



迴歸分析 (Regression Analysis)



FROM :

- 1-普物實驗課本附錄
- 2-中原大學楊仲準老師網頁
- 3-維基百科<http://zh.wikipedia.org/zh-tw/%E6%9C%80%E5%B0%8F%E4%BA%8C%E4%B9%98%E6%B3%95>

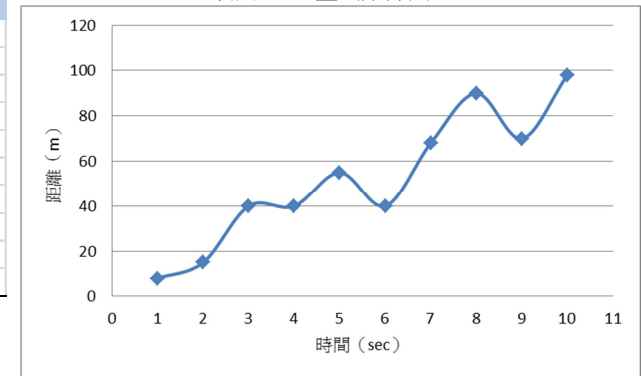
1



你有一組數據。

t(sec)	x(m)
1	8
2	15
3	40
4	40
5	55
6	40
7	68
8	90
9	70
10	98

利用EXCEL畫出數據圖。



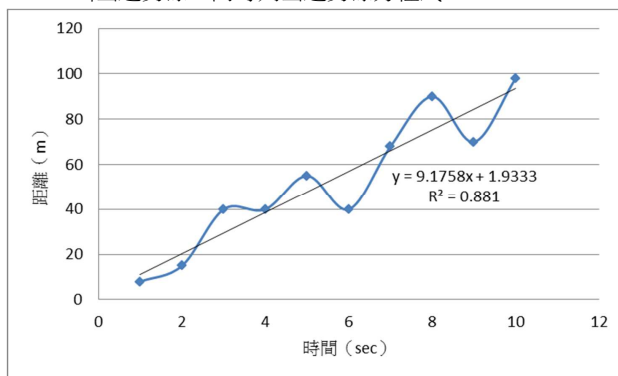
如何得到趨勢線？

2



最簡單的方式：

EXCEL叫出趨勢線，同時列出趨勢線方程式。



游標移到數據線上，
滑鼠按右鍵
選擇“加上趨勢線”

- 刪除(D)
- 重設以符合樣式(A)
- 變更數列圖表類型(M)...
- 選取資料(E)...
- 立體旋轉(R)...
- 新增資料標籤(B)
- 加上趨勢線(R)...
- 資料數列格式(O)...

3



可是...
畫完後...

什麼是趨勢線？

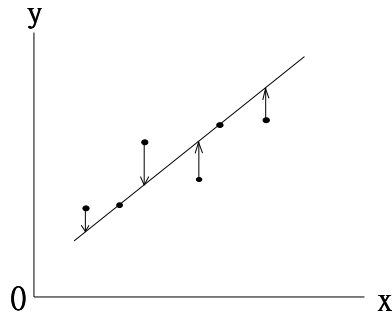
4



最小平方法

(Ordinary least square estimation, OLSE)

對於實驗數據圖，
最好的迴歸直線是每一點到這一條直線的“y方向偏離的平方和”為最小；
另一條最好的直線是每一點到這一條直線的“x方向偏離的平方和”為最小。



假設你有 n 對實驗值，

- (x1, y1) 、
- (x2, y2) 、
- (x3, y3) 、
- ... 、
- ... 、
- ... 、
- (xn, yn) 。

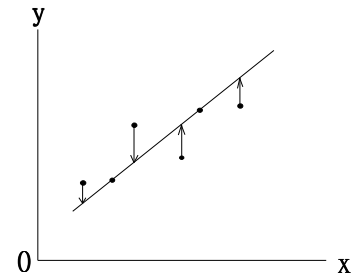
對於每一個已知的 x，我們要由所有的實驗值去預測出最好的 Y 值。

假設最適方程式的形式 $y = mx + b$

我們要求出 m 和 b，使得

$$\sum_1^n (y - Y)^2 = E$$

為最小！

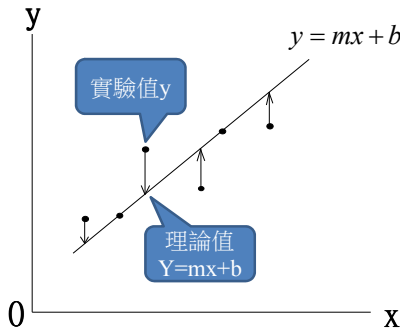


$$\sum_1^n (y - Y)^2 = E \text{ 為最小！}$$

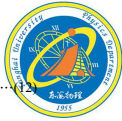
理論值
 $y - Y = y - (mx + b)$

實驗值

$$E = \sum_1^n [y - (mx + b)]^2$$



$$= \sum_1^n [y^2 - 2y(mx + b) + m^2b^2 + 2mbx + b^2]$$



要使 E 為最小，m 及 b 必須滿足下列方程式。

$$\frac{\partial E}{\partial b} = 0 = \sum_1^n [-2y + 2b + 2mx] \dots\dots\dots(3)$$

$$\therefore \sum_1^n y = \sum_1^n b + m \sum_1^n x = nb + m \sum_1^n x \dots\dots\dots(4)$$

$$\text{以及 } \frac{\partial E}{\partial m} = 0 = \sum_1^n [-2yx + 2bx + 2mx^2] \dots\dots\dots(5)$$

$$\therefore \sum_1^n xy = m \sum_1^n x^2 + b \sum_1^n x \dots\dots\dots(6)$$

$$b = \frac{\sum_1^n y}{n} - m \frac{\sum_1^n x}{n} = \bar{y} - m\bar{x} \dots\dots\dots(7)$$

$$\text{其中 } \bar{y} = \frac{\sum_1^n y}{n}, \bar{x} = \frac{\sum_1^n x}{n}$$

$$\Rightarrow \bar{y} = m\bar{x} + b \dots\dots\dots(8)$$

此式的意義是，最好的直線通過所有點座標的平均值。

將(7)式代入(6)式中

$$\sum_1^n xy = m \sum_1^n x^2 + \bar{y} \sum_1^n x - m\bar{x} \sum_1^n x \dots\dots\dots(9)$$

$$\sum_1^n xy = m \sum_1^n x^2 - m\bar{x} \sum_1^n x + \bar{y} \sum_1^n x \dots\dots\dots(10)$$

$$\sum_1^n xy = m \sum_1^n x^2 - \frac{m(\sum_1^n x)^2}{n} + \frac{\sum_1^n y \sum_1^n x}{n} \dots\dots\dots(11)$$

$$m = \frac{\sum_1^n xy - (\frac{\sum_1^n x \sum_1^n y}{n})}{\sum_1^n x^2 - \frac{(\sum_1^n x)^2}{n}} = \frac{n \sum_1^n xy - (\sum_1^n x)(\sum_1^n y)}{n \sum_1^n x^2 - (\sum_1^n x)^2} \dots\dots\dots(12)$$

$$\text{或 } m = \frac{\sum_1^n xy - n\bar{x}\bar{y}}{\sum_1^n x^2 - n\bar{x}^2} \dots\dots\dots(13)$$

亦可寫作

$$m = \frac{\sum_1^n (x - \bar{x})(y - \bar{y})}{\sum_1^n (x - \bar{x})^2} \dots\dots\dots(14)$$

$$\text{因為 } \sum_1^n (x - \bar{x})(y - \bar{y}) = \sum_1^n xy - \sum_1^n \bar{x}y - \sum_1^n x\bar{y} + \sum_1^n \bar{x}\bar{y} \dots\dots\dots(15)$$

$$= \sum_1^n xy - \bar{x} \sum_1^n y - \bar{y} \sum_1^n x + n\bar{x}\bar{y}$$

$$= \sum_1^n xy - n\bar{x}\bar{y} - n\bar{y}\bar{x} + n\bar{x}\bar{y}$$

$$= \sum_1^n xy - n\bar{x}\bar{y}$$

$$\sum_1^n (x - \bar{x})^2 = \sum_1^n x^2 - 2\sum_1^n x\bar{x} + \sum_1^n \bar{x}^2 \dots\dots\dots(16)$$

$$= \sum_1^n x^2 - 2\bar{x} \sum_1^n x + n\bar{x}^2 = \sum_1^n x^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_1^n x^2 - n\bar{x}^2$$



經過一連串繁複計算過程

.....

(計算過程在普物實驗講義附錄裡！)

9



得到 $y = mx + b$

其中

$$m = \frac{\sum_1^n xy - n\bar{x}\bar{y}}{\sum_1^n x^2 - n\bar{x}^2} = \frac{\sum_1^n (x-\bar{x})(y-\bar{y})}{\sum_1^n (x-\bar{x})^2} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}$$

$$\bar{x} = \frac{\sum_1^n x}{n}$$

$$b = \frac{1}{n} \left(\sum_1^n y - m \sum_1^n x \right) = \bar{y} - m\bar{x}$$

$$\bar{y} = \frac{\sum_1^n y}{n}$$

10



標準差

$$S(m) = \sqrt{\frac{1}{n-2} \left[\frac{n \sum_1^n y^2 - (\sum_1^n y)^2}{n \sum_1^n x^2 - (\sum_1^n x)^2} - m^2 \right]} = \sqrt{\frac{1}{n-2} (\overline{y^2} - \bar{y}^2 - m^2)}$$

迴歸係數誤差

$$\sigma_m = \frac{S(m)}{\sqrt{\sum_1^n (x-\bar{x})^2}}$$

$$\sigma_b = S(m) \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_1^n (x-\bar{x})^2}}$$

11



最適迴歸直線方程式

$$y = mx + b$$

最大上限(下限)統計誤差迴歸直線

$$y = (m \pm \sigma_m)x + (b \pm \sigma_b)$$

判斷係數 $R = \frac{\sum_1^n (x-\bar{x})(y-\bar{y})}{\sqrt{\sum_1^n (x-\bar{x})^2} \sqrt{\sum_1^n (y-\bar{y})^2}}$

(完美配適) $1 > R^2 > 0$ (無對應關係)

12



EXCEL列出表格
先將數據鍵入表格

A	B	C	D	E	F	G	H	I	J	K	L
x-時間(sec)	y-距離(m)	x-xbar	y-ybar	(x-xbar)(y-ybar)	(x-xbar) ²	(y-ybar) ²	[y-(mx+b)] ²			y1=10.4x+9.3	y2=8x-5.5
1	8	-4.5	-44.4	199.8	20.25	1971.36	9.67			19.70	2.52
2	15	-3.5	-37.4	130.9	12.25	1398.76	27.93			30.07	10.50
3	40	-2.5	-12.4	31	6.25	153.76	111.08			40.44	18.49
4	40	-1.5	-12.4	18.6	2.25	153.76	1.86			50.80	26.47
5	55	-0.5	2.6	-1.3	0.25	6.76	51.67			61.17	34.45
6	40	0.5	-12.4	-6.2	0.25	153.76	288.59			71.54	42.44
7	68	1.5	15.6	23.4	2.25	243.36	3.37			81.91	50.42
8	90	2.5	37.6	94	6.25	1413.76	214.93			92.28	58.40
9	70	3.5	17.6	61.6	12.25	309.76	210.69			102.64	66.39
10	98	4.5	45.6	205.2	20.25	2079.36	18.57			113.01	74.37
11											
12	平均值	5.5	52.4								
13	總和				757	82.5	7884.4	938.35			
14	m	9.18	σm	1.19							
15	b	1.93	σb	7.40							
16	s	10.83	R	0.94							
17			R ²	0.88							
18	趨勢線：y=9.18x+1.93										
19	統計誤差：y=(9.18±1.19)x+(1.93±7.4)										
20											



D2=B2-\$B\$12

D2 = =B2-\$B\$12
\$B\$12：鎖定B12不隨滑鼠拖曳變化。
\$B12：鎖定B不隨滑鼠拖曳變化。
B\$12：鎖定12不隨滑鼠拖曳變化。

A	B	C	D	E	F	G	H	I	J	K	L
x-時間(sec)	y-距離(m)	x-xbar	y-ybar	(x-xbar)(y-ybar)	(x-xbar) ²	(y-ybar) ²	[y-(mx+b)] ²			y1=10.4x+9.3	y2=8x-5.5
1	8	-4.5	-44.4	199.8	20.25	1971.36	9.67			19.70	2.52
2	15	-3.5	-37.4	130.9	12.25	1398.76	27.93			30.07	10.50
3	40	-2.5	-12.4	31	6.25	153.76	111.08			40.44	18.49
4	40	-1.5	-12.4	18.6	2.25	153.76	1.86			50.80	26.47
5	55	-0.5	2.6	-1.3	0.25	6.76	51.67			61.17	34.45
6	40	0.5	-12.4	-6.2	0.25	153.76	288.59			71.54	42.44
7	68	1.5	15.6	23.4	2.25	243.36	3.37			81.91	50.42
8	90	2.5	37.6	94	6.25	1413.76	214.93			92.28	58.40
9	70	3.5	17.6	61.6	12.25	309.76	210.69			102.64	66.39
10	98	4.5	45.6	205.2	20.25	2079.36	18.57			113.01	74.37
11											
12	平均值	5.5	52.4								
13	總和				757	82.5	7884.4	938.35			
14	m	9.18	σm	1.19							
15	b	1.93	σb	7.40							
16	s	10.83	R	0.94							
17			R ²	0.88							
18	趨勢線：y=9.18x+1.93										
19	統計誤差：y=(9.18±1.19)x+(1.93±7.4)										
20											

B12=AVERAGE(B2:B11)

F13=SUM(F2:F11)

B12 = =AVERAGE(B2:B11)：取B2-B11的平均值
記得要先鍵入“=”，才是執行方程式。

F13 = =SUM(F2:F11)：取F2-F11的總和。



F2=D2*E2

F2 = =D2*E2：D2和E2數值的乘績。

A	B	C	D	E	F	G	H	I	J	K	L
x-時間(sec)	y-距離(m)	x-xbar	y-ybar	(x-xbar)(y-ybar)	(x-xbar) ²	(y-ybar) ²	[y-(mx+b)] ²			y1=10.4x+9.3	y2=8x-5.5
1	8	-4.5	-44.4	199.8	20.25	1971.36	9.67			19.70	2.52
2	15	-3.5	-37.4	130.9	12.25	1398.76	27.93			30.07	10.50
3	40	-2.5	-12.4	31	6.25	153.76	111.08			40.44	18.49
4	40	-1.5	-12.4	18.6	2.25	153.76	1.86			50.80	26.47
5	55	-0.5	2.6	-1.3	0.25	6.76	51.67			61.17	34.45
6	40	0.5	-12.4	-6.2	0.25	153.76	288.59			71.54	42.44
7	68	1.5	15.6	23.4	2.25	243.36	3.37			81.91	50.42
8	90	2.5	37.6	94	6.25	1413.76	214.93			92.28	58.40
9	70	3.5	17.6	61.6	12.25	309.76	210.69			102.64	66.39
10	98	4.5	45.6	205.2	20.25	2079.36	18.57			113.01	74.37
11											
12	平均值	5.5	52.4								
13	總和				757	82.5	7884.4	938.35			
14	m	9.18	σm	1.19							
15	b	1.93	σb	7.40							
16	s	10.83	R	0.94							
17			R ²	0.88							
18	趨勢線：y=9.18x+1.93										
19	統計誤差：y=(9.18±1.19)x+(1.93±7.4)										
20											

G2=D2^2

G2 = =D2^2：D2數值的平方。
a^b = a^ab，如5²=5²。



I2=(C2-\$C\$14*B2-\$C\$15)^2

I2 = =(C2-\$C\$14*B2-\$C\$15)^2
每一組數據的誤差平方

A	B	C	D	E	F	G	H	I	J	K	L
x-時間(sec)	y-距離(m)	x-xbar	y-ybar	(x-xbar)(y-ybar)	(x-xbar) ²	(y-ybar) ²	[y-(mx+b)] ²			y1=10.4x+9.3	y2=8x-5.5
1	8	-4.5	-44.4	199.8	20.25	1971.36	9.67			19.70	2.52
2	15	-3.5	-37.4	130.9	12.25	1398.76	27.93			30.07	10.50
3	40	-2.5	-12.4	31	6.25	153.76	111.08			40.44	18.49
4	40	-1.5	-12.4	18.6	2.25	153.76	1.86			50.80	26.47
5	55	-0.5	2.6	-1.3	0.25	6.76	51.67			61.17	34.45
6	40	0.5	-12.4	-6.2	0.25	153.76	288.59			71.54	42.44
7	68	1.5	15.6	23.4	2.25	243.36	3.37			81.91	50.42
8	90	2.5	37.6	94	6.25	1413.76	214.93			92.28	58.40
9	70	3.5	17.6	61.6	12.25	309.76	210.69			102.64	66.39
10	98	4.5	45.6	205.2	20.25	2079.36	18.57			113.01	74.37
11											
12	平均值	5.5	52.4								
13	總和				757	82.5	7884.4	938.35			
14	m	9.18	σm	1.19							
15	b	1.93	σb	7.40							
16	s	10.83	R	0.94							
17			R ²	0.88							
18	趨勢線：y=9.18x+1.93										
19	統計誤差：y=(9.18±1.19)x+(1.93±7.4)										
20											

C14=F13/G13

C15=C12-C14*B12

$$m = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - m\bar{x}$$



A	B	C	D	E	F	G	H	I	J	K	L	
1	x	y		y-ybar	(x-xbar)	(y-ybar)	(x-xbar)					
2												
3												
4												
5												
6	5	55		4.5	-1.5	-37.4	130.9			19.70	2.52	
7				0.5	-1.5	-37.4	130.9			30.07	10.50	
8				1.5	-1.5	-37.4	130.9			40.44	18.49	
9				2.5	-1.5	-37.4	130.9			50.80	26.47	
10				3.5	-1.5	-37.4	130.9			61.17	34.45	
11				4.5	-1.5	-37.4	130.9			71.54	42.44	
12	平均值	5.5	52.4							81.91	50.42	
13	總和									92.28	58.40	
14	m	9.18								102.64	66.39	
15	b	1.93								113.01	74.37	
16	s	10.83										
17												
18	趨勢線：y=9.18x+1.93											
19	統計誤差：y=(9.18±1.19)x+(1.93±7.4)											
20												

$$\sigma_m = \frac{S(m)}{\sqrt{\sum_1^n (x - \bar{x})^2}}$$

$$\sigma_b = S(m) \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_1^n (x - \bar{x})^2}}$$

$$E14=C16/SQRT(G13)$$

$$E15=C16*SQRT(1/10+B12^2/G13)$$

$$E16=F13/(SQRT(G13)*SQRT(H13))$$

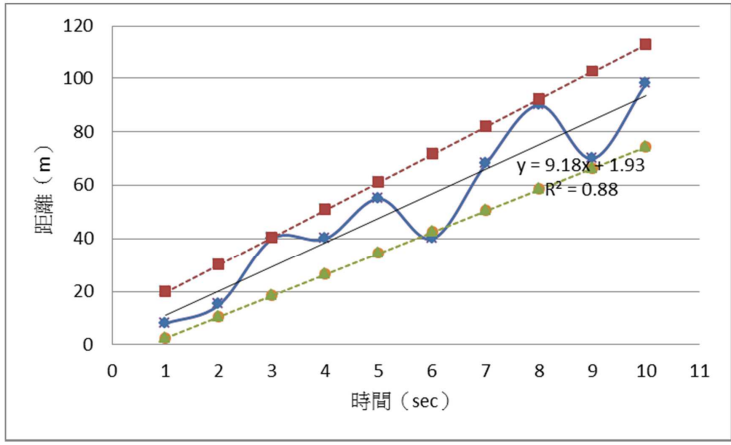
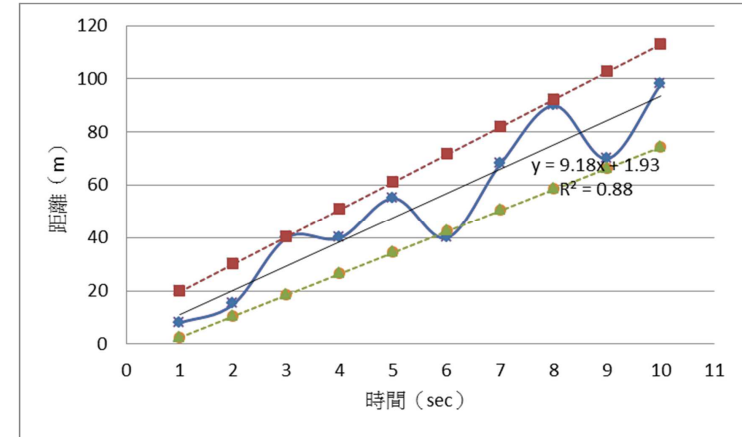
$$R = \frac{\sum_1^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_1^n (x - \bar{x})^2} \sqrt{\sum_1^n (y - \bar{y})^2}}$$



$$y = (9.2 \pm 1.2)x + (1.9 \pm 7.4)$$

$$y = (9.2 + 1.2)x + (1.9 + 7.4) = (10.4)x + (9.3)$$

$$y = (9.2 - 1.2)x + (1.9 - 7.4) = (8)x + (-5.5)$$



哪些數據點可以刪除不計？



我們沒有最好
只有追求更好

有空繼續補~~



東海大學應用物理學系
地址：40704台中市西屯區東海大學B0X803
電話：04-23590121*32100
網址：<http://physics.thu.edu.tw/>