H_1: $\beta_0 \neq 0$。代表當兩個自變數數值都等於 0 時，依變數數值不會等於 0。

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_0}{S_{b_0}} = \frac{77.1956}{36.7137} = 2.1026$$

E.左側臨界值-$t_{\frac{\alpha}{2},n-k-1}$ = -2.5706 < 檢定統計值 t = 2.1026 < 右側臨界值$t_{\frac{\alpha}{2},n-k-1}$ = 2.5706，檢定統計值位於接受虛無假設區域，推論接受虛無假設 H_0: $\beta_0 = 0$。代表當兩個自變數數值都等於 0 時，依變數數值等於 0。

答案：迴歸方程式中僅有參數 β_2 不等於 0

	係數	標準誤	t 統計	P-值	下限 95%	上限 95%
截距	77.1956	36.7137	2.1026	0.0895	-17.1801	171.5712
X 變數 1	0.2750	0.7858	0.3500	0.7406	-1.7450	2.2951
X 變數 2	6.1127	1.5365	3.9784	0.0105	2.1631	10.0623

多重共線性、線性重合(multicollinearity)

在複迴歸分析中，自變數之間可能並非相互獨立，而具有某種程度的相關性存在。利用自變數樣本觀測值之間的相關係數$r_{x_i x_j}$，代表兩自變數之間的相關程度。當兩自變數之間的相關係數$|r_{x_i x_j}|$大於 0.7 時，多重共線性的嚴重性就會很高，影響迴歸係數估計值的估算。

利用最小平方法估算迴歸模式參數的估計值時，若自變數之間有多重共線性的問題發生，可能會對於迴歸出來之係數數值產生很大的影響，甚至於會造成正的數值，變成負的數值。因此，當發現自變數之間有多重共線性時，利用最小平方法估算迴歸模式參數的估計值，其數值可信度，恐需大幅度的懷疑其準確性。

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions		
				(Constant)	X1	X2
1	1	2.950	1.000	0.00	0.00	0.00
	2	0.049	7.760	0.00	0.08	0.04
	3	0.001	56.276	1.00	0.92	0.96

[a] Dependent Variable: Y

在共線性診斷表中，當特徵值(eigenvalue)愈小、條件指標(condition index)愈大、變異數比率(variance proportions)愈大，表示迴歸模式具有共線性問題存在。

在逐步迴歸分析中，迴歸模式會依據自變數對依變數的影響程度，逐一挑選自變數放入迴歸模式中，因此，若利用逐步迴歸分析時，可以將共線性的問題排除。

15.6 運用迴歸參數進行估計和預測

在多元線性迴歸模式中，利用樣本資料以最小平方法，獲得估計線性迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} + b_3 \times x_{3i} + \ldots + b_k \times x_{ki}$。若檢定迴歸方程式中斜率 β_i 皆是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_i = 0$，接受對立假設 H_1: $\beta_i \neq 0$)，而調整複判定係數(adjusted coefficient of multiple determination) R_a^2數值高，顯示其適合度亦高。此估計複迴歸方程式可以應用於預測依變數 y_i。

點估計(point estimation)

阿花連鎖餐廳有 10 營業點，依據每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 的 10 個營業點樣本資料。獲得每個營業點個別的平均每日行銷費用 x_{1i} 和虱目魚套餐定價 x_{2i} 對平均每日販售虱目魚套餐數 y_i 的估計複迴歸方程式$\hat{y}_i = 183.0136 + 0.6488 \times x_{1i} - 0.8392 \times x_{2i}$。若平均每日行銷費用 x_{1i} 為 250 與虱目魚套餐定價 x_{2i} 為 350 時，每日販售虱目魚套餐數的預估值$\hat{y}_i = 183.0136 + 0.6488 \times 250 - 0.8392 \times 350 = 51.49$；平均每日行銷費用 x_{1i} 為 350 與虱目魚套餐定價 x_{2i} 為 150 時，每日販售虱目魚套餐數的預估值$\hat{y}_i = 183.0136 + 0.6488 \times 350 - 0.8392 \times 150 = 284.21$。

區間估計(interval estimation)

複迴歸分析的目的就是預測依變數 y_i，對依變數 y_i 的預測可以分為兩種：第一種是各自變數依序為一特定數值 $x_{10}, x_{20}, x_{30}, \ldots, x_{k0}$ 時，預測依變數之平均值$\bar{y}_0$、E(y₀)或 E(y|x₀)的信賴區間；第二種是各自變數依序為一特定數值 $x_{10}, x_{20}, x_{30}, \ldots, x_{k0}$ 時，預測依變數 y_i 之個別數值 y_0 的信賴區間。以下內容針對兩個自變數的一階複迴歸線性模式 $y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \varepsilon_i, i = 1, \ldots, n$

預測依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_{10},x_{20})$的信賴區間

在複迴歸方程式中，自變數 x_{10} 和 x_{20} 若屬於常態分布時，透過加法定理在複迴歸方程式($\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} + b_3 \times x_{3i} + \ldots + b_k \times x_{ki}$)中的依變數估計值$\hat{y}_0$亦屬於常態分布$\hat{y}_0 \sim N(E(y_0), \sigma_{\hat{y}_0}^2)$。

$$\sigma_{\hat{y}_0}^2 = \sigma^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2 \times \sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 + (x_{20}-\bar{x}_2)^2 \times \sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 - 2 \times (x_{10}-\bar{x}_1) \times (x_{20}-\bar{x}_2) \times \sum_{i=1}^n (x_{1i}-\bar{x}_1) \times (x_{2i}-\bar{x}_2)}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1) \times (x_{2i}-\bar{x}_2)\right]^2} + \frac{1}{n}\right]$$

若依變數母體變異數 σ^2 已知時，在特定自變數數值(x_{10} 和 x_{20})組合下，依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_{10},x_{20})$的信賴區間：

$$\hat{y}_0 \pm z_{\frac{\alpha}{2}} \times \sigma_{\hat{y}_0}$$

若依變數母體變異數 σ^2 未知時，在特定自變數數值(x_{10} 和 x_{20})組合下，依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_{10},x_{20})$的信賴區間：

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{\hat{y}_0}$$

在未知依變數母體變異數 σ² 時，可以利用依變數樣本變異數$S^2_{y|x_1x_2}$ =MSE = $\frac{\text{SSE}}{n-k-1}$ = $\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1}$估計依變數母體變異數 σ² 數值，由依變數估計值$\hat{y}_0$的變異數$S^2_{\hat{y}_0}$，再運用自由度 $n-k-1$ 的 t 分布，推估依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_{10},x_{20})$的信賴區間：

$$S^2_{\hat{y}_0} = S^2_{y|x_1x_2} \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right]$$

其中$S^2_y = S^2_{y|x_1x_2}$ = MSE = $\frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-k-1}$

範例 15.5 阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。透過估計複迴歸方程式$\hat{y}_i$ = b₀ + b₁×x_{1i} + b₂×x_{2i} 進行預測。請推估在每日行銷費用為 200 與虱目魚套餐定價為 150 時，平均販售虱目魚套餐數量$\bar{y}_0$在顯著水準 α = 0.05 的信賴區間。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191
10	275	96	280

題解：估計複迴歸方程式$\hat{y}_i$ = 183.0136 + 0.6488×x_{1i} – 0.8392×x_{2i}

點估計值$\hat{y}_0$ = 183.0136 + 0.6488×x_{1i} – 0.8392×x_{2i} = 183.0136 + 0.6488×200 – 0.8392×150 = 186.8936

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	150	150	156	154.451	1.5487	2.3984
2	160	140	180	169.331	10.6689	113.8251
3	179	125	190	194.246	-4.2459	18.0279
4	160	137	170	171.849	-1.8487	3.4177
5	190	120	198	205.579	-7.5785	57.4340
6	210	99	250	236.177	13.8227	191.0669
7	178	119	189	198.632	-9.6323	92.7818
8	160	145	168	165.135	2.8649	8.2075
9	180	124	191	195.734	-4.7339	22.4099
10	275	96	280	280.866	-0.8657	0.7495
合計	1842	1255	1972	1972.000	0.0000	510.3188
平均值	184.2	125.5	197.2			

$$S^2_y = S^2_{y|x_1x_2} = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{510.3188}{10-2-1} = 72.9027$$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	150	156	-34.20	24.50	-41.20	1169.64	600.25	-837.90
2	160	140	180	-24.20	14.50	-17.20	585.64	210.25	-350.90
3	179	125	190	-5.20	-0.50	-7.20	27.04	0.25	2.60
4	160	137	170	-24.20	11.50	-27.20	585.64	132.25	-278.30
5	190	120	198	5.80	-5.50	0.80	33.64	30.25	-31.90
6	210	99	250	25.80	-26.50	52.80	665.64	702.25	-683.70
7	178	119	189	-6.20	-6.50	-8.20	38.44	42.25	40.30
8	160	145	168	-24.20	19.50	-29.20	585.64	380.25	-471.90

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
9	180	124	191	-4.20	-1.50	-6.20	17.64	2.25	6.30
10	275	96	280	90.80	-29.50	82.80	8244.64	870.25	-2678.60
合計	1842	1255	1972	0.00	0.00	0.00	11953.60	2970.50	-5284.00
平均值	184.2	125.5	197.2						

$S_{\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)]^2}+\frac{1}{n}\right] = 72.9027 \times \left[\frac{(200-184.2)^2\times2970.50+(150-125.5)^2\times11953.60-2\times(200-184.2)\times(150-125.5)\times(-5284.00)}{11953.60\times2970.50-[-5284.00]^2}+\frac{1}{10}\right] = 72.9027 \times \left[\frac{12007577}{7587513}+\frac{1}{10}\right] = 122.6620$

$S_{\hat{y}_0} = \sqrt{S_{\hat{y}_0}^2} = \sqrt{122.6620} = 11.0753$

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{\hat{y}_0} = 186.8936 \pm t_{\frac{0.05}{2},10-2-1} \times 11.0753 = 186.8936 \pm 2.3646 \times 11.0753 = 186.8936 \pm 26.1886$

答案：在每日行銷費用為 200 與虱目魚套餐定價為 150 時，平均販售虱目魚套餐數量的信賴區間 160.70~213.08 套

練習 15.10 叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表，欲建立迴歸模式 $y_i = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \varepsilon_i$ 進行預測，其中 ε_i 為誤差項。請推估在每日行銷費用為 200 與兩公里內競爭咖啡店數量為 5 家時，平均每日販售咖啡杯數$\bar{y}_0$在顯著水準 $\alpha = 0.05$ 的信賴區間。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：估計複迴歸方程式$\hat{y}_i = 107.9864 + 0.5540\times x_{1i} - 6.4826\times x_{2i}$

點估計值$\hat{y}_0 = 107.9864 + 0.5540\times x_{1i} - 6.4826\times x_{2i} = 107.9864 + 0.5540\times200 - 6.4826\times5 = 186.3636$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	150	5	160	158.6661	1.3339	1.7793
2	160	3	180	177.1708	2.8292	8.0045
3	180	2	190	194.7324	-4.7324	22.3953
4	160	4	170	170.6882	-0.6882	0.4736
5	190	1	190	206.7545	-16.7545	280.7119
6	210	1	240	217.8335	22.1665	491.3548
7	140	9	130	127.1963	2.8037	7.8609
8	160	8	150	144.7579	5.2421	27.4800
9	180	7	160	162.3195	-2.3195	5.3799
10	200	6	170	179.8811	-9.8811	97.6352
合計	1730	46	1740	1740.0000	0.0000	943.0756
平均值	173.0	4.6	174.0			

$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{943.0756}{10-2-1} = 134.7251$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0
平均值	173.0	4.6	174.0								

$$S_{\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right] =$$

$$134.7251 \times \left[\frac{(200-173.0)^2\times 74.40+(5-4.6)^2\times 4610-2\times(200-173.0)\times(5-4.6)\times(-328.0)}{4610\times 74.40-[-328.0]^2}+\frac{1}{10}\right] = 134.7251 \times \left[\frac{62060}{235400}+\frac{1}{10}\right]$$

$$= 48.9910$$

$$S_{\hat{y}_0} = \sqrt{S_{\hat{y}_0}^2} = \sqrt{48.9910} = 6.9994$$

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{\hat{y}_0} = 186.3636 \pm t_{\frac{0.05}{2},10-2-1} \times 6.9994 = 186.3636 \pm 2.3646 \times 6.9994$$

$$= 186.3636 \pm 16.5507$$

答案：在每日行銷費用為 200 與兩公里內競爭咖啡店數量為 5 家時，平均每日販售咖啡杯數$\bar{y}_0$的信賴區間 169.81~202.91 杯

練習 15.11 某旅行社有 8 位業務員，上週工作時數 x_{1i}(單位：小時)、年資 x_{2i}(單位：年)和上週業績 y_i(單位：新台幣萬元)列於下表，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請推估在上周工作時數為 45 小時與年資 7 年時，平均上週業績$\bar{y}_0$在顯著水準 $\alpha = 0.05$ 的信賴區間。

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i
1	44	5	120
2	50	3	105
3	47	2	103
4	40	4	110
5	46	1	96
6	52	1	100
7	49	9	152
8	60	8	150

題解：利用表格協助公式計算出迴歸係數

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	44	5	120	-4.5	0.875	3	20.25	0.7656	-13.5	2.625	-3.9375
2	50	3	105	1.5	-1.125	-12	2.25	1.2656	-18.0	13.500	-1.6875
3	47	2	103	-1.5	-2.125	-14	2.25	4.5156	21.0	29.750	3.1875
4	40	4	110	-8.5	-0.125	-7	72.25	0.0156	59.5	0.875	1.0625
5	46	1	96	-2.5	-3.125	-21	6.25	9.7656	52.5	65.625	7.8125
6	52	1	100	3.5	-3.125	-17	12.25	9.7656	-59.5	53.125	-10.9375
7	49	9	152	0.5	4.875	35	0.25	23.7656	17.5	170.625	2.4375
8	60	8	150	11.5	3.875	33	132.25	15.0156	379.5	127.875	44.5625
合計	388	33	936	0.0	0.000	0	248.00	64.8750	439.0	464.000	42.5000
平均值	48.5	4.125	117.0								

$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{64.8750\times439-42.50\times464}{248\times64.875-[42.5]^2} = \frac{8760.13}{14282.8} = 0.61334$

$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{248\times464-42.50\times439}{248\times64.875-[42.5]^2} = \frac{96414.5}{14282.8} = 6.75042$

$b_0 = \bar{y}_i - b_1\times\bar{x}_1 - b_2\times\bar{x}_2 = 117.0 - 0.61334\times48.5 - 6.75042\times4.125 = 59.4077$

估計複迴歸方程式$\hat{y}_i = 59.4077 + 0.6133\times x_{1i} - 6.7504\times x_{2i}$

點估計值$\hat{y}_0 = 59.4077 + 0.6133\times x_{1i} - 6.7504\times x_{2i} = 59.4077 + 0.6133\times45 - 6.7504\times7 = 134.2608$

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	44	5	120	120.1466	-0.1466	0.0215
2	50	3	105	110.3258	-5.3258	28.3640
3	47	2	103	101.7354	1.2646	1.5993
4	40	4	110	110.9428	-0.9428	0.8890
5	46	1	96	94.3716	1.6284	2.6517
6	52	1	100	98.0516	1.9484	3.7962
7	49	9	152	150.2149	1.7851	3.1864
8	60	8	150	150.2112	-0.2112	0.0446
合計	388	33	936	936.0000	0.0000	40.5526
平均值	48.5	4.125	117.0			

$$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{40.5526}{8-2-1} = 8.1105$$

$S_{\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}\right] = 8.1105 \times \left[\frac{(45-48.5)^2\times64.875+(7-4.125)^2\times248-2\times(45-48.5)\times(7-4.125)\times(42.5)}{248\times64.875-[42.5]^2}+\frac{1}{8}\right] = 8.1105 \times \left[\frac{3699.906}{14282.8}+\frac{1}{8}\right] = 3.1148$

$S_{\hat{y}_0} = \sqrt{S_{\hat{y}_0}^2} = \sqrt{3.1148} = 1.7649$

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{\hat{y}_0} = 134.2608 \pm t_{\frac{0.05}{2},8-2-1} \times 1.7649 = 134.2608 \pm 2.5706 \times 1.7649 = 134.2608 \pm 4.5368$

答案：在上周工作時數為 45 小時與年資 7 年時，平均上週業績$\bar{y}_0$的信賴區間 129.72~138.80 萬元

預測依變數 y_0 的信賴區間

進行依變數 y_0 的信賴區間估算時，需先由 $y_0 - \hat{y}_0 = e_0$ 的抽樣分布切入。在複迴歸方程式中，自變數 x_{10} 和 x_{20} 若屬於常態分布時，透過加法定理在複迴歸方程式中的依變數估計值$\hat{y}_0$亦屬於常態分布 $\hat{y}_0\sim N(\text{E}(y_0),\sigma_{\hat{y}_0}^2)$。故 $y_0 - \hat{y}_0 = e_0$ 分布亦屬於常態分布 $y_0 - \hat{y}_0 = e_0\sim N(0,\sigma_{e_0}^2)$或 $y_0 - \hat{y}_0\sim N(0,\sigma_{y_0-\hat{y}_0}^2)$。

$\sigma_{e_0}^2 = \sigma_{y_0-\hat{y}_0}^2 = \sigma^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}+1\right]$

若依變數母體變異數 σ^2 已知時，依變數 y_0 的信賴區間

$$\hat{y}_0 \pm z_{\frac{\alpha}{2}} \times \sigma_{y_0-\hat{y}_0}$$

若依變數母體變異數 σ^2 未知時，依變數 y_0 的信賴區間

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{y_0-\hat{y}_0}$$

在未知依變數母體變異數 σ^2 時，可以利用依變數樣本變異數$S_{y|x_1x_2}^2$估計依變數母體變異數 σ^2 數值，由依變數與其估計值之差 $y_0 - \hat{y}_0$的變異數$S_{y_0-\hat{y}_0}^2$，再運用自由度 $n-k-1$ 的 t 分布，推估依變數 y_0 的信賴區間

$\sigma_{e_0}^2 = S_{y_0-\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}+1\right]$

其中$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-k-1}$

範例 15.6 阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。透過估計複迴歸方程式$\hat{y}_i = b_0 + b_1\times x_{1i} + b_2\times x_{2i}$ 進行預測。請推估在每日行銷費用為 200 與虱目魚套餐定價為 150 時，販售虱目魚套餐數量 y_0 在顯著水準 $\alpha = 0.05$ 的信賴區間。

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i
1	150	150	156
2	160	140	180
3	179	125	190
4	160	137	170
5	190	120	198
6	210	99	250
7	178	119	189
8	160	145	168
9	180	124	191
10	275	96	280

題解：估計複迴歸方程式$\hat{y}_i = 183.0136 + 0.6488\times x_{1i} - 0.8392\times x_{2i}$

$\hat{y}_0 = 183.0136 + 0.6488\times x_{1i} - 0.8392\times x_{2i} = 183.0136 + 0.6488\times 200 - 0.8392\times 150 = 186.8936$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	150	150	156	154.451	1.5487	2.3984
2	160	140	180	169.331	10.6689	113.8251
3	179	125	190	194.246	-4.2459	18.0279
4	160	137	170	171.849	-1.8487	3.4177
5	190	120	198	205.579	-7.5785	57.4340
6	210	99	250	236.177	13.8227	191.0669
7	178	119	189	198.632	-9.6323	92.7818
8	160	145	168	165.135	2.8649	8.2075
9	180	124	191	195.734	-4.7339	22.4099
10	275	96	280	280.866	-0.8657	0.7495
合計	1842	1255	1972	1972.000	0.0000	510.3188
平均值	184.2	125.5	197.2			

$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{510.3188}{10-2-1} = 72.9027$

營業點 i	行銷費用 x_{1i}	虱目魚套餐定價 x_{2i}	虱目魚套餐數 y_i	$x_{1i} - \bar{x}_1$	$x_{2i} - \bar{x}_2$	$y_i - \bar{y}$	$(x_{1i} - \bar{x}_1)^2$	$(x_{2i} - \bar{x}_2)^2$	$(x_{1i} - \bar{x}_1)\times(x_{2i} - \bar{x}_2)$
1	150	150	156	-34.20	24.50	-41.20	1169.64	600.25	-837.90
2	160	140	180	-24.20	14.50	-17.20	585.64	210.25	-350.90
3	179	125	190	-5.20	-0.50	-7.20	27.04	0.25	2.60
4	160	137	170	-24.20	11.50	-27.20	585.64	132.25	-278.30
5	190	120	198	5.80	-5.50	0.80	33.64	30.25	-31.90
6	210	99	250	25.80	-26.50	52.80	665.64	702.25	-683.70
7	178	119	189	-6.20	-6.50	-8.20	38.44	42.25	40.30
8	160	145	168	-24.20	19.50	-29.20	585.64	380.25	-471.90
9	180	124	191	-4.20	-1.50	-6.20	17.64	2.25	6.30
10	275	96	280	90.80	-29.50	82.80	8244.64	870.25	-2678.60
合計	1842	1255	1972	0.00	0.00	0.00	11953.60	2970.50	-5284.00
平均值	184.2	125.5	197.2						

$\sigma_{e_0}^2 = S_{y_0-\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times$
$\left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)]^2}+\frac{1}{n}+1\right] = 72.9027 \times$

$\left[\frac{(200-184.2)^2\times 2970.50+(150-125.5)^2\times 11953.60-2\times(200-184.2)\times(150-125.5)\times(-5284.00)}{11953.60\times 2970.50-[-5284.00]^2}+\frac{1}{10}+1\right]=72.9027\times\left[\frac{12007577}{7587513}+\frac{1}{10}+1\right]=195.5647$

$S_{y_0-\hat{y}_0}=\sqrt{S^2_{y_0-\hat{y}_0}}=\sqrt{195.5647}=13.9844$

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1}\times S_{y_0-\hat{y}_0}=186.8936\pm t_{\frac{0.05}{2},10-2-1}\times 13.9844=186.8936\pm 2.3646\times 13.9844=186.8936\pm 33.0676$

答案：在每日行銷費用為 200 與虱目魚套餐定價為 150 時，販售虱目魚套餐數量 y_0 的信賴區間 153.83~219.96 套

練習 15.12 叮叮連鎖咖啡店有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}(新台幣：元)、兩公里內競爭咖啡店數量 x_{2i}(單位：家)和平均每日販售咖啡杯數 y_i(單位：杯)列於下表，欲建立迴歸模式 $y_i=\beta_0+\beta_1\times x_{1i}+\beta_2\times x_{2i}+\varepsilon_i$ 進行預測，其中 ε_i 為誤差項。請推估在每日行銷費用為 200 與兩公里內競爭咖啡店數量為 5 家時，每日販售咖啡杯數 y_0 在顯著水準 $\alpha=0.05$ 的信賴區間。

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i
1	150	5	160
2	160	3	180
3	180	2	190
4	160	4	170
5	190	1	190
6	210	1	240
7	140	9	130
8	160	8	150
9	180	7	160
10	200	6	170

題解：估計複迴歸方程式 $\hat{y}_i=107.9864+0.5540\times x_{1i}-6.4826\times x_{2i}$

點估計值 $\hat{y}_0=107.9864+0.5540\times x_{1i}-6.4826\times x_{2i}=107.9864+0.5540\times 200-6.4826\times 5=186.3636$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$
1	150	5	160	158.6661	1.3339	1.7793
2	160	3	180	177.1708	2.8292	8.0045
3	180	2	190	194.7324	-4.7324	22.3953
4	160	4	170	170.6882	-0.6882	0.4736
5	190	1	190	206.7545	-16.7545	280.7119
6	210	1	240	217.8335	22.1665	491.3548
7	140	9	130	127.1963	2.8037	7.8609
8	160	8	150	144.7579	5.2421	27.4800
9	180	7	160	162.3195	-2.3195	5.3799
10	200	6	170	179.8811	-9.8811	97.6352
合計	1730	46	1740	1740.0000	0.0000	943.0756
平均值	173.0	4.6	174.0			

$S_y^2=S^2_{y|x_1x_2}=\text{MSE}=\frac{\text{SSE}}{n-k-1}=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1}=\frac{943.0756}{10-2-1}=134.7251$

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	150	5	160	-23.0	0.4	-14.0	529	0.16	322	-5.6	-9.2
2	160	3	180	-13.0	-1.6	6.0	169	2.56	-78	-9.6	20.8
3	180	2	190	7.0	-2.6	16.0	49	6.76	112	-41.6	-18.2
4	160	4	170	-13.0	-0.6	-4.0	169	0.36	52	2.4	7.8
5	190	1	190	17.0	-3.6	16.0	289	12.96	272	-57.6	-61.2
6	210	1	240	37.0	-3.6	66.0	1369	12.96	2442	-237.6	-133.2
7	140	9	130	-33.0	4.4	-44.0	1089	19.36	1452	-193.6	-145.2

營業點 i	行銷費用 x_{1i}	競爭咖啡店數量 x_{2i}	販售咖啡杯數 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
8	160	8	150	-13.0	3.4	-24.0	169	11.56	312	-81.6	-44.2
9	180	7	160	7.0	2.4	-14.0	49	5.76	-98	-33.6	16.8
10	200	6	170	27.0	1.4	-4.0	729	1.96	-108	-5.6	37.8
合計	1730	46	1740	0.0	0.0	0.0	4610	74.40	4680	-664.0	-328.0
平均值	173.0	4.6	174.0								

$$\sigma_{e_0}^2 = S_{y_0-\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}+1\right] = 134.7251$$

$$\times \left[\frac{(200-173.0)^2\times 74.40+(5-4.6)^2\times 4610-2\times(200-173.0)\times(5-4.6)\times(-328.0)}{4610\times 74.40-[-328.0]^2}+\frac{1}{10}+1\right] = 134.7251 \times \left[\frac{62060}{235400}+\frac{1}{10}+1\right]$$

$$= 183.7160$$

$$S_{y_0-\hat{y}_0} = \sqrt{S_{y_0-\hat{y}_0}^2} = \sqrt{183.7160} = 13.5542$$

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{y_0-\hat{y}_0} = 186.3636 \pm t_{\frac{0.05}{2},10-2-1} \times 13.5542 = 186.3636 \pm 2.3646 \times 13.5542 = 186.3636 \pm 32.0502$$

答案：在每日行銷費用為 200 與兩公里內競爭咖啡店數量為 5 家時，每日販售咖啡杯數 y_0 的信賴區間 154.31~218.41 杯

練習 15.13 某旅行社有 8 位業務員，上週工作時數 x_{1i}(單位：小時)、年資 x_{2i}(單位：年)和上週業績 y_i(單位：新台幣萬元)列於下表，欲建立複迴歸模式 $y_i = \beta_0 + \beta_1\times x_{1i} + \beta_2\times x_{2i} + \varepsilon_i$，其中 ε_i 為誤差項。請推估在上周工作時數為 45 小時與年資 7 年時，上週業績 y_0 在顯著水準 $\alpha = 0.05$ 的信賴區間。

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i
1	44	5	120
2	50	3	105
3	47	2	103
4	40	4	110
5	46	1	96
6	52	1	100
7	49	9	152
8	60	8	150

題解：利用表格協助公式計算出迴歸係數

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$x_{1i}-\bar{x}_1$	$x_{2i}-\bar{x}_2$	$y_i-\bar{y}$	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})$	$(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})$	$(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)$
1	44	5	120	-4.5	0.875	3	20.25	0.7656	-13.5	2.625	-3.9375
2	50	3	105	1.5	-1.125	-12	2.25	1.2656	-18.0	13.500	-1.6875
3	47	2	103	-1.5	-2.125	-14	2.25	4.5156	21.0	29.750	3.1875
4	40	4	110	-8.5	-0.125	-7	72.25	0.0156	59.5	0.875	1.0625
5	46	1	96	-2.5	-3.125	-21	6.25	9.7656	52.5	65.625	7.8125
6	52	1	100	3.5	-3.125	-17	12.25	9.7656	-59.5	53.125	-10.9375
7	49	9	152	0.5	4.875	35	0.25	23.7656	17.5	170.625	2.4375
8	60	8	150	11.5	3.875	33	132.25	15.0156	379.5	127.875	44.5625
合計	388	33	936	0.0	0.000	0	248.00	64.8750	439.0	464.000	42.5000
平均值	48.5	4.125	117.0								

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{64.8750\times 439-42.50\times 464}{248\times 64.875-[42.5]^2} = \frac{8760.13}{14282.8} = 0.61334$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})-\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} = \frac{248\times 464-42.50\times 439}{248\times 64.875-[42.5]^2} = \frac{96414.5}{14282.8} = 6.75042$$

$b_0 = \bar{y}_i - b_1 \times \bar{x}_1 - b_2 \times \bar{x}_2 = 117.0 - 0.61334 \times 48.5 - 6.75042 \times 4.125 = 59.4077$

估計複迴歸方程式$\hat{y}_i = 59.4077 + 0.6133 \times x_{1i} - 6.7504 \times x_{2i}$

點估計值$\hat{y}_0 = 59.4077 + 0.6133 \times x_{1i} - 6.7504 \times x_{2i} = 59.4077 + 0.6133 \times 45 - 6.7504 \times 7 = 134.2608$

業務員 i	上週工作時數 x_{1i}	年資 x_{2i}	上週業績 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	44	5	120	120.1466	-0.1466	0.0215
2	50	3	105	110.3258	-5.3258	28.3640
3	47	2	103	101.7354	1.2646	1.5993
4	40	4	110	110.9428	-0.9428	0.8890
5	46	1	96	94.3716	1.6284	2.6517
6	52	1	100	98.0516	1.9484	3.7962
7	49	9	152	150.2149	1.7851	3.1864
8	60	8	150	150.2112	-0.2112	0.0446
合計	388	33	936	936.0000	0.0000	40.5526
平均值	48.5	4.125	117.0			

$$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} = \frac{40.5526}{8-2-1} = 8.1105$$

$$\sigma_{e_0}^2 = S_{y_0-\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2+(x_{20}-\bar{x}_2)^2\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2-2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2-\left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}+\frac{1}{n}+1\right] = 8.1105 \times \left[\frac{(45-48.5)^2\times 64.875+(7-4.125)^2\times 248-2\times(45-48.5)\times(7-4.125)\times(42.5)}{248\times 64.875-[42.5]^2}+\frac{1}{8}+1\right] = 8.1105 \times \left[\frac{3699.906}{14282.8}+\frac{1}{8}+1\right] = 11.2253$$

$$S_{\hat{y}_0} = \sqrt{S_{\hat{y}_0}^2} = \sqrt{11.2253} = 3.3504$$

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-k-1} \times S_{\hat{y}_0} = 134.2608 \pm t_{\frac{0.05}{2},8-2-1} \times 3.3504 = 134.2608 \pm 2.5706 \times 3.3504 = 134.2608 \pm 8.6125$$

答案：在上周工作時數為 45 小時與年資 7 年時，上週業績 y_0 的信賴區間 125.65~142.87 萬元

15.7 質化自變數【選擇教材】

前面數個單元中，使用於建立迴歸模型的自變數與依變數皆屬於量化變數或定量變數(interval scale or ratio scale)。

範例 15.7　阿花連鎖餐廳有 10 營業點，每個營業點個別的平均每日行銷費用 x_{1i}、虱目魚套餐定價 x_{2i} 和平均每日販售虱目魚套餐數 y_i 列於下表。請利用 t 值法分別檢定迴歸方程式中斜率 β_1 和 β_2 是否等於 0。

營業點 i	行銷費用 x_{1i}	是否為自助式服務 x_{2i}	虱目魚套餐數 y_i
1	150	1	161
2	160	0	180
3	170	0	190
4	160	1	170
5	185	0	190
6	235	1	220
7	190	1	189
8	175	1	168
9	165	0	191
10	230	0	250

迴歸分析中若欲利用名目尺度(Nominal scale)或順序尺度(ordinal scale)變數為自變數時，必須轉換為虛

擬變數(Dummy variable)、**指標變數**或**指示變數**(indicator variable)。

在範例中，屬於二項式 nominal scale，其轉換比較簡單

$x_{2i} = 0$ 代表餐廳沒有提供自助式服務

$x_{2i} = 1$ 代表餐廳有提供自助式服務

SPSS 分析結果

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.963[a]	.928	.907	8.0777	

[a] Predictors: (Constant), X2, X1

ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.	
1	Regression	5862.154	2	2931.077	44.921	.000[a]	
	Residual	456.746	7	65.249			
	Total	6318.900	9				

[a] Predictors: (Constant), X2, X1
[b] Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics		
		B	Std. Error	Beta			Tolerance	VIF	
1	(Constant)	53.198	17.182		3.096	.017			
	X1	.808	.092	.889	8.751	.000	1.000	1.000	
	X2	-18.600	5.109	-.370	-3.641	.008	1.000	1.000	

[a] Dependent Variable: Y

獲得估計複迴歸方程式

$$\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} = 53.198 + 0.808 \times x_{1i} - 18.600 \times x_{2i}$$

參數解讀

在具有虛擬變數時，對於參數估計值的判讀，先依據虛擬變數的數值(類別)分類後，再探討其他自變數對於依變數的影響。

因此，當虛擬自變數(是否為自助式服務)屬於：0，代表餐廳沒有提供自助式服務時，其迴歸方程式

$E(y_i|$餐廳沒有提供自助式服務$) = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times 0 = 53.198 + 0.808 \times x_{1i}$

$E(y_i|$餐廳有提供自助式服務$) = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times 1 = 53.198 + 0.808 \times x_{1i} - 18.600 = 34.598 + 0.808 \times x_{1i}$

因此，在不管有沒有提供自助式服務時，每單位 x_{1i} 改變對於依變數期望值 $E(y_i)$的影響都是一樣，但是截距不同。沒有提供自助式服務者截距為 53.198，有提供自助式服務者 34.598。因此，有沒有提供自助式服務對於依變數期望值的影響，即為其參數數值 β_2，在此範例中為–18.600。

15.8 殘差分析

15.9 複迴歸分析應用【選擇教材】

15.9.1 研究問題

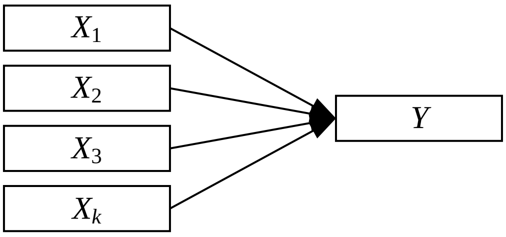
遊客的「服務滿意度指標」、「停留天數」、「同行人數」、「服務期望指標」是否可以預測或影響「消費金額」，其預測能力為何？

分析方法

多元迴歸分析法

研究問題中，預測變數(independent variable)包括「服務滿意度指標」、「停留天數」、「同行人數」、「服務期望指標」等四個，依變數(dependent variable)為「消費金額」變項一個，可採用**多元迴歸分析法**(multiple regression)或**複迴歸分析法**。

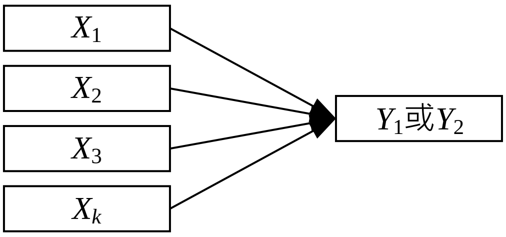
預測變數(k 個連續變項：等距和比例尺度) 自變數	校標變項(1 個連續變項：等距和比例尺度) 依變數



區分分析，對數式迴歸分析

若預測變項非連續變數(非等距變項、非比率變項而是名義變項或次序變項)，則此預測變數要化為虛擬變數(dummy variable)，如家庭狀況、性別、年齡組距等是類別變項，要納入為預測變項，其數據資料要化為「1」、「2」、「3」等，以虛擬變項方式轉化變項的方式，將此種類別變項作為一個預測變項。

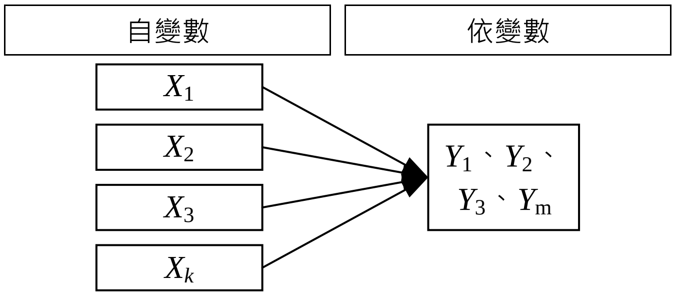
預測變數(k 個連續變項：等距和比例尺度) 自變數	校標變項【1 個(二分)分立變項：名目尺度】 依變數



區別分析(Discriminant Analysis)

進行複迴歸分析時，若依變數不是連續變項，而是二分類別變項或二分次序變項時，應以「區分分析」或對數式迴歸分析(Logistic regression analysis)，若依變數是多分類別變項或多分次序變項(水準數在三個以上)，則需進行區別分析。

預測變數(k 個連續變項：等距和比例尺度)	校標變項【1 個(多分)分立變項：名目尺度】



百分位數迴歸(Quantile regression)

共線性

複迴歸分析需要注意其「共線性」問題，即其由於自變數(自變項)間的相關性太高，造成迴歸分析之情境困擾。「共線性」問題，表示一個預測變項是其他自變項的線性組合，以二個自變項 X_1，X_2 為例，完全共線性代表的是 X_1 是 X_2 的線性函數，$X_1 = a + bX_2$，若模式中，有嚴重的共線性存在，則模式之參數就不能完全被估計出來。

自變數(自變項)間是否有共線性問題，可由下列指標判斷

1.容忍度(Tolerance)

容忍度等於 $1\text{-}R^2$，其中 R^2 是此自變數與其他自變數間的多元相關係數的平方，若 R^2 值太大，代表模式中其他自變數可以有效解釋此自變數。容忍度的值界於 0 與 1 間，若一個自變數的容忍度太小，表示此變項與其他自變項間有共線性問題；其值若接近 0，表示變項幾乎是其他變項的線性組合，此種情況下迴歸係數的估算值不夠穩定，而迴歸係數的計算值也會有很大誤差。

2.變異數膨脹因素(variance inflation factor; VIF)

變異數膨脹因素為容任度的倒數，VIF 的值愈大，表示自變數的容忍度愈小，愈有共線性的問題。

3.條件指標(Condition index; CI)

條件指標 CI 值愈大，愈有共線性問題。

在自變項相關矩陣之因素分析中，特徵值可作為變項間有多少層面(Dimension)的指標，若特徵值接近 0，表示原始變項間有高的內在相關存在，此組自變項間的相關矩陣就是一個「不佳的條件」(ill condition)，資料數值若稍微變動，即可能導致係數估計的大波動。

條件指標為最大特徵值與個別特徵值比例的平方根，條件指標若在 15 以上，表示可能有共線性問題，條件指標若在 30 以上，則表示有嚴重的共線性問題，CI 值愈大，愈有共線性問題。

在迴歸分析中，最好先呈現預測變數間相關矩陣，以探討變數間的相關情況，若某些自變數間的相關係數太高，可考量只挑選其中一個較重要的變項投入複迴歸分析

複迴歸分析基本假設

1.存在性(existence)

就自變數 X_1，X_2，X_3，…，X_k 的特殊組合而言，Y 變項(單變量)是一個隨機變數，具有某種機率分布情況，有一定的平均數與變異量

2.獨立性(independent)

每一個觀測值 Y 彼此間是統計獨立的，觀察值間沒有關聯。

3.直線性(linearity)關係

Y 變數(自變數 X_1，X_2，X_3，…，X_k 的線性組合)的平均數是 X_1，X_2，X_3，…，X_k 變項間的一個線性函數，此線性函數即為迴歸方程式。

4.變異數同質性

就 X_1，X_2，X_3，…，X_k 任何一個線性組合，依變數 Y 的變異數均相同。

5.常態性

就 X_1，X_2，X_3，…，X_k 任何一個線性組合而言，依變數 Y 的分布是常態性的。

複迴歸分析的原始迴歸方程式

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \ldots + B_kX_k$$

其中 B_0 為截距、B_k 為原始迴歸係數

標準化迴歸方程式

$$Z_Y = \beta_1 Z_{X1} + \beta_2 Z_{X2} + \beta_3 Z_{X3} + \ldots + \beta_k Z_{Xk}$$

其中β_k 為標準化迴歸係數

虛擬變數(dummy variable)的轉換，均需以「0」、「1」的方式表示，【虛擬變項數】等於【水準數】減一。若是二分變數(名目和順序尺度)，便以一個虛擬變數表示，此虛擬變項的兩個水準數值直接以「0」、「1」表示即可。若間斷變項(名目和順序尺度)有三個水準，則應以二個虛擬變數表示，二個虛擬變項的數值如下

虛擬變項 / 原變項	Homd1	Homd2	說明：1 表示是，0 表是否
單親家庭組 1	1	0	是單親家庭組，不是他人照顧組
他人照顧組 2	0	1	不是單親家庭組，是他人照顧組
雙親家庭組 3	0	0	不是單親家庭組，也不是他人照顧組，即為雙親家庭組

若間斷變數有四個水準，如地理位置變項(loc)，進行迴歸分析時，會有三個虛擬變項，三個虛擬變項如下

虛擬變項 / 原變項	Locd1	Locd2	Locd3	說明：1 表示是，0 表是否
北部 1	1	0	0	是北部，非中部，亦非南部
中部 2	0	1	0	是中部，非北部，亦非南部
南部 3	0	0	1	是南部，非北部，亦非中部
東部 4	0	0	0	非北部，非中部，亦非南部，因而是東部

除非是重要預測變項，否則不應輕易將間斷變項投入迴歸分析中。

複迴歸分析(強制選入法)SPSS 操作方法

1.欲投入迴歸分析的自變數若為間斷變項，應先將原始變數轉換為「**虛擬變數**」(dummy variables)。以性別為例，原始數值 1 代表男生、2 代表女生，需將其轉換為 0、1，亦即 0 代表男生、1 代表女生。使用 Transform(轉換) → Recode(重新編碼)→Into Different Variables…(成不同一變數…)等指令轉換。

2.Analyze/Statistics(統計分析) → Regression(迴歸分析) → Linear…(線性…)，即會出現 Linear Regression(線性迴歸)對話視窗。

3.將欲進行迴歸分析的依變數(因變項)：消費金額(cost_nr)自左邊的方塊中，點選進入右上角的 Dependent: (依變數)下方的方塊中。

4.將欲進行迴歸分析的自變數(預測變數、自變項)：服務滿意度指標(qulit_nr)、停留天數(day_n)、同行人數(p_number)和服務期望指標(ex_nr)自左邊的方塊中，點選進入右邊的 Independent(s): (自變數)下方的方塊中。

5.在 Independent(s): (自變數)下方 Methods: 下拉式選項中，選取「Enter」(強制選取法)。

6.接著按下方的 Statistics…(統計量)鈕，即會出現「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗。

7.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗中，左上角 Regression Coefficients(迴歸係數)方塊中，勾選□Estimates(估計值)和□Confidence intervals(信賴區間)選項。

8.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗中，勾選右邊的□Model fit(迴歸模式適合度檢定)、□R squared change(解釋量的改變量)、□Descriptive(統計量)、□Collinearity diagnostics(共線性診斷)等選項。

9.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗下面的 Residuals 方塊中，勾選□Durbin-Watson(顯示 Durbin-Watson 檢定統計量，標準化、非標準化之殘差和預測值之摘要統計量)選項。

10.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗，按 Continue 鈕，即會回到 Linear Regression(線性迴歸)對話視窗。

11.在 Linear Regression 對話視窗中，點選下面的 Plots…按鈕，會出現 Linear Regression: Plots 次對話視窗。

12.在 Linear Regression: Plots 次對話視窗中，左邊方塊內有數項資料名稱，其意義如下(*：表示暫時性的殘差變數)：

- 「DEPENDNT」: 為依變數(dependent variable)
- 「*ZPRED」: 標準化的迴歸預測值
- 「*ZRESID」: 標準化的殘差
- 「*DRESID」: 刪除型的殘差(deleted residual)
- 「*ADJPRED」: 調整的迴歸預測值(adjusted predicted values)
- 「*SRESID」: studentized 殘差
- 「*SDRESID」: studentized 刪除型的殘差

13.在 Linear Regression: Plots 次對話視窗中，點選 「*ZRESID」(標準化的殘差)進入 Y: (Y 軸)右邊的方塊，點選「DEPENDNT」(依變數)進入 X: (X 軸)右邊的方塊。

14.在 Linear Regression: Plots 次對話視窗下面的 Standardized Residuals Plots(標準化殘差圖)方塊中，勾選□Histogram(顯示標準化殘差值的次數分配圖，同時產生一常態分配曲線)和□Normal probability plot[顯示標準化殘差值的常態機率(P-P)圖]選項。

15.在 Linear Regression: Plots 次對話視窗右上角，按 Continues 鈕，回到 Linear Regression 對話視窗。

16.按 OK(確定)鈕，以執行複迴歸分析程序。

17.獲得以下複迴歸分析成果。

Descriptive Statistics

	Mean	Std. Deviation	N
消費金額	154.5152	52.4302	33
服務滿意度指標	15.5606	4.7265	33

	Mean	Std. Deviation	N
停留天數	1.4545	.5056	33
同行人數	3.2727	.6261	33
服務期望指標	2.2576	.5166	33

Correlations

		消費金額	服務滿意度指標	停留天數	同行人數	服務期望指標
Pearson Correlation	消費金額	1.000	.926	.473	.627	.812
	服務滿意度指標	.926	1.000	.524	.633	.764
	停留天數	.473	.524	1.000	.090	.770
	同行人數	.627	.633	.090	1.000	.501
	服務期望指標	.812	.764	.770	.501	1.000
Sig. (1-tailed)	消費金額	.	.000	.003	.000	.000
	服務滿意度指標	.000	.	.001	.000	.000
	停留天數	.003	.001	.	.310	.000
	同行人數	.000	.000	.310	.	.002
	服務期望指標	.000	.000	.000	.002	.
N	消費金額	33	33	33	33	33
	服務滿意度指標	33	33	33	33	33
	停留天數	33	33	33	33	33
	同行人數	33	33	33	33	33
	服務期望指標	33	33	33	33	33

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	服務期望指標, 同行人數, 服務滿意度指標, 停留天數[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: 消費金額

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.956[a]	.914	.901	16.4829	.914	73.945	4	28	.000	2.269

a. Predictors: (Constant), 服務期望指標, 同行人數, 服務滿意度指標, 停留天數

b. Dependent Variable: 消費金額

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	80358.565	4	20089.641	73.945	.000[a]
	Residual	7607.177	28	271.685		
	Total	87965.742	32			

a. Predictors: (Constant), 服務期望指標, 同行人數, 服務滿意度指標, 停留天數

b. Dependent Variable: 消費金額

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-26.178	17.998		-1.454	.157	-63.046	10.690		
	服務滿意度指標	8.092	1.077	.729	7.516	.000	5.887	10.298	.328	3.050
	停留天數	-32.228	10.758	-.311	-2.996	.006	-54.265	-10.190	.287	3.486
	同行人數	-6.023	7.077	-.072	-.851	.402	-20.520	8.474	.432	2.313
	服務期望指標	53.759	12.945	.530	4.153	.000	27.242	80.275	.190	5.268

a. Dependent Variable: 消費金額

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions				
				(Constant)	服務滿意度指標	停留天數	同行人數	服務期望指標
1	1	4.867	1.000	.00	.00	.00	.00	.00
	2	7.557E-02	8.025	.05	.00	.23	.05	.00
	3	4.024E-02	10.998	.24	.37	.04	.00	.00
	4	1.051E-02	21.519	.66	.62	.00	.44	.15
	5	6.546E-03	27.268	.06	.01	.72	.50	.85

a. Dependent Variable: 消費金額

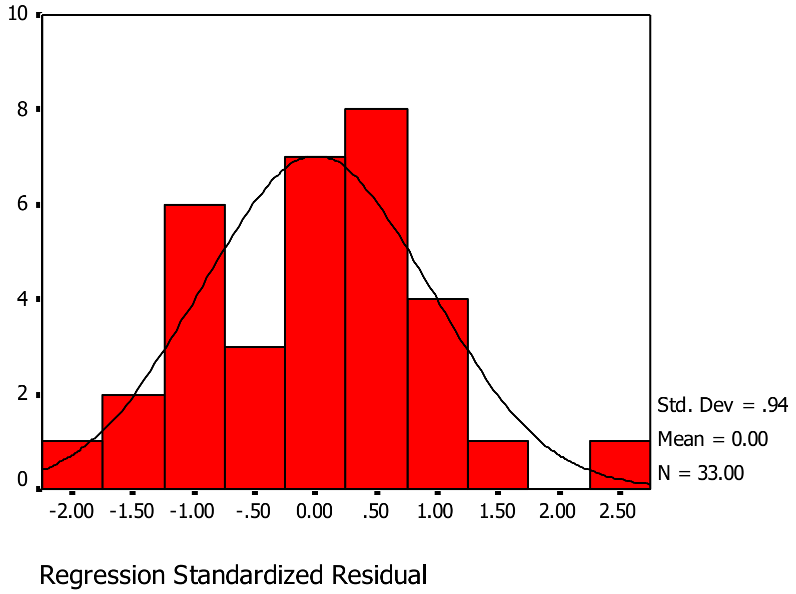
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N	
Predicted Value	31.8591	248.8520	154.5152	50.1119	33	
Residual	-36.5432	37.1409	2.670E-14	15.4183	33	
Std. Predicted Value	-2.448	1.883	.000	1.000	33	
Std. Residual	-2.217	2.253	.000	.935	33	

a. Dependent Variable: 消費金額

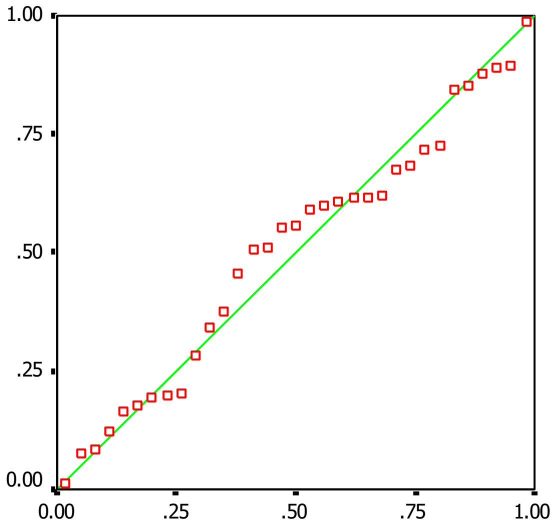
Histogram

Dependent Variable: 消費金額



Normal P-P Plot of Regression Standard

Dependent Variable: 消費金額



Observed Cum Prob

Scatterplot

Dependent Variable: 消費金額

消費金額

複迴歸分析(逐步迴歸法)SPSS 操作方法

1. Analyze/Statistics(統計分析) → Regression(迴歸分析) → Linear…(線性…)，即會出現 Linear Regression(線性迴歸)對話視窗。
2. 將欲進行迴歸分析的依變數：消費金額(cost_nr)自左邊的方塊中，點選進入右上角的 Dependent: (依變數)下方的方塊中。

3.將欲進行迴歸分析的自變數(預測變數、自變項)：服務滿意度指標(qulit_nr)、停留天數(day_n)、同行人數(p_number)和服務期望指標(ex_nr)自左邊的方塊中，點選進入右邊的 Independent(s): (自變數)下方的方塊中。

4.在 Independent(s): (自變數)下方 Methods: 下拉式選項中，選取「Stepwise」(逐步迴歸法)。

5.接著按下方的 Statistics…(統計量)鈕，即會出現「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗。

6.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗中，左上角 Regression Coefficients(迴歸係數)方塊中，勾選□Estimates(估計值)和□Confidence intervals(信賴區間)選項。

7.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗中，勾選右邊的□Model fit(迴歸模式適合度檢定)、□R squared change(解釋量的改變量)、□Descriptive(統計量)、□Collinearity diagnostics(共線性診斷)等選項。

8.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗下面的 Residuals 方塊中，勾選□Durbin-Watson(顯示 Durbin-Watson 檢定統計量，標準化、非標準化之殘差和預測值之摘要統計量)選項。

9.在「Linear Regression: Statistics」(線性迴歸：統計量)次對話視窗，按 Continue 鈕，即會回到 Linear Regression(線性迴歸)對話視窗。

10.在Linear Regression對話視窗中，點選下面的Plots…按鈕，會出現Linear Regression: Plots 次對話視窗。

11.在 Linear Regression: Plots 次對話視窗中，左邊方塊內有數項資料名稱，其意義如下(*：表示暫時性的殘差變數)：

- 「DEPENDNT」: 為依變數
- 「*ZPRED」: 標準化的迴歸預測值
- 「*ZRESID」: 標準化的殘差
- 「*DRESID」: 刪除型的殘差(deleted residual)
- 「*ADJPRED」: 調整的迴歸預測值(adjusted predicted values)
- 「*SRESID」: studentized 殘差
- 「*SDRESID」: studentized 刪除型的殘差

12.在 Linear Regression: Plots 次對話視窗中，點選 「*ZRESID」(標準化的殘差)進入 Y: (Y 軸)右邊的方塊，點選「DEPENDNT」(依變數)進入 X: (X 軸)右邊的方塊。

13.在 Linear Regression: Plots 次對話視窗下面的 Standardized Residuals Plots(標準化殘差圖)方塊中，勾選□Histogram(顯示標準化殘差值的次數分配圖，同時產生一常態分配曲線)和□Normal probability plot[顯示標準化殘差值的常態機率(P-P)圖]選項。

14.在 Linear Regression: Plots 次對話視窗右上角，按 Continues 鈕，回到 Linear Regression 對話視窗。

15.按 OK(確定)鈕，以執行複迴歸分析程序。

16.獲得以下複迴歸分析成果。

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	服務滿意度指標	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	服務期望指標	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	停留天數	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: 消費金額

選取變數的順序，進入模式的變數標準是F的機率≦0.05者，移除模式的變數標準是F的機率≧0.10者。

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.926[a]	.857	.853	20.1138	.857	186.433	1	31	.000	
2	.940[b]	.884	.876	18.4790	.026	6.728	1	30	.015	
3	.955[c]	.911	.902	16.4043	.028	9.068	1	29	.005	2.210

a. Predictors: (Constant), 服務滿意度指標

b. Predictors: (Constant), 服務滿意度指標, 服務期望指標

c. Predictors: (Constant), 服務滿意度指標, 服務期望指標, 停留天數

d. Dependent Variable: 消費金額

R為多元相關係數(Multiple correlation coefficient)

R Square(R^2)為多元決定係數(Multiple determination coefficient)

Adjusted R Square為調整後的決定係數

$$\text{Adjusted } R^2 = 1 - [(1 - R^2) \times \frac{n-1}{n-p-1}]$$

n：樣本總人數

p：迴歸方程式的自變數的個數

若自變數的個數很多，有時就要以調整後的決定係數代替原先的決定係數，因為增加新的自變數(預測變項、自變項)後，均會使R^2變大，此時以調整後的R^2表示較佳。

以樣本的R^2估計母群參數時，常會有高估的傾向，為避免高估之偏差情形發生，應採用調整後的R^2值，因為調整後的R^2是迴歸模式中變項數與樣本大小的函數，以調整後的R^2來估計母群性質，才不會有錯誤。

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75424.261	1	75424.261	186.433	.000[a]
	Residual	12541.482	31	404.564		
	Total	87965.742	32			
2	Regression	77721.555	2	38860.777	113.803	.000[b]
	Residual	10244.188	30	341.473		
	Total	87965.742	32			
3	Regression	80161.786	3	26720.595	99.295	.000[c]
	Residual	7803.956	29	269.102		
	Total	87965.742	32			

a. Predictors: (Constant), 服務滿意度指標

b. Predictors: (Constant), 服務滿意度指標, 服務期望指標

c. Predictors: (Constant), 服務滿意度指標, 服務期望指標, 停留天數

d. Dependent Variable: 消費金額

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-5.318	12.218		-.435	.666	-30.237	19.602		
	服務滿意度指標	10.272	.752	.926	13.654	.000	8.737	11.806	1.000	1.000
2	(Constant)	-29.674	14.635		-2.028	.052	-59.563	.215		
	服務滿意度指標	8.146	1.072	.734	7.598	.000	5.956	10.335	.416	2.406
	服務期望指標	25.442	9.809	.251	2.594	.015	5.409	45.474	.416	2.406

3	(Constant)	-36.541	13.191		-2.770	.010	-63.519	-9.564		
	服務滿意度指標	7.691	.964	.693	7.982	.000	5.720	9.662	.405	2.467
	服務期望指標	49.272	11.766	.485	4.188	.000	25.207	73.336	.228	4.394
	停留天數	-27.403	9.100	-.264	-3.011	.005	-46.015	-8.792	.397	2.518

a. Dependent Variable: 消費金額

B 為原始迴歸係數，Beta 為標準化的迴歸係數

從容忍度(Tolerance)指標看數學態度和探究動機數值較低(<0.4)，表示可能有共線性問題存在。

數學態度的容忍度數值為 0.257，表示模式中其餘七個自變數(預測變數、自變項)對數學態度變項的解釋量為 74.3%[(1-0.257)×100]。解釋量愈高，代表容忍度愈小，愈有共線性問題。

Excluded Variables[d]

Model		Beta In	T	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	停留天數	-.017[a]	-.212	.834	-.039	.725	1.379	.725
	同行人數	.068[a]	.775	.444	.140	.599	1.669	.599
	服務期望指標	.251[a]	2.594	.015	.428	.416	2.406	.416
2	停留天數	-.264[b]	-3.011	.005	-.488	.397	2.518	.228
	同行人數	.061[b]	.758	.455	.139	.599	1.671	.332
3	同行人數	-.072[c]	-.851	.402	-.159	.432	2.313	.190

a. Predictors in the Model: (Constant), 服務滿意度指標

b. Predictors in the Model: (Constant), 服務滿意度指標, 服務期望指標

c. Predictors in the Model: (Constant), 服務滿意度指標, 服務期望指標, 停留天數

d. Dependent Variable: 消費金額

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions			
				(Constant)	服務滿意度指標	服務期望指標	停留天數
1	1	1.958	1.000	.02	.02		
	2	4.194E-02	6.833	.98	.98		
2	1	2.944	1.000	.01	.00	.00	
	2	4.254E-02	8.319	.65	.33	.01	
	3	1.317E-02	14.954	.34	.67	.99	
3	1	3.898	1.000	.00	.00	.00	.00
	2	5.418E-02	8.482	.38	.01	.00	.42
	3	3.986E-02	9.889	.30	.52	.00	.12
	4	8.164E-03	21.851	.31	.47	1.00	.45

a. Dependent Variable: 消費金額

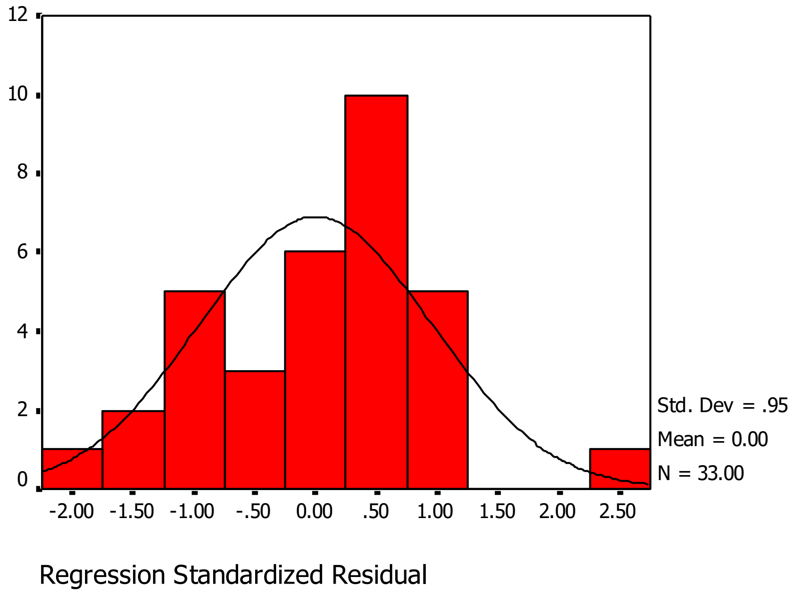
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	31.4748	248.7497	154.5152	50.0505	33
Residual	-35.3574	37.5252	1.938E-14	15.6165	33
Std. Predicted Value	-2.458	1.883	.000	1.000	33
Std. Residual	-2.155	2.288	.000	.952	33

a. Dependent Variable: 消費金額

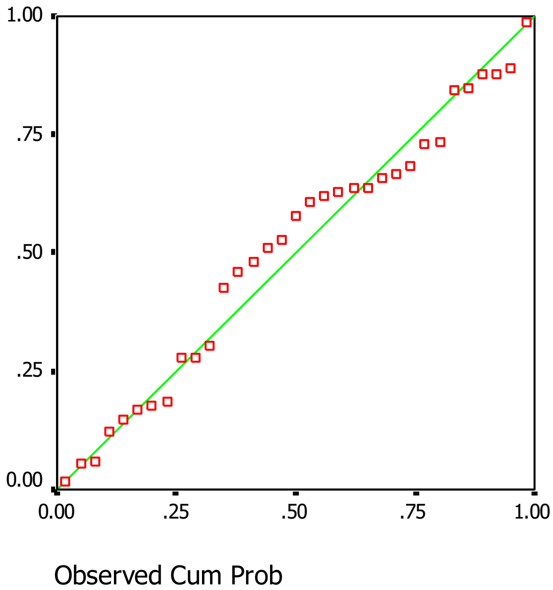
Histogram

Dependent Variable: 消費金額



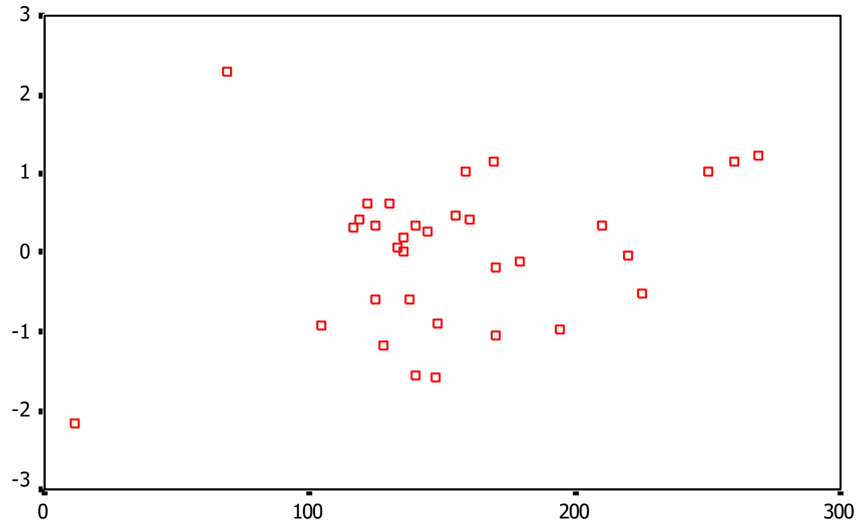
Normal P-P Plot of Regression Standard

Dependent Variable: 消費金額



Scatterplot

Dependent Variable: 消費金額



消費金額

表 數學態度、性別、自我投入、課堂焦慮、壓力、工作投入和探究動機預測數學成績之逐步複迴歸分析

選出的變數順序	多元相關係數 R	決定係數 R^2	R^2 改變量	F 值	F 值改變量	標準化迴歸係數 Beta
數學態度	0.412	0.170	0.170	60.877	60.877	0.501
SEXDV	0.445	0.198	0.029	36.711	10.587	0.155
自我投入	0.469	0.220	0.022	27.789	8.173	-0.173
課堂焦慮	0.486	0.237	0.017	22.853	6.497	0.307
壓力	0.505	0.255	0.018	20.128	7.280	-0.210
工作投入	0.519	0.269	0.014	17.986	5.676	0.176
探究動機	0.529	0.280	0.010	16.187	4.211	-0.173

13 個預測變項(自變數、自變項)預測效標變項(數學成績)時，進入迴歸分析方程式的顯著變項共有七個，多元相關係數為 0.529，其聯合解釋變異量為 0.280，即上述七個變項能聯合預測數學成績(依變數)28.0%的變異量。

就個別變項的解釋量以「數學態度」的預測力最佳，其解釋量為 17.0 %，其餘依次為「性別(sexdv)」、「自我投入」，其解釋量分別為 2.9 %、2.2 %，此前三個變項的聯合預測力達 22.0 %。

標準化迴歸方程式

數學成績 ＝0.501×數學態度 ＋0.155×SEXDV－0.173×自我投入 ＋0.307×課堂焦慮 －0.210×壓力 ＋0.176×工作投入 －0.173×探究動機

討論議題

1.非同步教學學習者之間討論議題：期中考分數複迴歸分析

【複迴歸分析】課程內容經過同步上課、上課練習與平常考之後，對【複迴歸分析】有初步認識，附加檔案位置提供一個附加 Excel 檔案，內含上課次數(蒞臨平台瀏覽數位教材)、張貼篇數(議題討論

和課程討論區)、討論次數(進入議題討論和課程討論區次數)、閱讀時數(開始上課中數位教材)、閱讀頁數(開始上課中瀏覽數位教材節點數量)、上課練習平均分數、Open 平常考平均分數、平常考 20180418(分數)【前述皆可視為自變數】和期中考 20180419(分數)【依變數】數值。第一回合請於 20180602 中午 12 點以前從「議題討論」區【張貼】標題：「分析意涵」，本文：請利用複迴歸分析解析其數值代表那些意涵，請具體敘述分享(50 個字以上)。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，檢視其他同學的回應。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳分析」，本文：哪一位同學詮釋得最好，具體說明其理由(30 個字以上)。期望透過複迴歸分析解析影響統計學(二)期中考成績的相關變數，理解複迴歸分析的運用價值。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

2.師生非同步教學討論議題：複迴歸分析應用

第一回合請於 D+3 日中午 12 點以前從「議題討論」區【張貼】標題：「應用情境」，本文：請針對第 15 章學習的【複迴歸分析】課程內容，陳述現在或未來最想運用到情境和目的。請具體標示出數個自變數(可操控變數：等距或比例尺度)和依變數(預測目標變數：等距或比例尺度)的具體名稱(變數)(20 個字以上)。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，請詳細檢視其他同學的回應內容。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳詮釋」，本文：哪一位同學詮釋得最具體與明確，請說明理由或者有哪些是值得讓自己學習之處？(20 個字以上)。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

3.師生非同步討論議題：期中考分數和期末考分數分析

附加檔案位置提供一個附加 Excel 檔案，依序內含上課次數(蒞臨平台瀏覽數位教材)、張貼篇數(議題討論和課程討論區)、討論次數(進入議題討論和課程討論區次數)、閱讀時數(開始上課中數位教材)、閱讀次數(開始上課中瀏覽數位教材節點次數)、上課練習平均分數、開書平常考平均分數、平常考 20200427(分數)、平常考 20200622(分數)、期中考 20200428(分數)和期末考 20200623(分數)數值。自行判斷哪些數值可以當自變數，那些數值可以當依變數。第一回合請於 20200629 中午 12 點以前從「議題討論」區【張貼】標題：「分析意涵」，本文：請利用簡單線性迴歸分析或複迴歸分析解析其數值代表那些意涵，請具體敘述討論分享(20 個字以上)。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，檢視其他同學的回應。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「最佳分析」，本文：哪一位同學詮釋得最佳，具體說明其理由(20 個字以上)。期望透過簡單線性迴歸分析或複迴歸分析解析影響本課程成績的相關變數，理解簡單線性迴歸分析或複迴歸分析的運用價值。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

4.學習者非同步教學討論議題：餐廳業績預測

第一回合請於 D+3 日中午 12 點以前從「議題討論」區【張貼】標題：「影響因素」，本文：請針對第 15 章學習的【複迴歸分析】課程內容與應用價值，在疫情期間三級警戒期間，餐廳無法提供內用服務，僅能夠外帶或外送營運。若自己是特定餐廳主管，最想了解哪些因素會影響餐廳營業額，可以納入複迴歸分析中，當成自變數分析使用。請具體列出標示出數個影響因素(自變數)名稱，並說明理由(10 個字以上)。

待有 30 篇第一回合【張貼】回應或第一回合【張貼】時間結束後，請詳細檢視其他同學的回應內

容。自己靜下心來，集思廣益，思考一下。第二回合【張貼】標題：「關鍵影響因素」，本文：透過歸納其他同學的分享，覺得最重要的關鍵影響因素有哪四項(可以當成自變數分析使用)？(10 個字以上)。第二回合【張貼】截止時間就是本議題在平台上的關閉時間。

重點整理

名稱	模式或方程式
確定性數學模式(Deterministic model)	$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki}$
複迴歸模式(Multiple regression model)	$y_i = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki} + \varepsilon_i$
複迴歸方程式(Multiple regression equation)	$E(y_i) = \beta_0 + \beta_1 \times x_{1i} + \beta_2 \times x_{2i} + \beta_3 \times x_{3i} + \ldots + \beta_k \times x_{ki}$
估計複迴歸方程式(Estimated multiple regression equation)	$\hat{y}_i = b_0 + b_1 \times x_{1i} + b_2 \times x_{2i} + b_3 \times x_{3i} + \ldots + b_k \times x_{ki}$

最小平方法數學法則

$$\text{Min SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_{1i} - b_2 \times x_{2i} - b_3 \times x_{3i} - \cdots - b_k \times x_{ki})^2$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 \times \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}$$

$$b_2 = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)\times(y_i-\bar{y}) - \sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\times\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(y_i-\bar{y})}{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 \times \sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2}$$

$$b_0 = \bar{y}_i - b_1 \times \bar{x}_1 - b_2 \times \bar{x}_2$$

複判定係數、多元判定係數(multiple coefficient of determination)或判定係數(coefficient of determination)即是<u>迴歸造成的平方和(sum of squares due to regression, SSR)</u>佔<u>總平方和(sum of squares total, SST)</u>的比例，常使用 R^2 或 r^2 符號代表。複判定係數代表迴歸方程式可以解釋(說明)變數 y 變異量的比例。R 稱為多元相關係數(multiple correlation coefficient)。

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$R = \sqrt{\frac{\text{SSR}}{\text{SST}}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \sqrt{1 - \frac{\text{SSE}}{\text{SST}}} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

調整判定係數、調整複判定係數或調整後的決定係數(adjusted coefficient of multiple determination; adjusted multiple coefficient of determination; adjusted R square, R_a^2, $\bar{R}^2$, or adjusted R^2)為評量複迴歸方程式的解釋能力。

$$R_a^2 = \bar{R}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1} = 1 - \frac{\text{SSE}}{\text{SST}} \times \frac{n-1}{n-k-1} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \times \frac{n-1}{n-k-1}$$

複迴歸變異數分析(anova)表

變異來源(source of variation)	平方和(sum of square)	自由度(degrees of freedom)	均方(mean square)	F
迴歸項(regression)	$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	k	$\text{MSR} = \frac{SSR}{k}$	$F = \frac{\text{MSR}}{\text{MSE}}$
誤差項(隨機項)(error)	$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-k-1$	$\text{MSE} = \frac{\text{SSE}}{n-k-1}$	
總變異(total)	$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$		

利用 F 值檢定的程序

A.設定顯著水準 α

B.虛無假設(null hypothesis) H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$

C.對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \ldots,$ or $\beta_k \neq 0$

D.計算檢定統計值—F 值

$$檢定統計值\ F = \frac{\text{MSR}}{\text{MSE}}$$

E.若檢定統計值 $F \le$ 臨界值 $F_{\alpha,k,n-k-1}$，接受虛無假設 H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$。

F.若檢定統計值 $F >$ 臨界值 $F_{\alpha,k,n-k-1}$，拒絕虛無假設 H_0: $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$，接受對立假設 H_1: $\beta_1 \neq 0$, $\beta_2 \neq 0$, $\beta_3 \neq 0, \ldots$, or $\beta_k \neq 0$。

依變數樣本變異數$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^n (y_i-\hat{y}_i)^2}{n-3} = \frac{\sum_{i=1}^n (y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-3} = \frac{\sum_{i=1}^n (y_i-\bar{y})^2 - b_1\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(y_i-\bar{y}) - b_2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)\times(y_i-\bar{y})}{n-3}$

運用依變數樣本變異數S_y^2估算依變數母體變異數σ_y^2時，相對應的 b_0、b_1 與 b_2 估算的變異數依序為：

$$S_{b_0}^2 = \left[\frac{\bar{x}_1^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 + \bar{x}_2^2\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 - 2\times\bar{x}_1\times\bar{x}_2\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} + \frac{1}{n}\right] \times S_y^2$$

$$S_{b_1}^2 = \frac{\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} \times S_y^2$$

$$S_{b_2}^2 = \frac{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} \times S_y^2$$

利用 t 值檢定的程序

A.設定顯著水準 α

B.虛無假設(null hypothesis) H_0: $\beta_i = 0$

C.對立假設(alternative hypothesis) H_1: $\beta_i \neq 0$

D.計算檢定統計值—t 值

$$檢定統計值\ t = \frac{b_i}{S_{b_i}}$$

E.若$-t_{\frac{\alpha}{2},n-k-1,}$(臨界值) $<$ 檢定統計值 $t <$ $t_{\frac{\alpha}{2},n-k-1,}$(臨界值)，接受虛無假設 H_0: $\beta_i = 0$。

F.若檢定統計值 $t < -t_{\frac{\alpha}{2},n-k-1,}$(臨界值)或檢定統計值 $t >$ $t_{\frac{\alpha}{2},n-k-1,}$(臨界值)，拒絕虛無假設 H_0: $\beta_i = 0$，接受對立假設(alternative hypothesis) H_1: $\beta_i \neq 0$。

若依變數母體變異數 σ^2 未知時，在特定自變數數值(x_{10} 和 x_{20})組合下，依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_{10},x_{20})$的信賴區間：

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-3} \times S_{\hat{y}_0}$$

$$S_{\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 + (x_{20}-\bar{x}_2)^2\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 - 2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} + \frac{1}{n}\right]$$

其中$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^n (y_i-\hat{y}_i)^2}{n-3} = \frac{\sum_{i=1}^n (y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-3}$

若依變數母體變異數 σ^2 未知時，依變數 y_0 的信賴區間

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-3} \times S_{y_0-\hat{y}_0}$$

$$\sigma_{e_0}^2 = S_{y_0-\hat{y}_0}^2 = S_{y|x_1x_2}^2 \times \left[\frac{(x_{10}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 + (x_{20}-\bar{x}_2)^2\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2 - 2\times(x_{10}-\bar{x}_1)\times(x_{20}-\bar{x}_2)\times\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)}{\sum_{i=1}^n (x_{1i}-\bar{x}_1)^2\times\sum_{i=1}^n (x_{2i}-\bar{x}_2)^2 - \left[\sum_{i=1}^n (x_{1i}-\bar{x}_1)\times(x_{2i}-\bar{x}_2)\right]^2} + \frac{1}{n} + 1\right]$$

其中$S_y^2 = S_{y|x_1x_2}^2 = \text{MSE} = \frac{\text{SSE}}{n-k-1} = \frac{\sum_{i=1}^n (y_i-\hat{y}_i)^2}{n-3} = \frac{\sum_{i=1}^n (y_i-b_0-b_1\times x_{1i}-b_2\times x_{2i})^2}{n-3}$

關鍵詞彙解釋

複判定係數(multiple coefficient of determination)

複判定係數、多元判定係數(multiple coefficient of determination)或判定係數(coefficient of determination)即是迴歸造成的平方和(sum of squares due to regression, SSR)佔總平方和(sum of squares total, SST)的比例，常使用 R^2 或 r^2 符號代表。複判定係數代表迴歸方程式可以解釋(說明)依變數 y 變異量的比例。R 稱為多元相關係數或複相關係數(multiple correlation coefficient)

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$R = \sqrt{\frac{\text{SSR}}{\text{SST}}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \sqrt{1 - \frac{\text{SSE}}{\text{SST}}} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

調整判定係數(adjusted coefficient of multiple determination)

利用複判定係數 R^2 評估複迴歸方程式的適合度時，當樣本數量較少或自變數較多，會使自由度降低，導致對複判定係數 R^2 有高估，自變數對可解釋之變異量(SSR)的影響。若在複迴歸模式中加入一些與模式無關的自變數，複判定係數 R^2 會增加，因此，無法客觀的代表複迴歸模式的解釋能力。故，建議使用調整判定係數、調整複判定係數或調整後的決定係數(adjusted coefficient of multiple determination; adjusted multiple coefficient of determination; adjusted R square, R_a^2, $\bar{R}^2$, or adjusted R^2)為評量複迴歸方程式的解釋能力。

$$R_a^2 = \bar{R}^2 = 1-(1-R^2)\times\frac{n-1}{n-k-1} = 1 - \frac{\text{SSE}}{\text{SST}}\times\frac{n-1}{n-k-1} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}\times\frac{n-1}{n-k-1}$$

其中 $\hat{y}_i$ = 依變數第 i 個觀測值的估計值 n = 觀測值數量 k = 自變數數量(個數)

從上述調整判定係數公式中，發現$\frac{n-1}{n-k-1} > 1$，因此複判定係數 R^2 > 調整判定係數R_a^2。當樣本數量 n 和自變數數量 k 在特定的比值下，可能出現$\frac{\text{SSE}}{\text{SST}} > \frac{n-k-1}{n-1}$，調整判定係數$R_a^2$會出現負值($R_a^2 < 0$)。當調整判定係數(adjusted R square)是負值時，一般統計軟體會出現調整判定係數(adjusted R square)等於 0 的結果。

多重共線性(multicollinearity)

在複迴歸分析中，自變數之間可能並非相互獨立，而具有某種程度的相關性存在。利用自變數樣本觀測值之間的相關係數$r_{x_i x_j}$，代表兩自變數之間的相關程度。當兩自變數之間的相關係數$|r_{x_i x_j}|$大於 0.7 時，多重共線性的嚴重性就會很高，影響迴歸係數估計值的估算。

簡單線性迴歸分析(simple linear regression analysis)

若欲分析的變數只有一個自變數(自變項)和另一個依變數(依變項)時，兩者的關係趨近於比例關係(線性關係、直線關係)時，歸類為簡單線性迴歸分析。

多元迴歸分析(multiple regression analysis)

在迴歸程序中若有超過兩個(含兩個)的自變數與一個依變數時，歸類為**多元迴歸分析**或**複迴歸分析**(multiple regression analysis)。

多變量迴歸分析(multi-variable regression analysis)

在迴歸關係中，若有多個自變數預測數個依變數，稱為**多變量迴歸分析**(multi-variable regression analysis)。