

迴歸分析

3/23

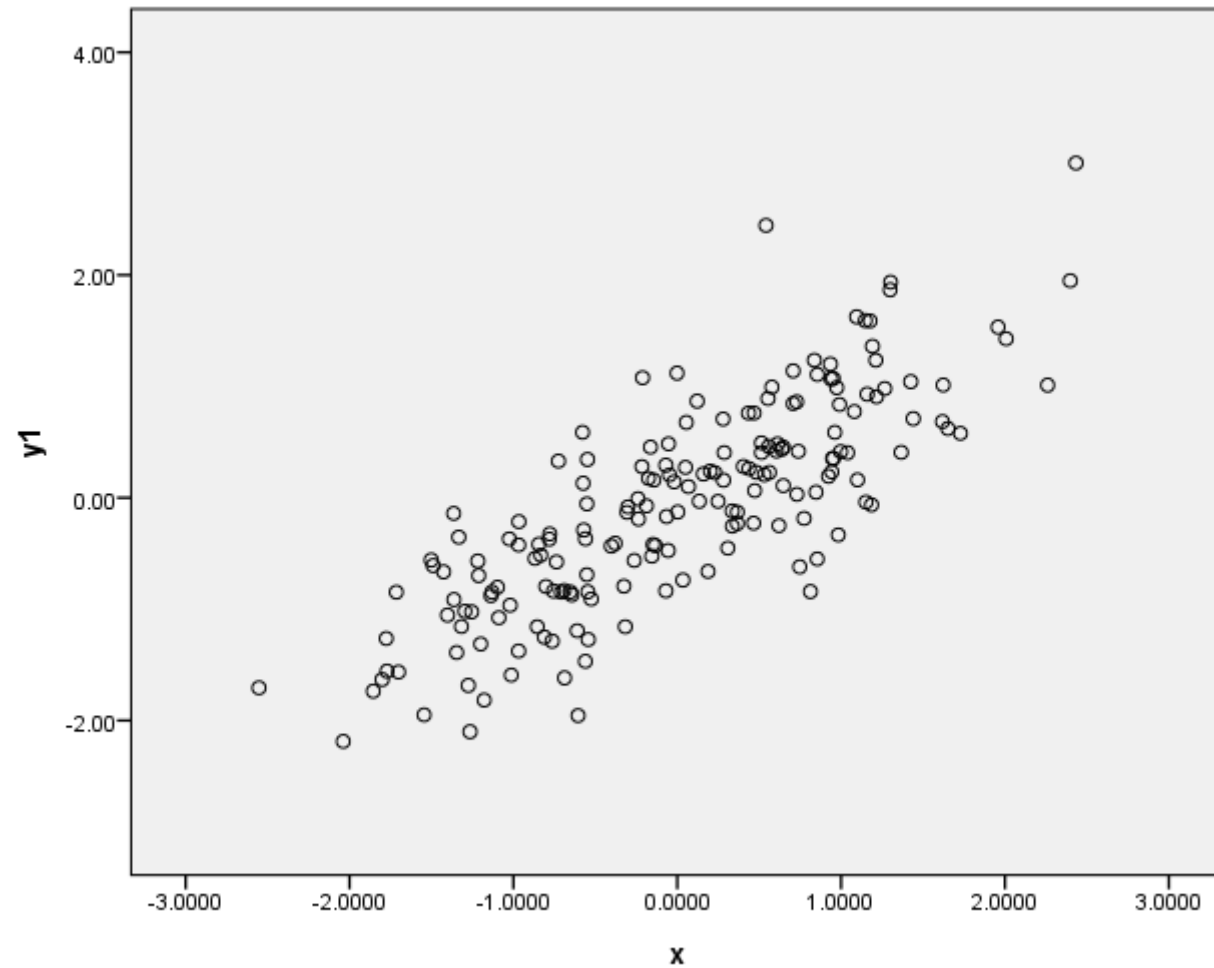
迴歸分析 regression

簡單來說迴歸分析是一個統計分析方法，
想要瞭解兩個研究變項，其中一個Y的變異
是否可由X去解釋，或是

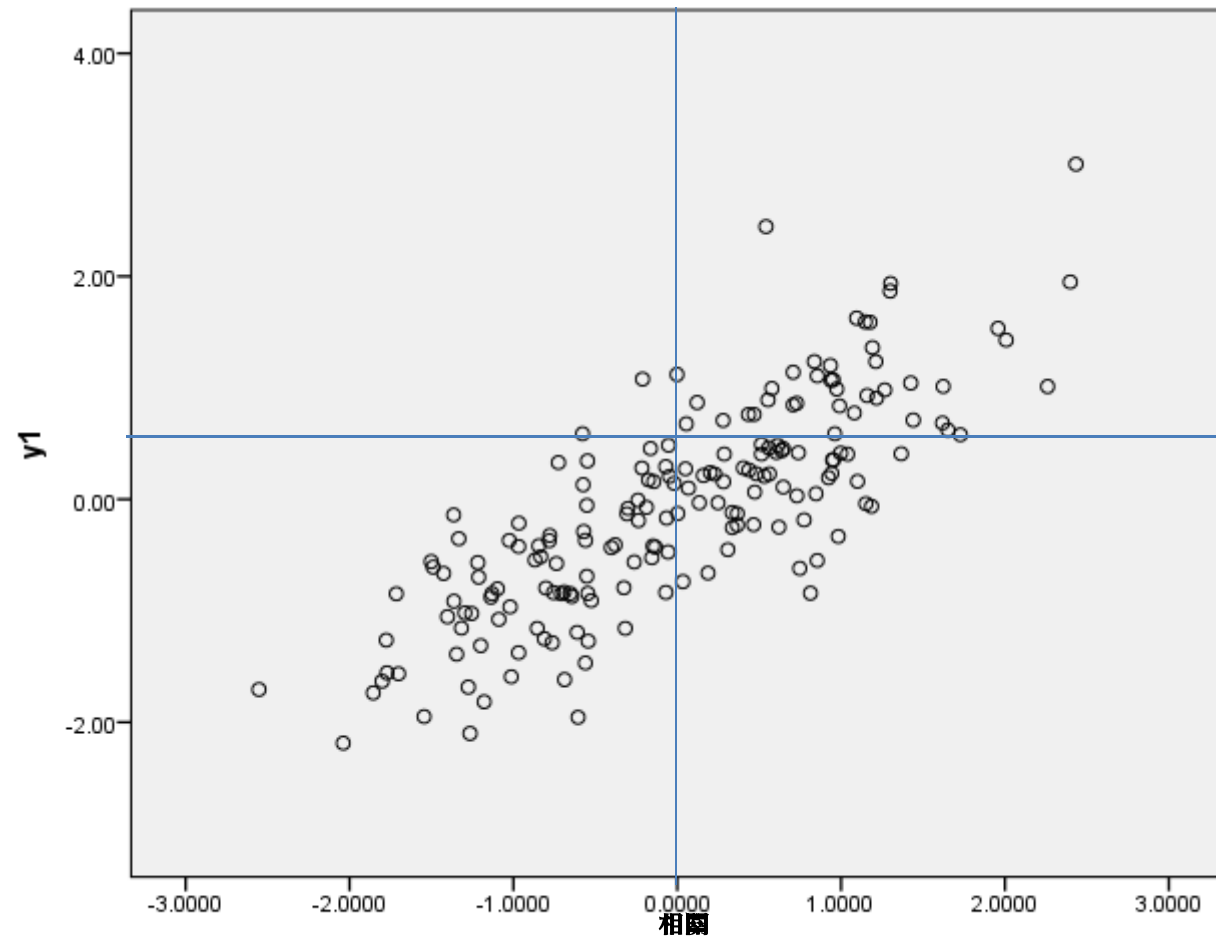
如何去估計X影響Y的程度

事實上regress是有典故的，父母身高的關係
記得，第一次作業我們先求散佈圖，再求
共變數(covariance)，如果這些散佈圖呈現
一個線性趨勢，是否可以用一條直線來描
述Y與X之間的關係，迴歸方程式

迴歸分析 regression-散佈圖



迴歸分析 regression-十字架



		Pearson 相關	顯著性 (雙尾)	叉積平方和	共變異數	個數
y1	y1	1		157.948	.863	184
	x	.807**	.000	134.471	.735	184
x	y1	.807**	.000	134.471	.735	184
	x	1		175.706	.960	184

** . 在顯著水準為0.01時 (雙尾), 相關顯著。

迴歸分析 regression-迴歸線

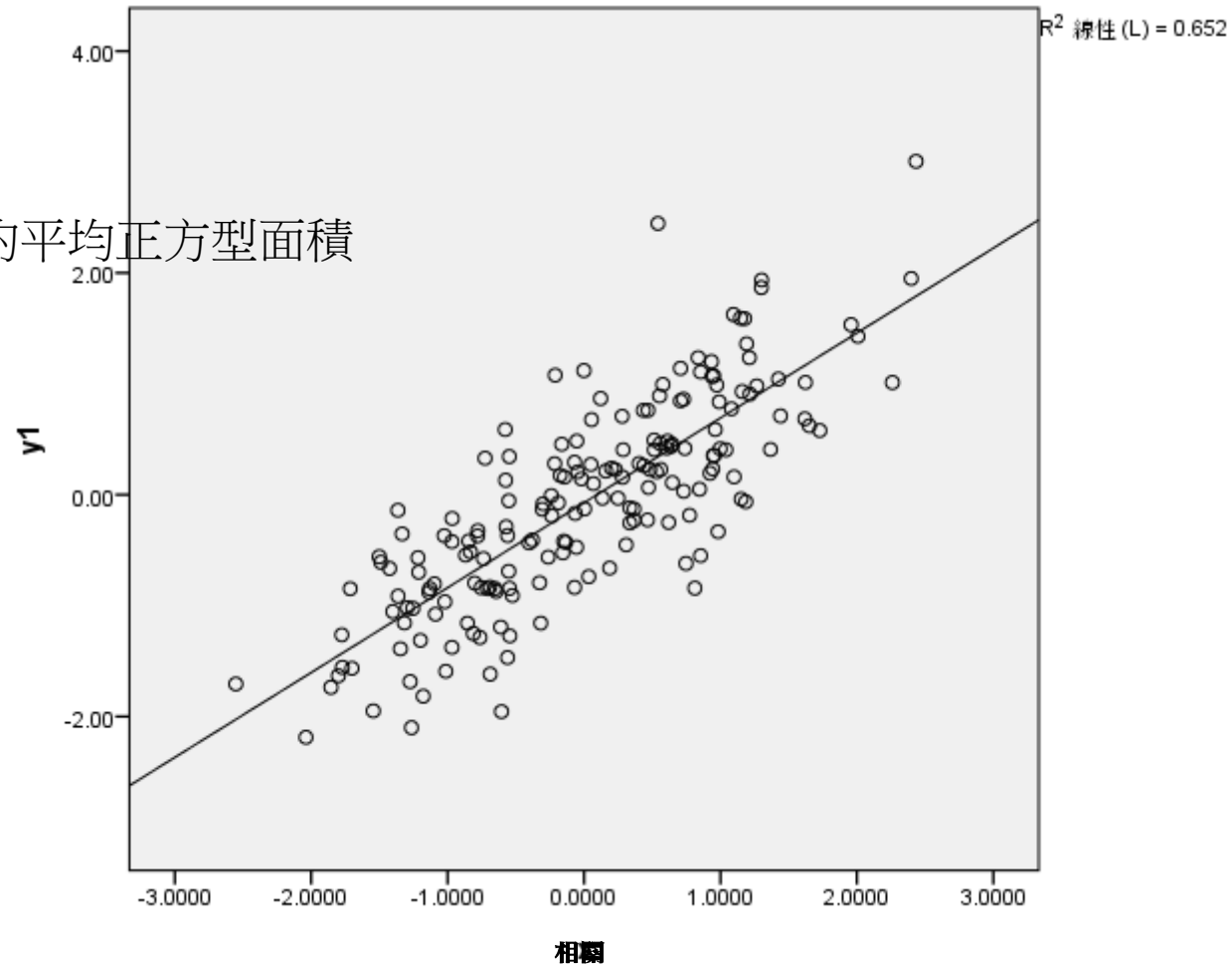
斜率

=共變數/變異數

=XY的平均長方形/X的平均正方形面積

=.735/.960

=.7656



		相關				
		Pearson 相關	顯著性 (雙尾)	叉積平方和	共變異數	個數
y1	y1	1		157.948	.863	184
	x	.807**	.000	134.471	.735	184
x	y1	.807**	.000	134.471	.735	184
	x	1		175.706	.960	184

** . 在顯著水準為0.01時 (雙尾), 相關顯著。

迴歸分析 regression-迴歸線

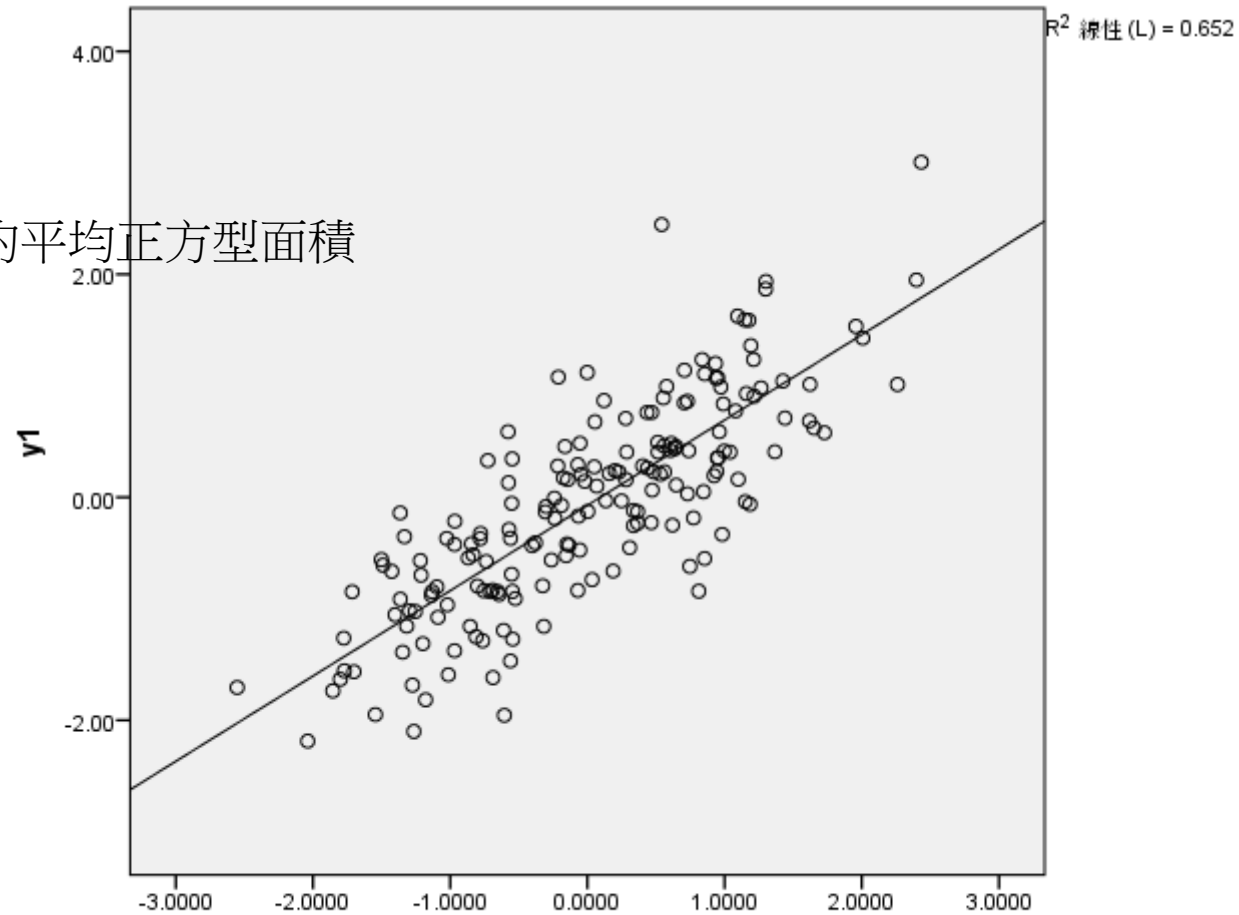
斜率

=共變數/變異數

=XY的平均長方形/X的平均正方形面積

=.735/.960

=.7656



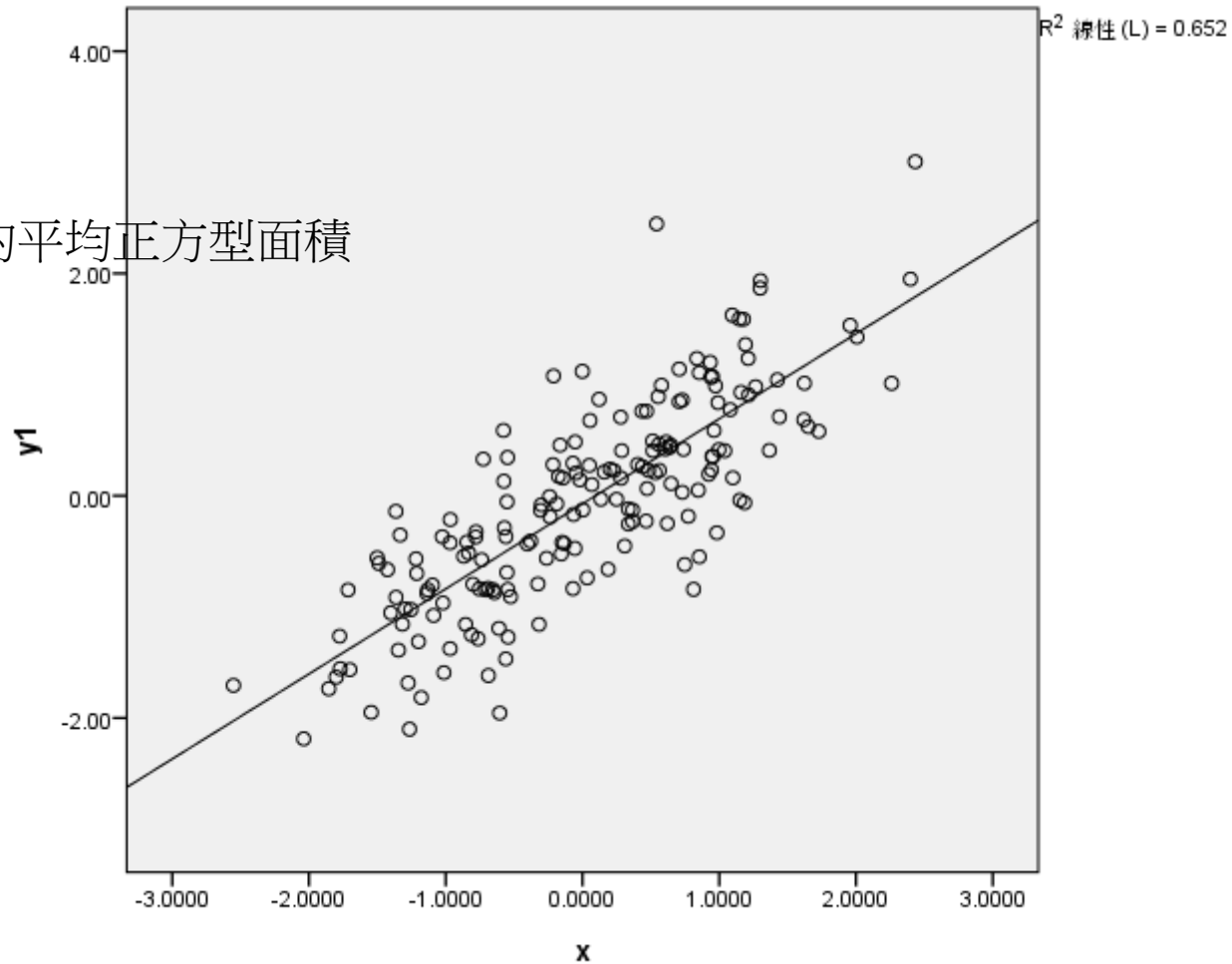
相關

		Pearson 相關	顯著性 (雙尾)	叉積平方和	共變異數	個數
y1	y1	1		157.948	.863	184
	x	.807**	.000	134.471	.735	184
x	y1	.807**	.000	134.471	.735	184
	x	1		175.706	.960	184

** . 在顯著水準為0.01時 (雙尾), 相關顯著。

迴歸分析 regression-

斜率
 =共變數/變異數
 =XY的平均長方形/X的平均正方形面積
 =.735/.960
 =.765
 =**只限**簡單迴歸分析



迴歸方程式
 $Y = -0.072 + .765 * X + e$

		係數 ^a				
模式		未標準化係數		標準化係數	t	顯著性
		B 之估計值	標準誤差	Beta 分配		
1	(常數)	-.072	.041		-1.777	.077
	x	.765	.041	.807	18.448	.000

a. 依變數: y1

迴歸分析 regression-

*輸出1 [文件1] - PASW Statistics Viewer

檔案(F) 編輯(E) 檢視(V) 插入(I) 樞軸(P) 格式(O) 分析(A) 統計圖(G) 效用值(U) 視窗(W) 說明(H)

輸出

- 圖形
 - 標題
 - 注意事項
 - 作用中資料集
 - y1 x 的散佈圖
- 相關
 - 標題
 - 注意事項
 - 作用中資料集
 - 相關
- 圖形
 - 標題
 - 注意事項
 - 作用中資料集
 - y1 x 的散佈圖
- 迴歸
 - 標題
 - 注意事項
 - 作用中資料集
 - 選入刪除的變數
 - 模式摘要
 - Anova
 - 係數

迴歸

[資料集1] C:\Documents and Settings

選入刪除的變數^b

模式	選入的變數	刪除的變數	方
1	x ^a	.	選入

a. 所有要求的變數已輸入。
b. 依變數: y1

模式摘要

模式	R	R 平方	調整後的 R 平方
1	.807 ^a	.652	

a. 預測變數 (常數), x

Anova

模式		平方和	df
1	迴歸	102.914	1
	殘差	55.035	182
	總數	157.948	183

a. 預測變數 (常數), x
b. 依變數: y1

係數

模式		未標準化係數	B 之估計值	標準誤差
1	(常數)	-.072		.041
	x	.765		.041

a. 依變數: y1

報表(P) 敘述統計(E) 表格(B) 比較平均數法(M) 一般線性模式(G) 概化線性模式(Z) 混合模式(X) 相關(C) 迴歸(R) 對數線性(O) 神經網路(W) 分類(Y) 維度縮減(D) 尺度(A) 無母數檢定(N) 預測(T) 存活分析(S) 複選題分析(U) 遺漏值分析(V)... 多個插補(I) 複合樣本(L) 品質控制(Q) ROC 曲線(V)... Amos 17...

線性(L)... 曲線估計(C)... 偏最小平方(S)... 二元 Logistic(G)... 多項式 Logistic(M)... 次序的(D)... Probit 分析(P)... 非線性(N)... 加權估計(W)... 二階最小平方方法(2)... 最適尺度 (CATREG)(O)...

線性(L)...

PASW Statistics 處理器已就緒 H: 145, W: 465 pt.

迴歸分析 regression-

線性迴歸

id
y
x
約為觀察值...
y1

依變數(D):

區塊1來自1

上一個 下一個(N)

自變數(I):

方法(M): 輸入

選擇變數(C): 規則(U)...

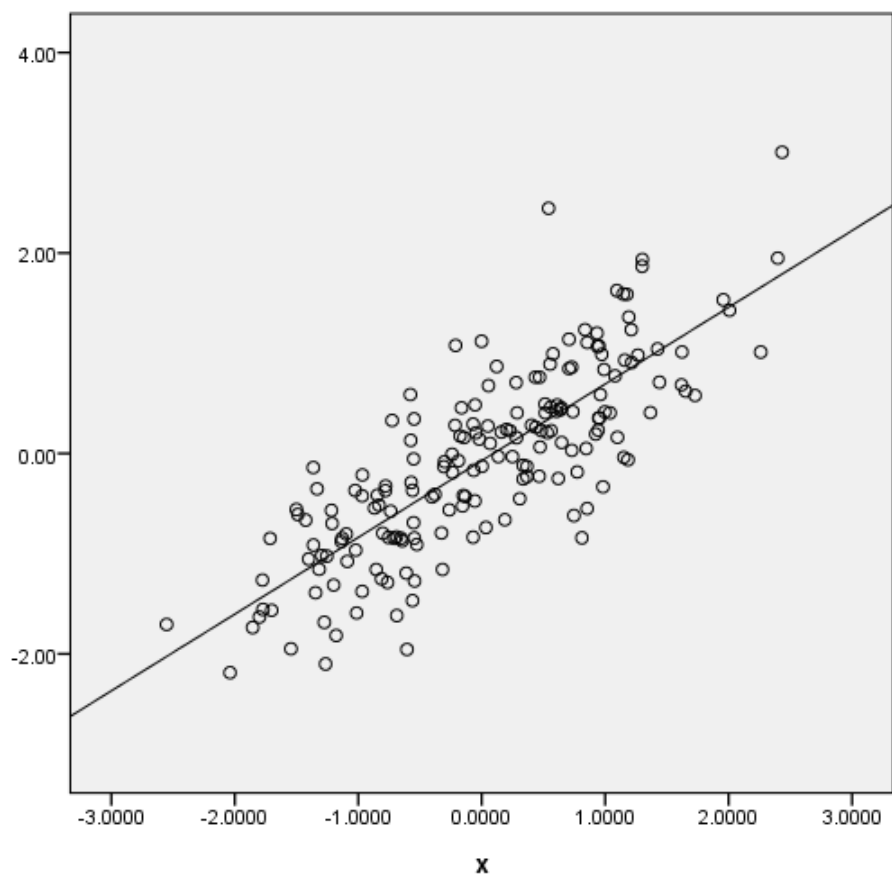
觀察值標記(C):

加權最小平方方法之權數(H):

統計量(S)...
圖形(T)...
儲存(S)...
選項(O)...
自助法(B)...

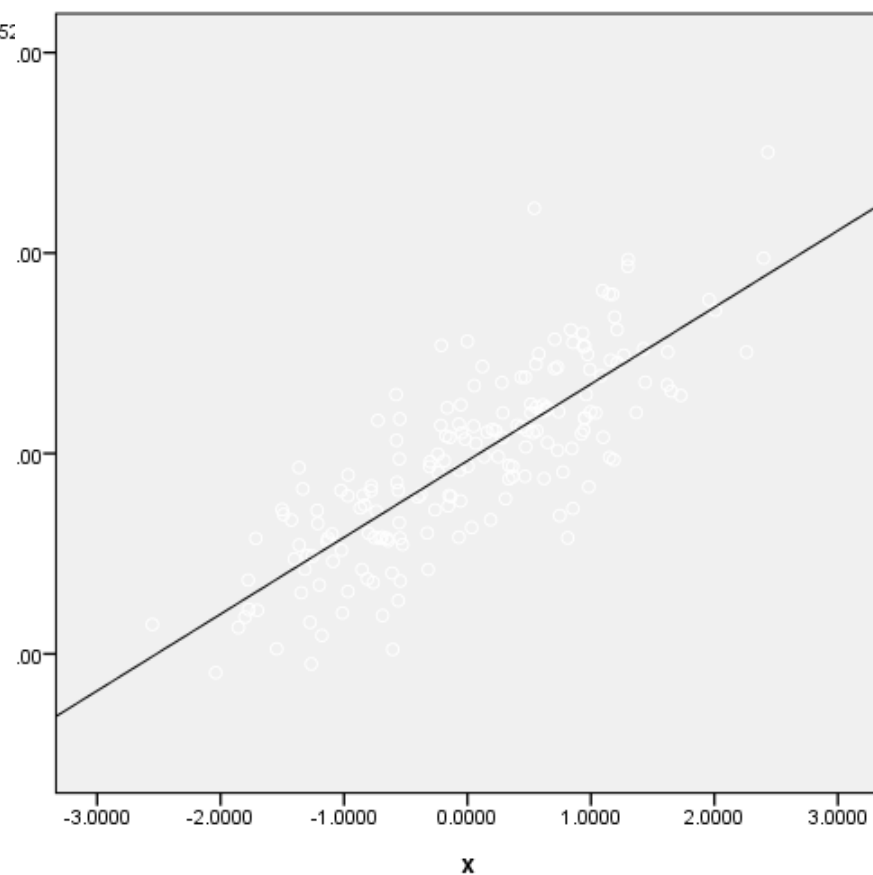
確定 貼上之後(P) 重設(R) 取消 輔助說明

迴歸方程式(統計關係)與函數關係



迴歸方程式

$$Y = -0.072 + 0.765X + e$$



$$Y = -0.072 + 0.765X$$

總成績計算方式(函數關係)

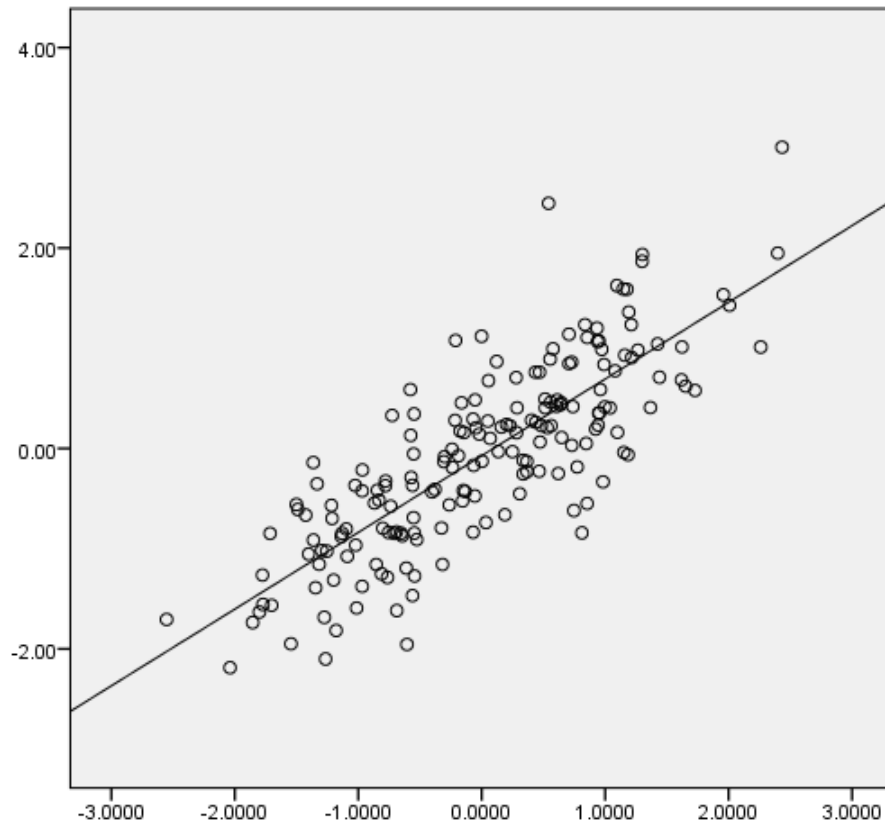
總成績=0.3作業+0.3出席+0.4期中考
作業、出席、期中考最高100、最低0

這些係數(斜率、權重、影響程度)是我給定的，
是函數關係，

總成績=20+0.3作業+0.3出席+0.4期中考
這20是？

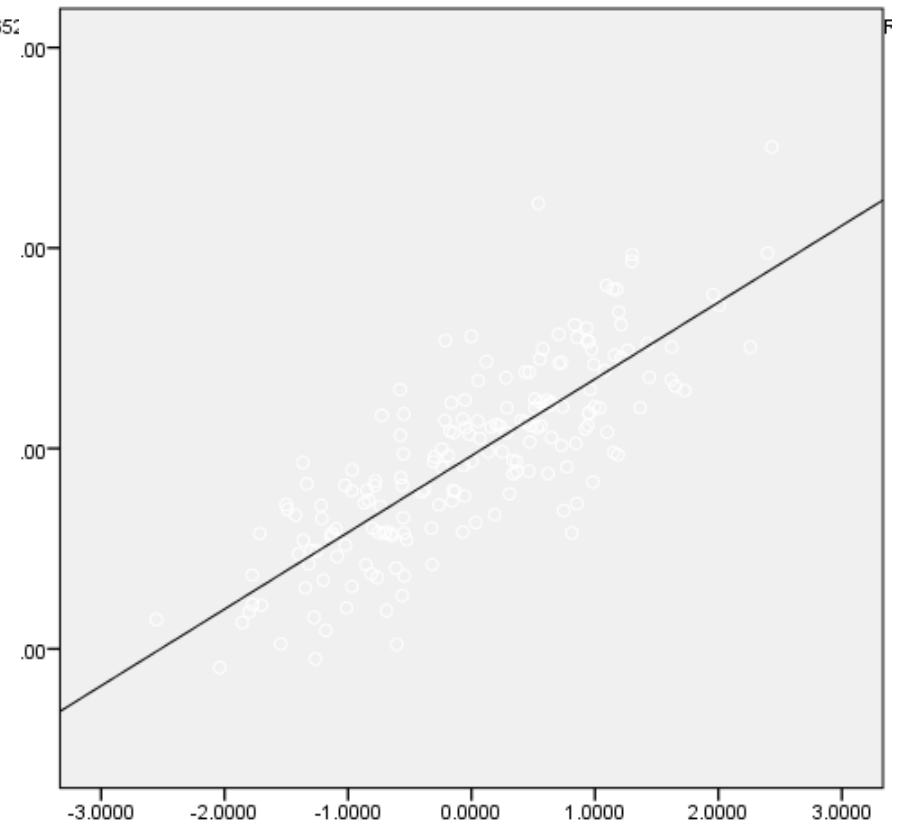
迴歸分析則是利用資料去找到這些係數!我們是用
估計這個字眼，利用最小平方法(hint平方就是正
方型面積)來估計。

迴歸方程式(統計關係)與函數關係



迴歸方程式

$$Y = -0.072 + 0.765X + e$$



$$Y = -0.072 + 0.765X$$

我們想用X來預測Y，所以從左圖中找到方程式 $Y = -0.072 + 0.765X$ ，但我們關心真正每個人的Y，這條方程式的估計值在線上，所以要加上測不準的誤差e，才是原來每個人的依變項，這條線就是我們捕捉到Y與X之間的關係。這條線的效力就是決定係數 R^2 ，簡單迴歸時 r^2 。