

CH14 迴歸分析 (Regression Analysis)



報告大綱

第一節 基本原理

壹、用途

貳、原理

一、簡單迴歸分析

二、多元迴歸分析

三、虛擬變項的迴歸分析

參、注意事項

一、注意基本假定

二、避免多元共線性產生

三、注意殘差值及偏離值

四、選定自變項投入模式的方法

第二節 操作與解說

壹、操作

貳、解說

參、操作的補充

肆、解說的補充

問題與操作

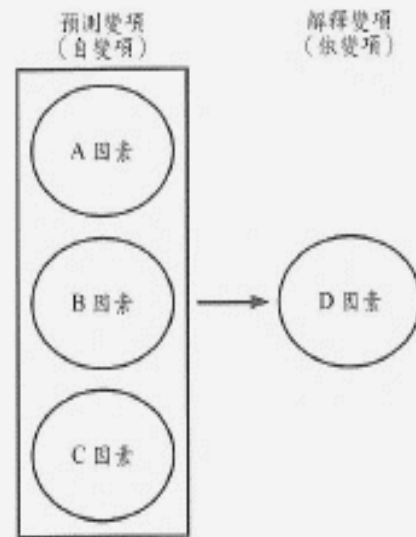


用途

使用時機

1. 社會科學研究主要有描述、解釋、預測及控制，對於社會現象預測的研究，可使用多元迴歸分析。
2. 迴歸分析提供研究者對於某一個社會現象的未來表現之預測。
3. 檢定一個或多個連續變項，和另一個連續變項之間是否具有顯著關係。

迴歸分析示意圖 (圖14-1) P266



1. A、B、C 稱為自變項或預測變項，D 為依變項或解釋變項。
2. 研究者想瞭解：
A、B、C 稱為三個自變項，哪一項對於變項 D 最有預測力，或 A、B、C 三個自變項對變項 D 的總預測力有多少？



用 途

迴歸分析大致區分(P265)

線性迴歸 (本節說明重點)	自變項與依變項形成直線關係的分析
非線性迴歸	自變項與依變項形成非直線關係的分析 如：二次曲線、三次曲線
邏輯式迴歸分析	依變項為兩類的名義變項， 自變項可以為連續變項與類別變項 與本節的迴歸分析之依變項為連續變項不同



原 理

區別項目	簡單迴歸分析	多元迴歸分析
使用時機	一個自變項與一個依變項所構成的線性關係，屬於單變項統計分析。	一個依變項，由多個自變項所形成的線性關係。
舉例說明	根據高中成績(連續變項)來預測大學聯考成績(連續變項)。 註： 高中成績(X)：預測變項 大學成績(Y)：效標變項	學生的 <u>家庭社經地位</u> 、 <u>課後補習時間</u> 、 <u>同儕關係</u> 、 <u>師生互動時間</u> 等4個自變項，對 <u>學習成就</u> (依變項)影響所形成的線性關係。



原理1-簡單迴歸分析

1.簡單迴歸分析的直線方程式：

$$\hat{Y} = a + bX \quad (14-1)$$

式中， \hat{Y} 為依變項（又稱為解釋變項）；

a 為截距；

b 為斜率： X 每變動一單位， Y 伴隨的變動量；

X 為自變項（又稱為預測變項）。



原理1-簡單迴歸分析

2.最小平方法(method of least square)：

(1)因為假設兩變項的函數關係為直線，故各點至此線之平行於Y軸的距離平方和 $\sum(Y-\hat{Y})^2$ 要最小。

(2)因為 $\sum(Y-\hat{Y})^2$ 要最小，經微分獲得a、b值的公式如下：

$$b_{y,x} = \frac{\sum XY - \left(\frac{\sum X \sum Y}{N}\right)}{\sum X^2 - \frac{(\sum X)^2}{N}} \quad (14-2)$$

$$a_{y,x} = \bar{Y} - b_{y,x}\bar{X} \quad (14-3)$$

公式中的N代表樣本數。



原理1-簡單迴歸分析

3.以CH9表9-1為例說明簡單迴歸分析之截距(a)與斜率(b)的計算過程：
瞭解10名高中生的高中成績是能預測大學聯考成績，高中及大學成績參考表14-1(P268)。將資料帶入a、b值的公式得到

$$b_{y,x} = \frac{63575 - \frac{770 \times 815}{10}}{60400 - \frac{(770)^2}{10}} \doteq 0.739$$

$$a_{y,x} = 81.5 - 0.739 \times 77 = 24.617$$

迴歸方程式為： $\hat{Y} = 24.617 + .739X$



原理1-簡單迴歸分析

表 14-1 10 名高中生的高中成績與大學聯考成績

學生	高中成績 (X)	大學聯考成績 (Y)	預測值 (\hat{Y})	原始值-平均値 ($Y - \bar{Y}$) = (SS_y)	預測值-平均値 ($\hat{Y} - \bar{Y}$) = (SS_{reg})	原始值-預測值 ($Y - \hat{Y}$) = (SS_{res})
A	85	90	87.40991	8.5	5.90991	2.59009
B	60	75	68.94144	-6.5	-12.55860	6.05856
C	75	85	80.02252	3.5	-1.47748	4.97748
D	80	85	83.71622	3.5	2.21622	1.28378
E	65	70	72.63514	-11.5	-8.86486	-2.63514
F	70	65	76.32883	-16.5	-5.17117	-11.32880
G	80	85	83.71622	3.5	2.21622	1.28378
H	95	90	94.79730	8.5	13.29730	-4.79730
I	90	95	91.10360	13.5	9.60360	3.89640
J	70	75	76.32883	-6.5	-5.17117	-1.32883
和 (Σ)	770	815	815	0	1E-05	-1E-05
平方和	60400	67275	67028.25	852.5	605.77	246.74
平均數	77.00 (\bar{X})	81.50 (\bar{Y})				
標準差	11.11 (s_1)	9.73 (s_2)				

註：表中-1E-05 代表小數點取五位。

總離均差
平方和
(總變異)

迴歸離均差
平方和
(被解釋變異)

殘差平方和
(誤差)



原理1-簡單迴歸分析

4.總變異、被解釋變異與非被解釋變異：

$$SS_T = SS_{reg} + SS_{res}$$

總變異 = 被解釋變異 + 非被解釋變異 (誤差)

$$\sum(Y - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y - \hat{Y})^2$$

總離均差平方和 = 迴歸離均差平方和 + 殘差平方和

以上例來說， $852.5 = 605.77 + 264.74$



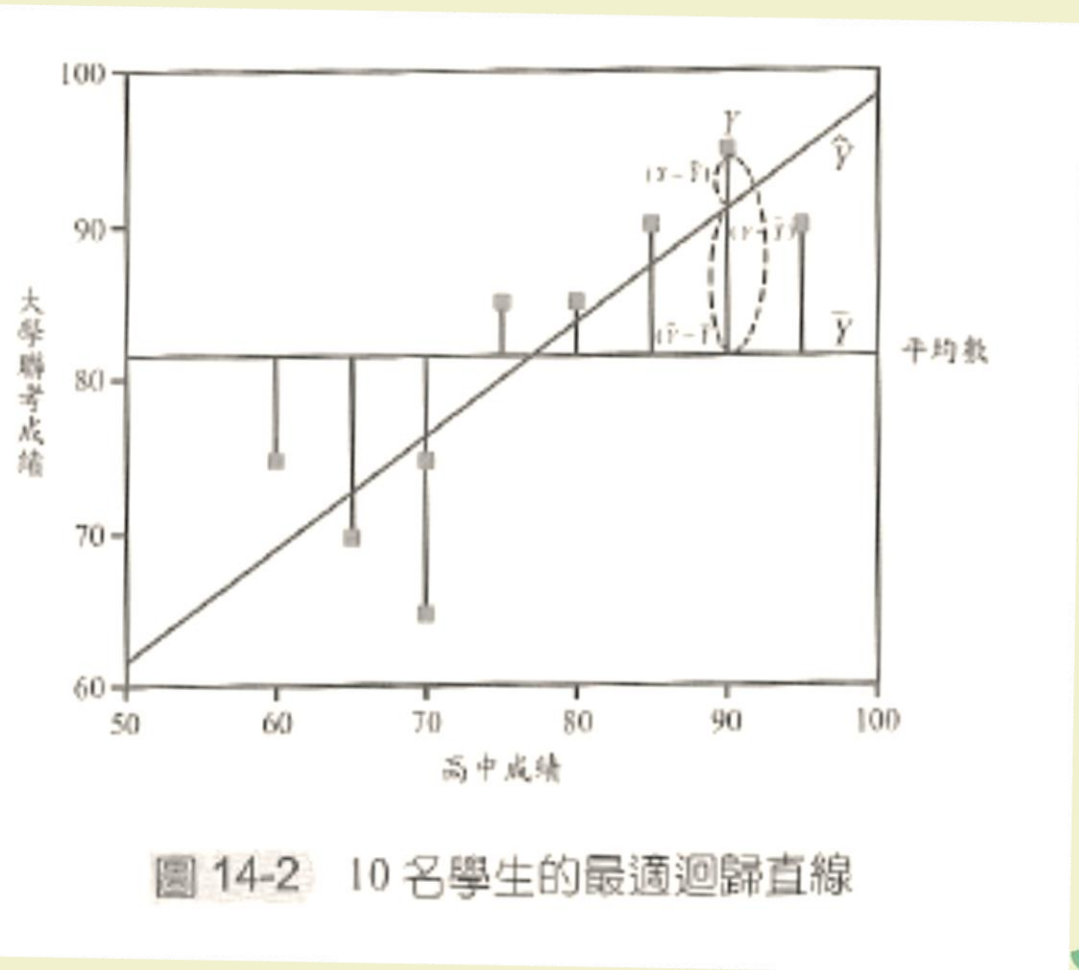
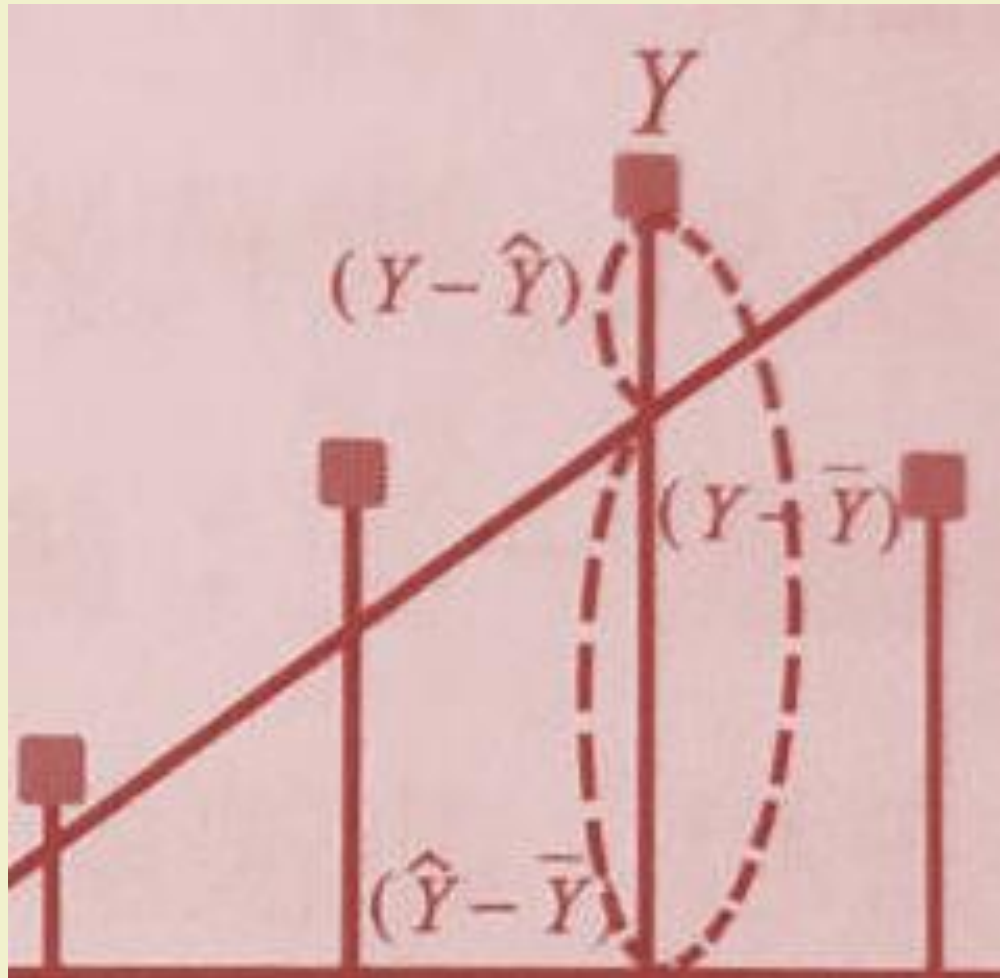
原理1-簡單迴歸分析

5.決定係數與疏離係數

	決定係數	疏離係數
作用	表示在效標變項(Y)的總變異中，由預測變項(X)解釋部分所佔的百分比(R^2)。	在變項X預測變項Y的迴歸分析中，用以表示點與迴歸直線的疏離程度的程度。
公式	$R^2 = 1 - \frac{SS_{res}}{SS_T} = \frac{SS_{reg}}{SS_T}$ <p>以上例來說，$R^2 = \frac{605.77}{852.5} = .711$。</p>	$\sqrt{1 - r^2}$
備註	迴歸離均差平方和愈大愈好， R^2 愈大，可以解釋的變異愈大，預測才正確。因此 R^2 愈接近1.0，代表愈有解釋力。	



原理1-簡單迴歸分析 (P270)



原理2-多元迴歸分析

(一)模式建立

迴歸分析自變項設定條件：

依據理論基礎	自變項與依變項的關係應該合理，在所建構研究架構中找出符合的理論基礎。
具實證經驗	過去的實證經驗提供選用自變項的重要依據，將過去研究已達顯著因素納入，可以減少嘗試錯誤或不知該如何選用自變項的困擾。
依邏輯推理	閱讀相關文獻後，選擇出可能影響的變項。
參考專家意見	找尋專家提供的建議，透過專業背景與過去對該議題的掌握，選擇出重要變項。



原理2-多元迴歸分析

多元迴歸分析模式：

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + \dots + b_nX_n + e \quad (14-5)$$

模式中「誤差值e」，必須符合三個條件：

- 常態分配(如果為偏態或不符合常態性，所檢定出的結果很難推論)
- 變異數同質性
- 相互獨立(誤差值之間不能有關聯性存在)



原理2-多元迴歸分析

迴歸方程式種類_(P271)

標準化迴歸方程式

$$y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

1. 受到納入變項標準化的影響，方程式中沒有常數(截距)項。
2. 各自變項的加權值 β (標準化迴歸係數) 介於-1.0與1.0之間，可以比較大小，數值愈大，表示該變項權重愈大。

實例說明：參考表14-2(P276)的 β 值

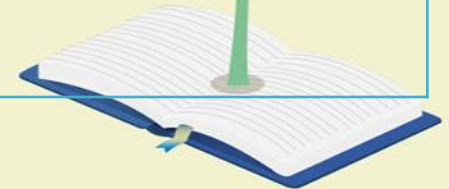
未標準化迴歸方程式

$$y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + \dots + b_n X_n$$

1. 方程式有截距項。
2. 各自變項的加權值 b (未標準化迴歸係數) 相互比較數值大小不具意義。

實例說明：參考表14-2(P276)的 b 值

在SPSS軟體中，迴歸分析結果檔兩種迴歸方程式都會出現。



原理2-多元迴歸分析-回歸分析步驟

步驟1.
提出假設性的迴
歸方程式

- 大量評閱文獻，找出合理的自變項和依變項
- 依據合理的論證定位自變項及依變項

步驟2.
評估資料符合基
本假定

- 自變項與依變項為線性關係
- 常態性
- 誤差獨立性
- 誤差等分散性

步驟3.
針對整體的迴歸方
程式及個別的迴歸
係數檢定

- 整體的迴歸方程式檢定：F值
- 個別的迴歸係數檢定：
對各個 β 、 e ，甚至每一個自變項對依變項的估計
檢定及預測分析



原理2-多元迴歸分析-迴歸分析步驟

步驟4

進行整體的迴歸
模式校正

- 瞭解自變項多元共線性的問題
- 透過資料的散佈圖可以瞭解資料是否有極端值
- 對於極端值及影響值的處理(刪除後再進行迴歸分析)
- 影響值的情形(圖14-4)
- 影響值與極端值(難辨別)：皆會影響直線的斜率
- SPSS實務操作：極端值與影響值的檢核

步驟5

對於各個變項間
的意義進行解釋

- 自變項以決定係數高低作為解釋依變項的程度
- 決定係數介於0~1.0之間，值越高可以預測依變項的程度愈高



原理2-多元迴歸分析-實例說明 (P276)

表 14-2 自變項對人力發展指數的迴歸分析摘要

變項	b	Std. Error	β	t	VIF
常數	-0.09**	.01		-8.18	
壽命	.01**	.00	.47**	28.90	2.43
識字	.00**	.00	.31**	17.94	2.83
教育	.00**	.00	.16**	7.86	3.70
所得	.00**	.00	.20**	14.00	1.91
Adj- R^2	.98				
F 值	$F_{(4,173)}$	= 2290**			

註： b 係數欄中有三個數值為 .00，達到 .01 的統計顯著水準，是因這三個變項的 b 係數都在小數點第四位之後才有數值，但它們有達到統計顯著水準，只是在小數點第二位無法看到更小的數值，因此仍將顯著符號標上，避免誤解，在此說明。

** $p < .01$.



原理3-虛擬變項的迴歸分析

- 一、若自變項中包括類別變項及連續變項，而依變項為連續變項時。
- 二、先將類別變項轉換為虛擬變項，進行虛擬編碼。
- 三、分為參照組及對照組。(原始變項為 N 個，新變項個數 $N-1$ 個)。
- 四、投入模式。
- 五、注意要掌握好每一變項的參照組，才可進行比較解說。



原理3-虛擬變項的迴歸分析

表 14-3 虛擬變項的結果

學生	原始資料				虛擬後的變項		
	成績	經濟	性別	社會階層	男性	高社會階層	中社會階層
1	85	20000	1	1	0	1	0
2	84	50000	2	2	1	0	1
3	81	45000	1	1	0	1	0
4	90	80000	2	3	1	0	0
5	70	30000	1	2	0	0	1
6	65	28000	1	1	0	1	0
7	75	34000	2	3	1	0	0
8	60	29000	2	1	1	1	0
9	88	72000	2	2	1	0	1
10	92	95000	1	1	0	1	0



注意事項

- 一、注意假設檢定(同一性或等分散性、自變項間的獨立性DW值)。
- 二、避免多元共線性產生，VIF指標判斷。
- 三、注意殘差值及偏離值。
- 四、選定自變項投入模式的方法(逐步法、強迫進入法)。

*逐步法:分為順向及反向選擇法，是一種探索式的分析。

(**順向**:先選取自變項與依變項相關係數最大—依變項中，淨相關最大的自變項一直到沒有顯著自變項為止；**反向**:所有自變項先納入—將自變項與依變項相關最小的移除一直到自變項都達顯著為止)

*強迫進入法:歸納的自變項強迫納入迴歸分析中，因為過去研究已掌握這些變項對依變項的重要因素。



操作與解說

Q:以校長的遴選功能、遴選影響及遴選指標對校長遴選問題進行多元回歸分析檢定。(強迫進入法)

表 14-4 自變項對校長遴選問題的迴歸分析摘要

變項	<i>b</i>	Std. Error	β	<i>t</i>	VIF
常數	.608**	.092		6.628	
功能	.195**	.032	.286**	6.203	1.246
影響	.065*	.026	.101*	2.438	1.005
指標	.338**	.042	.373**	8.093	1.242
Adj-R ²	.318				
F值	$F_{(3,396)}$	= 63.016**			

* $p < .05$. ** $p < .01$.



操作與解說

Q:以校長的遴選功能、遴選影響及遴選指標對校長遴選問題進行多元回歸分析檢定。(逐步迴歸分析法)

表 14-5 自變項對校長遴選問題的逐步迴歸分析摘要 (n = 400)

模式	變項	b	Std. Error	β	t	VIF
1	常數	.946**	.060		15.707	
	指標	.454**	.039	.500**	11.532	1.000
	功能	.190**	.032	.278	6.011	1.240
2	常數	.766**	.065		11.789	
	指標	.343**	.042	.378**	8.162	1.240
	功能	.190**	.032	.278	6.011	1.240
3	常數	.608**	.092		6.628	
	指標	.338**	.042	.373**	8.093	1.242
	功能	.195**	.032	.286**	6.203	1.246
	影響	.065**	.026	.101**	2.438	1.005
1	Adj-R ²	.249				
	F 值	132.99**				
2	Adj-R ²	.310				
	F 值	90.43**				
3	Adj-R ²	.318				
	F 值	63.02**				

**p < .01.

模型摘要^d

模型	R	R 平方	調整後 R 平方	標準標準誤	R 平方變更	F 值變更	變更統計量		顯著性 F 值變更
							自由度 1	自由度 2	
1	.500 ^a	.250	.249	.31916	.250	132.985	1	398	.000
2	.559 ^b	.313	.310	.30594	.063	36.131	1	397	.000
3	.568 ^c	.323	.318	.30405	.010	5.943	1	396	.015



問題一

Q:不同校別與職務對於校長遴選影響的迴歸分析。
(使用虛擬變項的迴歸分析)

表 14-6 不同職務對於校長遴選影響的迴歸分析

變項	<i>b</i>	Std. Error	β	<i>t</i>
常數	1.63**	.19		8.61
國小為參照組				
國中	.02	.06	.02	0.39
退休校長為參照組				
新職務一（教師）	.75**	.19	.65**	3.91
新職務二（主任或組長）	.85**	.19	.71**	4.40
新職務三（現任校長）	.87**	.22	.35**	3.92
Adj- R^2	.04			
<i>F</i> 值	$F_{(4,395)}$	= 5.214**		

註：虛擬變項投入迴歸分析之後，很容易有多元共線性，所以不呈現 VIF。

** $p < .01$.



問題二

Q:校別、職務(需要重新編碼)、指4、指5及指6對遴選影響的迴歸分析。

表 14-7 自變項對校長遴選影響的迴歸分析摘要 (n = 400)

變項	b	Std. Error	β	t
常數	1.660**	.212		7.834
國中 (國小為參照組)	.007	.062	.006	.120
教師 (退休校長為參照組)	.644**	.193	.554**	3.327
主任或組長 (退休校長為參照組)	.756**	.196	.628**	3.859
現任校長 (退休校長為參照組)	.793**	.222	.321**	3.578
指 4	-.032	.062	-.029	-.525
指 5	-.087	.056	-.090	-1.568
指 6	.191**	.059	.186**	3.238
Adj-R ²	.059			
F值	4.573**			

**p < .01.



感謝聆聽

