

Simple Regression Analysis and Correlation Analysis 簡單迴歸分析與相關分析

Jong Yih Kuo

jykuo@ntut.edu.tw

**Department of Computer Science and
Information Engineering
National Taipei University of Technology**

介紹

□ 迴歸分析目的

- 探究一個或數個自變數(Independent Variable) 和一個依變數(Dependent variable) 間的關係
 - 自變數(解釋變數)以 X 表示，依變數(被解釋變數)以 Y 表示。
 - 自變數 X 與依變數 Y 間的函數關係或數學方程式，稱迴歸模式。
- 建構數學方程式，解釋或預測因變數之值。

□ 案例

- 某股票分析師想建立臺灣某上市公司股價(Y)與該公司各項財務指標(X)之迴歸模式，以準確預測該公司股價。
- 某工程師想建立某種化學合成反應物含量(Y)與其合成時間(X)迴歸模式，預測該反應物含量。
- 大學附近某房屋仲介想藉由學生套房坪數大小(X)與套房月租(Y)之迴歸模式，預測學生套房之月租。

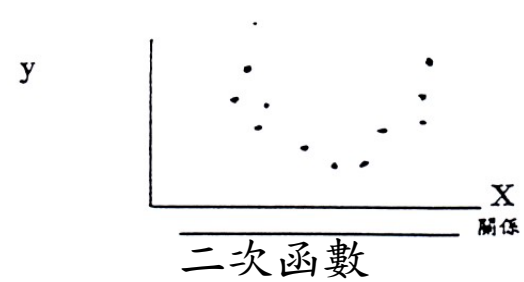
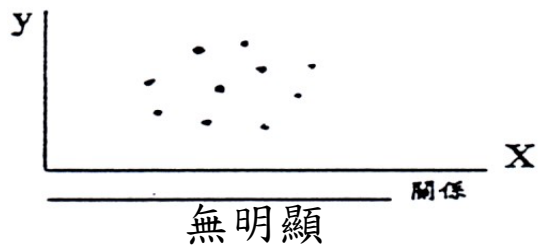
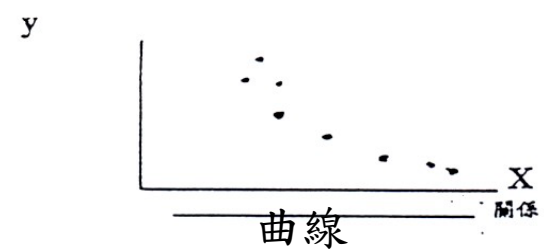
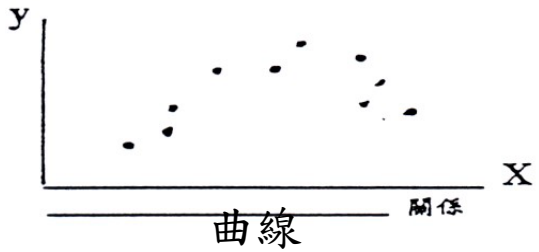
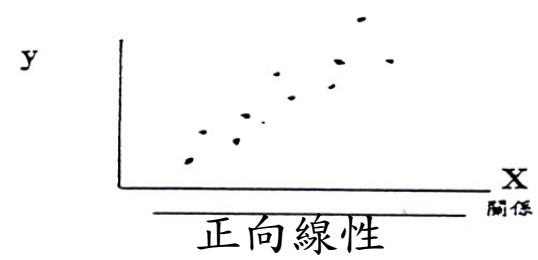
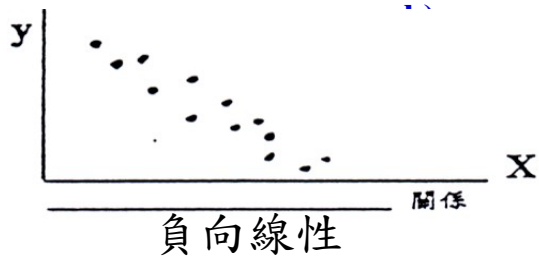
簡單迴歸與複迴歸

- 影響因變數之自變數通常不只一個，例如
 - 房價(Y) 可能與房子大小(x1)、房齡(x2)、離市區距離(x3)、有無空調(x4)等因素都有關，
 - 這些因素是否均會影響因變數？可利用迴歸分析找出影響房價最重要的因素，作為預測房價的重要根據。
- 簡單迴歸(Simple Regression)
 - 如何由房子大小預測房價
- 複迴歸或多元迴歸 (Multiple Regression)
 - 考慮多個自變數
 - 如何由房子大小、房齡...等預測房價？

散布圖

- ❑ 進行迴歸分析前，需先瞭解變數間呈現何種關係，才能建構適當數學方程式或迴歸模式。
- ❑ 利用散佈圖(Scatter Diagram)決定兩變數間的關係
 - 將X變數標示於橫座標，Y變數標示於縱座標，將各(X, Y)對應值點繪在X-Y二維座標，觀察點之變化。
 - 正相關(Positive Relationship)
 - 假如X增加Y增加；或X減少Y減少，稱為X與Y正相關。
 - 負相關(Negative Relationship)
 - 假如X增加Y減少；或X減少Y增加，稱為X與Y有負相關。
 - 不相關(No Relationship)
 - 散佈圖中之點大部份與水平軸平行，看不出任何特殊圖形。

散布圖



回歸分析步驟

- 迴歸分析資料收集時須注意
 - 資料須能代表所研究的系統或問題。
 - 迴歸分析前須先確定資料不含離群值。
- 迴歸模式用處
 - 描述資料。
 - 預測與估計依變數之值。

簡單迴歸分析

□ 決定簡單直線迴歸模式

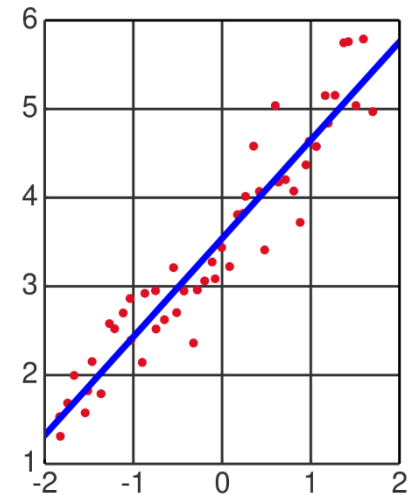
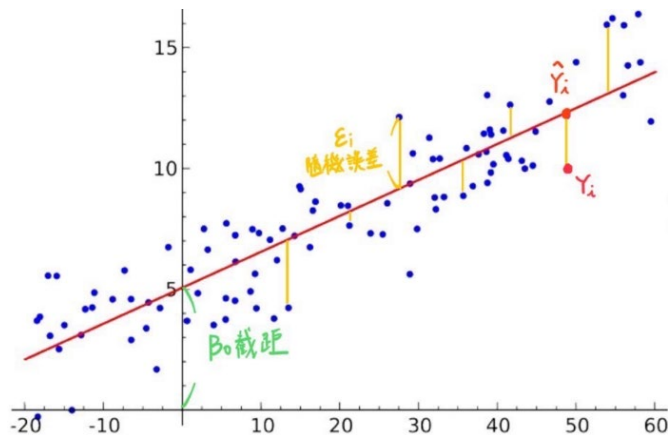
○ 由散佈圖看出自變數與依變數間關係。最簡單關係為直線。

□ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

○ Y_i 表第 i 個觀測值； X_i 表對應第 i 個觀測值之自變數值；

○ β_0 為截距； β_1 為斜率，表示自變數每增加一單位，依變數 Y 的改變量；

○ ε_i 表殘差項或隨機誤差。



簡單迴歸分析

□ 決定與觀測值最適配直線迴歸模式，方程式： $\hat{Y}_i = b_0 + b_1 X_i$

○ 最小平方方法(Least Squares Method)

Min SSE = $\min \sum (Y_i - \hat{Y}_i)^2 = \min \sum (Y_i - b_0 + b_1 X_i)^2$ ，利用微分方式獲得

- \hat{Y}_i 在自變數 X_i 時，依變數 Y_i 的估計值。
- b_0 與 b_1 分別為 β_0 與 β_1 之估計值。
- 找出 b_0 與 b_1 ，使 $\sum (Y_i - \hat{Y}_i)^2$ 為最小。
 - 殘差項(Residual) $\varepsilon_i = Y_i - \hat{Y}_i$ ， $\sum \varepsilon_i = 0$ (理想情況)

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^N x_i^2 - n\bar{x}^2}$$

(參考資料推導)

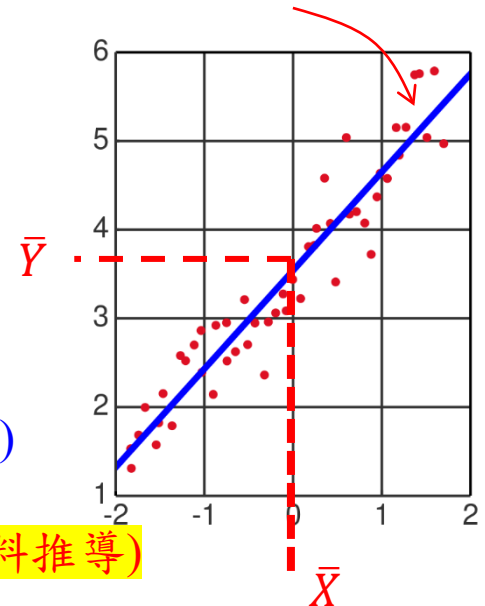
$$b_0 = \bar{Y} - b_1 \bar{X}$$

$$SS_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{1}{n-1} \left[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} \right]$$

(參考假設檢定講義，下一頁推導)

$$SS_x = S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1} = \frac{1}{n-1} \left[\sum_{i=1}^N x_i^2 - n\bar{x}^2 \right]$$

SS_{xy} : 樣本 XY 共變異數， SS_x : 樣本 X 變異數，樣本: $1/n-1$



公式推導 (參考)

$$SS_x = S^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n-1} = \frac{1}{n-1} [\sum_{i=1}^N x_i^2 - n\bar{x}^2]$$

$$\sum_{i=1}^N (x_i - \bar{x})^2 = \sum_{i=1}^N (x_i^2 - 2x_i\bar{x} - \bar{x}^2) = \sum_{i=1}^N x_i^2 - \sum_{i=1}^N x_i 2\bar{x} + \sum_{i=1}^N \bar{x}^2$$

$$= \sum_{i=1}^N x_i^2 - 2\bar{x} \sum_{i=1}^N x_i + n\bar{x}^2$$

$$= \sum_{i=1}^N x_i^2 - 2\bar{x}n\bar{x} + n\bar{x}^2$$

$$= \sum_{i=1}^N x_i^2 - n\bar{x}^2$$

$$= \sum_{i=1}^N x_i^2 - \frac{(\sum_{i=1}^N x_i)^2}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\Rightarrow \sum_{i=1}^n x_i = n\bar{x}$$

$$S_{xy} = SS_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{1}{n-1} [\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}]$$

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x} \bar{y})$$

$$= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \bar{x} y_i - \sum_{i=1}^n x_i \bar{y} + \sum_{i=1}^n \bar{x} \bar{y}$$

$$= \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} + \frac{n \times \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \times n}$$

$$= \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}$$

Exercise

- 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 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ，其中 ε_i 為誤差項。請利用最小平方法計算出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

評估迴歸模式－判定係數

□ 迴歸模式中，評估自變數預測因變數的能力：

○ 不可解釋變異量或隨機變異 $SSE = \sum (Y_i - \hat{Y}_i)^2$

$$\varepsilon_i = Y_i - \hat{Y}_i$$

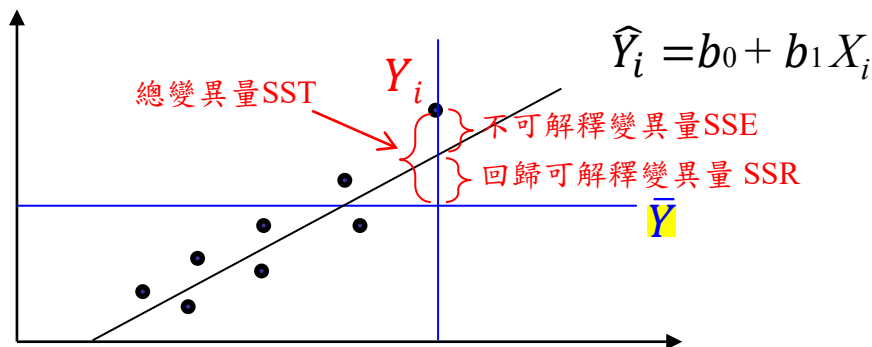
➢ 最小平方法，期望獲得殘差項之平方和是最小數值者，此數值為殘差項平方和 (Sum of Squares due to Error, SSE)。

○ 迴歸可解釋變異量 (Sum of Squares due to Regression, SSR) =

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 = \frac{[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}]^2}{\sum_{i=1}^N x_i^2 - n\bar{x}^2} = \frac{[(n-1) \times SS_{xy}]^2}{(n-1) \times SS_x} \quad (\text{參考資料推導})$$

○ 總變異量 (Sum of Squares Total, SST)

$$SST = \sum (Y_i - \bar{Y})^2 = \sum_{i=1}^N y_i^2 - \frac{(\sum_{i=1}^N y_i)^2}{n} = (n-1) \times SS_y \quad (\text{參考上一頁推導})$$



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

評估迴歸模式－判定係數

▣ 判斷X對預測Y提供有用資訊(或迴歸方程式是否顯著)

○ 由圖形判定(限簡單迴歸模式)

➢ 資料點與迴歸方程式越接近，表示迴歸模式越有用。

○ 判定係數 r^2 (Coefficient of Determination)

➢ 衡量自變數(X)所能解釋依變數(Y)之變異量占Y總變異量百分比

$$r^2 = \frac{\text{迴歸式可解釋的變異量}}{\text{總變異量}} = \frac{SSR}{SST} = \frac{\frac{((n-1) \times SS_{xy})^2}{(n-1) \times SS_x}}{(n-1) \times SS_y} = \frac{(SS_{xy})^2}{SS_x \times SS_y} \quad (\text{參考上一頁})$$

r^2 值越近 1 越好， $(0 \leq r^2 \leq 1)$

$r = 0$ 並不一定表示Y與X間沒有關係，僅表示Y與X間無線性關係。

○ 當兩個隨機變數的關係不獨立，並呈現線性相關，表達正負向關係和關係強弱，即為相關係數。 R_{xy}

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

(參考下一頁推導)

相關係數與判定係數推導(參考)

$$\begin{aligned} R_{xy} &= \frac{cov(x,y)}{SS_x \times SS_y} = \frac{SS_{xy}}{SS_x \times SS_y} = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \times \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\ &= \frac{\frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \times \sum_{i=1}^n y_i}{n}}{n}}{\sqrt{\sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n x_i^2}{n}} \times \sqrt{\sum_{i=1}^n y_i^2 - \frac{\sum_{i=1}^n y_i^2}{n}}} \end{aligned}$$

$$\text{判定係數 } r^2 = \frac{SSR}{SST} = \frac{\frac{((n-1) \times SS_{xy})^2}{(n-1) \times SS_x}}{(n-1) \times SS_y} = \frac{(SS_{xy})^2}{SS_x \times SS_y}$$

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

$$SSR = \sum (\hat{Y}_i - \bar{Y})^2 = \frac{[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}]^2}{\sum_{i=1}^N x_i^2 - n\bar{x}^2} = \frac{[(n-1) \times SS_{xy}]^2}{(n-1) \times SS_x}$$

Exercise

- 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

評估迴歸模式 – 統計檢定

□ 統計檢定

- 假設X與Y間完全無關(在預測Y值上，X幾乎未提供任何有用資訊)，則線性模式： $\bar{Y} = \beta_0 + \beta_1 X$ 中， β_1 值應0
- 檢定 β_1 。假如 $H_0: \beta_1 = 0$ 被拒絕，則可下結論認為有足夠證據顯示「X與Y間有顯著線性關係」或「X對預測Y提供有用資訊」。
- 迴歸分析之顯著性檢定須依下列殘差項 ε_i 的假設條件。
 - 各觀測點之 ε_i 平均值為0。
 - 各觀測點之 ε_i 之間相互獨立。
 - 共變異數 $Cov(\varepsilon_i, \varepsilon_j) = 0$, at $i \neq j$, $i, j = 1, \dots, n$ 。任兩殘差項不相關。
 - 各觀測點殘差項 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)。
 - 各觀測點之殘差項 ε_i 屬於常態分布 $\varepsilon_i \sim N(0, \sigma^2)$ 。

評估迴歸模式- ANOVA

NID : Normally and Independently Distributed

□ ANOVA的F檢定程序，假設： $\varepsilon_i \sim \text{NID}(0, \sigma^2)$

○ **H0**: $\beta_1=0$ (X與Y之間沒有線性關係；或對預測Y而言，迴歸模式無法提供有用之資訊)

○ **H1**: $\beta_1 \neq 0$ (X與Y之間有線性關係，即斜率不為0；或在預測Y上，迴歸模式有用)

○ 設定 α 值

○ 檢定值： $F = \frac{MSB}{MSW} = \frac{\frac{SS_B}{k-1}}{\frac{SS_w}{n-k}}$

$$SS_B = \sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2$$

$$SS_w = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2$$

○ 拒絕域：查F-表，自由度=(1, n-2)或計算p-值

○ 下結論

k = 2

變異來源	平方和	自由度	Mean Sum of Square	F
迴歸可解釋變異量	$SS_B = SSR$	$k-1=1$	$MSB = \frac{SSR}{1}$	$\frac{MSB}{MSW}$
殘差(不可解釋)變異量	$SS_w = SSE$	$n-k = n-2$	$MSW = \frac{SSE}{n-2}$	
總變異量	SST	$n-1$		

Exercise

- 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 $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ ，其中 ε_i 為誤差項。請利用ANOVA，設定 $\alpha = 0.05$ ，計算F值以及結論。

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

評估迴歸模式 – t 檢定

□ t 檢定程序，假設： $\varepsilon_i \sim \text{NID}(0, \sigma^2)$

○ $H_0: \beta_1 = \beta_1^*$ ， $H_1: \beta_1 \neq \beta_1^*$

○ 設定 α 值

○ 檢定值： $t = (b_1 - \beta_1^*) / \sqrt{MSE / SSX}$

○ 拒絕域：查 t-表，自由度=(1, n-2)或計算p-值

○ 下結論

範例一

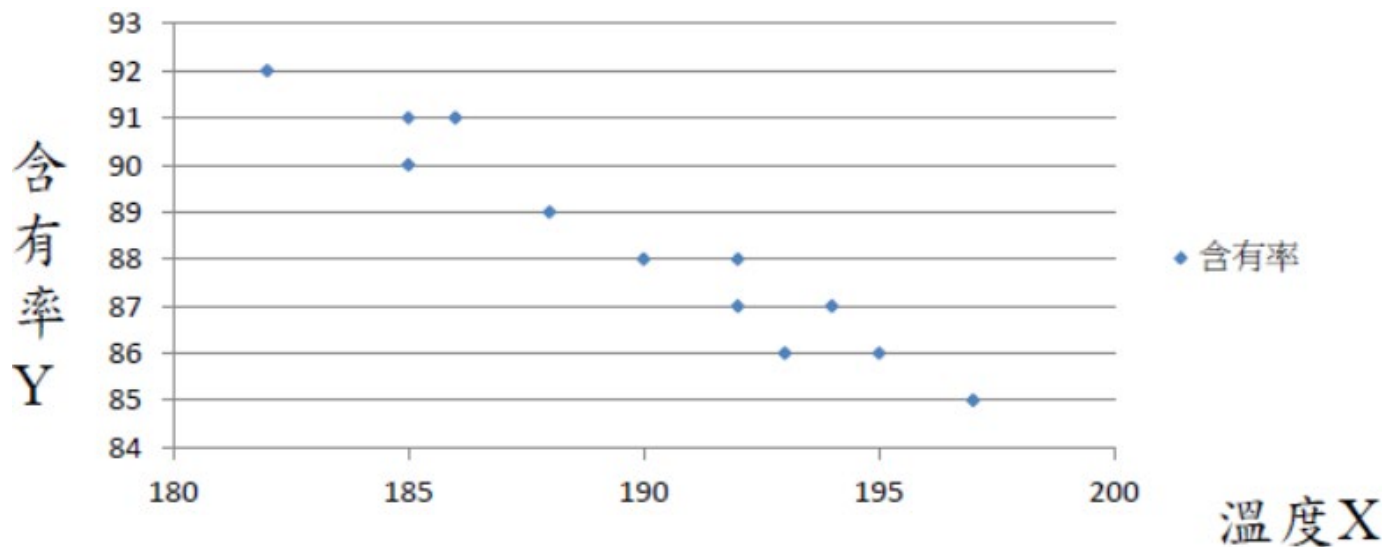
- 假設某產品之某種成份的含有率隨溫度變動而改變，12次實驗資料如下

溫度(F) (X)	182	185	186	188	190	193	194	195	192	197	192	185
含有率(%) (Y)	92	91	91	89	88	86	87	86	87	85	88	90

- 回答下列問題：
- a) 自變數與依變數各為何？
 - b) 畫出x-y散佈圖並判斷自變數與因變數之關係？
 - c) 找出迴歸方程式。
 - d) 判斷迴歸方程式是否適配原始資料？
 - e) 解釋迴歸係數 b_1 在本例中之意義為何？
 - f) 假設溫度為187°F，則估計之含有率為何？

範例一

- a) 本例中之自變數為 溫度 與因變數為 含有率。
- b) 畫出x-y散佈圖並判斷自變數與因變數之關係為何？
 - 自變數與因變數間之關係為 負向線性關係。



範例一

□ c) 找出迴歸方程式，令 $x^* = (x - 189)$ ， $y^* = (y - 89)$

○ 位移後， x^* 的變異數和 x 變異數相等，計算數字會變小

x	y	x^*	y^*	x^*y^*	x^{*2}	y^{*2}
182	92	-7	3	-21	49	9
185	91	-4	2	-8	16	4
186	91	-3	2	-6	9	4
188	89	-1	0	0	1	0
190	88	1	-1	-1	1	1
193	86	4	-3	-12	16	9
194	87	5	-2	-10	25	4
195	86	6	-3	-18	36	9
192	87	3	-2	-6	9	4
197	85	8	-4	-32	64	16
192	88	3	-1	-3	9	1
185	90	-4	1	-4	16	1
總計		$\Sigma x^* = 11$	$\Sigma y^* = -8$	$\Sigma x^*y^* = -121$	$\Sigma x^{*2} = 251$	$\Sigma y^{*2} = 62$

範例一

□ c) 找出迴歸方程式，令 $x^* = (x - 189)$ ， $y^* = (y - 89)$

$$n = 12, \Sigma x^* = 11, \Sigma y^* = -8, \Sigma x^* y^* = -121, \Sigma x^{*2} = 251, \Sigma y^{*2} = 62$$
$$SS_{xy} = SS_{x^*y^*} ; SS_x = n \times SS_{x^*} ; SS_y = n \times SS_{y^*}$$

$$(n - 1) \times SS_{x^*y^*} = \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} = -121 - \frac{11 \times (-8)}{12} = -113.667$$

$$(n - 1) \times SS_{x^*} = \sum_{i=1}^N x_i^2 - \frac{\sum_{i=1}^N x_i^2}{n} = 251 - \frac{11^2}{12} = 240.917$$

$$(n - 1) \times SS_{y^*} = \sum_{i=1}^N y_i^2 - \frac{\sum_{i=1}^N y_i^2}{n} = 62 - \frac{(-8)^2}{12} = 56.667$$

$$\hat{Y} = b_0 + b_1 X$$

$$b_1 = \frac{SS_{xy}}{SS_x} = \frac{(n-1) \times SS_{xy}}{(n-1) \times SS_x} = \frac{-113.667}{240.917} = -0.472$$

$$\bar{Y} = 89 + \frac{-8}{12} = 88.333$$

$$\bar{X} = 189 + \frac{11}{12} = 189.917$$

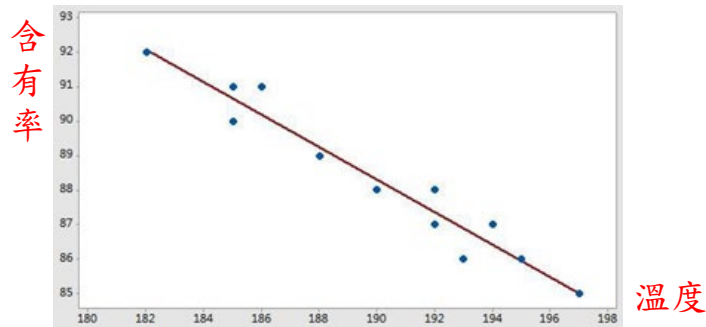
$$b_0 = \bar{Y} - b_1 \bar{X} = 88.333 - (-0.472)(189.917) = 177.974$$

$$\hat{Y} = b_0 + b_1 X = 177.974 - 0.472 X$$

評估迴歸模式－散佈圖

□ d)判斷迴歸方程式是否適配原始資料，分四部分說明：

○ 1) 將迴歸模式繪製原始資料散佈圖，判斷是否適配樣本點。



評估迴歸模式 – 判定係數

□ d) 判斷迴歸方程式是否適配原始資料，分四部分說明：

○ 2) 計算判定係數 r^2 。

➤ 迴歸式可解釋的變異量/總變異量 = $SSR/SST = 0.9464$

$$\begin{aligned} SSR = \sum (\hat{Y}_i - \bar{Y})^2 &= \frac{[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}]^2}{\sum_{i=1}^N x_i^2 - n\bar{x}^2} = \frac{[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}]^2}{\sum_{i=1}^N x_i^2 - \frac{\sum_{i=1}^N x_i^2}{n}} \\ &= \frac{[(n-1) \times SS_{xy}]^2}{(n-1) \times SS_x} = \frac{[-113.667]^2}{240.92} = 53.628 \end{aligned}$$

$$SST = \sum (Y_i - \bar{Y})^2 = \sum_{i=1}^N y_i^2 - \frac{(\sum_{i=1}^N y_i)^2}{n} = (n-1) \times SS_y = 56.667$$

➤ $r^2 \cong 0.9464$ ：溫度(X)變異會引起 94.64% 含有率變異。即溫度確實是影響含有率的重要負向因素。

○ 3) 計算相關係數 R_{xy} 。

➤ $R_{xy} = -\sqrt{r^2}$ (符號同 b_1 之符號) = -0.9728：溫度與含有率間有非常強的負向線性關係。此模式是一個好的預測模式。

評估迴歸模式 – ANOVA (Exercise)

○4) 利用ANOVA F test檢定溫度(x)與含有率(y)間是否有顯著的直線關係？

- 假設： $\epsilon_i \sim \text{NID}(0, \sigma^2)$
- $H_0: \beta_1 = 0$ (x與y之間沒有線性關係)
- $H_1: \beta_1 \neq 0$ (x與y之間有線性關係亦即斜率不為0)
- $\alpha = 0.05$ ，檢定值： $F = 53.628 / 0.3039 = 176.46$
- reject region：臨界值 $F(1, 10) = 4.965$
- 結論：溫度與含有率間有顯著的直線關係。

變異來源	Sum of Squares, SS	自由度	Mean Sum of Square, MS	F
迴歸可解釋變異量	$SS_B = SSR$	$k-1=1$	$MSB = \frac{SSR}{1}$	$\frac{MSB}{MSW}$
殘差(不可解釋)變異量	$SS_w = SSE$	$n-k = n-2$	$MSW = \frac{SSE}{n-2}$	
總變異量	SST	$n-1$		

	SS	自由度	MS	F
迴歸	53.628	1	53.628	176.46
殘差	3.039	10	0.3039	
總和	56.667	11		

$$k = 2, n = 12, k-1 = 1, n-k = 10$$

$$SSE = SST - SSR = 56.667 - 53.628 = 3.039$$

評估迴歸模式 – t 檢定(Exercise)

○ 利用t 檢定溫度(x)是否為含有率(y)的一個有用的預測變數？

➤ 假設： $\varepsilon_i \sim \text{NID}(0, \sigma^2)$

➤ $H_0: \beta_1 = 0$

➤ $H_1: \beta_1 \neq 0$

➤ 設定 α 值

➤ 檢定值： $t = -0.4718 / (0.3039 / 240.92)^{1/2} = -13.28$

➤ 拒絕域：臨界值 $t = +$ 或 -2.228

➤ 結論：溫度(x) 是 含有率(y)的一個有用的預測變數。

評估迴歸模式

□ e) 迴歸係數**b**在本例中之意義為何？

○ $b_1 = -0.4718$ ：當溫度每增加 1°F 時，含有率下降 0.4718% 。

□ f) 設溫度為**187**，則估計之含有率為何？

○ $\bar{Y} = 177.94 - 0.4718 \times (187) = 99.54\%$

□ 迴歸分析注意事項

○ 利用迴歸模式估計 y 時，所給定之 x 值須在樣本之 x 值範圍內， y 估計值才會準確。

➤ 上例中，當所給定之 X 值介於 182 與 197 間， Y 估計值才會準確。

迴歸模式報表(參考)

□ 回歸統計報表

R的倍數	0.9728
R平方	0.9464
調整的R平方	0.941
標準誤	0.551154
觀察值個數	12

□ ANOVA表

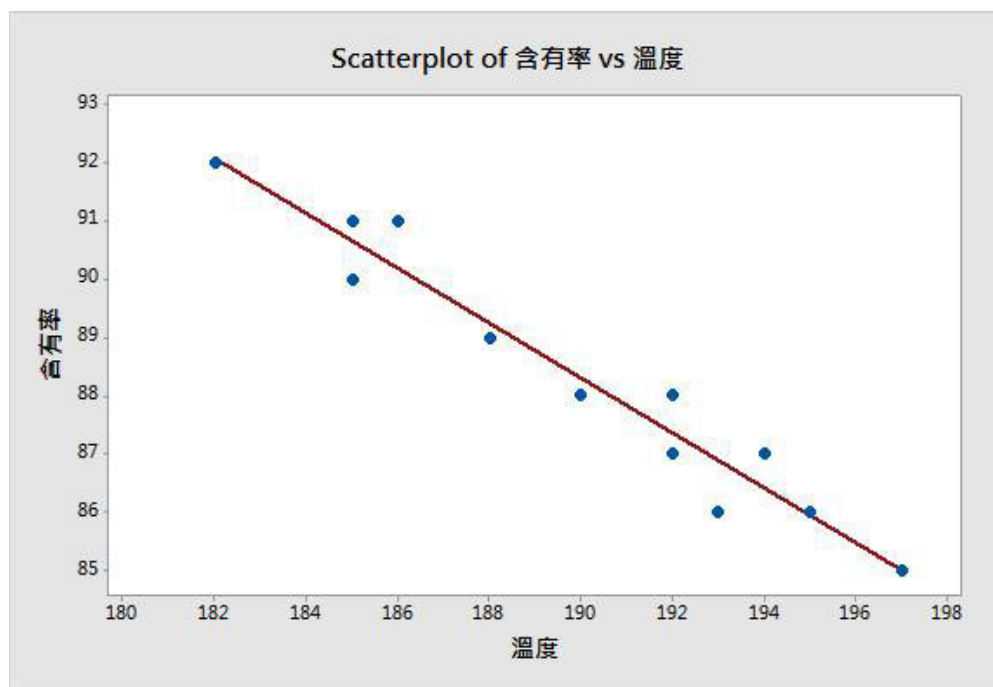
	自由度	SS	MS	F	顯著值
迴歸	1	53.62896	53.62896	176.5444	0
殘差	10	3.037703	0.30377		
總和	11	56.66667			

	係數	標準誤	t統計	P-值
截距	177.9377	6.74564	26.37818	0
溫度	-0.47181	0.035509	-13.287	0

迴歸模式報表(參考)

□ 殘差輸出

觀測值	含有率	殘差
1	90.65306	-0.65306
2	87.3504	0.649602
3	84.99135	0.008648
4	87.3504	-0.3504
5	85.93497	0.065029
6	86.40678	0.59322
7	86.87859	-0.87859
8	88.29402	-0.29402
9	89.23763	-0.23763
10	90.18125	0.818748
11	90.65306	0.346939
12	92.06849	-0.06849



範例二 (Homework)

- 某統計學課，隨機蒐查12位同學期中與期末考成績。適配期中、期末考成績簡單迴歸模式

○ $\bar{Y} = 20.50 + 0.737X$

Regression Analysis: Y versus X

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	1050.8	1050.83	40.44	0.000
X	1	1050.8	1050.83	40.44	0.000
Error	10	259.8	25.98		
Lack-of-Fit	7	149.3	21.33	0.58	0.752
Pure Error	3	110.5	36.83		
Total	11	1310.7			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
5.09738	80.18%	78.19%	72.51%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	20.50	8.59	2.39	0.038	
X	0.737	0.116	6.36	0.000	1.00

Regression Equation

$Y = 20.50 + 0.737 X$

期中考 成績(X)	56	74	90	63	91	53	81	65	74	90	77	63
期末考成績(Y)	64	80	82	68	89	61	76	70	80	93	71	58

多元 Linear regression (參考)

□ Example

○ 透過身高、性別等資料推測學生的體重。

$$y = b_0 + b_1x_1 + b_2x_2 + \dots$$

No	gender	height	weight
1	1	180	70
2	1	170	63
3	0	160	54
4

$$\begin{cases} 70 = \beta_0 + \beta_1 \times 180 + \beta_2 \times 1 \\ 63 = \beta_0 + \beta_1 \times 170 + \beta_2 \times 1 \\ 54 = \beta_0 + \beta_1 \times 160 + \beta_2 \times 0 \\ \dots \end{cases}$$

$$y = -58 + 0.7 \times x_0 + 2 \times x_2$$

