



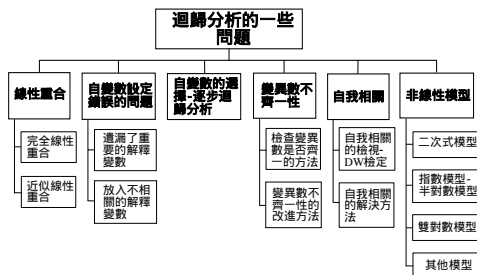
17 迴歸分析的一些問題

學習目的

1. 了解迴歸分析時常會碰到的一些問題如線性重合、字變數設定錯誤等問題。
2. 了解如何檢查近似線性重合。
3. 了解遺漏重要解釋變數的問題、及如何做逐步迴歸分析。
4. 了解變異數不齊一及如何改進。
5. 了解自我相關及其解決方法。
6. 討論非線性模型。



本章結構



線性重合

○ 線性重合

線性重合 (multicollinearity) 是指二個或多個自變數間具有高度線性相關的現象。

○ 完全線性重合

完全線性重合是指迴歸方程式中，解釋變數間具有完全線性相關。

○ 近似線性重合

近似線性重合是指迴歸方程式中，解釋變數間具有很高的線性相關。



線性重合

○ 近似線性重合的檢查方法

- ① 計算解釋變數間之相關係數
若 $r_{ij} < R^2$ (R^2 為原迴歸方程式的判定係數) 一般可判定無近似線性重合的現象。
- ② 利用 R^2
 R^2 為第 i 個解釋變數對其他解釋變數做迴歸的判定係數，又稱輔助的判定係數。若 R^2 大，會使 β 及 $V(\beta)$ 變大，而使 β 及 $V(\beta)$ 的計結果有問題。若 $R^2 < R^2$, $i=1, \dots, k$ ，則可判定無近似線性重合的情形。
- ③ 利用變異數膨脹因素 (variance inflation factor, VIF)。VIF 定義為：

$$VIF_i = \frac{1}{1 - R_i^2}$$

若 $VIF_i > 10$ ，則判定有線性重合的問題，則無線性重合的問題。



自變數設定錯誤的問題

○ 遺漏了重要的解釋變數

- (1) 估計的迴歸係數為 β 的偏誤估計式
- (2) 估計的變異數為 σ^2 的正偏誤估計式
- (3) 統計推論發生錯誤



自變數設定錯誤的問題

○ 放入不相關的解釋變數

- (1) $\hat{\beta}$ 仍為 β 的不偏估計式
- (2) 統計推論發生錯誤

逐步迴歸分析

○ 逐步迴歸分析

逐步迴歸分析是從眾多的自變數中，有系統的選擇較重要的變數放入迴歸模型。

變異數不齊一性

○ 變異數不齊一性的定義

若迴歸方程式中殘差項的變異數非為一固定常數，則稱發生變異數不齊一性，表為 $V(\varepsilon_i) = \sigma_i^2 \neq \sigma^2$ ，通常發生於橫斷面資料。當發生變異數不齊一時，最小平方估計式雖仍為一不偏估計式，但卻不是最佳線性不偏估計式 (BLUE)。

圖17.2 變異數不齊一性

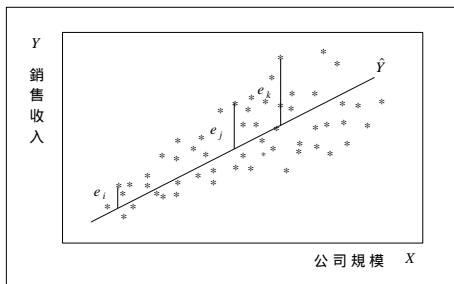


表17.20 調整變異數不齊一性的迴歸結果

	A	B	C	D	E
16		常數	標標準	t統計	P-值
17	$1/\sqrt{X_i}$	906.8412464	347.4098	2.610293	0.015645
18	$1/\sqrt{X_i}$	0.693948492	0.034713	47.1658	2.16E-24

自我相關

○ 自我相關的意義

當迴歸方程式的殘差項發生相關時，則稱發生自我相關，通常發生於時間數列資料。發生自我相關時，一般而言會使最小平方估計式發生偏誤，且不為最佳線性不偏估計式 (BLUE)。

圖17.5 無自我相關

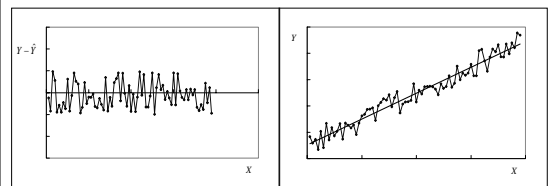
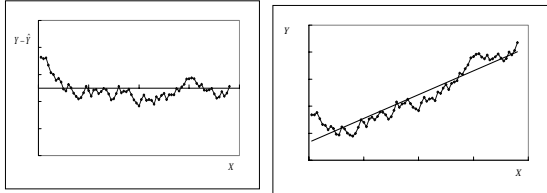


圖17.6 自我相關



自我相關

○ 自我相關的檢視 - DW 檢定

- ① 步驟1 利用OLS得估計迴歸方程式
- ② 步驟2 計算估計殘差值
- ③ 步驟3 檢定統計量DW
- ④ 步驟4 查DW值表

自我相關

○ 自我相關的判定法則

- ① 若 $DW < d_L$ ，則判定殘差值為真的自我相關。
- ② 若 $DW > 4 - d_L$ ，則判定為真的自我相關。
- ③ 若 $d_U < DW < 4 - d_U$ ，則判定不相關。
- ④ 若 $d_L < DW < d_U$ 或 $4 - d_U < DW < 4 - d_L$ ，則無法判定。

表17.24 薪資 DW 值的計算過程

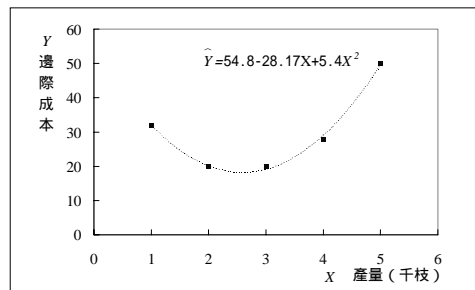
	D	C	D	D	F
1	薪資員工平均薪資指數Y	薪資工區工指數X	e_i	e_i^2	$e_i \cdot e_{i-1}$
2	25.20	46.1	1.877046	3.523081	
3	32	46.36	8.830386	77.77227	2.666272
4	32.96	38.1	1.622790	2.633588	4.111268
5	34.01	40.86	4.982149	24.82158	4.222104
6	37.6	35.12	-4.982149	24.82158	3.851643
7	41.4	34.87	-5.115039	26.15358	8.648952
8	43.46	35.24	-5.168812	26.71654	8.208004
9	36.28	45.42	3.944728	15.55997	8.206242
10	37.68	47.22	5.265050	27.72092	8.206242
11	43.2	32.28	-4.828627	23.31388	8.758421
12	43.47	38.11	-4.827686	23.31188	7.948743
13	38.18	33.12	-1.351736	1.827154	7.056494
14	41.26	44.42	-1.790444	3.205504	3.1194094
15	34	37.41	-1.011306	1.022731	1.8792804
16	36.88	44.42	-1.849870	3.421382	5.3829032
17	33.8	33.8	0.425049	0.180687	3.2584032
18	34.47	38.12	1.746723	3.051304	10.982204
19	47.54	32.28	6.427067	41.51682	7.317134
20			總計	11.7947584	42.267632
21			DW值		0.5484428

非線性模型

○ 二次式模型

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$$

圖17.8 邊際成本函數





非線性模型

○ 指數模型-半對數模型

$$P = ae^{bt} \varepsilon$$



圖17.9 人口成長模型-指數模型

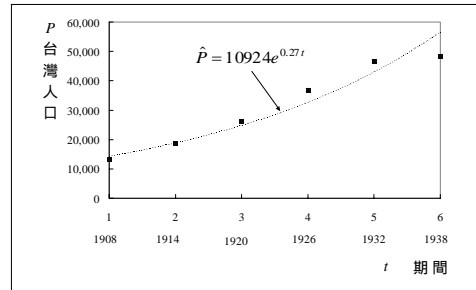
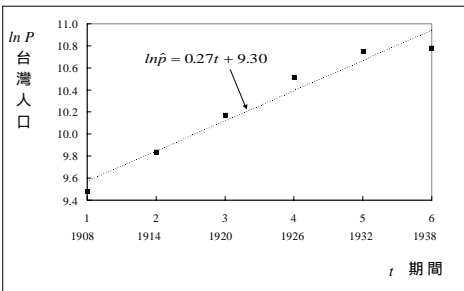


圖17.10 人口成長模型-半對數模型



非線性模型

○ 雙對數模型

$$Q = ALK^{\alpha} \varepsilon$$



圖17.11 供給函數

