十四、簡單線性迴歸分析

Chapter 14 Simple Linear Regression Analysis

目錄

學習目標

知識(認知)

1.可以清楚陳述簡單線性迴歸分析的意涵。

2.可以清楚陳述在迴歸分析中判定係數的意涵。

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，簡單線性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簡單線性迴歸分析程序。

態度(情意)

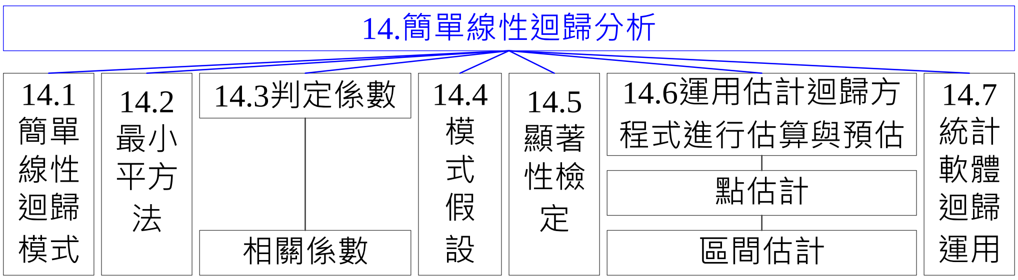
1.意識到在日常生活或未來工作環境中，簡單線性迴歸分析的重要性。

2.在各種情境下，依循簡單線性迴歸分析的程序，接受統計推論所傳達的意涵。

在迴歸分析程序中，將欲預測、推估或估計的變數視為依變數、因變數、應變數(response variable; regressand)、相依變數(dependent variable)或被解釋變數(explained variable)，欲操作、操縱的變數視為自變數、獨立變數(independent variable)、預測變數或解釋變數(explanatory variable; regressor)。迴歸分析的目的是期望瞭解自變數的數值或改變量對於依變數產生(影響程度)的數值或改變量。

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，歸類為多元迴歸分析或複迴歸分析(multiple 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)。

相關分析(correlation analysis)主要是敘述兩變數之間的關係之方向和關係之強度(程度之大小)的統計方法，兩個變數之間皆是屬於隨機變數。迴歸分析(regression analysis)主要是欲分析一個或一個以上自變數與依變數之間的影響程度，期望藉由瞭解自變數對依變數的影響程度，而達到預測依變數的目的。



章節結構圖

14.1 簡單線性迴歸模式

簡單線性迴歸模式(Simple linear regression model)是探討一個自變數和另一個依變數之間關係的統計法。自變數與依變數之間的關係可以分為**正向關係**、**負向關係**和**沒有(無)關係**三種。自變數與依變數之間的關係又可以分為**線性**(linear)和**非線性**(nonlinear)兩種。

設自變數為 X，例如餐廳的行銷費用；依變數為 Y，例如餐廳的營業額，兩者的關係為直線關係，可表示為確定模式或確定性數學模式(deterministic model)：

$$\boldsymbol{y_i = \beta_0 + \beta_1 \times x_i}\text{，其中 } \boldsymbol{i = 1, \ldots, n}$$

其中 $y_i =$ 依變數 Y 第 i 個觀測值的實際觀測值(變量)

$x_i =$ 自變數 X 第 i 個觀測值(變量)

$\beta_0 =$ 迴歸模式的參數(parameter)，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

$\beta_1 =$ 迴歸模式的參數(parameter)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

$n =$ 觀測值數量

代表依變數 Y 僅受到自變數 X 的影響，不受其他因素影響。故，只要確定自變數 X 數值，即可獲得依變數 Y 數值，自變數 X 與依變數 Y 之間有一對一的對應數值。但是，實務面，營業額還是會受到其他因素

的影響，如競爭對手的行銷費用、競爭對手的強弱、餐廳所在位置、餐廳目標消費者、…等因素。再考慮上述其他隨機(影響)因素後，可將確定模式(deterministic model)修正為機率模式(probabilistic model)、統計(機率)模式(statistical model)、迴歸模式、迴歸模型(regression model)或簡單線性迴歸模式(simple linear regression model)：

$$\boldsymbol{y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i}\text{，其中 } \boldsymbol{i = 1,\ldots, n}$$

其中 $y_i =$ 依變數 Y 第 i 個觀測值的實際觀測值(變量)

$x_i =$ 自變數 X 第 i 個觀測值(變量)

$\beta_0 =$ 迴歸模式的參數(parameter)，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

$\beta_1 =$ 迴歸模式的參數(parameter)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

$\varepsilon_i =$ 第 i 個觀測值的隨機變數，屬於隨機誤差(random error)，讀音 epsilon。有時亦可使用 e_i 符號代表。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 變異性或變動性。

$n =$ 觀測值數量

在簡單線性迴歸方程式中，假設誤差項 ε 的平均值或期望值為 0。因此，在迴歸模式中依變數 Y 的期望值 $E(y_i) = \beta_0 + \beta_1 \times x_i$，故依變數 Y 的期望值與自變數 X 屬於線性關係。敘述依變數 Y 的期望值 $E(y_i)$與自變數 X 關係的方程式，稱為迴歸方程式(regression equation)或預測方程式(prediction equation)。

簡單線性迴歸方程式(simple linear regression equation)

$$\mathbf{E}(y_i) = \mu_{y_i|x_i} = \boldsymbol{\beta_0 + \beta_1 \times x_i}\text{，其中 } \boldsymbol{i = 1,\ldots, n}$$

若上述迴歸方程式中參數 β_0 和 β_1 值已知時，可利用已知的自變數 x_i(變量)計算獲得依變數 y_i(變量)。但是，實際上參數 β_0 和 β_1 值未知，必須利用樣本資料進行估算。假設利用樣本統計值 b_0 和 b_1(有時使用$\hat{\beta}_0$和$\hat{\beta}_1$符號)作為迴歸參數 β_0 和 β_1 值的估計值，可獲得估計迴歸方程式(estimated regression equation)。

估計簡單線性迴歸方程式、樣本迴歸方程式、估計迴歸線(estimated regression line)或估計迴歸方程式(estimated regression equation)

$$\boldsymbol{\hat{y}_i = b_0 + b_1 \times x_i}\text{，其中 } \boldsymbol{i = 1,\ldots, n}$$

其中$\hat{y}_i$ = 在自變數為 x_i 時依變數 y_i 的估計值；依變數第 i 個觀測值的估計值或預測值(fitted value)。y_i'

$x_i =$ 自變數 X 第 i 個觀測值(變量)

$b_0 =$ 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 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-∞~+∞。

$b_1 =$ 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

名稱	模式或方程式
確定性數學模式(deterministic model)	$y_i = \beta_0 + \beta_1 \times x_i$
簡單線性迴歸模式(simple linear regression model)	$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$
簡單線性迴歸方程式(simple linear regression equation)	$E(y_i) = \beta_0 + \beta_1 \times x_i$
估計迴歸方程式(estimated regression equation)	$\hat{y}_i = b_0 + b_1 \times x_i$

樣本統計值 b_0 和 b_1(估計值)估算法可分為最小平方法(least squares method)和最大概似估計法(maximum likelihood method, ML)。

14.2 最小平方法

利用樣本資料中自變數 x_i 和依變數 y_i(實際觀測值)的對應數值，並使用自變數 x_i、截距 b_0 和斜率 b_1 推算依變數 y_i 的估計值$\hat{y}_i$，使得依變數 y_i 和其估計值$\hat{y}_i$的差(距)之平方和(sum square error, SSE)為最小數值，此為最小平方法(Least squares method)或普通最小平方法(ordinary least squares method, OLS)的特性。

最小平方法數學法則

$$\text{Min SSE} = \min \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i)^2$$

其中 y_i = 依變數 Y 第 i 個觀測值的實際觀測值(變量)

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

x_i = 自變數 X 第 i 個觀測值(變量)

b_0 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_0 的估計值(estimator)，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_1 的估計值(estimator)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

利用微分方式獲得估計迴歸方程式的斜率(slope)b_1 和截距(intercept)b_0 運算公式

$$\text{斜率 } b_1 = \frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{n \times \sum_{i=1}^{n}(x_i \times y_i) - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}[(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\frac{\sum_{i=1}^{n}[(x_i - \bar{x}) \times (y_i - \bar{y})]}{n-1}}{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \frac{S_{xy}}{S_x^2} = \frac{\text{自變數與依變數之共變異數}}{\text{自變數之變異數}}$$

$$\text{截距 } b_0 = \bar{y} - b_1 \times \bar{x} = \frac{\sum_{i=1}^{n} y_i \times \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n}(x_i \times y_i)}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

其中 y_i = 依變數 Y 第 i 個觀測值的實際觀測值(變量)

$\bar{y}$ = 依變數的樣本平均值

x_i = 自變數 X 第 i 個觀測值(變量)

$\bar{x}$ = 自變數的樣本平均值

b_0 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_0 的估計值(estimator)，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_1 的估計值(estimator)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

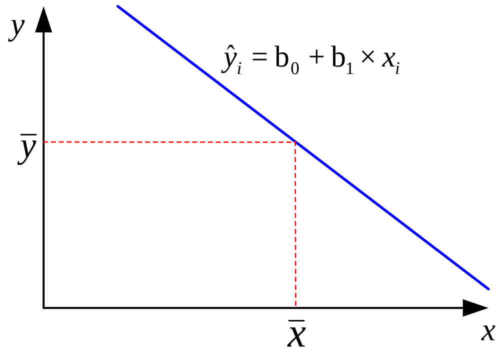
n = 觀測值數量

S_{xy} = 自變數 X 和依變數 Y 的樣本共變數

$S_x^2 = S_{xx}$ = 自變數 X 的樣本變異數

利用最小平方法估算依變數第 i 個觀測值之估計值$\hat{y}_i$的特徵

A.估計簡單線性迴歸方程式或樣本迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i$ 直線會通過自變數和依變數樣本平均值點$(\bar{x},\bar{y})$。



B.樣本殘差($\varepsilon_i = y_i - \hat{y}_i$)和為 0。$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i) = \sum_{i=1}^{n}\varepsilon_i = 0$。樣本殘差的期望值亦等於 0，$E(\varepsilon_i) = 0$。

C.樣本殘差 ε_i 與自變數 x_i 的共變數為 0，樣本殘差和自變數 X 無線性關係。$\mathrm{Cov}(x_i,\varepsilon_i) = 0 = E(x_i \times \varepsilon_i) - E(x_i) \times E(\varepsilon_i) = E(x_i \times \varepsilon_i) - E(x_i) \times 0 = E(x_i \times \varepsilon_i)$，故，$E(x_i \times \varepsilon_i) = 0$ 和$\sum_{i=1}^{n}(x_i \times \varepsilon_i) = 0$ 亦成立。

D.樣本殘差 ε_i 與依變數預估值$\hat{y}_i$的共變數為 0，樣本殘差 ε_i 與依變數預估值$\hat{y}_i$無線性關係。$\mathrm{Cov}(\hat{y}_i,\varepsilon_i) = 0 = E(\hat{y}_i \times \varepsilon_i) - E(\hat{y}_i) \times E(\varepsilon_i) = E(\hat{y}_i \times \varepsilon_i) - E(\hat{y}_i) \times 0 = E(\hat{y}_i \times \varepsilon_i)$，故，$E(\hat{y}_i \times \varepsilon_i) = 0$ 和$\sum_{i=1}^{n}(\hat{y}_i \times \varepsilon_i) = 0$ 亦成立。

範例 14.1 天空連鎖餐廳有 10 個營業點，每個營業點前一日行銷費用和前一日販售套餐數樣本資料依序列於下表。欲運用行銷費用 x_i 預測販售套餐數量 y_i，而建立估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i$。請利用最小平方法計算出估計簡單線性迴歸方程式的統計值：斜率與截距。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：

營業點 i	行銷費用 x_i	套餐數 y_i	$x_i \times y_i$	x_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x}) \times (y_i - \bar{y})$
1	150	156	23400	22500	-33	-41.2	1089	1359.6
2	160	180	28800	25600	-23	-17.2	529	395.6
3	180	190	34200	32400	-3	-7.2	9	21.6
4	160	170	27200	25600	-23	-27.2	529	625.6
5	190	198	37620	36100	7	0.8	49	5.6
6	210	250	52500	44100	27	52.8	729	1425.6
7	180	189	34020	32400	-3	-8.2	9	24.6
8	160	168	26880	25600	-23	-29.2	529	671.6
9	180	191	34380	32400	-3	-6.2	9	18.6
10	260	280	72800	67600	77	82.8	5929	6375.6
合計	1830	1972	371800	344300	0	0.0	9410	10924
平均值	183	197.2	37180	34430	0	0.0	941	1092.4

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{371800 - \frac{1830 \times 1972}{10}}{344300 - \frac{1830^2}{10}} = \frac{371800 - 360876}{344300 - 334890} = \frac{10924}{9410} = 1.1609$ 第一種公式計算方法

斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{10924}{9410} = 1.1609$ 第二種公式計算方法

截距 $b_0 = \bar{y} - b_1\times\bar{x} = 197.2 - 1.1609\times183 = 197.2 - 212.4433 = -15.2434$

行銷費用對銷售套餐數量的估計迴歸方程式 $\hat{y}_i = b_0 + b_1\times x_i = -15.2434 + 1.1609\times x_i$

答案：斜率 $b_1 = 1.1609$；截距 $b_0 = -15.2434$

練習 14.1 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1\times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用最小平方方法計算出估計簡單線性迴歸方程式的統計值：斜率與截距。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：透過表格的製作，協助理解運算過程

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$x_i\times y_i$	x_i^2	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	50	156	7800.0	2500.0	-10.8	-30.4	116.64	328.32
2	53	179	9487.0	2809.0	-7.8	-7.4	60.84	57.72
3	60	189	11340.0	3600.0	-0.8	2.6	0.64	-2.08
4	53	160	8480.0	2809.0	-7.8	-26.4	60.84	205.92
5	63	185	11655.0	3969.0	2.2	-1.4	4.84	-3.08
6	70	210	14700.0	4900.0	9.2	23.6	84.64	217.12
7	60	189	11340.0	3600.0	-0.8	2.6	0.64	-2.08
8	53	168	8904.0	2809.0	-7.8	-18.4	60.84	143.52
9	60	191	11460.0	3600.0	-0.8	4.6	0.64	-3.68
10	86	237	20382.0	7396.0	25.2	50.6	635.04	1275.12
合計	608	1864	115548.0	37992.0	0.0	0.0	1025.60	2216.80
平均值	60.8	186.4	11554.8	3799.2	0.0	0.0	102.56	221.68

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i\times y_i)-\frac{\sum_{i=1}^{n}x_i\times\sum_{i=1}^{n}y_i}{n}}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}} = \frac{115548-\frac{608\times1864}{10}}{37992-\frac{608^2}{10}} = \frac{115548-113331.2}{37992-36966.4} = \frac{2216.8}{1025.6} = 2.1615$ 第一種公式計算方法

斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{2216.80}{1025.60} = 2.1615$ 第二種公式計算方法

截距 $b_0 = \bar{y} - b_1\times\bar{x} = 186.4 - 2.1615\times60.8 = 186.4 - 131.4172 = 54.9828$

訓練費用對銷售咖啡杯數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1\times x_i = 54.9828 + 2.1615\times x_i$

答案：斜率 $b_1 = 2.1615$；截距 $b_0 = 54.9828$

練習 14.2 吻別連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天營業金額(單位：新台幣百元)依序列於下表。欲運用訓練費用 x_i 預測營業金額 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1\times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用最小平方方法計算出估計簡單線性迴歸方程式的統計值：斜率與截距。

營業點 i	訓練費用 x_i	營業金額 y_i
1	51	165
2	54	176
3	60	189
4	53	161
5	63	186
6	71	211
7	60	189
8	53	168
9	62	193
10	84	224

題解：透過表格的製作，協助理解運算過程

營業點 i	訓練費用 x_i	營業金額 y_i	$x_i \times y_i$	x_i^2	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})\times(y_i-\bar{y})$
1	51	165	8415.0	2601.0	-10.1	-21.2	102.01	214.12
2	54	176	9504.0	2916.0	-7.1	-10.2	50.41	72.42
3	60	189	11340.0	3600.0	-1.1	2.8	1.21	-3.08
4	53	161	8533.0	2809.0	-8.1	-25.2	65.61	204.12
5	63	186	11718.0	3969.0	1.9	-0.2	3.61	-0.38
6	71	211	14981.0	5041.0	9.9	24.8	98.01	245.52
7	60	189	11340.0	3600.0	-1.1	2.8	1.21	-3.08
8	53	168	8904.0	2809.0	-8.1	-18.2	65.61	147.42
9	62	193	11966.0	3844.0	0.9	6.8	0.81	6.12
10	84	224	18816.0	7056.0	22.9	37.8	524.41	865.62
合計	611	1862	115517.0	38245.0	0.0	0.0	912.90	1748.80
平均值	61.1	186.2	11551.7	3824.5	0.0	0.0	91.29	174.88

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{115517 - \frac{611 \times 1862}{10}}{38245 - \frac{611^2}{10}} = \frac{115517 - 113768}{38245 - 37332.1} = \frac{1748.8}{912.9} = 1.9157$ 第一種公式計算方法

斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{1748.80}{912.9} = 1.9157$ 第二種公式計算方法

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = 186.2 - 1.9157 \times 61.1 = 186.2 - 117.0464 = 69.1536$

訓練費用對營業金額的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = 69.1536 + 1.9157 \times x_i$

答案：斜率 $b_1 = 1.9157$；截距 $b_0 = 69.1536$

練習 14.3 小魚連鎖咖啡館蒐集 12 個營業點資料，以決定訓練費用 x_i(單位：新台幣百元)和銷售金額 y_i(單位：新台幣百元)的關係。獲得$\sum_{i=1}^{12} x_i = 80$、$\sum_{i=1}^{12} y_i = 1080$、$\sum_{i=1}^{12} x_i^2 = 680$、$\sum_{i=1}^{12} y_i^2 = 125080$和$\sum_{i=1}^{12} x_i \times y_i = 8510$。(A)請利用最小平方法計算出估計簡單線性迴歸方程式的統計值：斜率。(B)相關係數。(C)樣本殘差變異數(residual variance)MSE。

範例 14.2 小魚連鎖咖啡館蒐集 12 個營業點資料，利用訓練費用 x_i(單位：新台幣百元)預測銷售金額 y_i(單位：新台幣百元)建立迴歸關係。獲得估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i = 100.00 + 10.00 \times x_i$。若將原來訓練費用資料由單位是新台幣百元，調整為新台幣元；同時，銷售金額單位不變，請計算調整後的估計迴歸方程式。

題解：原來訓練費用資料由單位是新台幣百元，調整為新台幣元。若自變數 x_i 原來是 10(單位：新台幣百元)，調整單位後需要標示為 1000(單位：新台幣元)，數值才會等值。假設訓練費用調整後的變數數值 x_{in}(單位：新台幣元) = 100 × x_i(原始數值)，其調整後的平均值數值$\bar{x}_n$(單位：新台幣元) = 100 × $\bar{x}$(原始數值)。

原始估計迴歸方程式$\hat{y}_i$(單位：新台幣百元) = $b_0 + b_1 \times x_i$(單位：新台幣百元) = $100.00 + 10.00 \times x_i$

原始的斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 10.00$

調整後的斜率 $b_{1n} = \frac{\sum_{i=1}^{n}[(x_{in}-\bar{x}_n)\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_{in}-\bar{x}_n)^2} = \frac{100\times\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{100^2\times\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{100}{10000} \times \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{1}{100} \times b_1 = \frac{1}{100} \times 10.00 = 0.1000$

原始的截距 $b_0 = \bar{y} - b_1 \times \bar{x} = \bar{y} - 10.00 \times \bar{x} = 100.00$

調整後的截距 $b_{0n} = \bar{y} - b_{1n} \times \bar{x}_n = \bar{y} - \frac{1}{100} \times b_1 \times 100 \times \bar{x} = \bar{y} - b_1 \times \bar{x} = 100.00$

調整後的估計迴歸方程式$\hat{y}_{in}$(單位：新台幣百元) = $b_{0n} + b_{1n} \times x_{in}$(單位：新台幣元) = $100.00 + 0.1000 \times x_{in}$

答案：$\hat{y}_{in} = b_{0n} + b_{1n} \times x_{in} = 100.00 + 0.1000 \times x_{in}$

範例 14.3 小魚連鎖咖啡館蒐集 12 個營業點資料，利用訓練費用 x_i(單位：新台幣百元)預測銷售金額 y_i(單位：新台幣百元)建立迴歸關係。獲得估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i = 100.00 + 10.00 \times x_i$。若將原來銷售金額資料由單位是新台幣百元，調整為新台幣元；同時，訓練費用單位不變，請計算調整後的估計迴歸方程式。

題解：原來銷售金額資料由單位是新台幣百元，調整為新台幣元。若依變數 y_i 原來是 10(單位：新台幣百元)，調整單位後需要標示為 1000(單位：新台幣元)，數值才會等值。假設銷售金額調整後的變數數值 y_{in}(單位：新台幣元) = 100 × y_i(原始數值)，其調整後的平均值數值$\bar{y}_n$(單位：新台幣元) = 100 × $\bar{y}$(原始數值)。

原始估計迴歸方程式$\hat{y}_i$(單位：新台幣百元) = $b_0 + b_1 \times x_i$(單位：新台幣百元) = $100.00 + 10.00 \times x_i$

原始的斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 10.00$

調整後的斜率 $b_{1n} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_{in}-\bar{y}_n)]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{100\times\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 100 \times \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = 100 \times b_1 = 100 \times 10.00 = 1000.00$

原始的截距 $b_0 = \bar{y} - b_1 \times \bar{x} = \bar{y} - 10.00 \times \bar{x} = 100.00$

調整後的截距 $b_{0n} = \bar{y}_n - b_{1n} \times \bar{x} = 100 \times \bar{y} - 100 \times b_1 \times \bar{x} = 100 \times (\bar{y} - b_1 \times \bar{x}) = 100 \times 100 = 10000.00$

調整後的估計迴歸方程式$\hat{y}_{in}$(單位：新台幣元) = $b_{0n} + b_{1n} \times x_{in}$(單位：新台幣百元) = $10000.00 + 1000.00 \times x_{in}$

答案：$\hat{y}_{in} = b_{0n} + b_{1n} \times x_{in} = 10000.00 + 1000.00 \times x_{in}$

14.3 最大概似估計法(Maximum likelihood method, ML)【選擇教材】

最大概似估計法假設依變數 y_i 屬於常態分布，依變數 y_i 的平均值為 $b_0 + b_1 \times x_i$，依變數 y_i 的變異數 σ^2，可表示為 $y_i \sim N(b_0 + b_1 \times x_i, \sigma^2)$。

概似函數

$$L(\alpha,\beta,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\times\sigma}\times e^{-\frac{(y_i-b_0-b_1\times x_i)^2}{2\times\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}\times\sigma}\right)^n\times e^{-\frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{2\times\sigma^2}}$$

利用微分方式獲得估計迴歸方程式的斜率(slope)b_1、截距(intercept)b_0和變異數估計值$\hat{\sigma}^2$

斜率 $b_1=\dfrac{\sum_{i=1}^{n}(x_i\times y_i)-\frac{\sum_{i=1}^{n}x_i\times\sum_{i=1}^{n}y_i}{n}}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}=\dfrac{n\times\sum_{i=1}^{n}(x_i\times y_i)-\sum_{i=1}^{n}x_i\times\sum_{i=1}^{n}y_i}{n\times\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}=\dfrac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\dfrac{\frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{n-1}}{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\dfrac{S_{xy}}{S_x^2}$

截距 $b_0=\dfrac{\sum_{i=1}^{n}y_i\times\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\times\sum_{i=1}^{n}(x_i\times y_i)}{n\times\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}$

變異數估計值$\hat{\sigma}^2=\dfrac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{n}$

其中 y_i = 依變數第 i 個觀測值的實際觀測值

$\bar{y}$ = 依變數的樣本平均值

x_i = 自變數第 i 個觀測值

$\bar{x}$ = 自變數的樣本平均值

b_0 = 迴歸模式 $E(y_i)=\beta_0+\beta_1\times x_i$ 中，參數(parameter) β_0 的估計值(estimator)，截距(intercept)、常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 迴歸模式 $E(y_i)=\beta_0+\beta_1\times x_i$ 中，參數(parameter) β_1 的估計值(estimator)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

n = 觀測值數量

S_{xy} = 自變數 x 和依變數 y 的樣本共變數

S_x^2 = 自變數 x 的樣本變異數

若**依變數** y_i 和**隨機誤差** ε 屬於**常態分布**時，估計迴歸方程式的斜率(slope)b_1和截距(intercept)b_0與最小平方法的方程式完全相同。

範例 14.4 天空連鎖餐廳~~有數個營業點，每個營業點個別的平均每日行銷費用和平均每日販售套餐數依序列於下表。欲運用行銷費用預測販售套餐數量，而建立迴歸方程式。請利用**最大概似估計法**計算出估計簡單線性迴歸方程式的統計值：斜率與截距~~。

14.3 判定係數

估計迴歸方程式之適合度(goodness of fit)的測量值。在依變數 Y 的總變異中，可以由自變數 X 之變異解釋的部分。

透過樣本資料估算獲得估計迴歸方程式的斜率(slope)b_1和截距(intercept)b_0，在第 i 個樣本資料中，依變數的實際**觀測值** y_i 和**估計值**$\hat{y}_i$之間的差距，稱為第 i 項**殘差**(residual) ε_i。代表實際觀測值 y_i 與預估值$\hat{y}_i$之間的誤差量。在最小平方法中期望獲得殘差值之平方和是最小數值者，此數值屬於誤差造成的平方和、誤差項平方和、殘差平方和(sum of squares due to error, SSE)、不可解釋的變異或隨機變異。

$$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2=\sum_{i=1}^{n}y_i^2-b_0\times\sum_{i=1}^{n}y_i-b_1\times\sum_{i=1}^{n}x_i\times y_i$$

假設在未知母體總數 N 時，欲估計依變數 y_i。若沒有其他資料時，使用依變數 y_i 的樣本平均值$\bar{y}$作為第 i 項(任何)依變數 y_i 的估計值。$y_i-\bar{y}$的差距，即是使用樣本平均值$\bar{y}$作為第 i 項(任何)依變數 y_i 的估計值，所產生的差距。其對應的平方和稱為總平方和或總變異(sum of squares total, SST)。

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

為評量迴歸所產生依變數估計值$\hat{y}_i$和樣本平均值$\bar{y}$之間的差距$\hat{y}_i - \bar{y}$，其對應的平方和稱為迴歸造成的平方和、迴歸項平方和(sum of squares due to regression, SSR)或可解釋的變異。

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

依變數 y_i 的總差異 $y_i - \bar{y}$

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + y_i - \hat{y}_i$$

總差異 = 迴歸差異(可解釋的差異) + 誤差差異(誤差差異、不可解釋的差異)

依變數 y_i 的總變異$\sum_{i=1}^{n}(y_i - \bar{y})^2$

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

總平方和(總變異) = 迴歸造成的平方和(迴歸項平方和、可解釋的變異) + 誤差造成的平方和(誤差項平方和、不可解釋的變異、隨機變異)

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

判定係數或決定係數(coefficient of determination)即是迴歸造成的平方和(迴歸項平方和、可解釋的變異)佔總平方和(總變異)的比例，常使用 R^2 或 r^2 符號代表。R^2 數值範圍 0~1，愈靠近 1 迴歸方程式的適配度愈高，可以由自變數解釋依變數的變異量愈高。

$$\text{判定係數 } R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = \frac{\text{迴歸可解釋變異量}}{\text{總變異量}} = \frac{\text{SST}-\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

判定係數可以評量迴歸方程式的適配度，亦可評量迴歸方程式的解釋能力。

SST 和 SSR 計算公式

$$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left(y_i^2 - 2 \times y_i \times \bar{y} + \bar{y}^2\right) = \sum_{i=1}^{n} y_i^2 - 2 \times \sum_{i=1}^{n} y_i \times \bar{y} + \sum_{i=1}^{n} \bar{y}^2 = \sum_{i=1}^{n} y_i^2 - 2 \times n \times \bar{y} \times \bar{y} + n \times \bar{y} \times \bar{y} = \sum_{i=1}^{n} y_i^2 - n \times \bar{y}^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \frac{\left[\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}\right]^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

範例 14.5 天空連鎖餐廳有 10 營業點，每個營業點前一日行銷費用和前一日販售套餐數列於下表。請計算出估計簡單線性迴歸方程式的判定係數(coefficient of determination) R^2。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：行銷費用對銷售套餐數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$

(A)依據判定係數定義的計算方式

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	y_i^2	$x_i \times y_i$
1	150	156	158.89	-2.89	8.36	-41.20	1697.44	24336	23400
2	160	180	170.50	9.50	90.24	-17.20	295.84	32400	28800
3	180	190	193.72	-3.72	13.83	-7.20	51.84	36100	34200
4	160	170	170.50	-0.50	0.25	-27.20	739.84	28900	27200
5	190	198	205.33	-7.33	53.70	0.80	0.64	39204	37620
6	210	250	228.55	21.45	460.29	52.80	2787.84	62500	52500

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	y_i^2	$x_i\times y_i$
7	180	189	193.72	-4.72	22.27	-8.20	67.24	35721	34020
8	160	168	170.50	-2.50	6.25	-29.20	852.64	28224	26880
9	180	191	193.72	-2.72	7.39	-6.20	38.44	36481	34380
10	260	280	286.59	-6.59	43.44	82.80	6855.84	78400	72800
合計	1830	1972	1972.01	0.00	706.01	0.00	13387.60	402266	371800
平均值	183	197.2							

SSE $= \sum_{i=1}^{n} y_i^2 - b_0\times\sum_{i=1}^{n} y_i - b_1\times\sum_{i=1}^{n} x_i \times y_i = 402266 + 15.2434 \times 1972 - 1.1609 \times 371800 = 706.0085$ (第一種算法)

SSE $= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 706.01$ (第二種算法)

SST $= \sum_{i=1}^{n}(y_i - \bar{y})^2 = 13387.60$

SSR $= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 =$ SST – SSE = 13387.60 – 706.01 = 12681.59

$$\text{判定係數 } R^2 = \frac{SSR}{SST} = \frac{12681.59}{13387.60} = 0.9473$$

(B)依據 SST 和 SSR 公式的計算方式

營業點 i	行銷費用 x_i	套餐數 y_i	y_i^2	$x_i\times y_i$	x_i^2
1	150	156	24336	23400	22500
2	160	180	32400	28800	25600
3	180	190	36100	34200	32400
4	160	170	28900	27200	25600
5	190	198	39204	37620	36100
6	210	250	62500	52500	44100
7	180	189	35721	34020	32400
8	160	168	28224	26880	25600
9	180	191	36481	34380	32400
10	260	280	78400	72800	67600
合計	1830	1972	402266	371800	344300
平均值	183	197.2			

$$SST = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 402266 - \frac{1972^2}{10} = 402266 - 388878.4 = 13387.6$$

$$SSR = \frac{\left[\sum_{i=1}^{n}(x_i\times y_i) - \frac{\sum_{i=1}^{n} x_i\times\sum_{i=1}^{n} y_i}{n}\right]^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\left[371800 - \frac{1830\times1972}{10}\right]^2}{344300 - \frac{1830^2}{10}} = \frac{[371800-360876]^2}{344300-334890} = \frac{10924^2}{9410} = \frac{119333776}{9410} = 12681.59$$

$$\text{判定係數 } R^2 = \frac{SSR}{SST} = \frac{12681.59}{13387.6} = 0.9473$$

答案：判定係數 $R^2 = 0.9473$

練習 14.4 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1\times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請計算出估計簡單線性迴歸方程式的判定係數(coefficient of determination) R^2。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191

營業點 i	訓練費用 x_i	咖啡杯數 y_i
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式：$\hat{y}_i$ = $b_0 + b_1 \times x_i = 54.9828 + 2.1615 \times x_i$

(A)依據判定係數定義的計算方式

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	y_i^2	$x_i \times y_i$
1	50	156	163.06	-7.06	49.79	-30.40	924.16	24336	7800
2	53	179	169.54	9.46	89.48	-7.40	54.76	32041	9487
3	60	189	184.67	4.33	18.74	2.60	6.76	35721	11340
4	53	160	169.54	-9.54	91.02	-26.40	696.96	25600	8480
5	63	185	191.16	-6.16	37.89	-1.40	1.96	34225	11655
6	70	210	206.29	3.71	13.80	23.60	556.96	44100	14700
7	60	189	184.67	4.33	18.74	2.60	6.76	35721	11340
8	53	168	169.54	-1.54	2.37	-18.40	338.56	28224	8904
9	60	191	184.67	6.33	40.06	4.60	21.16	36481	11460
10	86	237	240.87	-3.87	14.97	50.60	2560.36	56169	20382
合計	608	1864	1864.00	0.00	376.86	0.00	5168.40	352618	115548
平均值	60.8	186.4							

SSE = $\sum_{i=1}^{n} y_i^2$ $- b_0 \times \sum_{i=1}^{n} y_i$ $- b_1 \times \sum_{i=1}^{n} x_i \times y_i$ = 352618 – 54.9828 × 1864 – 2.1615 × 115548 = 376.86(第一種算法)

SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ = 376.86(第二種算法)

SST = $\sum_{i=1}^{n}(y_i - \bar{y})^2$ = 5168.40

SSR = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ = SST – SSE = 5168.40 – 376.86 = 4791.54

判定係數 $R^2 = \frac{SSR}{SST} = \frac{4791.54}{5168.40} = 0.9271$

(B)依據 SST 和 SSR 公式的計算方式

營業點 i	訓練費用 x_i	咖啡杯數 y_i	y_i^2	$x_i \times y_i$	x_i^2
1	50	156	24336	7800	2500.0
2	53	179	32041	9487	2809.0
3	60	189	35721	11340	3600.0
4	53	160	25600	8480	2809.0
5	63	185	34225	11655	3969.0
6	70	210	44100	14700	4900.0
7	60	189	35721	11340	3600.0
8	53	168	28224	8904	2809.0
9	60	191	36481	11460	3600.0
10	86	237	56169	20382	7396.0
合計	608	1864	352618	115548	37992.0
平均值	60.8	186.4			

$$SST = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = 352618 - \frac{1864^2}{10} = 352618 - 347449.6 = 5168.4$$

$$SSR = \frac{\left[\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}\right]^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\left[115548 - \frac{608 \times 1864}{10}\right]^2}{37992.0 - \frac{608^2}{10}} = \frac{[115548 - 113331.2]^2}{37992.0 - 36966.4} = \frac{2216.8^2}{1025.6} = \frac{4914202.24}{1025.6} = 4791.54$$

判定係數 $R^2 = \frac{SSR}{SST} = \frac{4791.54}{5168.4} = 0.9271$

答案：判定係數 $R^2 = 0.9271$

相關係數(correlation coefficient)

當兩個隨機變數的關係屬於不獨立(有相互關係)時，並呈現線性相關，表達正負向關係和關係強弱者，即為相關係數。

樣本相關係數(Sample correlation coefficient) R_{xy}、r_{xy} 或 γ_{xy}

相關係數 $R_{xy} = r_{xy} =$ (b_1 的正負符號) × $\sqrt{\text{判定係數}}$ = (b_1 的正負符號) × $\sqrt{R^2}$

相關係數 $R_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\times\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\times\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} =$

$$\frac{\sum_{i=1}^{n} x_i \times y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \times \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

其中 b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

R^2 = 判定係數(coefficient of determination)。數值範圍 0~1。

範例 14.6　天空連鎖餐廳有 10 營業點，每個營業點前一日行銷費用和前一日販售套餐數列於下表。請計算出估計簡單線性迴歸方程式的相關係數(correlation coefficient) R_{xy}。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：依據判定係數定義的計算方式

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	150	156	158.89	-2.89	8.36	-41.20	1697.44
2	160	180	170.50	9.50	90.24	-17.20	295.84
3	180	190	193.72	-3.72	13.83	-7.20	51.84
4	160	170	170.50	-0.50	0.25	-27.20	739.84
5	190	198	205.33	-7.33	53.70	0.80	0.64
6	210	250	228.55	21.45	460.29	52.80	2787.84
7	180	189	193.72	-4.72	22.27	-8.20	67.24
8	160	168	170.50	-2.50	6.25	-29.20	852.64
9	180	191	193.72	-2.72	7.39	-6.20	38.44
10	260	280	286.59	-6.59	43.44	82.80	6855.84
合計	1830	1972	1972.01	0.00	706.01	0.00	13387.60
平均值	183	197.2					

行銷費用對銷售套餐數量的估計迴歸方程式　$\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$

$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 706.01$；$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 13387.60$

$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = SST - SSE = 13387.60 - 706.01 = 12681.59$

判定係數 $R^2 = \frac{SSR}{SST} = \frac{12681.59}{13387.60} = 0.9473$

估計迴歸方程式中 b_1 屬於正值故在相關係數中為+

相關係數 R_{xy} = (b_1 的正負符號)×$\sqrt{R^2}$ = $+\sqrt{0.9473}$ = 0.9733

答案：相關係數 R_{xy} = 0.9733

練習 14.5　小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販

售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請計算出估計簡單線性迴歸方程式的相關係數(correlation coefficient) R_{xy}。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = 54.9828 + 2.1615 \times x_i$

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	y_i^2	$x_i \times y_i$
1	50	156	163.06	-7.06	49.79	-30.40	924.16	24336	7800
2	53	179	169.54	9.46	89.48	-7.40	54.76	32041	9487
3	60	189	184.67	4.33	18.74	2.60	6.76	35721	11340
4	53	160	169.54	-9.54	91.02	-26.40	696.96	25600	8480
5	63	185	191.16	-6.16	37.89	-1.40	1.96	34225	11655
6	70	210	206.29	3.71	13.80	23.60	556.96	44100	14700
7	60	189	184.67	4.33	18.74	2.60	6.76	35721	11340
8	53	168	169.54	-1.54	2.37	-18.40	338.56	28224	8904
9	60	191	184.67	6.33	40.06	4.60	21.16	36481	11460
10	86	237	240.87	-3.87	14.97	50.60	2560.36	56169	20382
合計	608	1864	1864.00	0.00	376.86	0.00	5168.40	352618	115548
平均值	60.8	186.4							

SSE $= \sum_{i=1}^{n} y_i^2 - b_0 \times \sum_{i=1}^{n} y_i - b_1 \times \sum_{i=1}^{n} x_i \times y_i = 352618 - 54.9828 \times 1864 - 2.1615 \times 115548 = 376.86$(第一種算法)

SSE $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 376.86$(第二種算法)

SST $= \sum_{i=1}^{n} (y_i - \bar{y})^2 = 5168.40$

SSR $= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \text{SST} - \text{SSE} = 5168.40 - 376.86 = 4791.54$

判定係數 $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{4791.54}{5168.40} = 0.9271$

估計迴歸方程式中 b_1 屬於正值故在相關係數中為+

相關係數 $R_{xy} = (b_1$ 的正負符號$) \times \sqrt{R^2} = +\sqrt{0.9271} = 0.9629$

答案：相關係數 $R_{xy} = 0.9629$

練習 14.6 美美連鎖咖啡館行銷經理欲運用每月異業行銷費用 x_i(單位：新台幣千元)預測販售咖啡數量 y_i(單位：百杯)，建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。隨機抽出 5 個分店，每個營業點前一個月的異業行銷費用和前一個月販售咖啡數量依序列於下表。請計算出估計簡單線性迴歸方程式的相關係數(correlation coefficient) R_{xy}。

分店 i	行銷費用 x_i	咖啡數量 y_i
1	50	106
2	53	102
3	60	119
4	54	116
5	63	125

題解：

分店 i	行銷費用 x_i	咖啡數量 y_i	$x_i \times y_i$	x_i^2
1	50	106	5300	2500
2	53	102	5406	2809
3	60	119	7140	3600
4	54	116	6264	2916
5	63	125	7875	3969
合計	280	568	31985	15794
平均	56	113.6	6397	3158.8

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{31985 - \frac{280\times568}{5}}{15794 - \frac{280^2}{5}} = \frac{31985-31808}{15794-15680} = \frac{177}{114} = 1.5526$

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = 113.6 - 1.5526 \times 56 = 113.6 - 86.9474 = 26.6526$

亦可運用 Excel 軟體中資料分析的回歸指令，可以獲得異業行銷費用對銷售咖啡數量的估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i = 26.6526 + 1.5526 \times x_i$

分店 i	行銷費用 x_i	咖啡數量 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$y_i-\bar{y}$	$(y_i-\bar{y})^2$	y_i^2	$x_i \times y_i$
1	50	106	104.28	1.72	2.94	-7.60	57.76	11236	5300
2	53	102	108.94	-6.94	48.19	-11.60	134.56	10404	5406
3	60	119	119.81	-0.81	0.66	5.40	29.16	14161	7140
4	54	116	110.49	5.51	30.31	2.40	5.76	13456	6264
5	63	125	124.47	0.53	0.28	11.40	129.96	15625	7875
合計	280	568	568.00	0.00	82.38	0.00	357.20	64882	31985
平均值	56	113.6							

SSE $= \sum_{i=1}^{n} y_i^2 - b_0 \times \sum_{i=1}^{n} y_i - b_1 \times \sum_{i=1}^{n} x_i \times y_i = 64882 - 26.6526 \times 568 - 1.5526 \times 31985 = 82.38$(第一種算法)

SSE $= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 82.38$(第二種算法)

SST $= \sum_{i=1}^{n}(y_i - \bar{y})^2 = 357.20$

SSR $= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = SST - SSE = 357.20 - 82.38 = 274.82$

判定係數 $R^2 = \frac{SSR}{SST} = \frac{274.82}{357.20} = 0.7694$

估計迴歸方程式中 b_1 屬於正值故在相關係數中為+

相關係數 $R_{xy} = (b_1$ 的正負符號$) \times \sqrt{R^2} = +\sqrt{0.7694} = 0.8771$

答案：相關係數 $R_{xy} = 0.8771$

範例 14.7 美美連鎖咖啡館行銷經理與了解各分店每月營業額 y_i(單位：新台幣千元)與周遭服務人口 x_i(單位：百人)之間的迴歸關係，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。隨機抽取 5 個分店資料，獲得平均月營業額$\bar{y} = 350$ 千元，平均服務人口$\bar{x} = 50$ 百人，$\sum_{i=1}^{n=5}(x_i - \bar{x})^2 = 2200$，$\sum_{i=1}^{n=5}(y_i - \bar{y})^2 = 300000$，$\sum_{i=1}^{n=5}(x_i - \bar{x}) \times (y_i - \bar{y}) = 25000$。請計算出估計簡單線性迴歸方程式判定係數和相關係數(correlation coefficient) R_{xy}。

題解：

斜率 $b_1 = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})\times(y_i-\bar{y})]}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{25000}{2200} = 11.3636 =$ 斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = 350 - 11.3636 \times 50 = 350 - 568.182 = -218.182$

服務人口數對每月營業額的估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i = -218.182 + 11.3636 \times x_i$

SST $= \sum_{i=1}^{n}(y_i - \bar{y})^2 = 300000$

$\sum_{i=1}^{n=5}(x_i - \bar{x})^2 = 2200 = \sum_{i=1}^{n=5} x_i^2 - \frac{\left(\sum_{i=1}^{n=5} x_i\right)^2}{n=5}$

$$SSR = \frac{\left[\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}\right]^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\left[\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}\right]^2}{\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]^2} \times \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] =$$

$$\left[\frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\right]^2 \times \left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = [\text{斜率 } b_1]^2 \times \sum_{i=1}^{n=5}(x_i - \bar{x})^2 = 11.3636^2 \times 2200 = 284090.909$$

$$\text{判定係數 } R^2 = \frac{SSR}{SST} = \frac{284090.909}{300000} = 0.9470$$

估計迴歸方程式中 b_1 屬於正值故在相關係數中為+

$$\text{相關係數 } R_{xy} = (b_1 \text{ 的正負符號}) \times \sqrt{R^2} = +\sqrt{0.9470} = 0.9731$$

答案：判定係數 $R^2 = 0.9470$；相關係數 $R_{xy} = 0.9731$

判定係數(coefficient of determination)的數值範圍 0~1。相關係數(correlation coefficient)的數值範圍–1~+1。**相關係數**(correlation coefficient) R_{xy} 運用上僅限定於兩個變數之間屬於線性關係者；**判定係數**(coefficient of determination) R^2 可以使用於線性關係、非線性關係或兩個以上自(獨立)變數的關係。故，判定係數運用範圍比較大。

14.4 模式假設

簡單線性迴歸分析的**機率模式**(probabilistic model)或**迴歸模式(型)**(regression model)：

$$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i \text{，其中 } i = 1, \ldots, n$$

其中 y_i = 依變數第 i 個觀測值的實際觀測值

x_i = 自變數第 i 個觀測值

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

β_1 = 迴歸模式的參數，迴歸係數(regression coefficient)或斜率(slope)。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 變動性。

n = 觀測值數量

估計簡單線性迴歸方程式或樣本迴歸方程式

$$\hat{y}_i = b_0 + b_1 \times x_i \text{，其中 } i = 1, \ldots, n$$

其中 $\hat{y}_i$ = 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_i = 自變數第 i 個觀測值

b_0 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 參數(parameter) β_0 的估計值，截距(intercept)、常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

利用判定係數(coefficient of determination, R^2)估計迴歸方程式之適合度(goodness of fit)，其必須先經過顯著性檢定(斜率 β_1 是否等於 0 的假設檢定)。當未達顯著性水準時，判定係數 R^2 數值高低不具任何意義；達到顯著性水準時，判定係數 R^2 數值高低才具有代表性意義。

迴歸分析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.各觀測點之誤差項 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。$\mathrm{Cov}(\varepsilon_i,\varepsilon_j) = 0$ at $i \neq j$, $i, j = 1,\ldots, n$。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。$V(\varepsilon_i) = \sigma^2$。殘差項之變異數具有均一性(齊一性)。

D.各觀測點之誤差項 ε_i 屬於常態分布，$\varepsilon_i \sim N(0,\sigma^2)$。

14.5 斜率顯著性檢定

簡單線性迴歸方程式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，若斜率 β_1 等於 0 時，依變數 $E(y_i)$和自變數 x_i 之間沒有關係存在；當斜率 β_1 不等於 0 時，代表依變數 $E(y_i)$和自變數 x_i 之間有相關性存在。故在檢定依變數 $E(y_i)$和自變數 x_i 之間的迴歸關係時，必須進行斜率 β_1 是否等於 0 的假設檢定，此稱為顯著性檢定(Testing for significance)。

進行斜率 β_1 是否等於 0 的假設檢定時，可以利用 t 值檢定和 F 值檢定。在進行斜率 β_1 是否等於 0 的假設檢定前，需要先估計迴歸模式中誤差項或殘差項 ε_i 的變異數 σ^2。

誤差項 ε_i 的變異數 σ^2 估算

在自變數 x_i 與依變數 y_i 的迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$ 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)，可由誤差項平方和或殘差平方和(sum square error or sum of squares for error, SSE)除以其自由度(degree of freedom, df)獲得。在計算誤差項平方和或殘差平方和(sum square error, SSE)時，需先估算迴歸模式的兩個參數(β_0 和 β_1)，因此誤差項平方和或殘差平方和的自由度為 $n-2$。

$$\mathrm{MSE} = S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{n-2} = \frac{\sum_{i=1}^{n}y_i^2 - \mathrm{b}_0\times\sum_{i=1}^{n}y_i - \mathrm{b}_1\times\sum_{i=1}^{n}x_i\times y_i}{n-2}$$

誤差項 ε_i 之標準(偏)差 σ 的估計值 S 稱為估計值的標準(偏)差(standard error of the estimate)

$$S = \sqrt{\mathrm{MSE}} = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}y_i^2 - \mathrm{b}_0\times\sum_{i=1}^{n}y_i - \mathrm{b}_1\times\sum_{i=1}^{n}x_i\times y_i}{n-2}}$$

範例 14.8 天空連鎖餐廳有 10 營業點，每個營業點前一日行銷費用和前一日販售套餐數列於下表。試利用行銷費用 x_i 預測販售套餐數量 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請計算出估計簡單線性迴歸方程式誤差項 ε_i 之變異數 σ^2 和標準(偏)差 σ 的估計值。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：行銷費用對銷售套餐數量的估計迴歸方程式：$\hat{y}_i = \mathrm{b}_0 + \mathrm{b}_1 \times x_i = -15.2433 + 1.1609 \times x_i$

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	150	156	158.89	-2.89	8.36
2	160	180	170.50	9.50	90.24
3	180	190	193.72	-3.72	13.83
4	160	170	170.50	-0.50	0.25
5	190	198	205.33	-7.33	53.70
6	210	250	228.55	21.45	460.29

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
7	180	189	193.72	-4.72	22.27
8	160	168	170.50	-2.50	6.25
9	180	191	193.72	-2.72	7.39
10	260	280	286.59	-6.59	43.44
合計	1830	1972	1972.01	0.00	706.01
平均值	183	197.2			

殘差平方和(sum square error) SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ = 706.01

誤差平均平方和(mean square error) MSE = $S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{706.01}{10-2} = 88.2513$

誤差項標準差之估計值 $S = \sqrt{\text{MSE}} = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{88.2513} = 9.3942$

答案：誤差項 ε_i 變異數之估計值 $S^2 = 88.2513$；標準(偏)差之估計值 $S = 9.3942$

練習 14.7 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請計算出估計簡單線性迴歸方程式誤差項 ε_i 之變異數 σ^2 和標準(偏)差 σ 的估計值。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = 54.9828 + 2.1615 \times x_i$

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	50	156	163.06	-7.06	49.79
2	53	179	169.54	9.46	89.48
3	60	189	184.67	4.33	18.74
4	53	160	169.54	-9.54	91.02
5	63	185	191.16	-6.16	37.89
6	70	210	206.29	3.71	13.80
7	60	189	184.67	4.33	18.74
8	53	168	169.54	-1.54	2.37
9	60	191	184.67	6.33	40.06
10	86	237	240.87	-3.87	14.97
合計	608	1864	1864.00	0.00	376.86
平均值	60.8	186.4			

殘差平方和(sum square error) SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ = 376.86

誤差平均平方和(mean square error) MSE = $S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{376.86}{10-2} = 47.1075$

$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{47.1075} = 6.8635$

答案：誤差項 ε_i 之變異數之估計值 $S^2 = 47.1075$；標準(偏)差之估計值 $S = 6.8635$

t 值檢定

利用樣本資料檢定迴歸方程式中斜率 β_1 是否等於 0，設立假設

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

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

經過統計驗證的結果顯示，若接受虛無假設(null hypothesis) H_0: $\beta_1 = 0$ 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: $\beta_1 \neq 0$ 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_1 之估計值 b_1 之抽樣分布資料。

迴歸方程式斜率 β_1 之估計值 b_1 之抽樣分布

b_1 期望值 $E(b_1) = \beta_1$

b_1 標準(偏)差$\sigma_{b_1} = \dfrac{\sigma}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$

分布方式屬於常態分布。

若誤差項 ε_i 之標準(偏)差 σ 未知時，可以利用誤差項 ε_i 之標準(偏)差 σ 的估計值 S 取代標準(偏)差，以獲得 b_1 標準(偏)差σ_{b_1}的估計值S_{b_1}。

b_1 標準(偏)差σ_{b_1}的估計值$S_{b_1} = \dfrac{S}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$

檢定統計值 $t = \dfrac{b_1}{S_{b_1}}$

雙尾檢定(a two-tailed test)

若左側臨界值$-t_{\frac{\alpha}{2},n-2} \leq$ 檢定統計值 $t \leq$ 右側臨界值$t_{\frac{\alpha}{2},n-2}$，接受虛無假設(null hypothesis) H_0: $\beta_1 = 0$。

若檢定統計值 $t <$ 左側臨界值$-t_{\frac{\alpha}{2},n-2}$或檢定統計值 $t >$ 右側臨界值$t_{\frac{\alpha}{2},n-2}$，拒絕虛無假設(null hypothesis) H_0: $\beta_1 = 0$，接受對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$。

其中$t_{\frac{\alpha}{2},n-2}$：為右尾機率$\frac{\alpha}{2}$，自由度 $n-2$ 的 t 分布數值。

範例 14.9　天空連鎖餐廳有 10 營業點，每個營業點前一日行銷費用和前一日販售套餐數列於下表。欲運用行銷費用 x_i 預測販售套餐數量 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用 t 值法檢定迴歸方程式中斜率 β_1 是否等於 0。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	x_i^2
1	150	156	158.89	-2.89	8.36	22500
2	160	180	170.50	9.50	90.24	25600
3	180	190	193.72	-3.72	13.83	32400
4	160	170	170.50	-0.50	0.25	25600

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
5	190	198	205.33	-7.33	53.70	36100
6	210	250	228.55	21.45	460.29	44100
7	180	189	193.72	-4.72	22.27	32400
8	160	168	170.50	-2.50	6.25	25600
9	180	191	193.72	-2.72	7.39	32400
10	260	280	286.59	-6.59	43.44	67600
合計	1830	1972	1972.01	0.00	706.01	344300
平均值	183	197.2				

行銷費用對銷售套餐數量的估計迴歸方程式：$\hat{y}_i$ = b$_0$ + b$_1$×x_i = –15.2433 + 1.1609×x_i

A.設定顯著水準 α = 0.01，左側臨界值 $-t_{\frac{\alpha}{2},n-2} = -t_{\frac{0.01}{2},10-2} = -t_{0.005,8}$ = -3.3554，右側臨界值 $t_{\frac{\alpha}{2},n-2} = t_{\frac{0.01}{2},10-2}$ = $t_{0.005,8}$ = 3.3554(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　　殘差平方和(sum square error) SSE = $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ = 706.01

誤差平均平方和(mean square error) MSE = $S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{706.01}{10-2}$ = 88.2513

$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{88.2513}$ = 9.3942

$S_{b_1} = \frac{S}{\sqrt{\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}}} = \frac{9.3942}{\sqrt{344300-\frac{1830^2}{10}}} = \frac{9.3942}{\sqrt{344300-334890}} = \frac{9.3942}{97.0052}$ = 0.0968

檢定統計值 $t = \frac{b_1}{S_{b_1}} = \frac{1.1609}{0.0968}$ = 11.9875

E.因檢定統計值 t = 11.9875 > 右側臨界值 $t_{\frac{\alpha}{2},n-2}$ = 3.3554。接受對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

答案：t 值法檢定迴歸方程式中斜率 β_1 不等於 0

練習 14.8　小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用 t 值法檢定迴歸方程式中斜率 β_1 是否等於 0。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式　$\hat{y}_i$ = b$_0$ + b$_1$×x_i = 54.9828 + 2.1615×x_i

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
1	50	156	163.06	-7.06	49.79	2500.0
2	53	179	169.54	9.46	89.48	2809.0
3	60	189	184.67	4.33	18.74	3600.0
4	53	160	169.54	-9.54	91.02	2809.0

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
5	63	185	191.16	-6.16	37.89	3969.0
6	70	210	206.29	3.71	13.80	4900.0
7	60	189	184.67	4.33	18.74	3600.0
8	53	168	169.54	-1.54	2.37	2809.0
9	60	191	184.67	6.33	40.06	3600.0
10	86	237	240.87	-3.87	14.97	7396.0
合計	608	1864	1864.00	0.00	376.86	37992.0
平均值	60.8	186.4				

A.設定顯著水準 $\alpha = 0.01$，左側臨界值 $-t_{\frac{\alpha}{2},n-2} = -t_{\frac{0.01}{2},10-2} = -t_{0.005,8} = -3.3554$，右側臨界值 $t_{\frac{\alpha}{2},n-2} = t_{\frac{0.01}{2},10-2} = t_{0.005,8} = 3.3554$(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

殘差平方和(sum square error) SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 376.8612$

誤差平均平方和(mean square error) MSE = $S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{376.8612}{10-2} = 47.1076$

$$S = \sqrt{S^2} = \sqrt{MSE^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{47.1076} = 6.8635$$

$$S_{b_1} = \frac{S}{\sqrt{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = \frac{6.8635}{\sqrt{37992-\frac{608^2}{10}}} = \frac{6.8635}{\sqrt{37992-36966.4}} = \frac{6.8635}{32.0250} = 0.2143$$

$$\text{檢定統計值 } t = \frac{b_1}{S_{b_1}} = \frac{2.1615}{0.2143} = 10.0854$$

E.檢定統計值 $t = 10.0854 > $ 右側臨界值 $t_{\frac{\alpha}{2},n-2} = 3.3554$。接受對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

答案：t 值法檢定迴歸方程式中斜率 β_1 不等於 0

F 值檢定

利用 F 機率分布以樣本資料檢定迴歸方程式中斜率 β_1 是否等於 0，使用於驗證迴歸關係的顯著性。若只有一個自變數 x_i(簡單線性迴歸分析)時，利用 t 值檢定和 F 值檢定的結果相同。當有兩個(含)以上自變數 x_i 時，僅可以使用 F 值檢定法，以驗證全部自變數 x_i 與依變數 y_i 之間關係的顯著性。若欲檢定迴歸方程式中斜率 β_1 是否等於特定數值(C)或進行左右尾檢定時，皆必須使用 t 值檢定法，無法使用 F 值檢定法。

使用情境	t 值檢定	F 值檢定
一個自變數	●	●
兩個(含)以上自變數	X	●
檢定斜率 β_1 是否等於特定數值(C)	●	X
左右尾檢定	●	X

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, *df*)獲得。迴歸自由度(regression degrees of freedom, *df*)等於自變數之個數。

$$\text{MSR} = \frac{SSR}{df} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{df} = \frac{\text{迴歸項平方和}}{\text{迴歸自由度}}$$

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: $\beta_1 = 0$。

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

D.計算檢定統計值—F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{迴歸造成的均方}}{\text{誤差均方}}$$

E.若檢定統計值 $F \leq$ 臨界值 $F_{\alpha,1,n-2}$，接受虛無假設 H_0: $\beta_1 = 0$。

F.若檢定統計值 $F >$ 臨界值 $F_{\alpha,1,n-2}$，拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$。

其中 $F_{\alpha,1,n-2}$：系分子自由度 1，分母自由度 $n-2$ 的右尾機率 α 的 F 分布數值。n 為樣本數量。

簡單線性迴歸變異數分析(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)	$\text{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	1	$\text{MSR} = \frac{SSR}{1}$	$F = \frac{\text{MSR}}{\text{MSE}}$
誤差項(隨機項)(error)	$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-2$	$\text{MSE} = \frac{SSE}{n-2}$	
合計(total)	$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$		

範例 14.10 天空連鎖餐廳有 10 營業點，每個營業點前一日行銷費用和前一日販售套餐數列於下表。欲運用行銷費用 x_i 預測販售套餐數量 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用 F 值法檢定迴歸方程式中斜率 β_1 是否等於 0。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$
1	150	156	158.89	-2.89	8.36	-41.20	1697.44	1467.61
2	160	180	170.50	9.50	90.24	-17.20	295.84	712.92
3	180	190	193.72	-3.72	13.83	-7.20	51.84	12.13
4	160	170	170.50	-0.50	0.25	-27.20	739.84	712.92
5	190	198	205.33	-7.33	53.70	0.80	0.64	66.04
6	210	250	228.55	21.45	460.29	52.80	2787.84	982.45
7	180	189	193.72	-4.72	22.27	-8.20	67.24	12.13
8	160	168	170.50	-2.50	6.25	-29.20	852.64	712.92
9	180	191	193.72	-2.72	7.39	-6.20	38.44	12.13
10	260	280	286.59	-6.59	43.44	82.80	6855.84	7990.35
合計	1830	1972	1972.01	0.00	706.01	0.00	13387.60	12681.59
平均值	183	197.2						

行銷費用對銷售套餐數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$

$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 706.01$　　$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 13387.60$

$\text{SSR} = \text{SST} - \text{SSE} = 13387.60 - 706.01 = 12681.59$

SSR = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ = 12681.59

MSE = $S^2 = \frac{SSE}{n-2} = \frac{706.01}{10-2}$ = 88.2513 MSR = $\frac{SSR}{df} = \frac{12681.59}{1}$ = 12681.59

A.設定顯著水準 α = 0.01，臨界值 $F_{\alpha,1,n\text{-}2} = F_{0.01,1,10\text{-}2} = F_{0.01,1,8}$ = 11.2586(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

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

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

D.計算檢定統計值─F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{12681.59}{88.2513} = 143.6986$$

E.檢定統計值 F = 143.6986 > 臨界值 $F_{\alpha,1,n\text{-}2}$ = 11.2586，拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

答案：F 值法檢定迴歸方程式中斜率 β_1 不等於 0

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)	12681.59	1	12681.59	143.6986
誤差項(隨機項)(error)	706.01	8	88.25	
總變異(total)	13387.60	9		

練習 14.9 某研究團隊欲評估石鱉背殼寬度 x_i (單位：公分)與長度 y_i (單位：公分)的關係，研究結果獲得下列簡單線性迴歸模式 $\hat{y}_i$ = 1.400+1.500×x_i 及變異數分析表如下：請先計算變異數分析有缺值的欄位與計算決定係數(coefficient of determination)。

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	p 值
迴歸模型(regression)		1			<0.001
誤差項(隨機項)(error)	0.399		0.031		
總和(total)	5.229	14			

題解：依據簡單線性迴歸變異數分析表

簡單線性迴歸變異數分析(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 = $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	1	MSR = $\frac{SSR}{1}$	$F = \frac{\text{MSR}}{\text{MSE}}$
誤差項(隨機項)(error)	SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-2$	MSE = $\frac{SSE}{n-2}$	
總變異(total)	SST = $\sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$		

題目中已知誤差項和總和的平方和數值，透過 SST = SSR + SSE，欲計算迴歸項平方和 SSR = SST – SSE = 5.229 - 0.399 = 4.900。

已知迴歸項和總和的自由度數值。總和自由度 n - 1 = 14，n = 15，透過 n – 1 = 1 + n – 2，欲計算誤差項自由度 n – 2 = n – 1 – 1 = 14 – 1 = 13。

前面已經計算出回歸項平方和 SSR = 4.900，同時已知迴歸項自由度 1，故，迴歸項均方 MSR = $\frac{SSR}{df} = \frac{4.900}{1}$ = 4.900。

題目中已知誤差項平方和 SSE = 0.399，前面也已經算出誤差項自由度 = 13，亦可透過 MSE = $\frac{SSE}{n-2} = \frac{0.399}{13}$ = 0.0306923，以驗證題目提供的誤差項均方 MSE = 0.031 數值的準確性。

$F = \frac{\text{MSR}}{\text{MSE}} = \frac{4.900}{0.0306923} = 159.65$

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	p 值
迴歸模型(regression)	4.900	1	4.900	159.65	<0.001
誤差項(隨機項)(error)	0.399	13	0.031		
總和(total)	5.229	14			

決定係數 $R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{4.900}{5.229} = 0.9371$

A.設定顯著水準 $\alpha = 0.05$，臨界值 $F_{\alpha,1,n-2} = F_{0.05,1,15-2} = F_{0.05,1,13} = 4.6672$(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

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

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

D.計算檢定統計值—F 值

$$\text{檢定統計值 } F = \frac{\text{MSR}}{\text{MSE}} = \frac{4.900}{0.0306923} = 159.65$$

E.檢定統計值 $F = 159.65$ > 臨界值 $F_{\alpha,1,n-2} = 4.6672$，拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

答案：F 值法檢定迴歸方程式中斜率 β_1 不等於 0

練習 14.10 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請利用 F 值法檢定迴歸方程式中斜率 β_1 是否等於 0。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = 54.9828 + 2.1615 \times x_i$

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	50	156	163.06	-7.06	49.79	-30.40	924.16
2	53	179	169.54	9.46	89.48	-7.40	54.76
3	60	189	184.67	4.33	18.74	2.60	6.76
4	53	160	169.54	-9.54	91.02	-26.40	696.96
5	63	185	191.16	-6.16	37.89	-1.40	1.96
6	70	210	206.29	3.71	13.80	23.60	556.96
7	60	189	184.67	4.33	18.74	2.60	6.76
8	53	168	169.54	-1.54	2.37	-18.40	338.56
9	60	191	184.67	6.33	40.06	4.60	21.16
10	86	237	240.87	-3.87	14.97	50.60	2560.36
合計	608	1864	1864.00	0.00	376.86	0.00	5168.40
平均值	60.8	186.4					

$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 376.86$　　$\text{SST} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 5168.40$

SSR = SST – SSE = 5168.40 – 376.86 = 4791.54

MSE = $S^2 = \frac{SSE}{n-2} = \frac{376.86}{10-2}$ = 47.1075 MSR = $\frac{SSR}{df} = \frac{4791.54}{1}$ = 4791.54

A.設定顯著水準 α = 0.01，臨界值 $F_{\alpha,1,n\text{-}2} = F_{0.01,1,10\text{-}2} = F_{0.01,1,8}$ = 11.2586(使用 Excel 軟體 F.INV.RT 函數查詢獲得)。

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

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

D.計算檢定統計值–F 值

$$檢定統計值\ F = \frac{\text{MSR}}{\text{MSE}} = \frac{4791.54}{47.1075} = 101.7150$$

E.檢定統計值 F = 101.7150 > 臨界值 $F_{\alpha,1,n\text{-}2}$ = 11.2586。接受對立假設(alternative hypothesis)H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

答案：F 值法檢定迴歸方程式中斜率 β_1 不等於 0

若檢定迴歸方程式中斜率 β_1 是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$)，而判定係數(coefficient of determination) R^2 數值高時，解釋迴歸方程式時不會有疑慮；若判定係數 R^2 數值高時，β_1 是顯著性的等於 0(接受虛無假設 H_0: $\beta_1 = 0$)或 β_1 是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$)，而判定係數 R^2 數值低時，在解釋和運用迴歸方程式應特別謹慎。

一般若要運用自變數對依變數進行預測時，會比較看重判定係數，判定係數高時自變數對依變數的預測才會準確。此外，若欲探索自變數對依變數是否有影響或影響程度，則著重於檢定迴歸方程式中斜率 β_1 是否等於 0。

	斜率 $\beta_1 = 0$	斜率 $\beta_1 \neq 0$
R^2 數值高	自變數對依變數沒有意義	自變數對依變數預測能力佳，具有意義
R^2 數值低	自變數對依變數沒有意義	自變數對依變數預測能力弱，具有意義，影響力小

若檢定迴歸方程式中斜率 β_1 是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$)，獲得自變數 x_i 與依變數 y_i 存在顯著關係的推論，不能單因此統計檢定的證據就斷然判定自變數 x_i 與依變數 y_i 存在因果關係，還是需要具有相關的理論基礎為佐證，才能較客觀的判定自變數 x_i 與依變數 y_i 存在因果關係。

若檢定迴歸方程式中斜率 β_1 是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$)，獲得自變數 x_i 與依變數 y_i 存在顯著關係的推論，但是不能據此推論自變數 x_i 與依變數 y_i 存在線性關係，僅能夠推論自變數 x_i 與依變數 y_i 存在顯著關係。欲驗證自變數 x_i 與依變數 y_i 存在線性關係時，可以利用樣本相關係數 r_{xy} 進行檢定。母體相關係數 ρ_{xy} (讀音 rho)假設：

A.虛無假設(null hypothesis) H_0: $\rho_{xy} = 0$。

B.對立假設(alternative hypothesis) H_1: $\rho_{xy} \neq 0$。

若拒絕虛無假設(null hypothesis) H_0: $\rho_{xy} = 0$，接受對立假設(alternative hypothesis) H_1: $\rho_{xy} \neq 0$，可以驗證自變數 x_i 與依變數 y_i 存在線性關係。

論證	斜率 $\beta_1 \neq 0$
因果關係	需要具有相關的理論基礎為佐證
線性關係	利用樣本相關係數 r_{xy} 進行檢定

練習 14.11 天空連鎖餐廳有數個營業點，每個營業點個別的平均每月行銷費用(x_i，單位：萬元)和營業額(y_i，單位：十萬元)，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。現今隨機抽出 6 個營業

點，其行銷費用與營業額如下所示。

營業點 i	1	2	3	4	5	6
行銷費用 x_i	2	3	8	9	12	10
營業額 y_i	4	5	8	10	13	10

(A)試計算估計迴歸方程式$\hat{y}_i = b_0 + b_1 \times x_i$之統計值。

(B)以顯著水準 $\alpha = 0.05$ 檢定 x_i 對 y_i 是否有顯著性影響？

(C)計算斜率係數的 95 %信賴區間。

(D)計算 x_i 與 y_i 的相關係數。

題解：

營業點 i	行銷費用 x_i	營業額 y_i	$x_i \times y_i$	x_i^2	y_i^2
1	2	4	8	4	16
2	3	5	15	9	25
3	8	8	64	64	64
4	9	10	90	81	100
5	12	13	156	144	169
6	10	10	100	100	100
合計	44	50	433	402	474

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n}x_i \times \sum_{i=1}^{n}y_i}{n}}{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}} = \frac{433-\frac{44\times50}{6}}{402-\frac{44^2}{6}} = \frac{433-366.6667}{402-322.6667} = \frac{66.3333}{79.3333} = 0.8361$

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = \frac{50}{6} - 0.8361 \times \frac{44}{6} = 8.3333 - 6.1314 = 2.2017$

行銷費用對營業額的估計迴歸方程式　$\hat{y}_i = b_0 + b_1 \times x_i = 2.2017 + 0.8361 \times x_i$

營業點 i	行銷費用 x_i	營業額 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	2	4	3.8739	0.1261	0.0159	-5.3333	28.4444
2	3	5	4.7101	0.2899	0.0841	-4.3333	18.7778
3	8	8	8.8908	-0.8908	0.7934	0.6667	0.4444
4	9	10	9.7269	0.2731	0.0746	1.6667	2.7778
5	12	13	12.2353	0.7647	0.5848	4.6667	21.7778
6	10	10	10.5630	-0.5630	0.3170	2.6667	7.1111
合計	44	50	50.0000	0.0000	1.8697	0.0000	79.3333

A.設定顯著水準 $\alpha = 0.05$，臨界值$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},6-2} = t_{0.025,4} = 2.7764$(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　SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i)^2 = 1.8697$

$MSE = S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{1.8697}{6-2} = 0.4674$

$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{0.4674} = 0.6837$

$S_{b_1} = \frac{S}{\sqrt{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = \frac{0.6837}{\sqrt{402-\frac{44^2}{6}}} = \frac{0.6837}{\sqrt{402-322.6667}} = \frac{0.6837}{8.9069} = 0.0768$

檢定統計值 $t = \frac{b_1}{S_{b_1}} = \frac{0.8361}{0.0768} = 10.8867$

E.檢定統計值 $t = 10.8867 >$ 臨界值$t_{\frac{\alpha}{2},n-2} = 2.7764$。接受對立假設(alternative hypothesis)H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

斜率係數 β_1 的 95 %信賴區間：$b_1 \pm t_{\frac{\alpha}{2},n-2} \times \sqrt{\frac{MSE}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \rightarrow 0.8361 \pm 2.7764 \times \sqrt{\frac{0.4674}{79.3333}} \rightarrow$

$0.8361 \pm 2.7764 \times 0.0768 \rightarrow 0.8361 \pm 0.2131 \rightarrow$ 信賴區間為 0.6230~1.0492

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i \times y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \times \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} = \frac{433 - \frac{44 \times 50}{6}}{\sqrt{402 - \frac{44^2}{6}} \times \sqrt{474 - \frac{50^2}{6}}} = \frac{433 - 366.6667}{\sqrt{402 - 322.6667} \times \sqrt{474 - 416.6667}} = \frac{66.3333}{67.4422} = 0.9836$$

答案：(A)斜率 $b_1 = 0.8361$；截距 $b_0 = 2.2017$；(B)t 值法檢定迴歸方程式中斜率 β_1 不等於 0，x_i 對 y_i 有顯著性影響；(C)斜率係數的 95 %信賴區間 0.6230~1.0492；(D) $r_{xy} = 0.9836$

14.6 運用估計迴歸方程式進行估算與預估

在簡單線性迴歸模式中，利用樣本資料以最小平方法，獲得估計線性迴歸方程式。若檢定迴歸方程式中斜率 β_1 是顯著性的不等於 0(拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$)，而判定係數(coefficient of determination) R^2 數值高，顯示其適合度亦高。此迴歸方程式可以應用於預測依變數 y_i。

點估計(point estimation)

天空連鎖餐廳有 10 營業點，依據每個營業點個別的平均每日行銷費用 x_i 和平均每日販售套餐數 y_i10 個營業點樣本資料。獲得行銷費用 x_i 對銷售套餐數量 y_i 的估計迴歸方程式為$\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$。若行銷費用 x_i 為 250 時，販售套餐數量的預估值$\hat{y}_i = -15.2433 + 1.1609 \times 250 = 274.9798$；行銷費用 x_i 為 350 時，販售套餐數量的預估值$\hat{y}_i = -15.2433 + 1.1609 \times 350 = 391.0691$。

區間估計(interval estimation)

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：第一種是自變數為一特定數值 x_0 時，預測依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_0)$的信賴區間；第二種是自變數為一特定數值 x_0 時，預測依變數 y_i 之個別數值 y_0 的信賴區間。

預測依變數之平均值的信賴區間

一般情況下估計迴歸方程式依變數的估計值$\hat{y}_0 = b_0 + b_1 \times x_0$ 很少剛好等於期望值 $E(y_0)$，若欲瞭解估計值$\hat{y}_0$與期望值 $E(y_0)$之間的差距時，需依據估計迴歸方程式中依變數預估值$\hat{y}_0$的變異數$S_{\hat{y}_0}^2$。

$$Var(\hat{y}_0) = Var(b_0 + b_1 \times x_0) = Var(\bar{y} - b_1 \times \bar{x} + b_1 \times x_0) = Var[\bar{y} + b_1 \times (x_0 - \bar{x})] = Var(\bar{y}) + (x_0 - \bar{x})^2 Var(b_1) + 2 \times (x_0 - \bar{x}) \times Cov(\bar{y}, b_1) = \frac{\sigma^2}{n} + (x_0 - \bar{x})^2 \times \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} + 0 = \sigma^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\right]$$

$$Cov(\bar{y}, b_1) = \frac{\sum_{i=1}^{N}\{[\bar{y} - E(\bar{y})] \times [b_1 - E(b_1)]\}}{N} = \frac{\sum_{i=1}^{N}\{[\bar{y} - E(\bar{y})] \times [b_1 - \beta_1]\}}{N} = 0$$

母體變異數 σ^2 未知，以樣本變異數 S^2 取代，即為 MSE 數值

依變數預估值$\hat{y}_0$的變異數$S_{\hat{y}_0}^2 = S^2 \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\right] = MSE \times \left[\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}\right]$

估計迴歸方程式中依變數預估值$\hat{y}_0$的標準(偏)差$S_{\hat{y}_0}$

依變數預估值$\hat{y}_0$的標準(偏)差$S_{\hat{y}_0} = S \times \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$

迴歸方程式中預測依變數之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_0)$的信賴區間為

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

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S \times \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$$

其中$t_{\frac{\alpha}{2},n-2}$：右尾機率$\frac{\alpha}{2}$，自由度 $n-2$ 的 t 分布數值。

S 稱為估計值的標準(偏)差(standard error of the estimate)：誤差項 ε_i 之標準(偏)差 σ 的估計值。$S = \sqrt{S^2} = \sqrt{\text{MSE}} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{n-2}}$

範例 14.11 天空連鎖餐廳有 10 營業點，依據每個營業點前一日行銷費用 x_i 和前一日販售套餐數 y_i 十個營業點樣本資料。欲運用行銷費用 x_i 預測販售套餐數量 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請分別計算在行銷費用 x_0 為 250 和 350 時，販售套餐數量之平均值$\bar{y}_0$的信賴區間。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
1	150	156	158.89	-2.89	8.36	22500
2	160	180	170.50	9.50	90.24	25600
3	180	190	193.72	-3.72	13.83	32400
4	160	170	170.50	-0.50	0.25	25600
5	190	198	205.33	-7.33	53.70	36100
6	210	250	228.55	21.45	460.29	44100
7	180	189	193.72	-4.72	22.27	32400
8	160	168	170.50	-2.50	6.25	25600
9	180	191	193.72	-2.72	7.39	32400
10	260	280	286.59	-6.59	43.44	67600
合計	1830	1972	1972.01	0.00	706.01	344300
平均值	183.0	197.2				

行銷費用對銷售套餐數量的估計迴歸方程式 $\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$

若行銷費用 x_0 為 250 元時，販售套餐數量的預估值$\hat{y}_0 = -15.2433 + 1.1609 \times 250 = 274.9798$ 套，其販售套餐數量之平均值$\bar{y}_0$的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{\hat{y}_0} = 274.9798 \pm 2.3060 \times 7.1362 = 274.9798 \pm 16.4561 \rightarrow$ 信賴區間為 258.5~291.4 套

其中顯著水準 $\alpha = 0.05$，自由度 $df = n-2 = 10-2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

$S_{\hat{y}_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \times \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} = \sqrt{\frac{706.01}{10-2}} \times \sqrt{\frac{1}{10} + \frac{(250-183)^2}{344300-\frac{1830^2}{10}}} = 9.3942 \times$

$\sqrt{0.1+\frac{67^2}{344300-334890}}=9.3942\times\sqrt{0.1+\frac{4489}{9410}}=9.3942\times0.7596=7.1362$

行銷費用 x_0 為 350 元時，販售套餐數量的預估值 $\hat{y}_0$ = -15.2433 + 1.1609 × 350 = 391.0691 套。其販售套餐數量之平均值 $\bar{y}_0$ 的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2}\times S_{\hat{y}_0} = 391.0691 \pm 2.3060\times16.4432 = 391.0691 \pm 37.9181 \rightarrow$ 信賴區間為 353.2~429.0 套

$$S_{\hat{y}_0}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}}\times\sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}}=\sqrt{\frac{706.01}{10-2}}\times\sqrt{\frac{1}{10}+\frac{(350-183)^2}{344300-\frac{1830^2}{10}}}=9.3942\times\sqrt{0.1+\frac{167^2}{344300-334890}}=$$

$$9.3942\times\sqrt{0.1+\frac{27889}{9410}}=9.3942\times1.7504=16.4432$$

答案：行銷費用 250 元，信賴區間 258.5~291.4 套；行銷費用 350 元，信賴區間為 353.2~429.0 套

練習 14.12 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i=\beta_0+\beta_1\times x_i+\varepsilon_i$，其中 ε_i 為誤差項。請分別計算在訓練費用 x_0 為 55 和 85(單位：新台幣百元)時，販售咖啡杯數量之平均值 $\bar{y}_0$、E(y_0)或 E($y|x_0$)的信賴區間。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式 $\hat{y}_i$ = $b_0+b_1\times x_i=54.9828+2.1615\times x_i$

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
1	50	156	163.06	-7.06	49.79	2500.0
2	53	179	169.54	9.46	89.48	2809.0
3	60	189	184.67	4.33	18.74	3600.0
4	53	160	169.54	-9.54	91.02	2809.0
5	63	185	191.16	-6.16	37.89	3969.0
6	70	210	206.29	3.71	13.80	4900.0
7	60	189	184.67	4.33	18.74	3600.0
8	53	168	169.54	-1.54	2.37	2809.0
9	60	191	184.67	6.33	40.06	3600.0
10	86	237	240.87	-3.87	14.97	7396.0
合計	608	1864	1864.00	0.00	376.86	37992.0
平均值	60.8	186.4				

若訓練費用 x_0 為 55(單位：新台幣百元)時，販售咖啡杯數的預估值 $\hat{y}_0$ = 54.9828 + 2.1615×55 = 173.8635 杯，其販售咖啡杯數之平均值 $\bar{y}_0$ 的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2}\times S_{\hat{y}_0} = 173.8635 \pm 2.3060\times2.5012 = 173.8635 \pm 5.7677 \rightarrow$ 信賴區間為 168.1~179.6 杯

其中顯著水準 $\alpha=0.05$，自由度 $df=n-2=10-2=8$，$t_{\frac{\alpha}{2},n-2}=t_{\frac{0.05}{2},10-2}=t_{0.025,8}=2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

$$S_{\hat{y}_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \times \sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = \sqrt{\frac{376.86}{10-2}} \times \sqrt{\frac{1}{10}+\frac{(55-60.8)^2}{37992-\frac{608^2}{10}}} = 6.8635\times\sqrt{0.1+\frac{-5.8^2}{37992-36966.4}} =$$

$$6.8635\times\sqrt{0.1+\frac{33.64}{1025.6}} = 6.8635\times0.3644 = 2.5012$$

若訓練費用 x_0 為 85(單位：新台幣百元)時，販售咖啡杯數的預估值$\hat{y}_0$ $= 54.9828 + 2.1615\times85 = 238.7075$ 杯，其販售咖啡杯數之平均值$\bar{y}_0$的信賴區間：

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2}\times S_{\hat{y}_0} = 238.7075 \pm 2.3060\times5.6223 = 238.7075 \pm 12.9650$ → 信賴區間為 225.7~251.7 杯

其中顯著水準 $\alpha = 0.05$，自由度 $df = n-2 = 10-2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

$$S_{\hat{y}_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \times \sqrt{\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = \sqrt{\frac{376.86}{10-2}} \times \sqrt{\frac{1}{10}+\frac{(85-60.8)^2}{37992-\frac{608^2}{10}}} = 6.8635\times\sqrt{0.1+\frac{24.2^2}{37992-36966.4}} =$$

$$6.8635\times\sqrt{0.1+\frac{585.64}{1025.6}} = 6.8635\times0.8192 = 5.6223$$

答案：訓練費用 55(單位：新台幣百元)，販售咖啡杯數之平均值的信賴區間 168.1~179.6 杯；訓練費用 85(單位：新台幣百元)，信賴區間 225.7~251.7 杯

當自變數數值 x_0 離自變數樣本平均值$\bar{x}$愈遠時，依變數預估值$\hat{y}_0$之變異數$S_{\hat{y}_0}^2$和標準(偏)差$S_{\hat{y}_0}$的數值愈大，依據公式中其分子愈大所致，同時使預測依變數之平均值$\bar{y}_0$的信賴區間愈大。

預測依變數 y_i 之個別數值 y_0 的信賴區間

在迴歸方程式中利用自變數 x_0 預測依變數數值 y_0 的單一數值(實際數值)時，需要瞭解依變數 y_0 的信賴區間或預測區間(prediction interval)。

為獲得依變數 y_0 的信賴區間，必須依據自變數為 x_0 時，其依變數 y_0 的變異數$S_{y_0}^2$。

依變數 y_0 的變異數$S_{y_0}^2 = S^2 + S_{\hat{y}_0}^2 = S^2 + S^2\times\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}\right] = S^2\times\left[1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}\right]$

在自變數為 x_0 時，其依變數 y_0 的標準(偏)差S_{y_0}。

依變數 y_0 的標準(偏)差$S_{y_0} = S\times\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}}$

迴歸方程式中預測依變數 y_0 的信賴區間或預測區間為

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

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2}\times S\times\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}}$$

其中$t_{\frac{\alpha}{2},n-2}$：右尾機率$\frac{\alpha}{2}$，自由度 $n-2$ 的 t 分布數值。

S 稱為估計值的標準(偏)差(standard error of the estimate)：誤差項 ε_i 之標準(偏)差 σ 的估計值。$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-b_0-b_1\times x_i)^2}{n-2}}$

範例 14.12 天空連鎖餐廳有 10 營業點，依據每個營業點前一日行銷費用 x_i 和前一日販售套餐數 y_i 十個營業點樣本資料。請分別計算在行銷費用 x_0 為 250 和 350 時，販售套餐數量 y_0 的信賴區間。

營業點 i	行銷費用 x_i	套餐數 y_i
1	150	156

營業點 i	行銷費用 x_i	套餐數 y_i
2	160	180
3	180	190
4	160	170
5	190	198
6	210	250
7	180	189
8	160	168
9	180	191
10	260	280

題解：

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$(y_i-\hat{y}_i)^2$	x_i^2
1	150	156	158.89	-2.89	8.36	22500
2	160	180	170.50	9.50	90.24	25600
3	180	190	193.72	-3.72	13.83	32400
4	160	170	170.50	-0.50	0.25	25600
5	190	198	205.33	-7.33	53.70	36100
6	210	250	228.55	21.45	460.29	44100
7	180	189	193.72	-4.72	22.27	32400
8	160	168	170.50	-2.50	6.25	25600
9	180	191	193.72	-2.72	7.39	32400
10	260	280	286.59	-6.59	43.44	67600
合計	1830	1972	1972.01	0.00	706.01	344300
平均值	183.0	197.2				

行銷費用對銷售套餐數量的估計迴歸方程式：$\hat{y}_i = b_0 + b_1 \times x_i = -15.2433 + 1.1609 \times x_i$

若行銷費用 x_0 為 250 元時，販售套餐數量的預估值$\hat{y}_0 = -15.2433 + 1.1609 \times 250 = 274.9798$ 套，當行銷費用 x_0 為 250 元時，販售套餐數量 y_0 的標準(偏)差S_{y_0}。

$$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} = \sqrt{\frac{706.01}{10-2}} = \sqrt{88.2513} = 9.3942$$

$$S_{y_0} = S \times \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = 9.3942 \times \sqrt{1+\frac{1}{10}+\frac{(250-183)^2}{344300-\frac{1830^2}{10}}} = 9.3942 \times \sqrt{1.1+\frac{67^2}{344300-334890}}$$

$$= 9.3942 \times \sqrt{1.1+\frac{4489}{9410}} = 9.3942 \times 1.2558 = 11.7973$$

若顯著水準 $\alpha = 0.05$，自由度 $df = n-2 = 10-2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

販售套餐數量 y_0 的信賴區間：

$$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{y_0} = 274.9798 \pm 2.3060 \times 11.7973 = 274.9798 \pm 27.2046 \rightarrow \text{信賴區間 } 247.8\text{~}302.2 \text{ 套}$$

販售套餐數量 y_0 的信賴區間或預測區間 247.8~302.2，比販售套餐數量之平均值$\bar{y}_0$、$E(y_0)$或 $E(y|x_0)$的信賴區間 258.5~291.4 更大，因此利用平均值的估計法比利用個別數值更為精準。

行銷費用 x_0 為 350 元時，販售套餐數量的預估值$\hat{y}_0 = -15.2433 + 1.1609 \times 350 = 391.0691$ 套。當行銷費用 x_0 為 350 時，販售套餐數量 y_0 的標準(偏)差S_{y_0}

$$S_{y_0} = S \times \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = 9.3942 \times \sqrt{1+\frac{1}{10}+\frac{(350-183)^2}{344300-\frac{1830^2}{10}}} = 9.3942 \times \sqrt{1.1+\frac{167^2}{344300-334890}}$$

$$= 9.3942 \times \sqrt{1.1+\frac{27889}{9410}} = 9.3942 \times 2.0159 = 18.9376$$

若顯著水準 $\alpha = 0.05$，自由度 $df = n-2 = 10-2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟

體 T.INV 函數查詢獲得)。

販售套餐數量 y_0 的信賴區間：

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{y_0} = 391.0691 \pm 2.3060 \times 18.9376 = 391.0691 \pm 43.6700$ → 信賴區間 347.4~434.7 套

答案：行銷費用 250 元，販售套餐數量 y_0 的信賴區間 247.8~302.2 套；行銷費用 350 元，販售套餐數量 y_0 的信賴區間為 347.4~434.7 套

個別數值和平均值信賴區間預估時，皆在自變數 $x_0 = \bar{x}$，信賴區間最小，精準度最高。

練習 14.13 小美連鎖咖啡館有 10 個營業點，每個營業點前一天的訓練費用(單位：新台幣百元)和前一天販售咖啡杯數量(單位：杯)依序列於下表。欲運用訓練費用 x_i 預測販售咖啡杯數 y_i，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。請分別計算在訓練費用 x_0 為 55 和 85(單位：新台幣百元)時，販售咖啡杯數量 y_0 的信賴區間。

營業點 i	訓練費用 x_i	咖啡杯數 y_i
1	50	156
2	53	179
3	60	189
4	53	160
5	63	185
6	70	210
7	60	189
8	53	168
9	60	191
10	86	237

題解：訓練費用對銷售咖啡杯數量的估計迴歸方程式 $\hat{y}_i = b_0 + b_1 \times x_i = 54.9828 + 2.1615 \times x_i$

營業點 i	訓練費用 x_i	咖啡杯數 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	x_i^2
1	50	156	163.06	-7.06	49.79	2500.0
2	53	179	169.54	9.46	89.48	2809.0
3	60	189	184.67	4.33	18.74	3600.0
4	53	160	169.54	-9.54	91.02	2809.0
5	63	185	191.16	-6.16	37.89	3969.0
6	70	210	206.29	3.71	13.80	4900.0
7	60	189	184.67	4.33	18.74	3600.0
8	53	168	169.54	-1.54	2.37	2809.0
9	60	191	184.67	6.33	40.06	3600.0
10	86	237	240.87	-3.87	14.97	7396.0
合計	608	1864	1864.00	0.00	376.86	37992.0
平均值	60.8	186.4				

若訓練費用 x_0 為 55(單位:新台幣百元)時，販售咖啡杯數的預估值 $\hat{y}_0 = 54.9828 + 2.1615 \times 55 = 173.8635$ 杯，其販售咖啡杯數 y_0 的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{y_0} = 173.8635 \pm 2.3060 \times 7.3050 = 173.8635 \pm 16.8454$ → 信賴區間為 157.0~190.7 杯

其中顯著水準 $\alpha = 0.05$，自由度 $df = n - 2 = 10 - 2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

$$S_{y_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \times \sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}} = \sqrt{\frac{376.86}{10-2}} \times \sqrt{1+\frac{1}{10}+\frac{(55-60.8)^2}{37992-\frac{608^2}{10}}} = 6.8635 \times$$

$$\sqrt{1+0.1+\frac{-5.8^2}{37992-36966.4}} = 6.8635 \times \sqrt{1+0.1+\frac{33.64}{1025.6}} = 6.8635 \times 1.0643 = 7.3050$$

若訓練費用 x_0 為 85(單位：新台幣百元)時，販售咖啡杯數的預估值 $\hat{y}_0 = 54.9828 + 2.1615 \times 85 = 238.7075$ 杯，其販售咖啡杯數 y_0 的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{y_0} = 238.7075 \pm 2.3060 \times 8.8723 = 238.7075 \pm 20.4595 \rightarrow$ 信賴區間為 218.2~259.2 杯

其中顯著水準 $\alpha = 0.05$，自由度 $df = n - 2 = 10 - 2 = 8$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},10-2} = t_{0.025,8} = 2.3060$(使用 Excel 軟體 T.INV 函數查詢獲得)。

$$S_{y_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} \times \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} = \sqrt{\frac{376.86}{10-2}} \times \sqrt{1 + \frac{1}{10} + \frac{(85-60.8)^2}{37992 - \frac{608^2}{10}}} = 6.8635 \times \sqrt{1 + 0.1 + \frac{24.2^2}{37992 - 36966.4}} = 6.8635 \times \sqrt{1 + 0.1 + \frac{585.64}{1025.6}} = 6.8635 \times 1.2927 = 8.8723$$

答案：訓練費用 55(單位：新台幣百元)，販售咖啡杯數的信賴區間 157.0~190.7 杯；訓練費用 85(單位：新台幣百元)，信賴區間 218.2~259.2 杯

練習 14.14 天空連鎖餐廳有數個營業點，每個營業點個別的平均每月行銷費用(x_i，單位：萬元)和營業額(y_i，單位：十萬元)，欲建立迴歸模式 $y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$，其中 ε_i 為誤差項。現今隨機抽出 5 個營業點，其行銷費用與營業額如下所示。

營業點 i	1	2	3	4	5
行銷費用 x_i	6	10	8	9	12
營業額 y_i	8	12	8	10	13

(A)試計算估計迴歸方程式 $\hat{y}_i = b_0 + b_1 \times x_i$ 之統計值。

(B)以顯著水準 $\alpha = 0.05$ 檢定迴歸模式是否斜率等於 0。

(C)假設行銷費用為 9 萬元，試計算營業額的平均值 $\mu_{y_i|x_i}$ 的 95 %信賴區間。

題解：

營業點 i	行銷費用 x_i	營業額 y_i	$x_i \times y_i$	x_i^2	y_i^2
1	6	8	48	36	64
2	10	12	120	100	144
3	8	8	64	64	64
4	9	10	90	81	100
5	12	13	156	144	169
合計	45	51	478	425	541

斜率 $b_1 = \frac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{478 - \frac{45 \times 51}{5}}{425 - \frac{45^2}{5}} = \frac{478 - 459}{425 - 405} = \frac{19}{20} = 0.9500$

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = \frac{51}{5} - 0.9500 \times \frac{45}{5} = 10.2 - 8.55 = 1.65$

行銷費用對營業額的估計迴歸方程式 $\hat{y}_i = b_0 + b_1 \times x_i = 1.65 + 0.95 \times x_i$

營業點 i	行銷費用 x_i	營業額 y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	6	8	7.3500	0.6500	0.4225
2	10	12	11.1500	0.8500	0.7225
3	8	8	9.2500	-1.2500	1.5625
4	9	10	10.2000	-0.2000	0.0400
5	12	13	13.0500	-0.0500	0.0025
合計	45	51	51.0000	0.0000	2.7500

A.設定顯著水準 $\alpha = 0.05$，臨界值 $t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},5-2} = t_{0.025,3} = 3.1824$(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。SSE = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i)^2 = 2.7500$

$$MSE = S^2 = \frac{SSE}{n-2} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2} = \frac{2.7500}{5-2} = 0.9167$$

$$S = \sqrt{S^2} = \sqrt{\frac{SSE}{n-2}} = \sqrt{0.9167} = 0.9574$$

$$S_{b_1} = \frac{S}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}} = \frac{0.9574}{\sqrt{425-\frac{45^2}{5}}} = \frac{0.9574}{\sqrt{425-405}} = \frac{0.9574}{4.4721} = 0.2141$$

$$檢定統計值\ t = \frac{b_1}{S_{b_1}} = \frac{0.9500}{0.2141} = 4.4372$$

E.檢定統計值 t = 4.4372 > 臨界值$t_{\frac{\alpha}{2},n-2}$ = 3.1824。接受對立假設(alternative hypothesis) H_1: $\beta_1 \neq 0$，因此迴歸方程式中斜率 β_1 不等於 0。自變數與依變數之間的迴歸關係達到顯著性相關，迴歸方程式具有解釋(預測)能力。

若行銷費用 x_0 為 9 萬元時，營業額的預估值$\hat{y}_0$ = 1.65 + 0.95 × x_i = 1.65 + 0.95 × 9 = 10.2 拾萬元，其營業額之平均值$\bar{y}_0$的信賴區間。

$\hat{y}_0 \pm t_{\frac{\alpha}{2},n-2} \times S_{\hat{y}_0}$ = 10.2 ± 3.1824 × 0.4282 = 10.2 ± 1.3627 → 信賴區間為 8.8373~11.5627 拾萬元

其中顯著水準 α = 0.05，自由度 $df = n - 2 = 5 - 2 = 3$，$t_{\frac{\alpha}{2},n-2} = t_{\frac{0.05}{2},5-2} = t_{0.025,3}$ = 3.1824(使用 Excel 軟體 T.INV 函數查詢獲得)。

$$S_{\hat{y}_0} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \times \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}}} = \sqrt{\frac{2.75}{5-2}} \times \sqrt{\frac{1}{5} + \frac{(9-9)^2}{425-\frac{45^2}{5}}} = 0.9574 \times \sqrt{0.2 + \frac{0^2}{425-405}} = 0.9574 \times \sqrt{0.2 + \frac{0}{20}} = 0.9574 \times 0.4472 = 0.4282$$

答案：(A)斜率 b_1 = 0.95；截距 b_0 = 1.65；(B)t 值法檢定迴歸方程式中斜率 β_1 不等於 0；(C)營業額的平均值 $\mu_{y_i|x_i}$的 95 %信賴區間 8.8373~11.5627 拾萬元

14.7 統計軟體迴歸運用

利用 Excel 2007 協助線性迴歸資料分析，在 Excel 視窗中勾選工具(T)→選取增益集(I)…→勾選分析工具箱→按確定按鈕。回到 Excel 視窗中勾選工具(T)→選取資料分析(D)…→在資料分析視窗中選取迴歸後按確定。在迴歸視窗中，輸入 Y 範圍(Y):依變數在 excel 視窗中的位置；輸入 X 範圍(X):自變數在 excel 視窗中的位置。勾選信賴度(O) 95 %。勾選殘差(R)和標準化殘差(T)。按確定。

利用 Excel 2010 協助線性迴歸資料分析，在 Excel 視窗工具列中勾選檔案→選取選項→出現 Excel 選項小視窗勾選增益集→選取增益集中的分析工具箱-VBA→按執行(G)…→出現增益集小視窗，勾選分析工具箱-VBA 後按確定按鈕。回到 Excel 視窗中工具列勾選資料→選取資料分析→在資料分析視窗中選取迴歸後按確定。在迴歸視窗中，輸入 Y 範圍(Y):依變數在 excel 視窗中的位置；輸入 X 範圍(X):自變數在 excel 視窗中的位置。勾選信賴度(O) 95 %。勾選殘差(R)和標準化殘差(T)。按確定。

摘要輸出

迴歸統計	
R 的倍數	0.9733
R 平方	0.9473

迴歸統計	
調整的 R 平方	0.9407
標準誤	9.3942
觀察值個數	10

ANOVA

	自由度	SS	MS	F	顯著值
迴歸	1	12681.59	12681.59	143.70	0.0000
殘差	8	706.01	88.25		
總和	9	13387.60			

	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	-15.2434	17.9694	-0.8483	0.4209	-56.6809	26.1942	-56.6809	26.1942
X 變數 1	1.1609	0.0968	11.9875	0.0000	0.9376	1.3842	0.9376	1.3842

殘差輸出

觀察值	預測 Y	殘差	標準化殘差
1	158.89	-2.89	-0.33
2	170.50	9.50	1.07
3	193.72	-3.72	-0.42
4	170.50	-0.50	-0.06
5	205.33	-7.33	-0.83
6	228.54	21.46	2.42
7	193.72	-4.72	-0.53
8	170.50	-2.50	-0.28
9	193.72	-2.72	-0.31
10	286.59	-6.59	-0.74

利用 **SPSS** 分析結果

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method	
1	VAR00001[a]	.	Enter	

[a] All requested variables entered.
[b] Dependent Variable: VAR00002

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	
1	.973[a]	.947	.941	9.3942	

[a] Predictors: (Constant), VAR00001
[b] Dependent Variable: VAR00002

ANOVA[b]

Model		Sum of Squares	df	Mean Square	*F*	Sig.	
1	Regression	12681.591	1	12681.591	143.699	.000[a]	
	Residual	706.009	8	88.251			
	Total	13387.600	9				

[a] Predictors: (Constant), VAR00001
[b] Dependent Variable: VAR00002

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			
		B	Std. Error	Beta	*t*	Sig.	
1	(Constant)	-15.243	17.969		-.848	.421	
	VAR00001	1.161	.097	.973	11.987	.000	

[a] Dependent Variable: VAR00002

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N	
Predicted Value	158.891	286.589	197.200	37.538	10	
Std. Predicted Value	-1.021	2.381	0.000	1.000	10	
Standard Error of Predicted Value	2.985	8.027	3.949	1.512	10	
Adjusted Predicted Value	159.686	304.410	198.690	42.042	10	
Residual	-7.326	21.456	0.000	8.857	10	
Std. Residual	-0.780	2.284	0.000	0.943	10	
Stud. Residual	-1.350	2.518	-0.050	1.095	10	
Deleted Residual	-24.409	26.085	-1.490	13.014	10	
Stud. Deleted Residual	-1.437	5.174	0.221	1.854	10	
Mahal. Distance	0.009	5.671	0.900	1.713	10	
Cook's Distance	0.000	2.465	0.336	0.777	10	
Centered Leverage Value	0.001	0.630	0.100	0.190	10	

[a] Dependent Variable: VAR00002

14.8 殘差分析：驗證模型假設【選擇教材】

殘差分析(residual analysis)可**協助判斷迴歸模式中假設的適切程度**。

第 i 項樣本觀測值 y_i 的殘差：即等於迴歸分析中的誤差項 ε_i

$$\varepsilon_i = y_i - \hat{y}_i$$

其中 $y_i =$ 依變數第 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；依變數第 i 個觀測值的估計值

$\varepsilon_i =$ 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 變動性。

營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	殘差 $y_i-\hat{y}_i$
1	150	156	158.89	-2.89
2	160	180	170.50	9.50
3	180	190	193.72	-3.72
4	160	170	170.50	-0.50
5	190	198	205.33	-7.33
6	210	250	228.55	21.45
7	180	189	193.72	-4.72
8	160	168	170.50	-2.50
9	180	191	193.72	-2.72
10	260	280	286.59	-6.59
合計	1830	1972	1972.01	0.00
平均值	183	197.2		

迴歸分析之顯著性檢定必須依據下列**誤差項** ε_i 的**假設**。

A.各觀測點之誤差項 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。

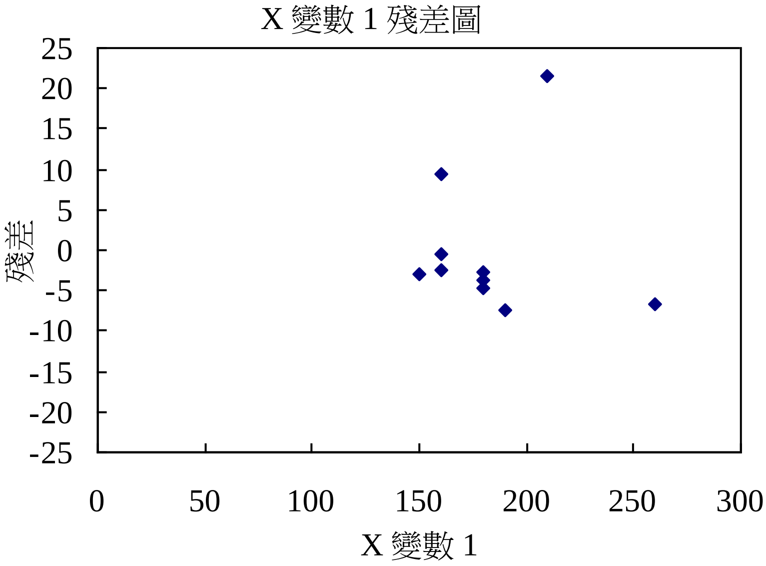
C.各觀測點之誤差項 ε_i 之變異數 σ^2 皆相等。$V(\varepsilon_i) = \sigma^2$。殘差項之變異數具有均一性(齊一性)。

D.各觀測點之誤差項 ε_i 屬於常態分布的隨機變數。

基於上述假設為基礎，方能夠利用樣本觀測值，推論自變數 x 與依變數 y 之間的關係，應用 t 值檢定和 F 值檢定以驗證自變數 x 與依變數 y 之間的關係是存在；並且在自變數 x 與依變數 y 之間具有線性關係前提下，推論特定自變數數值 x_0 時，依變數 y_0 的信賴區間。**若上述的假設有問題時，自變數 x 與依變數 y 之間的關係驗證和依變數 y_0 之信賴區間的推論，皆屬於無效推論**。

殘差分析可以提供誤差項 ε_i 是否符合上述假設的判斷資訊。殘差分析主要是依據圖形進行研判。

自變數 x(X 軸)與殘差值(Y 軸)圖



自變數 x(X 軸)與殘差值(Y 軸)圖可能出現的三種狀況：

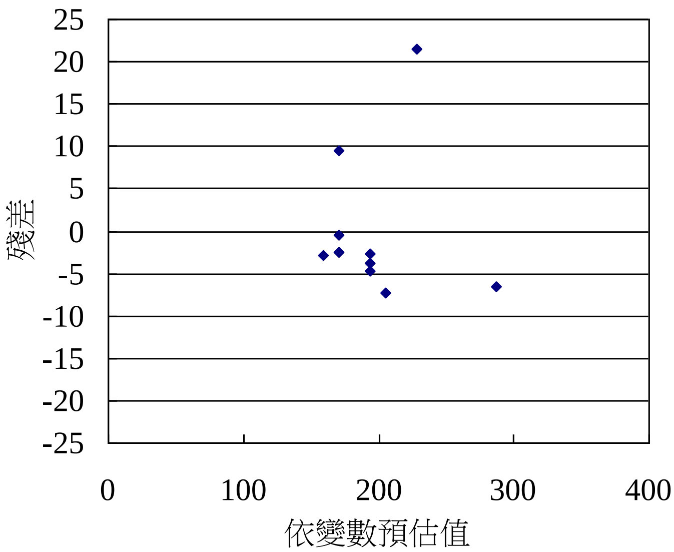
A. 符合假設『各觀測點之誤差項 ε_i 之變異數 σ^2 皆相等。$V(\varepsilon_i) = \sigma^2$。殘差項之變異數具有均一性(齊一性)。』，對所有的自變數 x_i 而言，誤差項 ε_i 的變異數皆 σ^2 相等，代表分散程度均等。故在自變數 x(X 軸)與殘差值(Y 軸)圖中，應該所有樣本觀測值之點(X 軸)，皆分布在以殘差值為 0 之中心的水平帶狀(長條狀)區域中，在 X 軸方向上下分散程度均等。

B. 若假設『各觀測點之誤差項 ε_i 之變異數 σ^2 皆相等。$V(\varepsilon_i) = \sigma^2$。殘差項之變異數具有均一性(齊一性)。』無法成立時，可能出現自變數 x_i 愈大而變異數 σ^2 愈大或愈小的情況。若是自變數 x_i 愈大而變異數 σ^2 愈大時，應該所有樣本觀測值之點(X 軸)，水平分布愈往右邊上下分散程度愈大；自變數 x_i 愈大而變異數 σ^2 愈小時，應該所有樣本觀測值之點(X 軸)，水平分布愈往左邊上下分散程度愈大。

C. 欲擬分析迴歸模式無法符合(解釋)樣本觀測值的分布情況時，可能是所有樣本觀測值之點，在 X 軸分向上下分散程度均等，但是並非水平狀況，可能是 V、U、M、︿、﹏狀態，迴歸分析模式可能是曲線或多變數迴歸模式。

依變數預估值 $\hat{y}_i$(X 軸)與殘差值(Y 軸)圖

在簡單(單變數)迴歸分析中，依變數預估值 $\hat{y}_i$(X 軸)與殘差值(Y 軸)圖與自變數 x(X 軸)與殘差值(Y 軸)圖提供判斷誤差項 ε_i 是否符合假設的資訊相同。在多元(多變數)迴歸分析中，使用一個以上自變數，故較常使用依變數預估值 $\hat{y}_i$(X 軸)與殘差值(Y 軸)圖進行研判。



標準殘差、標準化殘差(standardized residual, SR)

為了減少測量單位對殘差數值的影響，因此，大部分統計分析軟體皆使用標準殘差數值。

利用最小平方法估算迴歸方程式參數時，其殘差值和為 0，殘差值之平均值亦為 0。因此，將殘差值 $\varepsilon_i = y_i - \hat{y}_i$ 除以其標準(偏)差 $S_{y_i-\hat{y}_i}$ 即可獲得標準殘差。

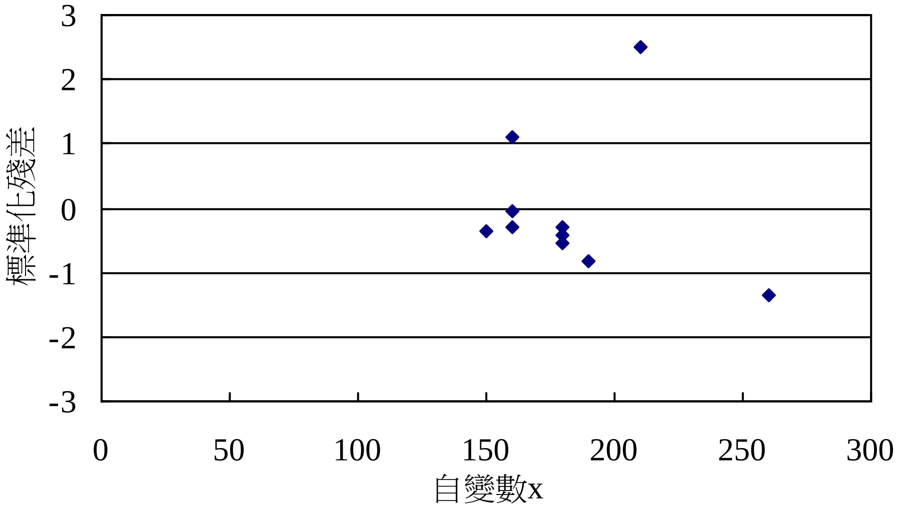
$$標準殘差 = \frac{y_i-\hat{y}_i}{S_{y_i-\hat{y}_i}}$$

第 i 項觀測值之殘差的標準(偏)差

$$S_{y_i-\hat{y}_i} = S \times \sqrt{1-\frac{1}{n}-\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

其中 S = 依變數估計值之標準(偏)差

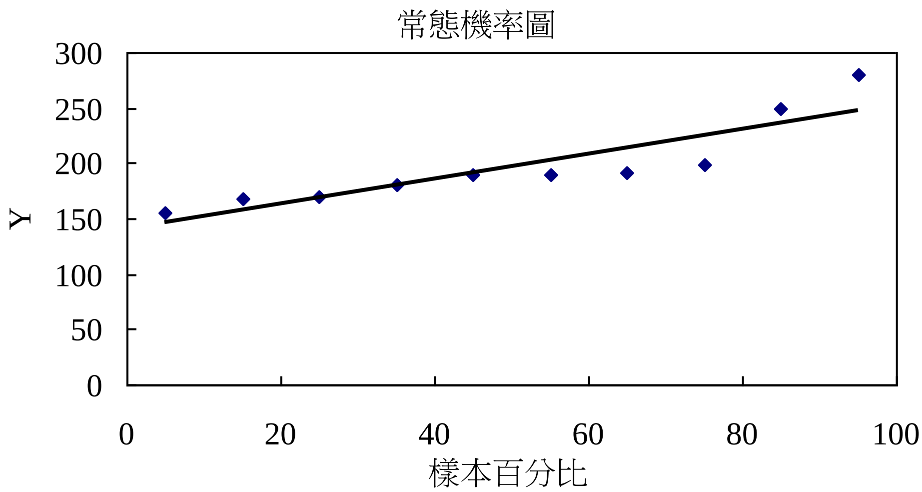
營業點 i	行銷費用 x_i	套餐數 y_i	$\hat{y}_i$	$y_i-\hat{y}_i$	$x_i-\bar{x}$	$(x_i-\bar{x})^2$	$S_{y_i-\hat{y}_i}$	標準殘差
1	150	156	158.89	-2.89	-33	1089	8.3194	-0.3474
2	160	180	170.50	9.50	-23	529	8.6293	1.1009
3	180	190	193.72	-3.72	-3	9	8.9074	-0.4176
4	160	170	170.50	-0.50	-23	529	8.6293	-0.0579
5	190	198	205.33	-7.33	7	49	8.8863	-0.8249
6	210	250	228.55	21.45	27	729	8.5199	2.5176
7	180	189	193.72	-4.72	-3	9	8.9074	-0.5299
8	160	168	170.50	-2.50	-23	529	8.6293	-0.2897
9	180	191	193.72	-2.72	-3	9	8.9074	-0.3054
10	260	280	286.59	-6.59	77	5929	4.8807	-1.3502
合計	1830	1972	1972.01	0.00	0	9410		
平均值	183	197.2						



若欲符合『各觀測點之誤差項 ε_i 屬於常態分布的隨機變數。』的假設條件，在標準殘差圖中標準殘差點的分布以 Y 軸為基準，應屬於以 0 為中心上下對稱的常態分布型態。期望有 95 %以上的標準殘差點落在-2 到+2 區間，若可以達到此條件，可以判定符合『各觀測點之誤差項 ε_i 屬於**常態分布**的隨機變數。』的假設條件。

常態機率圖(normal probability plot)

常態機率圖亦是判斷『各觀測點之誤差項 ε_i 屬於常態分布的隨機變數。』假設條件的工具之一。



14.9 殘差分析：離群值和其他影響高的觀測值(Residual analysis: Outliers and influential)【選擇教材】

14.10 簡單線性迴歸分析統計軟體運用【選擇教材】

14.10.1 研究問題

欲分析遊客滿意度，是否具有預測消費金額的能力？

分析方法

自變數(自變項)為遊客的滿意度，屬於等距變項(scale variable)，為連續變項，依變數(因變項)消費金額為連續變項

自變數 連續變項		依變數 連續變項
X_1	⇨	Y_1

簡單線性迴歸分析

$$Y(\text{依變數}) = b_0 + b_1X(\text{自變數})$$

Durbin-Watson 統計量

- 檢定相鄰殘差項間是否相關的一種統計量，若相鄰殘差項間是相關，則其總差異必小或大，因此可用相鄰殘差項間的總差異，來判斷殘差項是否相關或獨立。

$$DW = \frac{\sum_{i=2}^{n}(E_i - E_{i-1})^2}{\sum_{i=1}^{n} E_i^2}$$

其中：$E_i = Y_i - \hat{y}_i \quad , i = 1, 2, 3, \ldots, n$

$Y_i = i$ 位置的實際值

$\hat{y}_i$ = 為 $E(Y_i)$的不偏估計統計量

- 若殘差項間是正相關時，則其差異必小，反之若殘差項間是負相關，則差異必大。故若 DW 值小時，表示殘差是正相關；若大時，表示負相關。
- 當 DW 值愈接近 2 時，殘差項間愈無相關。
- 當 DW 值愈接近 0 時，殘差項間正相關愈強。
- 當 DW 值愈接近 4 時，殘差項間負相關愈強。

迴歸模式的基本假設(迴歸模式的殘差分析)

1. 每一誤差變數(變項)均具有常態分配(常態分布)，其期望值為 0，變異數為 σ^2，即 $N(0,\sigma^2)$。
2. 誤差變項彼此間是獨立。
3. 迴歸共線假設(迴歸線性假設)。

14.10.2 簡單線性迴歸分析 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. 在 Linear Regression(線性迴歸)對話視窗中，點選左邊方塊中的依變數(因變項)：消費金額[cost_in]進入右邊的 Dependent: (因變項)下面方塊中，點選左邊方塊中的自變項：服務滿意度指標[qulit_in]進入右邊的 Independent(s): (自變數、獨立變數)下面的方塊中。
3. 在右邊中間的 Method:選項中，共有五種篩選變數的選擇：
   - 「Enter」: 設定強制一次進入的迴歸分析方法，此為預定(內定)方法。強制選取法。簡單線性迴歸均選取此種方法。
   - 「Stepwise」: 設定逐步迴歸選取法。將前向選取法和後向選取法的綜合，首先依據前向選取法的方式，一次考慮一個自變數，判斷其貢獻是否已達設定的標準，若是則將其「納入」多元迴歸方程式中。之後，在依據後向選取法的方式，檢驗目前多元迴歸方程式中的所有自變數，一一評估是否應被剔除，在分析過程中，被剔除的自變數無法再進入多元迴歸方程式中。使用於多元(複)迴歸分析中。

- 「Remove」: 設定強制一次移除的迴歸分析法。
- 「Backward」: 設定反向移除式的迴歸分析法，後向(向後)選取法，先將全部的自變數納入多元迴歸方程式中，然後一次評估一個自變數，判斷其貢獻是否無法達到設定的標準，若是則將其自迴歸方程式中「剔除」之，此法在分析過程中，並不再考慮任何已被剔除的自變數。
- 「Forward」: 設定正向選取式的迴歸分析法，前向(向前)選取法，一次考慮一個自變數，判斷其貢獻是否已達設定的標準，若是則將其「納入」複(多元)迴歸方程式中。此分析法在選取過程中並不剔除任何已在迴歸方程式中的自變數。
- 上述各種選取方法所獲得的迴歸模式並不能保證是「最好的(Best)」，他只能說在所採用的方法內，是「最佳的(Optimum)」。故要評估各種方法所獲得的迴歸模式，何者是最好的，有必要再進一步分析比較，並檢驗是否合乎各種迴歸模式的假設。

4.點選下面的 Statistics… 按鈕，會出現 Linear Regression: Statistics 對話方塊，在左上角 Regression Coefficients 方塊中勾選□Estimates(此為預定選項，用以顯示迴歸係數、迴歸係數的標準差、標準化的迴歸係數、t 值和 t 分布的雙尾機率)和□Confidence intervals(顯示每個非標準化迴歸係數之 95%的信賴區間)選項。

5.在 Linear Regression: Statistics 對話方塊右上角勾選□Model fit(此為預定選項，顯示相關係數 R、判定係數 R^2、調整之判定係數、標準誤和 Anova 表)、□Descriptives(顯示平均值、標準差、單尾機率之相關係數矩陣)和□Collinearity diagnostics[執行共線性的診斷，顯示變異數擴張因子(Variance inflation factor: VIF)、交互離差矩陣(cross-product deviation matrix)、條件指標(Condition indices)和變異數分解比例。亦顯示在迴歸方程式中之變異數的寬容度(Tolerance)；不在迴歸方程式中之變數，則顯示若在下一步中，它要進入方程式時，該變數的寬容度]選項。

6.在 Linear Regression: Statistics 對話方塊下面的 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。

7.在 Linear Regression: Statistics 對話方塊中，點選右上角的 Continue 回到 Linear Regression 對話方塊。

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

9.在 Linear Regression: Plots 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(因變數)
- 「*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

10.在 Linear Regression: Plots 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/因變數)進入 X: (X 軸)右邊的方塊。

11.在 Linear Regression: Plots 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。

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

13.在 Linear Regression 對話視窗中按 OK 鈕，執行簡單線性迴歸程序。

14.獲得以下分析成果

Descriptive Statistics

	Mean	Std. Deviation	N
消費金額	2971.43	1978.82	40
服務滿意度指標	180.90	16.45	40

①.兩變數的相關係數為 0.986

Correlations

		消費金額	服務滿意度指標	
Pearson Correlation	消費金額	1.000	.986	
	服務滿意度指標	.986	1.000	
Sig. (1-tailed)	消費金額	.	.000	
	服務滿意度指標	.000	.	
N	消費金額	40	40	
	服務滿意度指標	40	40	

Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	服務滿意度指標	.	Enter

a. All requested variables entered.

b. Dependent Variable: 消費金額

②.判定係數(R2)為 0.973，此模式的解釋能力(預估能力)相當高，達 97.3 %。

③.統計量 Durbin-Watson 為 0.187，相當接近於 0，可判定殘差彼此間相關性相當強。

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	.986[a]	.973	.972	330.83	.973	1357.320	1	38	.000	.187

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額

④.適合性檢定：從變異數分析中顯示 P 值(Sig.)為 0.000，達到顯著水準 0.05，表示此模式適合利用服務滿意度指標(自變數)來解釋(預估)消費金額(依變數/因變數)。

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148554157.246	1	148554157.246	1357.320	.000[a]
	Residual	4158974.529	38	109446.698		
	Total	152713131.775	39			

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額

⑤.簡單線性迴歸方程式：Y(消費金額) = -18497.196 + 118.677*X(服務滿意度指標)

⑥.方程式的斜率與截距的檢定 P 值(Sig.)為 0.000 達到 0.05 顯著水準，故斜率與截距在方程式中均是存在，不為數值 0。

⑦.因為獨立變數只有一個，故 Tolerance 和 VIF 值均為 1.000。因此，當獨立變數只有一個時，其 Tolerance

和 VIF 值均沒有意義。The tolerance statistics, which assess the potential impact of multicollinearity among the independent variables on the accuracy of regression coefficients, were all higher than 0.1, thereby indicating that the results were reliable.

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
	(Constant)	-18497.196	585.067		-31.616	.000	-19681.603	-17312.789		
	服務滿意度指標	118.677	3.221	.986	36.842	.000	112.156	125.198	1.000	1.000

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
1	1	1.996	1.000	.00	.00
	2	4.005E-03	22.325	1.00	1.00

a. Dependent Variable: 消費金額

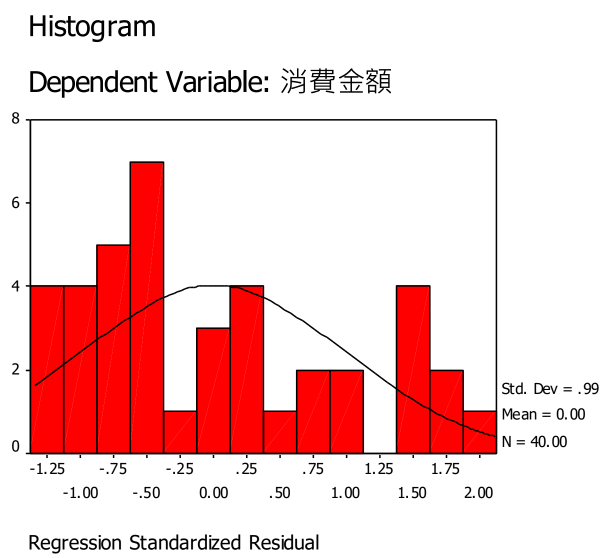
⑧.利用服務滿意度指標來預測消費金額是相當合適的一個變數(因素)，而且所建立的模式也令人滿意，惟不代表此時可用該模式進行預測。尚須進行評估該模式是否符合簡單線性迴歸模式的假設，故尚須進行殘差分析。

Residuals Statistics

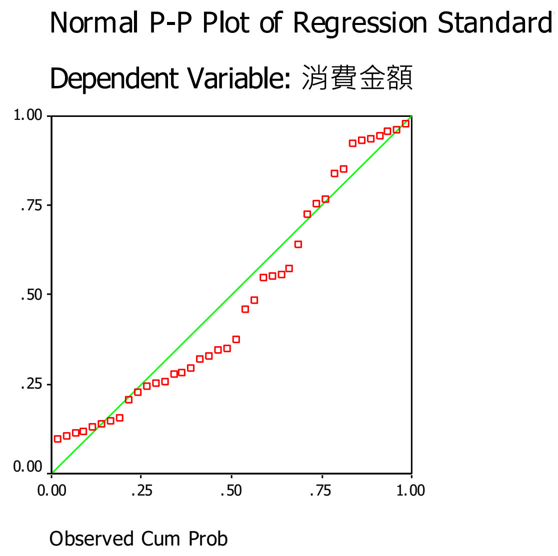
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-339.66	5831.53	2971.43	1951.69	40
Residual	-425.94	669.66	8.19E-13	326.56	40
Std. Predicted Value	-1.697	1.465	.000	1.000	40
Std. Residual	-1.287	2.024	.000	.987	40

a. Dependent Variable: 消費金額

⑨.標準化殘差次數分配表顯示殘差的機率分配不接近常態機率分布。屬向左偏斜態勢。



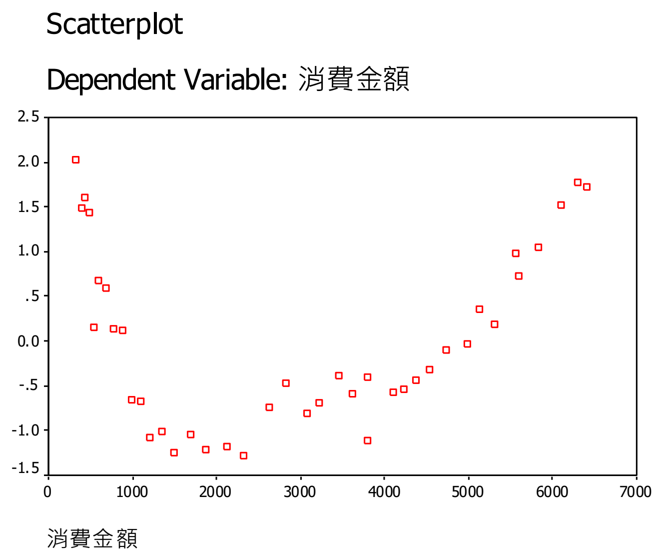
⑩.標準化殘差常態機率分配 P-P 圖，不接近常態機率分配。



⑪.標準化殘差對依變數散佈圖：顯示明顯的圖樣，故此迴歸模式並不適當。

⑫.結論：從基本的迴歸分析，可獲得此迴歸模式均符合所需的檢定，且判定係數很高。惟從殘差分析中，獲得此迴歸模式並不符合假設，故此模式並不適當。然而此模式的判定係數高達 97.3%，顯示利用服務滿意度指標來解釋(預測)消費金額是很恰當，不需要再考慮其他的變數。

⑬.以下將利用變數轉換的方法，繼續探討一較合適的迴歸模式。考慮的變數轉換方法：將消費金額(依變數/因變數)取對數、開平方和平方，再利用服務滿意度指標(自變數)進行迴歸分析。



14.10.3 簡單線性迴歸分析變數轉換 SPSS 操作方法

1. 針對依變數(因變數)消費金額進行數值對數轉換，在 SPSS 軟體中選取 Tranform→Compute…，即會出現 Compute Variable 對話方塊。
2. 在 Compute Variable 對話視窗中左上角 Target Variable: 下面的空格中輸入數值轉換後的變數名稱 cost_log，在右邊的 Functions: 下面選項中選取 LG10(numexpr)函數進入上面的 Numeric Expression: 空格中，再將欲轉換的依變數：消費金額(cost_in)選入 Numeric Expression: 空格中，其方程式表示為 LG10(cost_in)。

3. 在 Compute Variable 對話視窗中，勾選下面的 OK 按鈕，以執行數值轉換程序，即會在 SPSS Data Editor 視窗中出現 cost_log 的變數欄位。
4. 針對依變數(因變數)消費金額進行數值開平方(開根號)轉換，在 SPSS 軟體中選取 Tranform→Compute…，即會出現 Compute Variable 對話視窗。
5. 在 Compute Variable 對話視窗中左上角 Target Variable: 下面的空格中輸入數值轉換後的變數名稱 cost_ns，在右邊的 Numeric Expression: 空格中，將欲轉換的依變數：消費金額(cost_in)選入，其方程式表示為 cost_in**0.5。
6. 在 Compute Variable 對話視窗中，勾選下面的 OK 按鈕，以執行數值轉換程序，即會在 SPSS Data Editor 視窗中出現 cost_ns 的變數欄位。

14.10.4 對數轉換依變數(消費金額)的分析結果

Descriptive Statistics

	Mean	Std. Deviation	N
消費金額取對數	3.3336	.3954	40
服務滿意度指標	180.90	16.45	40

①.兩變數的相關係數為 0.984

Correlations

		消費金額取對數	服務滿意度指標	
Pearson Correlation	消費金額取對數	1.000	.984	
	服務滿意度指標	.984	1.000	
Sig. (1-tailed)	消費金額取對數	.	.000	
	服務滿意度指標	.000	.	
N	消費金額取對數	40	40	
	服務滿意度指標	40	40	

②.判定係數(R^2)為 0.969，此模式的解釋能力(預估能力)相當高，達 96.9 %。

③.統計量 Durbin-Watson 為 0.142，相當接近於 0，可判定殘差彼此間相關性相當強。

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	.984[a]	.969	.968	7.030E-02	.969	1196.057	1	38	.000	.142

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額取對數

④.適合性檢定：從變異數分析中顯示 P 值(Sig.)為 0.000，達到顯著水準 0.05，表示此模式適合利用服務滿意度指標(自變數)來解釋(預估)消費金額取對數(依變數/因變數)。

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.911	1	5.911	1196.057	.000[a]
	Residual	.188	38	4.942E-03		
	Total	6.099	39			

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額取對數

⑤.簡單線性迴歸方程式：log(Y)(消費金額取對數) = -0.949 + $2.367*10^{-2}$*X(服務滿意度指標)

⑥.方程式的斜率與截距的檢定 P 值(Sig.)為 0.000 達到 0.05 顯著水準，故斜率與截距在方程式中均是存在，不為數值 0。

⑦.因為獨立變數只有一個，故 Tolerance 和 VIF 值均為 1.000。因此，當獨立變數只有一個時，其 Tolerance 和 VIF 值均沒有意義。

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)	-.949	.124		-7.632	.000	-1.200	-.697		
	服務滿意度指標	2.367E-02	.001	.984	34.584	.000	.022	.025	1.000	1.000

a. Dependent Variable: 消費金額取對數

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	服務滿意度指標
1	1	1.996	1.000	.00	.00
	2	4.005E-03	22.325	1.00	1.00

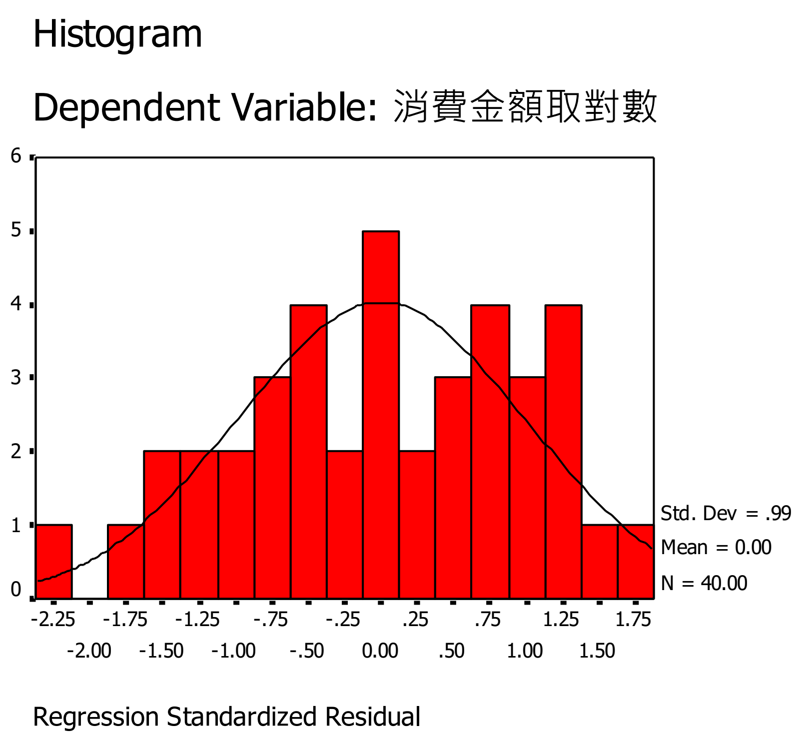
a. Dependent Variable: 消費金額取對數

Residuals Statistics[a]

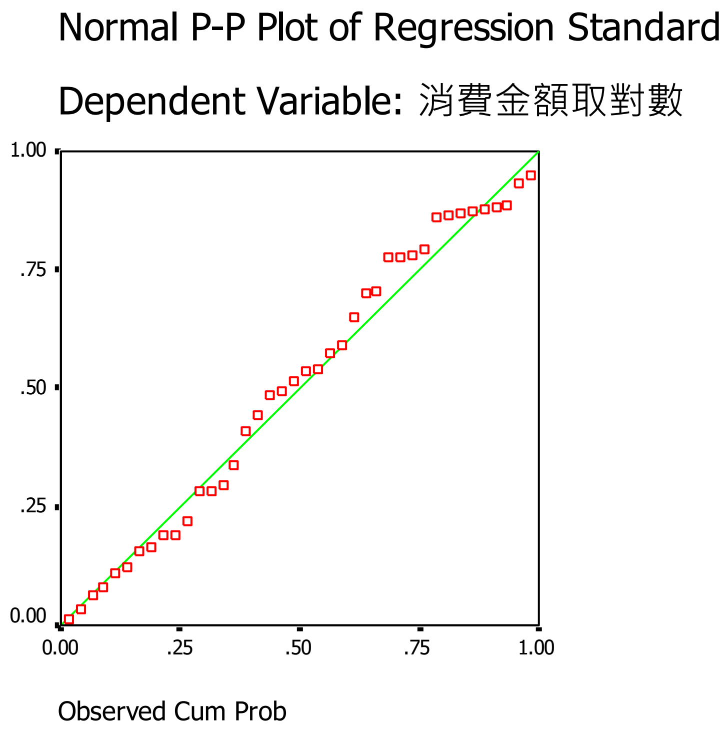
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	2.6732	3.9041	3.3336	.3893	40
Residual	-.1546	.1158	6.661E-16	6.939E-02	40
Std. Predicted Value	-1.697	1.465	.000	1.000	40
Std. Residual	-2.200	1.647	.000	.987	40

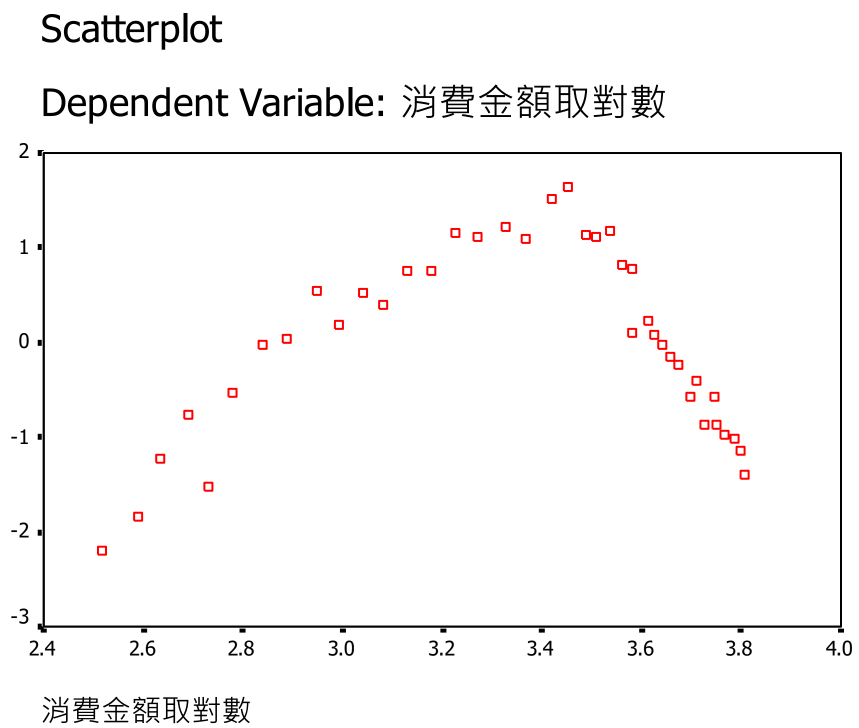
a. Dependent Variable: 消費金額取對數

⑧.標準化殘差次數分配表顯示殘差的機率分配不接近常態機率分布。屬稍向右偏斜態勢。惟與原始消費金額所獲得的機率分布圖相比，消費金額經取對數後分析所獲得的機率分布比較趨近常態分布。



⑨.標準化殘差常態機率分配 P-P 圖，比較接近常態機率分布。





⑩.標準化殘差對依變數散佈圖：顯示明顯的圖樣，故此迴歸模式並不適當。

⑪.結論：從基本的迴歸分析，可獲得此迴歸模式均符合所需的檢定，且判定係數很高。惟從殘差分析中，獲得此迴歸模式並不符合假設，故此模式並不適當。

14.10.5 開平方轉換依變數(消費金額)的分析結果

Descriptive Statistics

	Mean	Std. Deviation	N
消費金額開平方	50.8487	19.8928	40
服務滿意度指標	180.90	16.45	40

①.消費金額開平方與服務滿意度指標的相關係數為 0.999。

Correlations

		消費金額開平方	服務滿意度指標
Pearson Correlation	消費金額開平方	1.000	.999
	服務滿意度指標	.999	1.000
Sig. (1-tailed)	消費金額開平方	.	.000
	服務滿意度指標	.000	.
N	消費金額開平方	40	40
	服務滿意度指標	40	40

②.判定係數(R^2)為 0.997，此模式的解釋能力(預估能力)最高，達 99.7 %。

③.統計量 Durbin-Watson 為 1.333，比較接近於 2，可判定殘差彼此間相關性相對較低。

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	.999[a]	.997	.997	1.0196	.997	14808.173	1	38	.000	1.333

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額開平方

④.適合性檢定：從變異數分析中顯示 P 值(Sig.)為 0.000，達到顯著水準 0.05，表示此模式適合利用服務滿意度指標(自變數)來解釋(預估)消費金額開平方(依變數/因變數)。

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	15393.787	1	15393.787	14808.173	.000[a]
	Residual	39.503	38	1.040		
	Total	15433.290	39			

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額開平方

⑤.簡單線性迴歸方程式：$\sqrt{Y}$(消費金額開平方) = -167.693 + 1.208*X(服務滿意度指標)

⑥.方程式的斜率與截距的檢定 P 值(Sig.)為 0.000 達到 0.05 顯著水準，故斜率與截距在方程式中均是存在，不為數值 0。

⑦.因為獨立變數只有一個，故 Tolerance 和 VIF 值均為 1.000。因此，當獨立變數只有一個時，其 Tolerance 和 VIF 值均沒有意義。

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	95% Confidence Interval for B		Collinearity Statistics	
Model		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-167.693	1.803		-93.001	.000	-171.343	-164.043		
	服務滿意度指標	1.208	.010	.999	121.689	.000	1.188	1.228	1.000	1.000

a. Dependent Variable: 消費金額開平方

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	服務滿意度指標
1	1	1.996	1.000	.00	.00
	2	4.005E-03	22.325	1.00	1.00

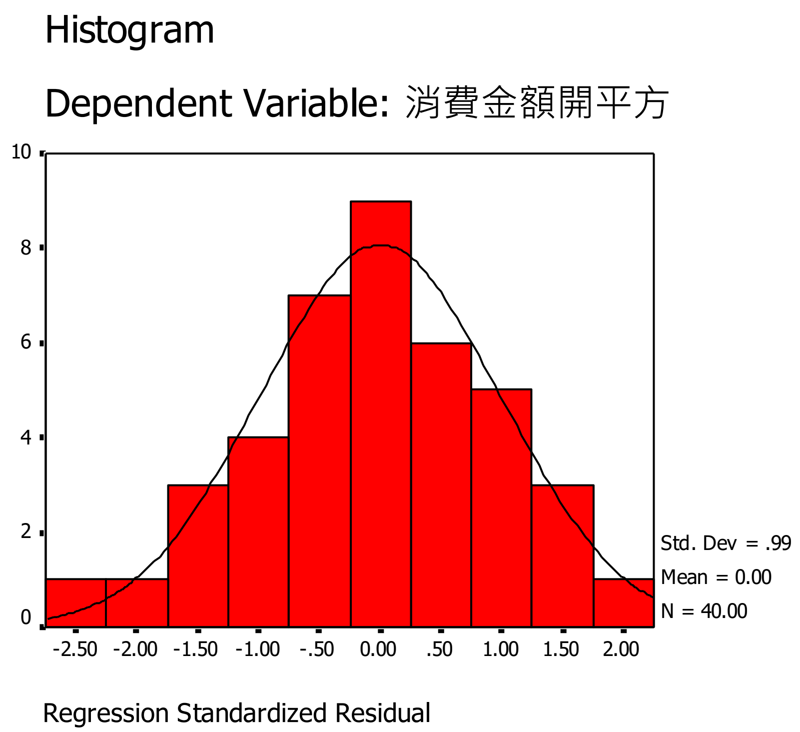
a. Dependent Variable: 消費金額開平方

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	17.1433	79.9635	50.8487	19.8674	40
Residual	-2.3619	2.2282	-6.5725E-15	1.0064	40
Std. Predicted Value	-1.697	1.465	.000	1.000	40
Std. Residual	-2.317	2.185	.000	.987	40

a. Dependent Variable: 消費金額開平方

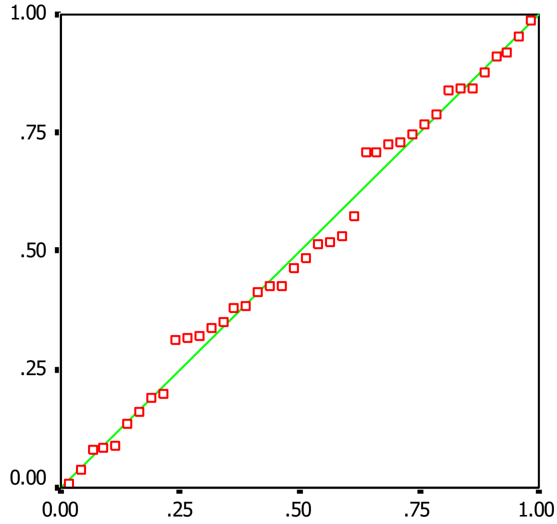
⑧.標準化殘差次數分配表顯示殘差的機率分配較接近常態機率分布。



⑨.標準化殘差常態機率分配 P-P 圖，較接近常態機率分配。

Normal P-P Plot of Regression Standard

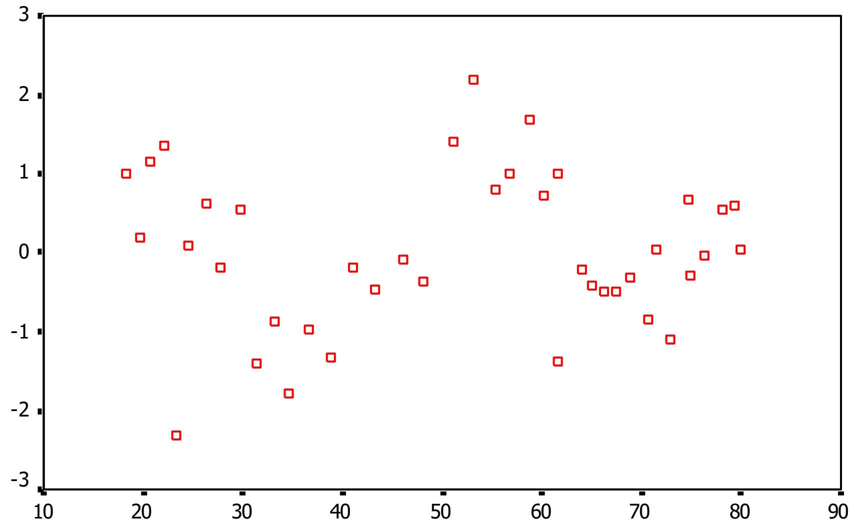
Dependent Variable: 消費金額開平方



Observed Cum Prob

Scatterplot

Dependent Variable: 消費金額開平方



消費金額開平方

⑭.標準化殘差對依變數散佈圖：顯示沒有明顯的圖樣，故此迴歸模式較適當。

⑮.結論：從基本的迴歸分析，可獲得此迴歸模式均符合所需的檢定，且判定係數很高。從殘差分析中，獲得此迴歸模式並較符合假設，故此模式最適當。

14.10.6 平方轉換依變數(消費金額)的分析結果

Descriptive Statistics

	Mean	Std. Deviation	N
消費金額平方	12647194.8250	12856899.1387	40

	Mean	Std. Deviation	N
服務滿意度指標	180.90	16.45	40

Correlations

		消費金額平方	服務滿意度指標
Pearson Correlation	消費金額平方	1.000	.923
	服務滿意度指標	.923	1.000
Sig. (1-tailed)	消費金額平方	.	.000
	服務滿意度指標	.000	.
N	消費金額平方	40	40
	服務滿意度指標	40	40

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	.923[a]	.853	.849	5001691.887 1	.853	219.693	1	38	.000	.105

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額平方

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.496E+15	1	5.496E+15	219.693	.000[a]
	Residual	9.506E+14	38	2.502E+13		
	Total	6.447E+15	39			

a. Predictors: (Constant), 服務滿意度指標

b. Dependent Variable: 消費金額平方

Coefficients

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)	-1.18E+08	8.85E+06		-13.333	.000	-1.36E+08	-1.00E+08		
	服務滿意度指標	7.22E+05	4.87E+04	.923	14.822	.000	6.23E+05	8.20E+05	1.000	1.000

a. Dependent Variable: 消費金額平方

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions	
				(Constant)	服務滿意度指標
1	1	1.996	1.000	.00	.00
	2	4.005E-03	22.325	1.00	1.00

a. Dependent Variable: 消費金額平方

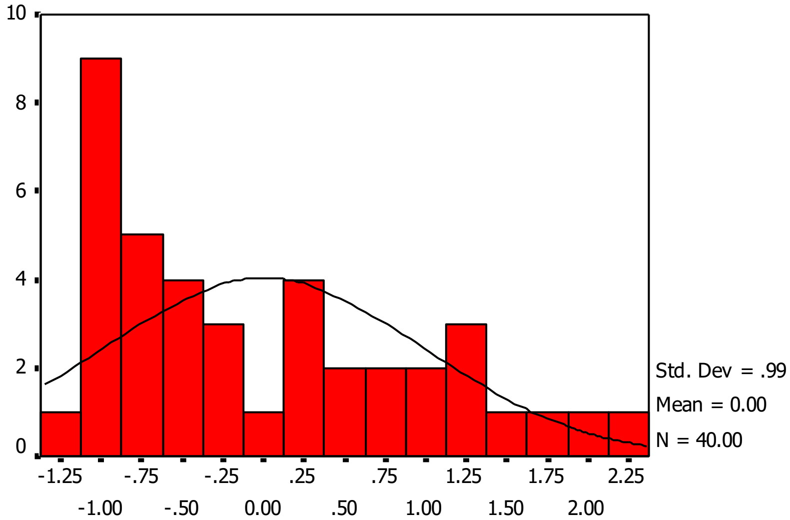
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-7.49E+06	3.00E+07	1.26E+07	1.19E+07	40
Residual	-5.89E+06	1.09E+07	4.66E-10	4.94E+06	40
Std. Predicted Value	-1.697	1.465	.000	1.000	40
Std. Residual	-1.178	2.182	.000	.987	40

a. Dependent Variable: 消費金額平方

Histogram

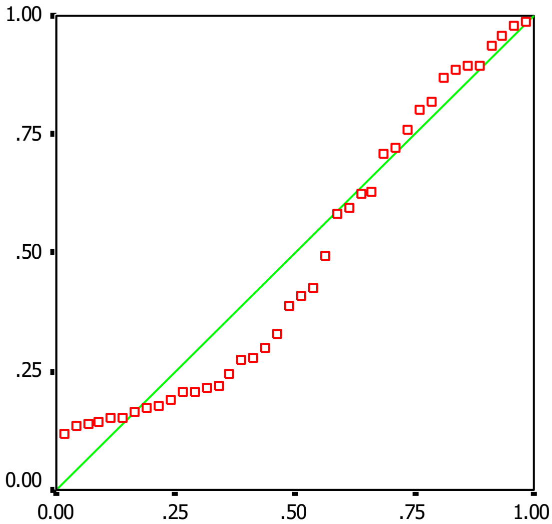
Dependent Variable: 消費金額平方



Regression Standardized Residual

Normal P-P Plot of Regression Standard

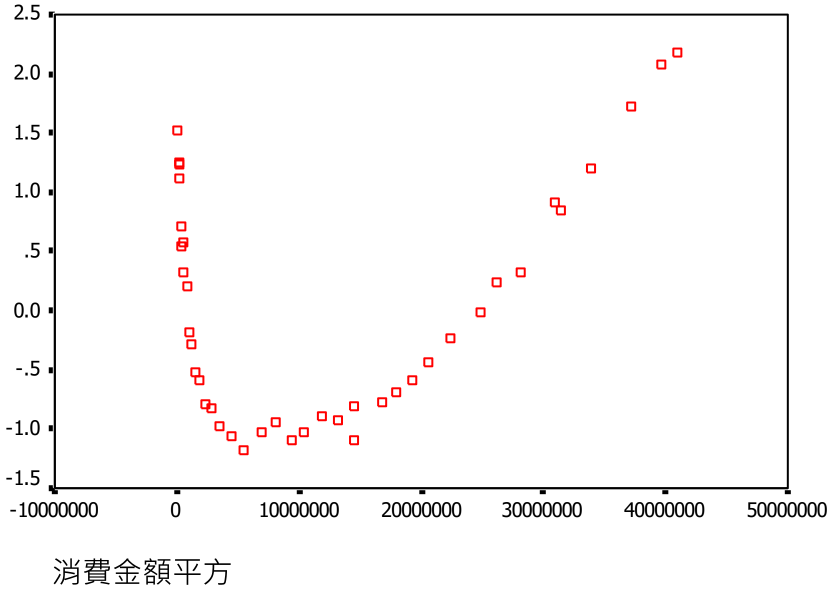
Dependent Variable: 消費金額平方



Observed Cum Prob

Scatterplot

Dependent Variable: 消費金額平方



討論議題

1.師生非同步教學討論議題：簡單線性迴歸分析應用

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 14 章學習的【簡單線性迴歸分析】課程內容，陳述現在或未來最想運用到情境和理由。請具體的標示您想運用的情境下，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)分別的具體名稱(變數)(20 個字以上)。

待有25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。

2.學習者非同步教學討論議題：期末考分數預測

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：「最重要因素」，本文：請針對第 14 章學習的【簡單線性迴歸分析】課程內容與應用價值，在疫情期間三級警戒期間，無法到校進行期末考評量，僅能夠在宿舍或住家同步進行。自己認為哪一個因素是影響期末考分數的最關鍵因素，此因素可以納入在簡單線性迴歸分析中，當成自變數分析使用。請具體列出標示最關鍵的影響因素(自變數)，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：「關鍵因素」，本文：透過相互分享討論比較後，自己最後認為哪一個是影響期末考分數的關鍵因素？並詮釋接下來的因應作為(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)	$y_i = \beta_0 + \beta_1 \times x_i$
簡單線性迴歸模式(simple linear regression model)	$y_i = \beta_0 + \beta_1 \times x_i + \varepsilon_i$
迴歸方程式(regression equation)	$E(y_i) = \beta_0 + \beta_1 \times x_i$
估計迴歸方程式(estimated regression equation)	$\hat{y}_i = b_0 + b_1 \times x_i$

最小平方法數學法則

$$\text{Min SSE} = \min \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i)^2$$

利用微分方式獲得估計迴歸方程式的**斜率**(slope)b_1 和**截距**(intercept)b_0

斜率 $b_1 = \dfrac{\sum_{i=1}^{n}(x_i \times y_i) - \frac{\sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \dfrac{n \times \sum_{i=1}^{n}(x_i \times y_i) - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n} y_i}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \dfrac{\sum_{i=1}^{n}[(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \dfrac{\frac{\sum_{i=1}^{n}[(x_i - \bar{x}) \times (y_i - \bar{y})]}{n-1}}{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \dfrac{S_{xy}}{S_x^2}$

截距 $b_0 = \bar{y} - b_1 \times \bar{x} = \dfrac{\sum_{i=1}^{n} y_i \times \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \times \sum_{i=1}^{n}(x_i \times y_i)}{n \times \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$

判定係數(coefficient of determination)即是迴歸造成的平方和(迴歸項平方和、可解釋的變異)佔總平方和(總變異)的比例，常使用 R^2 或 r^2 符號代表。R^2 數值範圍 0~1，愈靠近 1 迴歸方程式的適配度愈高。

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

樣本相關係數(Sample correlation coefficient) R_{xy}、r_{xy} 或 γ_{xy}

$$R_{xy} = r_{xy} = (b_1\text{ 的正負符號}) \times \sqrt{\text{判定係數}} = (b_1\text{ 的正負符號}) \times \sqrt{R^2}$$

$$r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x}) \times (y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \times \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} =$$

$$\frac{\sum_{i=1}^{n} x_i \times y_i - \frac{\sum_i^n x_i \times \sum_i^n y_i}{n}}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \times \sqrt{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}}$$

迴歸方程式斜率 β_1 之估計值 b_1 之抽樣分布

b_1 期望值 $E(b_1) = \beta_1$

b_1 標準(偏)差$\sigma_{b_1} = \dfrac{\sigma}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$

分布方式屬於常態分布。

若誤差項 ε_i 之標準(偏)差 σ 未知時，可以利用誤差項 ε_i 之標準(偏)差 σ 的估計值 S 取代標準(偏)差，以獲得 b_1 標準(偏)差σ_{b_1}的估計值S_{b_1}。

b_1 標準(偏)差σ_{b_1}的估計值$S_{b_1} = \dfrac{s}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$

$$t\text{ 檢定統計值 } t = \frac{b_1}{S_{b_1}}$$

	t 值檢定	F 值檢定
一個自變數	●	●
兩個(含)以上自變數	X	●
檢定斜率 β_1 是否等於特定數值(C)	●	X
左右尾檢定	●	X

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: $\beta_1 = 0$。

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

D.計算檢定統計值—F 值

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\text{迴歸造成的均方}}{\text{誤差均方}}$$

E.若檢定統計值 $F <$ 臨界值 $F_{\alpha,1,n\text{-}2}$，接受虛無假設 H_0: $\beta_1 = 0$。

F.若檢定統計值 $F >$ 臨界值 $F_{\alpha,1,n\text{-}2}$，拒絕虛無假設 H_0: $\beta_1 = 0$，接受對立假設 H_1: $\beta_1 \neq 0$。

	斜率 $\beta_1 = 0$	斜率 $\beta_1 \neq 0$
R^2 數值高	自變數對依變數沒有意義	自變數對依變數預測能力佳，具有意義
R^2 數值低	自變數對依變數沒有意義	自變數對依變數預測能力弱，具有意義，影響力小

迴歸方程式中預測依變數之平均值$\bar{y}_0$的信賴區間為

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

$$\hat{y}_0 \pm t_{n-2,\frac{\alpha}{2}} \times S \times \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$$

迴歸方程式中預測依變數 y_0 的信賴區間為

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

$$\hat{y}_0 \pm t_{n-2,\frac{\alpha}{2}} \times S \times \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}$$

關鍵詞彙解釋

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)。

最小平方法(Least squares method)

利用樣本資料中的自變數 x_i 和依變數 y_i(實際觀測值)的對應數值，並使用自變數 x_i、截距 b_0 和斜率 b_1 推算依變數 y_i 的估計值$\hat{y}_i$，使得依變數 y_i 和其估計值$\hat{y}_i$的差(距)之平方和(sum square error, SSE)為最小數值，此為最小平方法(Least squares method)或普通最小平方法(ordinary least squares method, OLS)的特性。

最小平方法數學法則

$$\text{Min SSE} = \min \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \min \sum_{i=1}^{n}(y_i - b_0 - b_1 \times x_i)^2$$

其中 $y_i =$ 依變數第 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(estimator)；依變數第 i 個觀測值的估計值

x_i = 自變數第 i 個觀測值

b_0 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_0 的估計值(estimator)，截距(intercept)或常數項(constant term)。數值可能範圍-∞~+∞。

b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中，參數(parameter) β_1 的估計值(estimator)，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

判定係數(coefficient of determination)

判定係數或決定係數(coefficient of determination)即是迴歸造成的平方和(迴歸項平方和、可解釋的變異)佔總平方和(總變異)的比例，常使用 R^2 或 r^2 符號代表。R^2 數值範圍 0~1，愈靠近 1 迴歸方程式的適配度愈高。

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} = \frac{\text{迴歸可解釋變異量}}{\text{總變異量}} = \frac{SST-SSE}{SST} = 1-\frac{SSE}{SST} = 1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

判定係數可以評量迴歸方程式的**適配度**，亦可評量迴歸方程式的**解釋能力**。

相關係數(correlation coefficient)

當兩個隨機變數的關係屬於不獨立(有相互關係)時，並呈現線性相關，表達正負向關係和關係強弱者，即為相關係數。

樣本相關係數(Sample correlation coefficient) R_{xy}、r_{xy} 或 γ_{xy}

$$R_{xy} = r_{xy} = (b_1\text{的正負符號}) \times \sqrt{\text{判定係數}} = (b_1\text{的正負符號}) \times \sqrt{R^2}$$

$$r_{xy} = \frac{cov(x,y)}{S_x \times S_y} = \frac{S_{xy}}{S_x \times S_y} = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\times\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\times\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{\sum_{i=1}^{n}x_i \times y_i - \frac{\sum_{i}^{n}x_i \times \sum_{i}^{n}y_i}{n}}{\sqrt{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}\times\sqrt{\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}}}$$

其中 b_1 = 迴歸模式 $E(y_i) = \beta_0 + \beta_1 \times x_i$ 中參數(parameter) β_1 的估計值，迴歸係數(regression coefficient)或斜率(slope)。數值可能範圍-∞~+∞。

R^2 = 判定係數(coefficient of determination)。數值範圍 0~1。