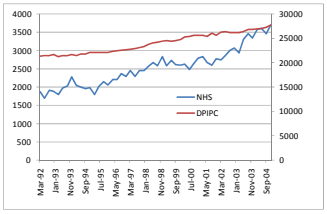
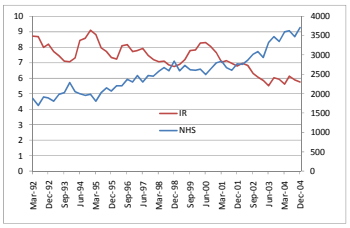
Multiple Regression Procedure

1. 多重迴歸分析的第一步是逐一檢視每個可能納入迴歸分析模式的因素
2. 用correlation相關係數或散佈圖檢查銷售量與因素間之關係，如果關係不明顯，就應質疑所設定之多重迴歸模式是否適當。

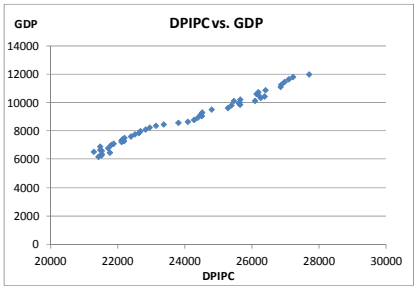




Department of Logistics Management

Multiple Regression Procedure

3. 檢視因素之間是否有multi-collinearity的問題。

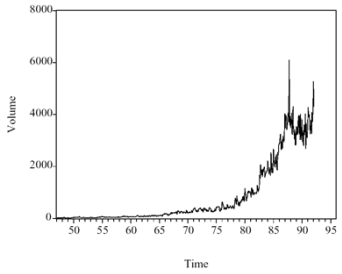


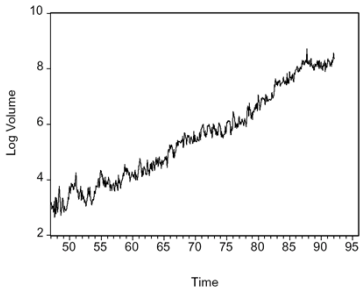
篩選多餘的因素：逐一將每個因素當成應變數，和其他因素(自變數)做迴歸分析，如 $x_{2i} \approx c_0 + c_1 x_{1i} + c_3 x_{3i} + \ldots + c_k x_{ki}$

Multiple Regression Procedure

4. 進一步探索各因素對銷售量的影響程度，如果因素與銷售量間並非線性關係時，除了考慮平方與指數關係，也可將因素進行數學轉換，如對數轉換(log transformation) 。

$$y_t \approx \beta_0 e^{\beta_1 \text{TIME}_t} \quad \Rightarrow \ln y_t \approx \ln \beta_0 + \beta_1 \text{TIME}_t$$





3

Department of Logistics Management

Multiple Regression Procedure

5. 如果不同區域會影響銷售，則將區域設為0-1變數。例如分區為北、中、南、東，則設定北區、中區及南區三個0-1變數，三者皆為0則代表東部。季節性因素的影響也可以同樣方式處理。有k個類別，就設定k-1個0-1變數。

6. 評估迴歸結果的第一步是先檢視整體模式之適合度。R^2的意義是所有因素解釋了多少比例的變異量。第二步是逐一檢視各因素之迴歸係數是否顯著，這要看每一迴歸係數的*t*-test值(絕對值應大於1.645)及p值(應小於0.05)。

4

Multiple Regression for Seasonality

假設銷售量只受到季節性影響

s is the number of seasons, s=4, 6, 12, or 52

Define s-1 seasonal dummy variables

如果是第1季，D_1 =1, D_2 =0, D_3 =0

如果是第2季，D_1 =0, D_2 =1, D_3 =0

如果是第3季，D_1 =0, D_2 =0, D_3 =1

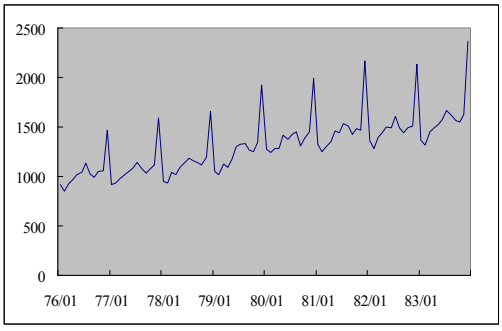
如果是第4季，D_1 =0, D_2 =0, D_3 =0

第s季的影響

Sales at each period → $y=\gamma_0+\gamma_1 D_1+\ldots+\gamma_{s-1}D_{s-1}+\varepsilon$ ← errors

當季與第s季的差距

Example: Liquor Sales



Liquor: 1976.01-1983.12

Setting up the Data

$$Y = \gamma_0 + \beta \cdot time + \sum_{i=1}^{s-1} \gamma_i D_i + \varepsilon$$ additive model

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Month	Liquor Sales	time	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11
107	75/10	1010		0	0	0	0	0	0	0	0	0	1	0
108	75/11	1016		0	0	0	0	0	0	0	0	0	0	1
109	75/12	1378		0	0	0	0	0	0	0	0	0	0	0
110	76/1	915	1	1	0	0	0	0	0	0	0	0	0	0
111	76/2	854	2	0	1	0	0	0	0	0	0	0	0	0
112	76/3	922	3	0	0	1	0	0	0	0	0	0	0	0
113	76/4	965	4	0	0	0	1	0	0	0	0	0	0	0
114	76/5	1014	5	0	0	0	0	1	0	0	0	0	0	0
115	76/6	1040	6	0	0	0	0	0	1	0	0	0	0	0
116	76/7	1137	7	0	0	0	0	0	0	1	0	0	0	0
117	76/8	1026	8	0	0	0	0	0	0	0	1	0	0	0
118	76/9	992	9	0	0	0	0	0	0	0	0	1	0	0
119	76/10	1052	10	0	0	0	0	0	0	0	0	0	1	0
120	76/11	1056	11	0	0	0	0	0	0	0	0	0	0	1
121	76/12	1469	12	0	0	0	0	0	0	0	0	0	0	0
122	77/1	916	13	1	0	0	0	0	0	0	0	0	0	0
123	77/2	934	14	0	1	0	0	0	0	0	0	0	0	0
124	77/3	987	15	0	0	1	0	0	0	0	0	0	0	0

D_{12}?

Department of Logistics Management

Regression with Seasonal Factors

$Y=\gamma_0+\gamma_1 D_1+\gamma_2 D_2+\ldots+\gamma_{11} D_{11}+\varepsilon$ 不考慮趨勢

	-569.9	-612	-641.8	-581.8	-537.4	-608.8	-636	-706.4	-724.6	-810.3	-769.3	1913.9
標準差	108.23	108.23	108.23	108.23	108.23	108.23	108.23	108.23	108.23	108.23	108.23	76.531
R^2=0.4876	216.46											
F值	7.266	84										
	4E+06	4E+06										

s1	s2	s3	s4	s5	s6
1144.6	1103.6	1189.3	1207.5	1277.9	1305.1
s7	s8	s9	s10	s11	s12
1376.5	1332.1	1272.1	1301.9	1344.0	1913.9

Combining Trend and Seasonal Factors

$Y=\gamma_0 + \beta \cdot time + \gamma_1 D_1 + \gamma_2 D_2 + \ldots + \gamma_{11} D_{11} + \varepsilon$ 考慮季節與趨勢

	-562.7	-597.7	-620.4	-553.2	-501.7	-566	-586.1	-649.3	-660.5	-739	-690.8	7.129	1528.911
標準差	27.30	27.30	27.30	27.31	27.32	27.32	27.33	27.35	27.36	27.37	27.39	0.20	22.19
R^2=	0.9678	54.596											
F值	207.82	83											
	7E+06	247397											

β=7.129

s1	s2	s3	s4	s5	s6
838.1	790.0	868.4	879.6	942.8	962.9
s7	**s8**	**s9**	**s10**	**s11**	**s12**
1027.2	975.7	908.5	931.2	966.2	1528.9

Forecasting Future Sales

1984.1⇒ time=97, D1=1, D2=D3=…=D11=0

第12季　　第1季與第12季的差距

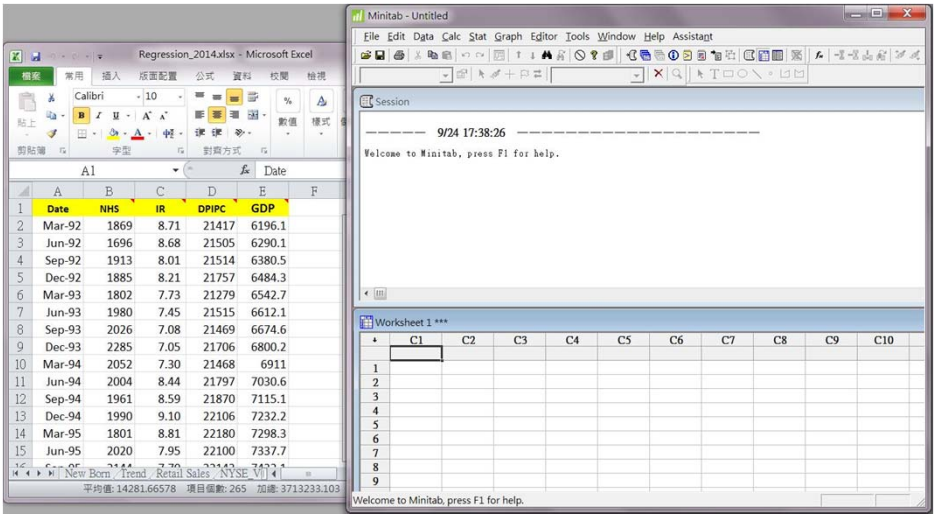
$Forecast=\gamma_0 + \beta \cdot 97 + \gamma_1 \cdot 1$

$=1528.9+7.129 \times 97+(-690.8)$

$=838.1+7.129 \times 97$

$=1529.613$

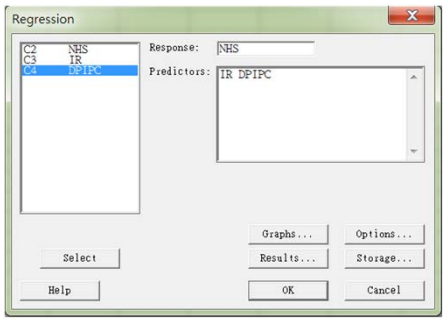
Running Regression on Minitab

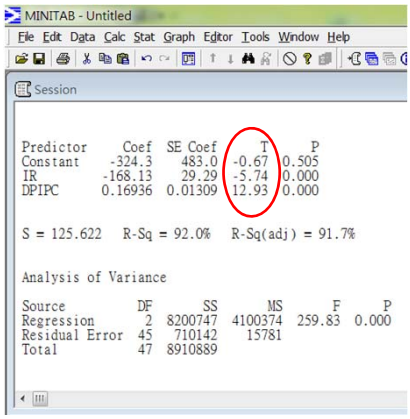


	Date	NHS	IR	DPIPC	GDP
2	Mar-92	1869	8.71	21417	6196.1
3	Jun-92	1696	8.68	21505	6290.1
4	Sep-92	1913	8.01	21514	6380.5
5	Dec-92	1885	8.21	21757	6484.3
6	Mar-93	1802	7.73	21279	6542.7
7	Jun-93	1980	7.45	21515	6612.1
8	Sep-93	2026	7.08	21469	6674.6
9	Dec-93	2285	7.05	21706	6800.2
10	Mar-94	2052	7.30	21468	6911
11	Jun-94	2004	8.44	21797	7030.6
12	Sep-94	1961	8.59	21870	7115.1
13	Dec-94	1990	9.10	22106	7232.2
14	Mar-95	1801	8.81	22180	7298.3
15	Jun-95	2020	7.95	22100	7337.7

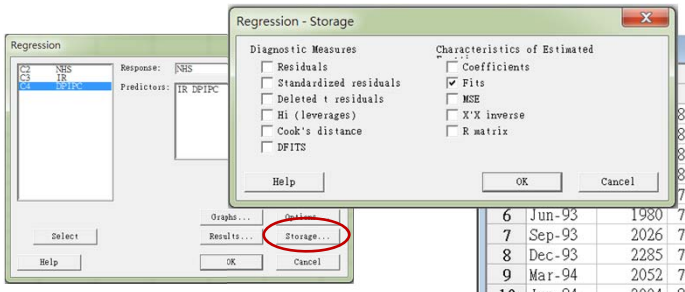
Running Regression on Minitab

Stat>Regression>Regression



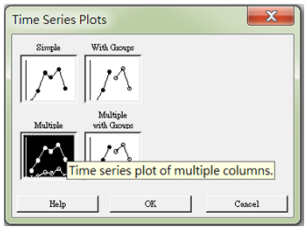


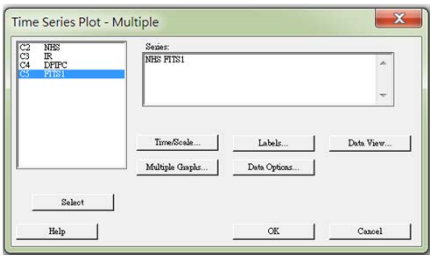
T-test絕對值>1.645代表影響顯著



	C3	C4	C5
	IR	DPIPC	FITS1
	8.71000	21417	1838.41
	8.67667	21505	1858.92
	8.01000	21514	1972.53
	8.20667	21757	1980.62
	7.73333	21279	1979.24
6	Jun-93	1980	7.45333	21515	2066.29
7	Sep-93	2026	7.08000	21469	2121.26
8	Dec-93	2285	7.05333	21706	2165.88
9	Mar-94	2052	7.29667	21468	2084.67
10	Jun-94	2004	8.44000	21797	1948.16
11	Sep-94	1961	8.58667	21870	1935.87
12	Dec-94	1990	9.10000	22106	1889.53

Graph>Time Series Plot

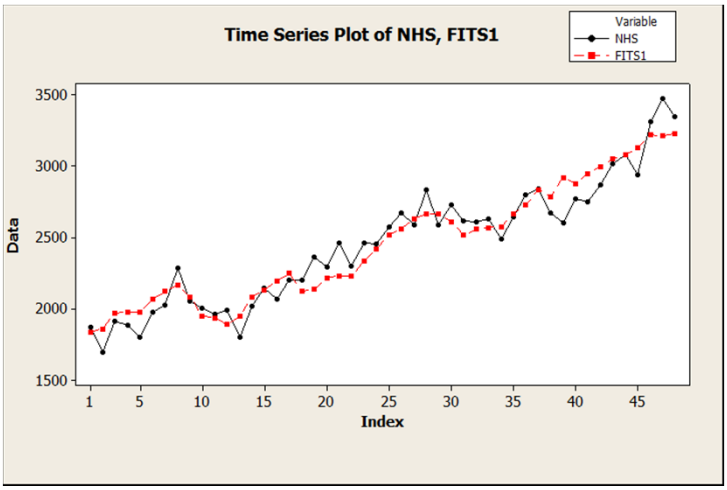




13

New Houses Sold vs. Regression Model

$$NHSF1_i = -324.33 + 0.17(DPIPC)_i - 168.13(IR)_i$$



14

Multicollinearity

$$NHSF1_i = -324.33 + 0.17(DPIPC)_i - 168.13(IR)_i$$

$$NHSF2_i = 1884 - 0.01(DPIPC)_i + 0.23(GDP)_i - 147.82(IR)_i$$

- 新加入的因素(GDP)可能與原有因素(DPIPC)高度相關，互相重疊的結果是不合理的迴歸係數。
- 因素(自變數)越多，迴歸的結果不見得會更好

	DPIPC	GDP	IR
DPIPC	1		
GDP	***0.99***	1	
IR	−0.65	−0.67	1

15

Department of Logistics Management

MINITAB - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help

Session

Predictor	Coef	SE Coef	T	P
Constant	-324.3	483.0	-0.67	0.505
IR	-168.13	29.29	-5.74	0.000
DPIPC	0.16936	0.01309	12.93	0.000

S = 125.622 R-Sq = 92.0% R-Sq(adj) = 91.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	8200747	4100374	259.83	0.000
Residual Error	45	710142	15781		
Total	47	8910889			

Worksheet 1 ***

	C1-T	C2	C3	C4	C5
	Date	NHS	IR	DPIPC	
1	Mar-92	1869	8.71000	21417	
2	Jun-92	1696	8.67667	21505	
3	Sep-92	1913	8.01000	21514	
4	Dec-92	1885	8.20667	21757	
5	Mar-93	1802	7.73333	21279	
6	Jun-93	1980	7.45333	21515	
7	Sep-93	2026	7.08000	21469	
8	Dec-93	2285	7.05333	21706	

MINITAB - Untitled

File Edit Data Calc Stat Graph Editor Tools Window Help

Session

Predictor	Coef	SE Coef	T	P
Constant	1884.2	901.5	2.09	0.042
IR	-147.82	28.18	-5.25	0.000
DPIPC	-0.01378	0.06593	-0.21	0.835
GDP	0.23224	0.08218	2.83	0.007

S = 116.876 R-Sq = 93.3% R-Sq(adj) = 92.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	8309850	2769950	202.78	0.000
Residual Error	44	601038	13660		
Total	47	8910889			

Worksheet 1 ***

	C1-T	C2	C3	C4	C5
	Date	NHS	IR	DPIPC	GDP
1	Mar-92	1869	8.71000	21417	6196.1
2	Jun-92	1696	8.67667	21505	6290.1
3	Sep-92	1913	8.01000	21514	6380.5
4	Dec-92	1885	8.20667	21757	6484.3
5	Mar-93	1802	7.73333	21279	6542.7
6	Jun-93	1980	7.45333	21515	6612.1
7	Sep-93	2026	7.08000	21469	6674.6
8	Dec-93	2285	7.05333	21706	6800.2

Forecasting with Multiple Regression

- 迴歸模式估計銷售量(y)與可控制因素或可觀察因素(x)之線性關係

$$y=f(x_1,\ldots,x_k)+\varepsilon \quad \text{神秘的真實關係}$$

$$y\approx\beta_0+\beta_1x_1+\ldots+\beta_kx_k+\varepsilon \quad \text{假設是線性關係}$$

$$\text{業績}\approx\text{基本營業額}+\beta_1\text{景氣}+\beta_2\text{行銷}\ldots+\beta_k\text{品質}+\text{運氣}$$

- 下一期的行銷費用與產品品質皆可控制。
- 如何得知下一期的景氣，以預測下一期的銷售？

Choosing Valid Indicators

1. 選擇領先指標

本期已發生的指數可預測下一期的銷售

生產量、出貨量、庫存量

氣候、原物料盤價

消費者信心指數、股市指數、房地產交易量

2. 利用過去紀錄以估計可觀察因素的未來值。

Case: Multiple Regression Forecasting

1. X公司行銷部門要求預測某產品線各種規格之銷售量。
2. 消費者產品生命週期長，預測期間為未來3至12個月。
3. 由於品項眾多，預測人員決定專注於總銷售量之預測。
4. 因準備時間有限，目標為10%的預測誤差。
5. 與行銷部門協議使用下列影響銷售的因素

   Time

   Average retail price 平均零售價格

   National advertising expenditures 廣告費用

   Nielsen shipment data 零售業出貨指數(景氣)

   Direct mail coupon 郵寄折價券

   FSI 1, 2, 3, 4, 5 五種報章雜誌廣告與折價券

多重迴歸結果

TABLE 5.11 Variables and Statistics of Consumption Model

R-squared = 0.96 *F*-stat = 24.55
Adj. *R*-squared = 0.92 DW = 2.26

Variable	*t*-Stat
1. Time	−0.72
2. Average retail price	−2.70
3. National advertising expenditures	2.52
4. Nielsen shipment data in units	6.48
5. FSI 1	4.38
6. FSI 2	2.31
7. Direct mail coupon	1.93
8. FSI 3	1.25
9. FSI 4	2.15
10. FSI 5	2.81

Notes: (1) Dummy variables are used to capture the effects of variables that have no quantitative data.
(2) FSI stands for "free-standing insert."

不一定要排除t值小於2的因素

TABLE 5.12 Forecasts versus Actuals: Consumptio Model
Brand X

Month	Actuals	Forecasts	Absolute Error
January	2,578	2,563	1%
February	2,788	2,783	0%
March	2,957	2,957	0%
April	2,670	2,758	3%
May	2,447	2,466	1%
June	3,016	3,016	0%

Note: Mean absolute percentage error (MAPE) = 0.81%

進行預測的困難

- 廣告費用與折價券種類為可控制因素
- 無法控制或預知 Nielsen 出貨指數
- 發展第二個多重迴歸模式以預測Nielsen出貨指數

R-squared = 0.96 F-stat = 36.93
Adj. *R*-squared = 0.93 DW = 2.41

Variable	*t*-Stat
1. Time	2.40
2. Retail consumption	3.59
3. Trade inventory	1.87
4. Trade price	−1.55
5. Trade promotion 1 early shipment	5.82
6. Trade promotion 1 sell in	16.01
7. Trade promotion 1 post shipment	4.19
8. Trade promotion 2 early shipment	9.57
9. Trade promotion 2 sell in	1.18
10. Trade promotion 3 early shipment	2.62
11. Trade promotion 3 sell in	7.29
12. Trade promotion 3 post shipment	13.55

Month	Actuals	Forecasts	Absolute Error
January	69,158	69,190	0%
February	45,927	47,216	3%
March	40,183	40,183	0%
April	56,427	54,841	3%
May	81,854	72,788	12%
June	50,505	52,726	4%
July	37,064	36,992	0%
August	58,212	57,347	2%
September	96,566	95,112	2%

Note: Mean absolute percentage error (MAPE) = 3.0%

21

Department of Logistics Management

預測未來銷售量

- 影響出貨指數的因素也是只能觀察。
- 根據過去數據，利用外插法估計影響出貨指數的因素之未來值。例如以過去各期的業界平均價格估計未來的業界平均價格。
- 將因素的估計值代入第二個多重迴歸模式，以估計Nielsen出貨指數 的未來值。
- 將估計的Nielsen出貨指數與其他可控制因素代入第一個多重迴歸模式，以預測計產品線的未來總銷售量。
- 根據過去六個月的銷售比例，將總銷售量分割為各種規格的銷售量。

22

Conclusion

- 線性迴歸可用於趨勢分析與簡單的因果關係分析
- 多重迴歸分析可以有效分析多重因素對於銷售量的影響。
- 慎選影響銷售量的因素，因素越多不代表未來預測會更準確。
- 小心解讀迴歸分析的結果，利用圖形輔助分析。