



## 逢甲大學學生報告 ePaper

報告題名：影響手機價格高低不同的因素

作者：李俊穎、廖晟淵、吳昀珊、王良禎、蔡承翰、安大成

系級：統計三乙

學號：D9420926、D9452980、D9551726、D9588178、D9590834、D9593379

開課老師：陳婉淑

課程名稱：迴歸分析

開課系所：統計系

開課學年：96 學年度 第1 學期

## 目次

摘要.....	2-
研究動機.....	3-
研究目的.....	4-
選擇變數.....	5-
第一章 資料分析.....	6-
第一節 一般敘述統計	
第二節 散佈圖	
第三節 建立模式.....	13-
第二章 殘差分析	
第一節 檢查殘差平均是否為 0.....	21-
第二節 檢查殘差是否來自常態.....	21-
第三節 殘差圖.....	22-
第四節 檢查影響點與異常點.....	25-
第三章 刪除異常點後之迴歸分析.....	28-
第四章 結論.....	33-
參考文獻.....	34-

## 摘要

手機已經成為現今最主要的通訊產品之一，除此之外手機的外型也是顯現個人身份地位的象徵，不同的造型、功能以及配備都凸顯著每個人不同的品味，工作環境的不同都造成選擇上的差異，像是黑莓機風靡歐美黑莓機，能夠掌握最新資訊處理重要急件，成為商務人士的最愛，結合照相功能的手機，讓消費者只須花一筆錢就擁有兩種功能以及音樂播放軟體，讓手機不單單只是聯絡的工具，還多了娛樂的功能，也因此功能不同也造成手機的價格不一，Nokia 運用獨占性的領到策略，提供穩定的零間支援及提高主力機種效率化的生產體制造就在手機品牌上難以攻破的地位，所以我們以 Nokia 的手機作為比較，例如相機已經幾乎成為手機必備的功能之一，但是不同的畫素和照像功能就變成影響手機價個的因素，不同功能不同等級所耗費的製造成本也會有所不同，我們想從中探討出這些因素影響價格的關係和程度，我們移除變數  $X_5$  (手機通話時間)，選取  $X_1$  (手機內建相機畫素)、 $X_2$  (手機介面畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度) 五個變數留在模式中。

**關鍵字：** 向前選取法、向後消去法、逐步迴歸法、散佈圖

**最終模式的配適迴歸線：**

$$\hat{y}_i = -3434.1506 + 14.2755x_1 + 1.2354x_2 + 209.2073x_3 - 20.2042x_4 - 351.7187x_6$$

## 研究動機

科技日新月異，手機人手多隻，價格從零元到三萬元不等，甚至還有更昂貴的手機，功能不一，從一開始只是用來通話的黑金剛到現在的高規格高畫素多功能手機，譬如：照相功能、手機遊戲、藍牙傳輸、鈴聲和絃、MP3 …等等，只能基本功能用來說電話的手機現在被迫只能 0 元銷售，而多功能的手機越來越多，所以我們想研究這些功能到底對於價格的影響有多高，現在手機業者競爭激烈，已經看不到以前單純的只是為了通話存在的手機，連沒照相功能的手機都被視為異物，數量越來越少，照相功能已經變成手機必備了，消費者追求的是可不可以跟照起來跟相機一樣清晰，功能越來越多像是 3G、上網、用手機聊 MSN 或是發送 e-mail 甚至是玩 online games 都有，以及跟衛星導航的結合，所以這些因素引起我們的興趣，這些功能會對價格產生多少影響力。

## 研究目的

我們找了是否有相機功能、相機畫素、手機重量、待機時間、通話時間、手機的厚度、螢幕尺寸、藍芽 …等變數，想知道廠商在決定價格的時候，有沒有把這些因素考慮進去，影響的成份大不大，像是某些功能會造成成本偏高，為了避免不同廠商有不同的定價模式，為了消除這個無法看出的變數，我們決定拿同一家廠牌公司的手機來做比較，看這家公司在決定自家廠牌手機的價格是如何訂定價格。

我們最後選擇了 NOKIA，因為他是全球領導行動通訊業者又是全球最大的行動電話以及行動、寬頻、與 IP 網路的領導供應商，

廠牌	出貨量	佔有率	成長率
Nokia	44.2	28.9%	11.9%
Motorola	25.1	16.4%	50.3%
Samsung	19.1	12.5%	55.3%
Siemens	12.3	8.0%	41.4%
SONY Ericsson	8.5	5.6%	57.4%
LG	8.1	5.3%	44.6%
其他	35.7	23.3%	37.8%

由此可見 NOKIA 在市場上的佔有率，相信他在價格上有一定的公信力，讓消費者願意花此價格購買手機，同一廠牌的規格以及品質會比較一致，我們再針對不同的功能進行比對分析。

影響手機價格高低不同的因素

## 選擇變數

$Y$ : 手機價格(元)

$X_1$ : 手機內建相機畫素(百萬) - 畫素越高製作成本越高。

$X_2$ : 手機介面畫素(百萬)

$X_3$ : 手機重量(g) - 越精密的零件所耗成本越多。

$X_4$ : 待機時間(小時) - 電池的成本。

$X_5$ : 通話時間(小時) - 電池的成本。

$X_6$ : 手機厚度(公厘mm) - 要把東西放在體積越小的容器裡，需要

比較精密的零件，相對的作業工程也較繁雜。

因為藍芽幾乎是目前手機必備的，螢幕大小也都大同小異所以這兩個變數最後我們沒有採取。

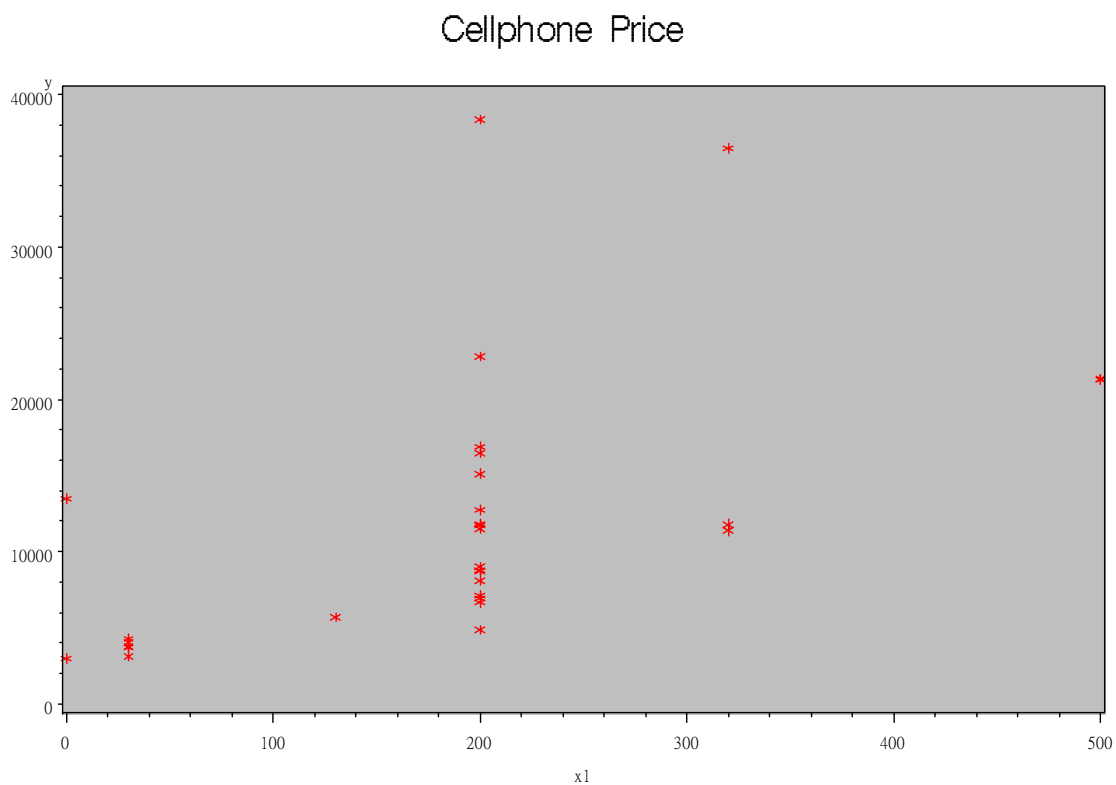
## 第一章 資料分析

### 第一節 一般敘述統計

	價格	相機畫素 (萬)	手機畫素(萬)	手機重量(g)	待機時間 (小時)	通話時間 (小時)	厚度(mm)
平均數	12053.30	188.00	955.01	108.37	285.47	4.18	16.98
標準誤	1617.69	22.70	151.83	5.18	10.37	0.27	0.71
中間值	10208	200	1670	104	284	3.6	17.3
眾數	-	200	1670	115	300	3	20.7
標準差	8860.45	124.33	831.63	28.35	56.81	1.47	3.88
變異數	78507535.80	15457.93	691600.81	803.48	3227.77	2.16	15.06
峰度	2.90	1.19	-2.06	4.59	1.95	2.46	-0.69
偏態	1.66	0.66	-0.28	1.59	1.10	1.41	-0.41
範圍	35399	500	1663.45	144	266	6.5	13.5
最小值	3011	0	6.55	66	190	2.5	9.5
最大值	38410	500	1670	210	456	9	23
總和	361599	5640	28650.2	3251	8564	125.35	509.4
個數	30	30	30	30	30	30	30

## 第二節 散佈圖

圖一  $Y$ (手機價格)對  $X_1$ (手機內建相機畫素)的散佈圖

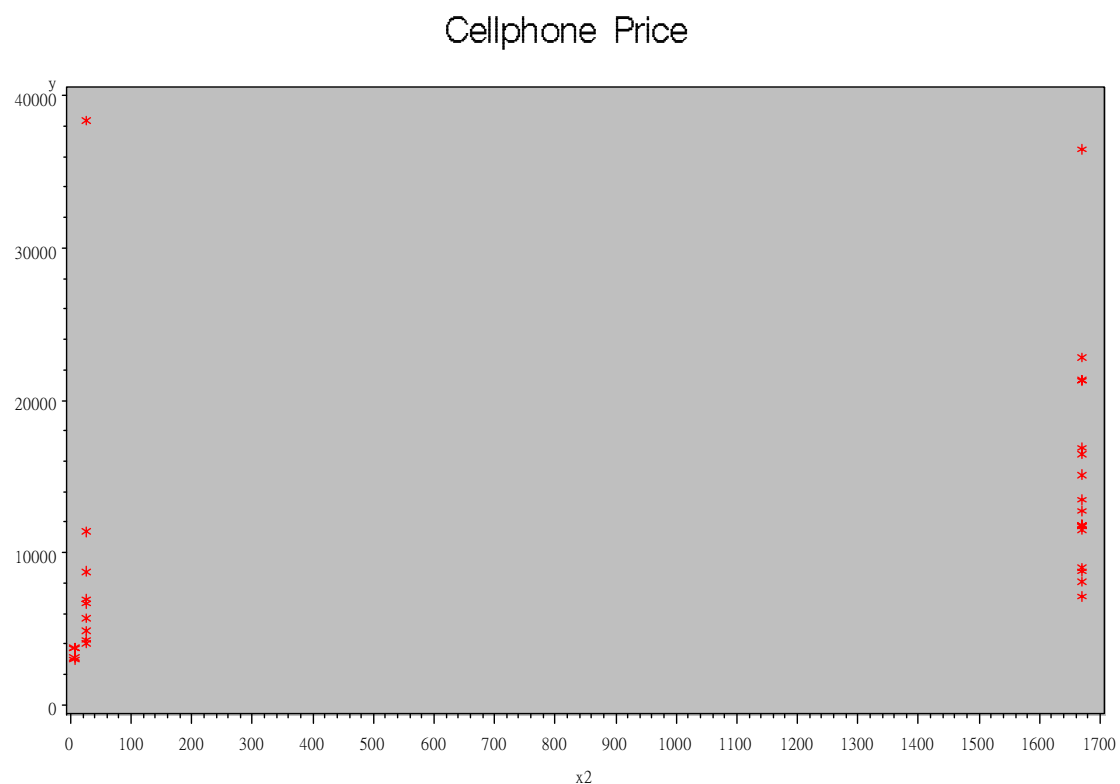


*Correlation Coefficient= 0.47867*

由圖一看出  $Y$  與  $X_1$  無線性關係，得知手機內建相機畫素對手機價格無明顯的關係。



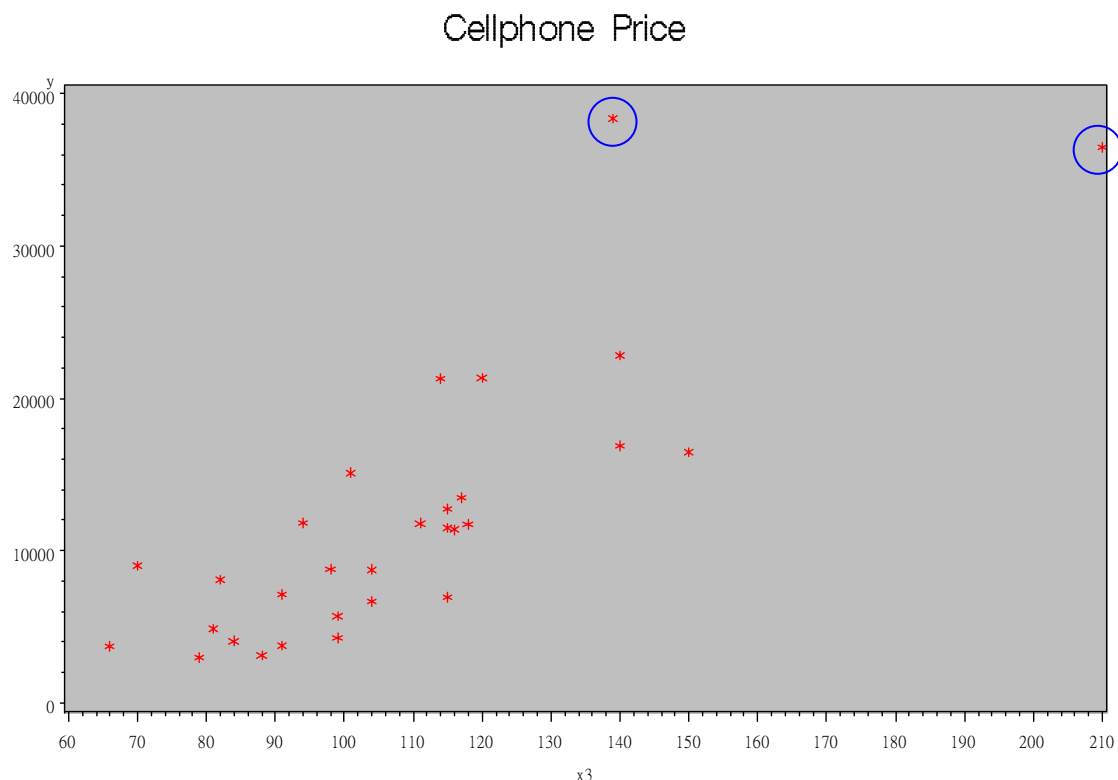
圖二  $Y$ (手機價格)對  $X_2$ (手機介面畫素)的散佈圖



*Correlation Coefficient= 0.25534*

由圖二看出  $Y$  與  $X_2$  無線性關係，可知道手機介面畫素對手機價格無明顯的關係。

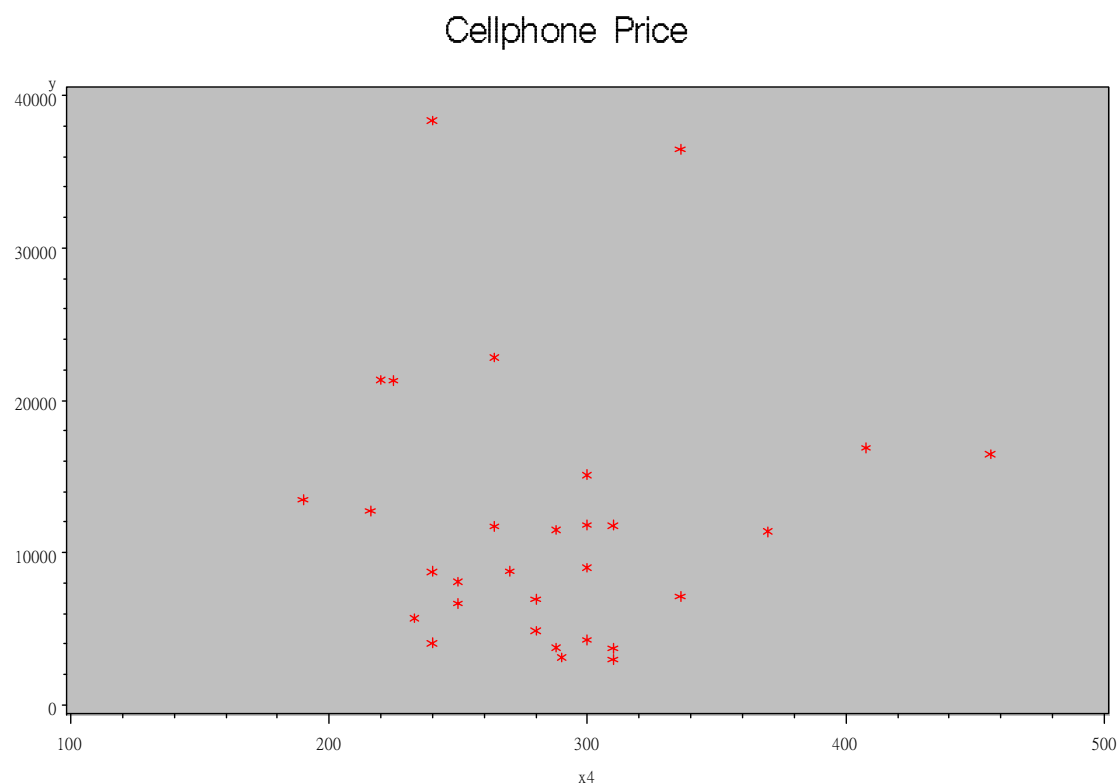
圖三  $Y$ (手機價格)對  $X_3$ (手機重量)的散佈圖



*Correlation Coefficient= 0.82375*

由圖三看出  $Y$  與  $X_3$  有線性關係，有顯著的線性相關。表示手機的價格會隨著手機的重量而呈現正相關。此時發現似乎有異常點出現，右方的點是一款結合折疊式小筆記型電腦的手機 nokiaE90，由於結合筆電的功能，故其重量比僅為手機的其他款式來的重；而上方的點是一款外殼為 18K 鍍金的手機 nokia8800 故其價格上比其他款手機來的昂貴，在後面的分析我們會深入討論。

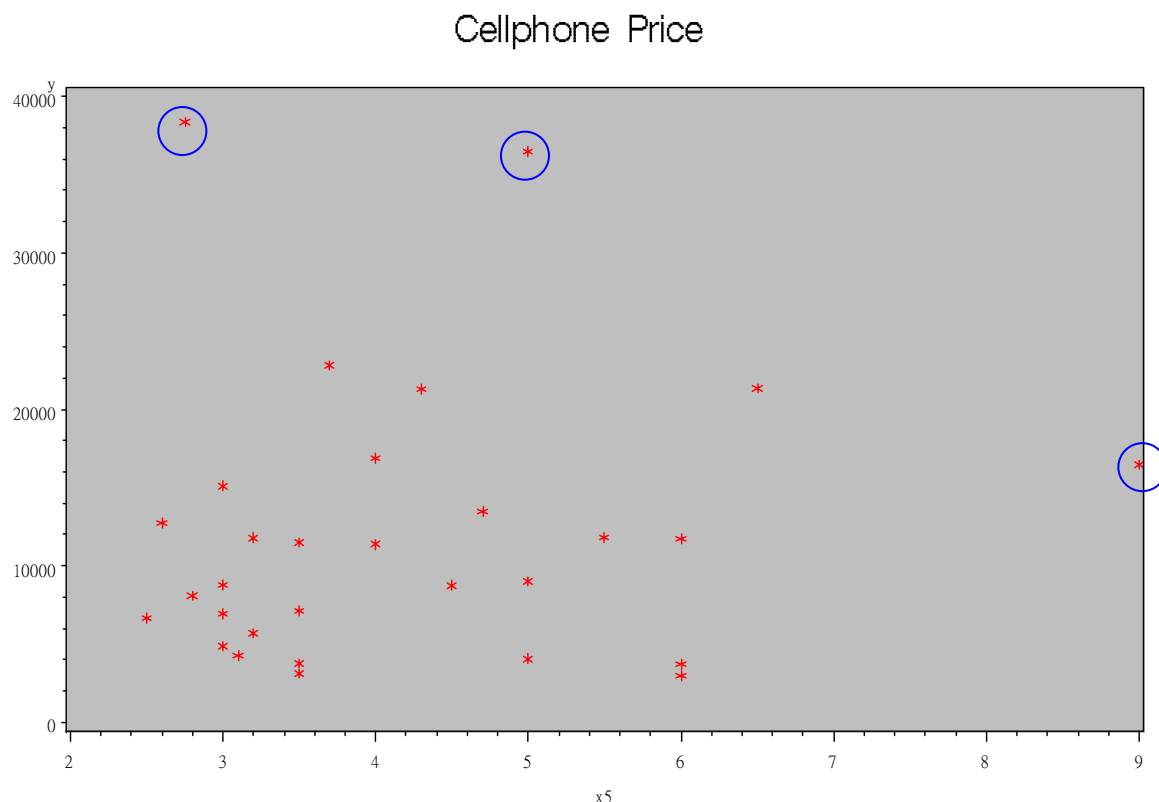
圖四  $Y$  (手機價格)對  $X_4$  (待機時間)的散佈圖



*Correlation Coefficient* =  $-0.00779$

由圖四看出  $Y$  與  $X_4$  無線性關係，點約集中在某個範圍。所以手機價格與待機時間沒有很明顯的關聯。亦即現在手機的待機時間約略相同。

圖五  $Y$  (手機價格)對  $X_5$  (通話時間)的散佈圖

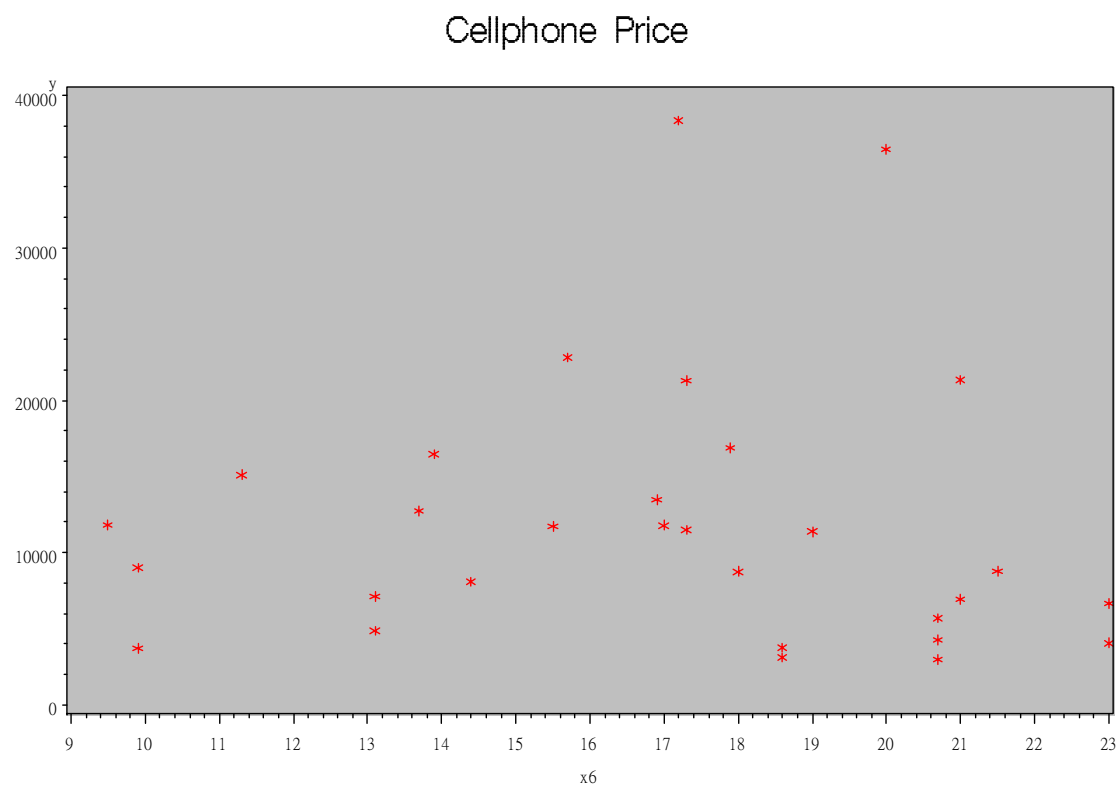


*Correlation Coefficient* =  $-0.00430$

由圖五看出  $Y$  與  $X_5$  無線性關係，點約集中在某個範圍。

所以手機價格與待機時間沒有很明顯的關聯。亦即現在手機的通話時間約略相同，此時也發現似乎有異常點出現，右方的點為一款以通話時間為主打的手機 nokiaE61i；左上方的點為一款外殼以 18K 鍍金的手機 nokia8800，故其價格比其他手機來的昂貴；上方的點為一款折疊式小型筆記型電腦的手機 E90，故其價格比其他手機來的昂貴，在後面的分析我們會加以討論。

圖六  $Y$ (手機價格)對  $X_6$ (手機厚度)的散佈圖



*Correlation Coefficient= 0.03419*

由圖六看出  $Y$  與  $X_6$  無線性關係，點分布在各個範圍中，沒有明顯的關係。所以手機價格與手機厚度沒有很明顯的關聯。

### 第三節 建立模式

#### 一、模式一：Full model

表一		Parameter Estimates					
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation
Intercept	Intercept	1	3990.1757	6666.3360	0.60	0.5553	0
x1	camera pixels	1	15.6155	7.8272	2.00	0.0580	1.4482
x2	cell phone pixels	1	-1.2931	1.3313	-0.97	0.3415	1.8745
x3	weight	1	282.2412	37.2636	7.57	<.0001	1.7061
x4	await time	1	-41.7583	16.4609	-2.54	0.0184	1.3375
x5	communticate time	1	60.7892	605.7830	0.10	0.9209	1.2146
x6	thickness	1	-739.4930	266.5273	-2.77	0.0108	1.6356
$R^2 = 0.8084 \quad R_{adj}^2 = 0.7584 \quad \hat{\sigma} = 4354.75864$							

由表一可知配適迴歸線為：

$$\hat{y}_i = 3990.1757 + 15.6155x_1 - 1.2931x_2 + 282.2412x_3 - 41.7583x_4 + 60.7892x_5 - 739.4930x_6$$

所以由表一中得知  $R^2 = 0.8084$  及  $R_{adj}^2 = 0.7584$  表示用此配適迴歸線所能解釋到

$Y$  的變異有 80.84%，因此解釋能力相當好。並從中得知  $X_3$  (手機重量)、 $X_4$  (待機時間)  $X_6$  (手機厚度) 有顯著的效果，而  $X_1$  (內建相機畫素) 也有不錯的顯著效果，其餘的變數皆無顯著效果。

影響手機價格高低不同的因素

## 二、模式二：選擇重要變數

### (一) 向前選取法(FORWARD selection)：

在每一次選擇的步驟中，選出一個變項，對模式的貢獻最大者，進入迴歸方程式中，並對尚未進入迴歸程式的預測變項加以考驗，以決定某一個預測變項是否有資格被納入迴歸模式中。

表二		Summary of Forward Selection						
Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x3	weight	1	0.6621	0.6621	14.5633	54.87	<.0001
2	x1	camera pixels	2	0.0501	0.7122	10.5480	4.70	0.0391
3	x6	thickness	3	0.0341	0.7463	8.4535	3.50	0.0728
4	x4	await time	4	0.0542	0.8006	3.9435	6.80	0.0152
5	x2	cell phone pixels	5	0.0078	0.8083	5.0101	0.97	0.3336
R-Square = 0.8083 and C(p) = 5.0101								

所以由表四得知，最後選取的變數有  $X_1$  (手機內建相機畫素)、 $X_2$  (手機介面畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度)。

影響手機價格高低不同的因素

(二)向後消去法(BACKWARD selection)：

首先將所有預測變項放入迴歸方程式中，而後在每一次淘汰的步驟中，剔出一個變項，對模式的貢獻最小者，並對留在迴歸方程式中的預測變項加以考驗，以決定某一個預測變項是否應繼續被保留在迴歸模式中。

Step	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x5	communticate time	5	0.0001	0.8083	5.0101	0.01	0.9209
2	x2	cell phone pixels	4	0.0078	0.8006	3.9435	0.97	0.3336

由表二可知消去的變數有  $x_2$ 、 $X_5$ (手機厚度)，最後我們留下

Variable $X_2$ 、 $X_5$ Removed : R-Square = 0.8006 and C(p) = 3.9435					
Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	2008.77152	6092.23281	1974618	0.11	0.7444
x1	12.88339	7.14834	58996292	3.25	0.0836
x3	269.20460	33.85158	1148635065	63.24	<.0001
x4	-39.50821	15.15368	123456336	6.80	0.0152
x6	-604.95162	220.60514	136579347	7.52	0.0111

所以由表三得知，我們最後選取的變數為  $X_1$ (手機內建相機畫素)、 $X_3$ (手機重量)、 $X_4$ (待機時間)、 $X_6$ (手機厚度)。



(三) 逐步迴歸法 (STEPWISE selection) :

是傾向選擇法與反向淘汰法的綜合。首先模式中不包含任何預測變項。然後採順向選擇法，根據對模式的貢獻最大者，挑選預測變項進入迴歸模式中。而在每一步驟中，已被納入模式的預測變項則必須再經過反向淘汰法的考驗，以決定該變項要被淘汰亦或留下。

由表五得知，最後留下的變數有  $X_1$  (手機內建相機畫素)、 $X_2$  (手機介面畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度)。

表五		Summary of Stepwise Selection							
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr>F
1	x3		weight	1	0.6621	0.6621	14.5633	54.87	<.0001
2	x1		camera pixels	2	0.0501	0.7122	10.5480	4.70	0.0391
3	x6		thickness	3	0.0341	0.7463	8.4535	3.50	0.0728
4	x4		await time	4	0.0542	0.8006	3.9435	6.80	0.0152
R-Square = 0.8006 and C(p) = 3.9435									

(四). 校正後的複判定係數法( $R_{adj}^2$  selection):

校正後的複判定係數法( $R_{adj}^2$ ): 估算全部可能的迴歸模式之 $R_{adj}^2$ 值, 相互比較, 以選取最大之 $R_{adj}^2$ 為最佳最有效的迴歸模式。

表六 Adjusted R-Square Selection Method				
Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
4	0.7687	0.8006	3.9435	x1 x3 x4 x6
5	0.7684	0.8083	5.0101	x1 x2 x3 x4 x6
5	0.7590	0.8006	5.9435	x1 x3 x4 x5 x6
6	0.7584	0.8084	7.0000	x1 x2 x3 x4 x5 x6
3	0.7486	0.7747	5.0544	x3 x4 x6
4	0.7393	0.7753	6.9811	x2 x3 x4 x6
4	0.7386	0.7747	7.0544	x3 x4 x5 x6
5	0.7285	0.7753	8.9801	x2 x3 x4 x5 x6
3	0.7171	0.7463	8.4535	x1 x3 x6
4	0.7119	0.7516	9.8204	x1 x3 x5 x6
$R^2 = 0.8006$ $R_{adj}^2 = 0.7687$				

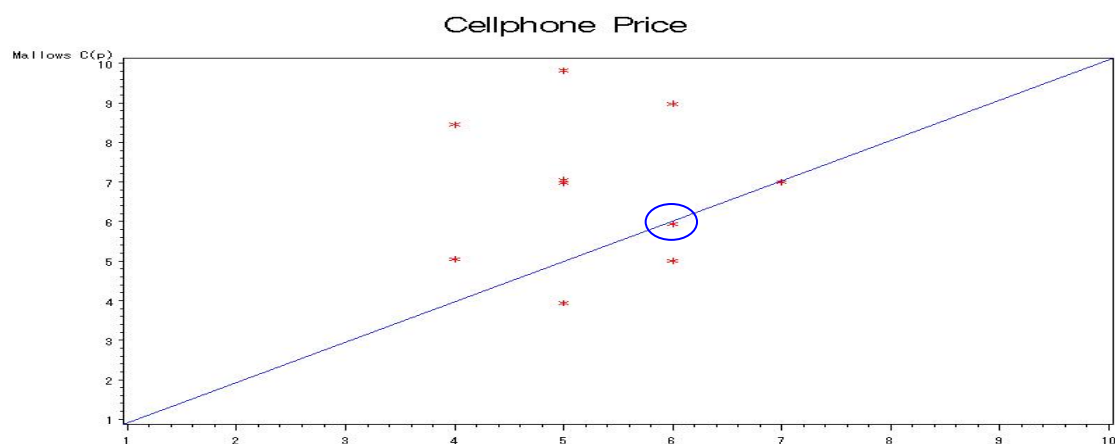
由表六得知, 最後所選取的變數為 $X_1$ (手機內建相機畫素)、 $X_3$ (手機重量)、 $X_4$ (待機時間)、 $X_6$ (手機厚度)。

(五). CP 選取法：

Mallows(1973)的  $C_p$  法：估算全部的迴歸模式之  $C_p$  值，相互比較，以選取最小之值  $C_p$  及  $C_p - P$  小為最佳、最有效的迴歸模式。

表七 Adjusted R-Square Selection Method				
Number in Model	Adjusted R-Square	R-Square	C(p)	Variables in Model
4	0.7687	0.8006	3.9435	x1 x3 x4 x6
5	0.7684	0.8083	5.0101	x1 x2 x3 x4 x6
5	0.7590	0.8006	5.9435	x1 x3 x4 x5 x6
6	0.7584	0.8084	7.0000	x1 x2 x3 x4 x5 x6
3	0.7486	0.7747	5.0544	x3 x4 x6
4	0.7393	0.7753	6.9811	x2 x3 x4 x6
4	0.7386	0.7747	7.0544	x3 x4 x5 x6
5	0.7285	0.7753	8.9801	x2 x3 x4 x5 x6
3	0.7171	0.7463	8.4535	x1 x3 x6
4	0.7119	0.7516	9.8204	x1 x3 x5 x6

由下圖得知，最後所選取的變數為  $X_1$  (手機內建相機畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_5$  (通話時間)、 $X_6$  (手機厚度)。



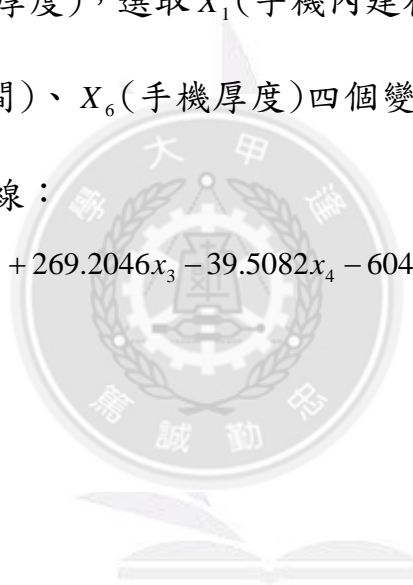
### 三、最終模式

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t-Value	Pr >  t	Variance Inflation
Intercept	1	2008.7715	6092.2328	0.33	0.7444	0
X <sub>1</sub>	1	12.8834	7.1483	1.80	0.0836	1.2612
X <sub>3</sub>	1	269.2046	33.8516	7.95	<.0001	1.4701
X <sub>4</sub>	1	-39.5082	15.1537	-2.61	0.0152	1.1835
X <sub>6</sub>	1	-604.9516	220.6051	-2.74	0.0111	1.1700

綜合上述五種選擇重要變數方法之交集，我們移除變數 X<sub>2</sub> (手機介面畫素)、X<sub>5</sub> (手機厚度)，選取 X<sub>1</sub> (手機內建相機畫素)、X<sub>3</sub> (手機重量)、X<sub>4</sub> (待機時間)、X<sub>6</sub> (手機厚度) 四個變數留在模式中。

最終模式的配適迴歸線：

$$\hat{y}_i = 2008.7715 + 12.8834x_1 + 269.2046x_3 - 39.5082x_4 - 604.9516x_6$$



#### 四、檢測VIF(多重共線性問題)

多重共線性(Multicollinearity)：在迴歸模式中，某些自變數或所有自變數之間具有高度線性相關的現象。

判斷準則：Variance Inflation >10

Parameter Estimates							
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Variance Inflation
Intercept	Intercept	1	2008.77152	6092.23281	0.33	0.7444	0
x1	camera pixels	1	12.88339	7.14834	1.80	0.0836	1.26120
x3	weight	1	269.20460	33.85158	7.95	<.0001	1.47013
x4	await time	1	-39.50821	15.15368	-2.61	0.0152	1.18348
x6	thickness	1	-604.95162	220.60514	-2.74	0.0111	1.17001

各變數的VIF值皆小於10故無多重共線性發生。

## 第二章 殘差分析

### 第一節 檢查殘差平均是否為0

虛無假設 $H_0: \mu = 0$

對立假設 $H_1: \mu \neq 0$

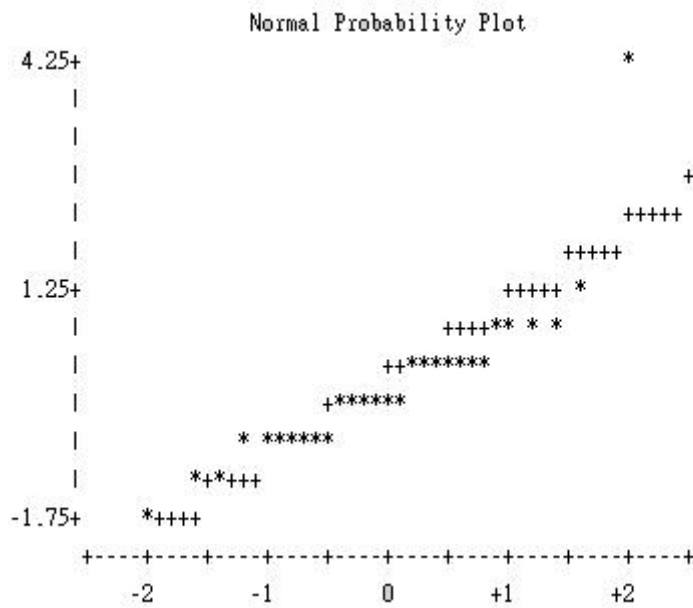
Test	Statistic	p Value
Student's t	t 0.0372	Pr >  t  0.9706
Sign	M - 2	Pr >=  M  0.5847
Signed Rank	S -37.5	Pr >=  S  0.4499

由此表可看出p-value的值皆明顯 $>0.05$ ，所以接受 $H_0$ ，即殘差平均數為0，符合基本假設 $E(\epsilon_i) = 0$ 。

### 第二節 檢查殘差是否來自常態

Stem Leaf	#	Boxplot
4 1	1	*
3		
3		
2		
2		
1		
1 2	1	
0 55669	5	
0 001444	6	+---+---+
-0 4321111	7	*-----*
-0 8766655	7	+-----+
-1 22	2	
-1 5	1	
-----+-----+-----+		

影響手機價格高低不同的因素

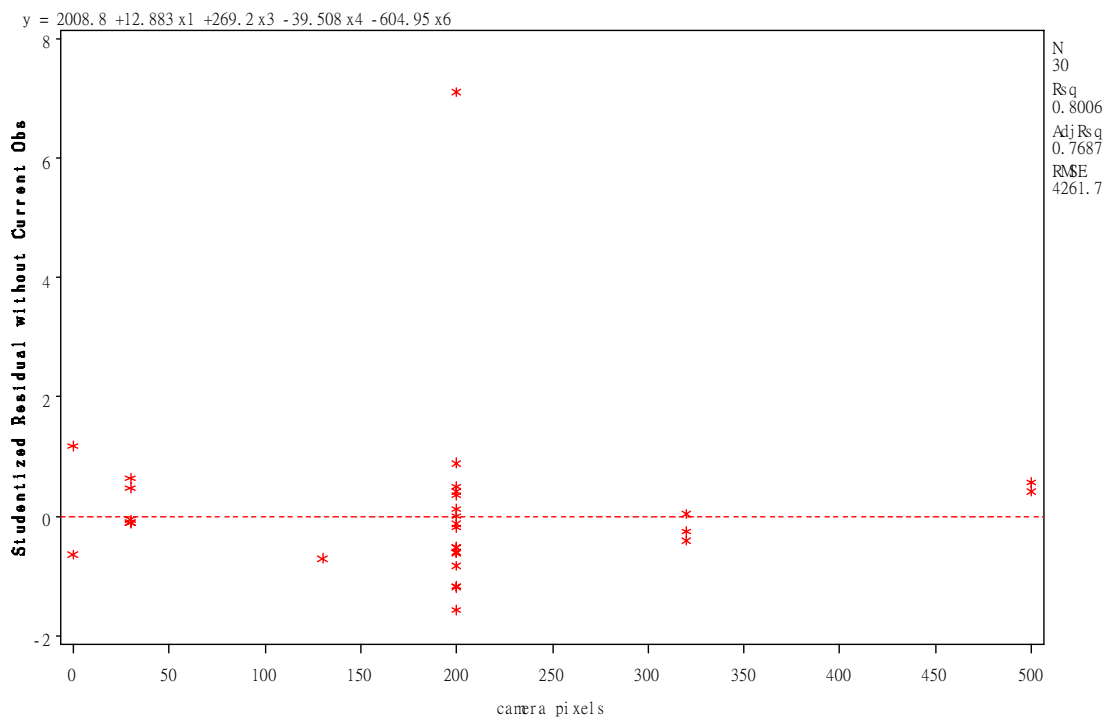


由以上兩個圖可看出，殘差來自常態。

### 第三節 R-student殘差圖

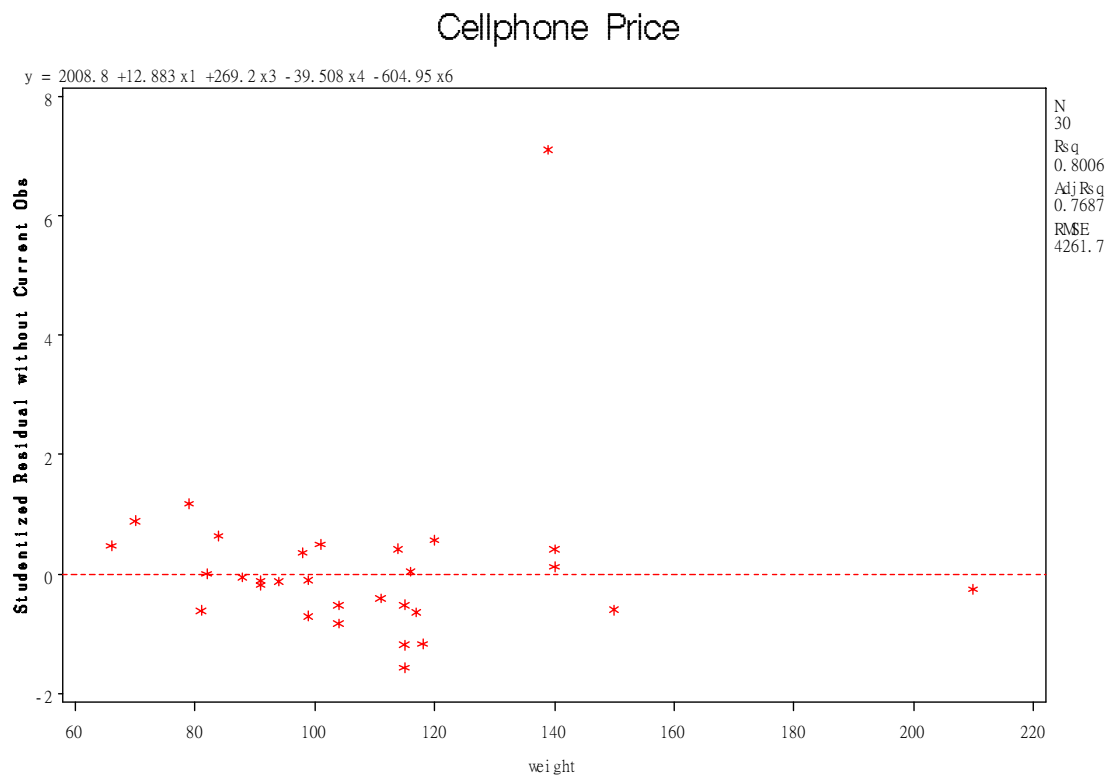
1. Y(手機價格)和 $X_1$ 手機內建相機畫素

Cellphone Price

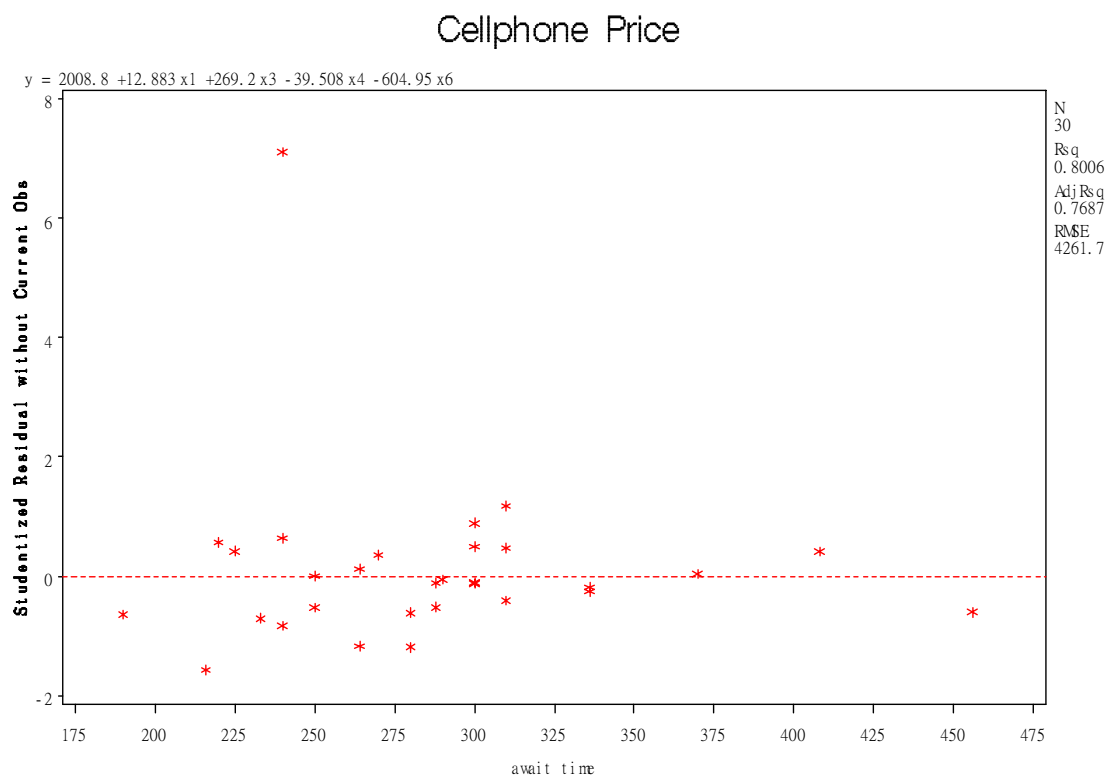


影響手機價格高低不同的因素

## 2. Y手機價格和X<sub>3</sub>手機重量



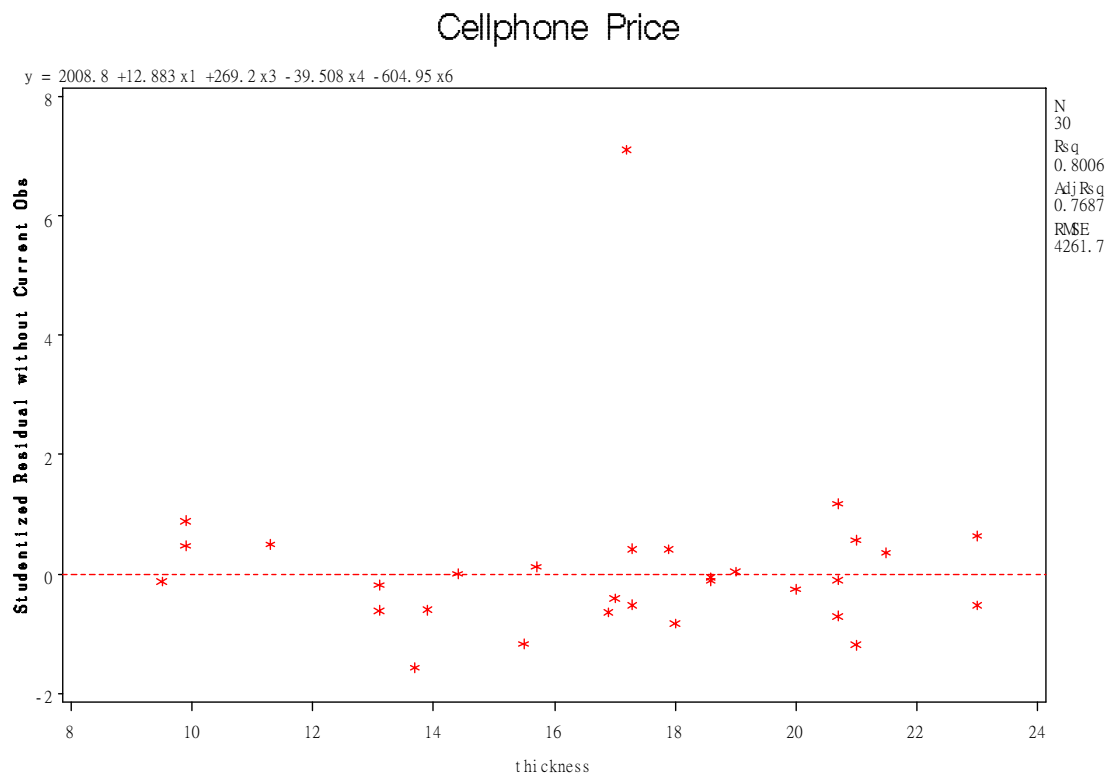
## 3. Y手機價格和X<sub>4</sub>待機時間



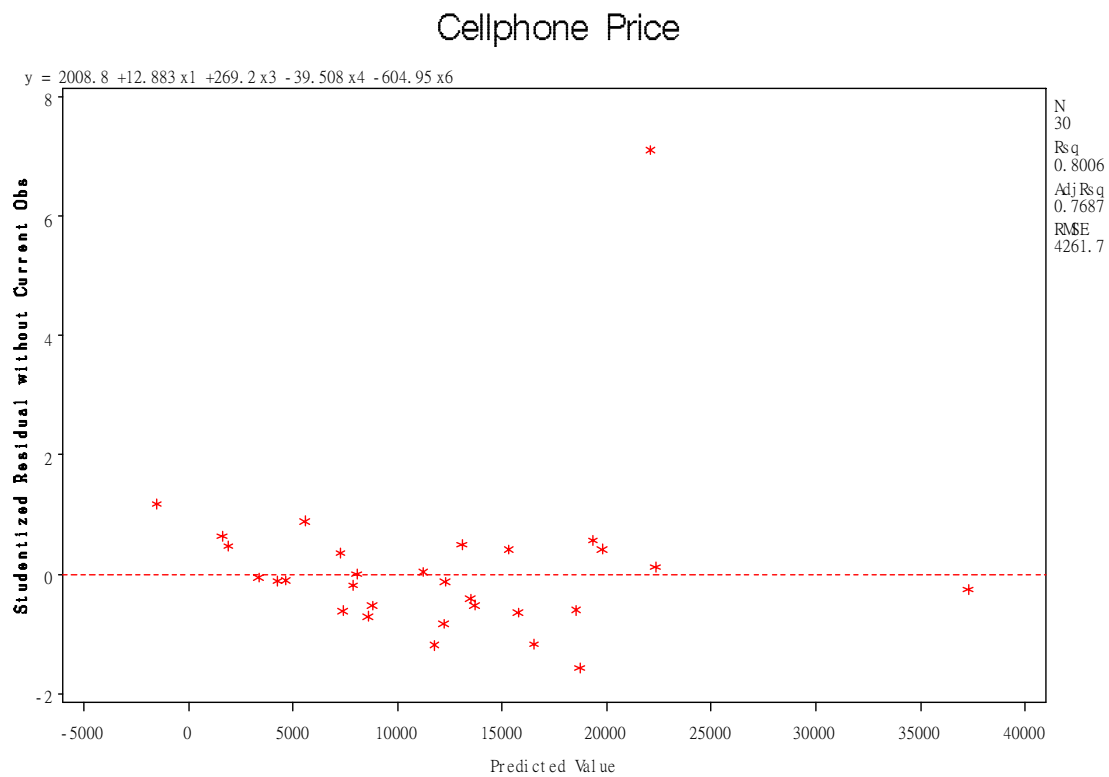


影響手機價格高低不同的因素

#### 4. Y手機價格和X<sub>6</sub>手機厚度



#### 5. 殘差預測圖



## 第四節 檢查影響點與異常點

### 一. 影響點偵測

#### (一) DFFITS

判斷準則： $|DFFITS_i| > 1$

<u>Obs</u>	<u>DFFITS</u>
13	2.8447

觀察值13的 $|DFFITS_i|$ 大於1，故觀察值13為影響點。

#### (二) The hat matrix elements $h_{ii}$

判斷準則： $h_{ii} > 2 * p/n$

$$2 * p/n = 2 * 5/30 = 0.33$$

<u>Obs</u>	<u>Hat Diag H</u>
6	0.4853
15	0.3729
24	0.3375

觀察值6、15、24的 $h_{ii}$ 值大於0.33故為影響點。

#### (三) Cook's distance statistic $D_i$

判斷準則： $D_i > 1$

無觀察值大於1，故無影響點。

#### (四) DFBETAS

判斷準則： $|DFBETA_i| > 1$

<u>DFBETAS</u>					
<u>obs</u>	<u>Intercept</u>	<u>X<sub>1</sub></u>	<u>X<sub>3</sub></u>	<u>X<sub>4</sub></u>	<u>X<sub>6</sub></u>
13	0.8848	-0.8800	2.1910	-1.8293	-0.8659

由此發現觀察值13為影響點。

### (五)COVRATIO

判斷準則： $1 - 3 * p/n < COVRATIO_i < 1 + 3 * p/n$

$$1 - 3 * p/n = 1 - 3 * 5/30 = 0.5$$

$$1 + 3 * p/n = 1 + 3 * 5/30 = 1.5$$

<u>Obs</u>	<u>CovRatio</u>
1	1.7002
3	1.5432
4	1.5106
6	2.3509
7	1.6671
11	1.5013
13	0.0049
15	1.8161
24	1.7295

觀察值1、3、4、6、7、11、15、24的COVRATIO<sub>i</sub>值>1.5與觀

察值13的COVRATIO<sub>i</sub>值<0.5，故此數點為影響點。

## 二. 異常點偵測

### (一)studentized Residuals

判斷準則： $|r_i| > 3$

<u>Obs</u>	<u>Student Residual</u>
13	4.117

由此發現該觀察值13為異常點。

### (二)Studentized Deleted Residuals

判斷準則： $|t_i| > 3$

<u>Obs</u>	<u>RStudent</u>
13	7.1110

由此發現該觀察值13為異常點。

綜合上述各檢測方法，得下表：

<u>Methods</u>	<u>obs</u>
DFFITS	13
The hat matrix elements $h_{ii}$	6、15、24
Cook's distance statistic $D_i$	-
DFBETAS	13
COVRATIO	1、3、4、6、7、11、 13、15、24
studentized Residuals	13
Studentized Deleted Residuals	13

得知觀察值13為影響點及異常點，我們回至原始資料檢查，非人為誤差，該觀察值13是一款外殼為18K鍍金的手機，其價格及重量會比其他手機高，故決定將該觀察值13移除再做一次檢測(得下表)。

<u>Methods</u>	<u>obs</u>
DFFITS	6
The hat matrix elements $h_{ii}$	6、7、14
Cook's distance statistic $D_i$	-
DFBETAS	6
COVRATIO	6、7、14
studentized Residuals	-
Studentized Deleted Residuals	-

由上表得知觀察值6為影響點，但是無觀察值為影響點，若再移除觀察值6其判定係數 $R^2$  及修正後判定係數 $R_a^2$ 反而降低，故不需再移除任何觀察值。經由移除觀察值後其判定係數 $R^2$  及修正後判定係數 $R_a^2$ 的變動如下表：

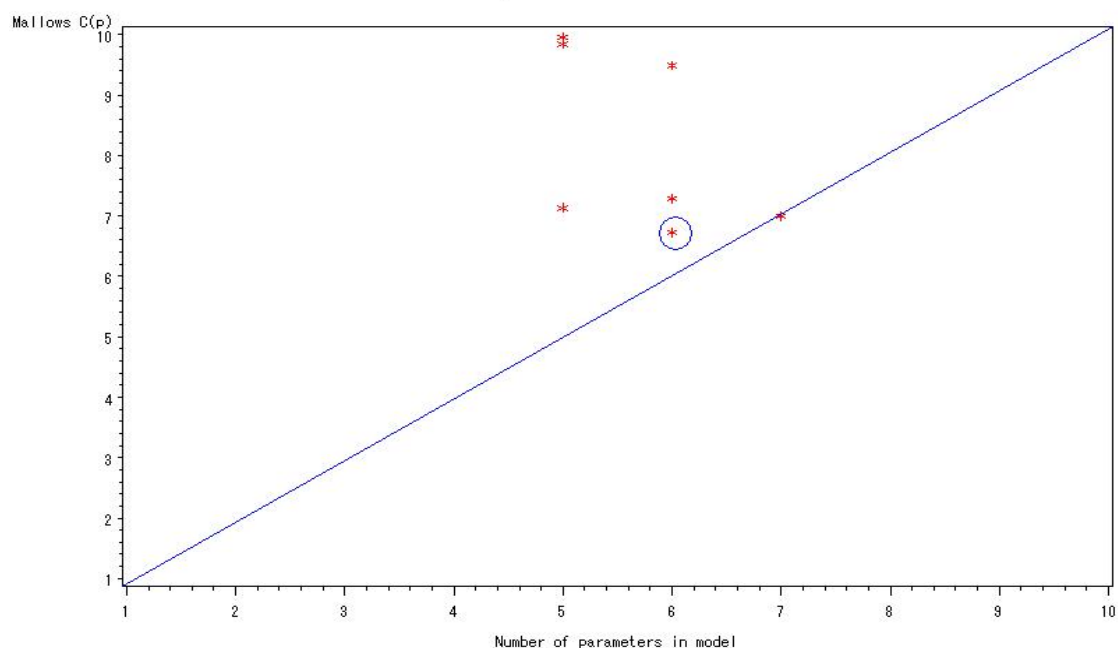
	未移除觀察值	移除觀察值13
$R^2$	0.8006	0.9062
$R_a^2$	0.7687	0.8906

### 第三章 刪除異常點後之迴歸分析

#### 一、選取重要變數

Methods	Variables
Forward selection	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$
Backward selection	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$
Stepwise selection	$X_1$ 、 $X_2$ 、 $X_3$
$R^2_{adj}$ selection	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_5$ 、 $X_6$
Cp selection	$X_1$ 、 $X_2$ 、 $X_3$ 、 $X_4$ 、 $X_6$

Cellphone Price



#### 二、最終模式

Parameter Estimates						
Variable	DF	Parameter Estimates	Standard Error	t-Value	Pr >  t	Variance Inflation
Intercept	1	-3434.1506	3778.6896	-0.91	0.3729	0
$X_1$	1	14.2755	4.3181	3.31	0.0031	1.4486
$X_2$	1	1.2354	0.8084	1.53	0.1401	2.1709
$X_3$	1	209.2073	22.8688	9.15	<.0001	2.0245
$X_4$	1	-20.2042	9.0713	-2.23	0.0360	1.3048
$X_6$	1	-351.7187	156.1882	-2.25	0.0342	1.8464

綜合上述五種選擇重要變數方法之交集，我們移除變數  $X_5$  (手機厚度)，選取  $X_1$  (手機內建相機畫素)、 $X_2$  (手機介面畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度) 五個變數留在模式中。

最終模式的配適迴歸線：

$$\hat{y}_i = -3434.1506 + 14.2755x_1 + 1.2354x_2 + 209.2073x_3 - 20.2042x_4 - 351.7187x_6$$

### 三、殘差分析

#### (一) 檢查殘差平均是否為0

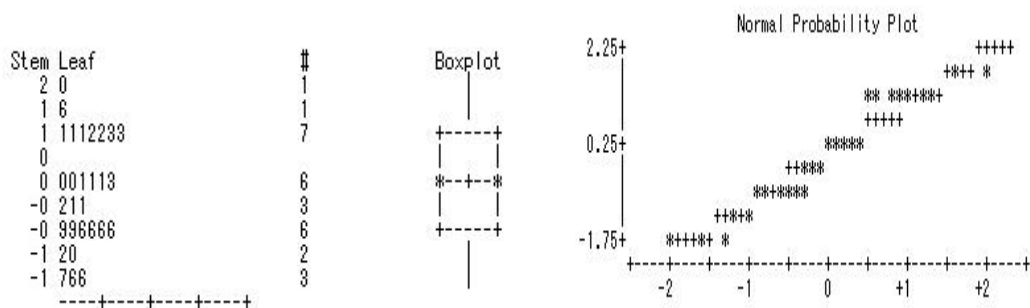
虛無假設  $H_0: \mu = 0$

對立假設  $H_1: \mu \neq 0$

Test	Statistic	p Value
Student's t	t 0.1797	Pr >  t  0.8587
Sign	M 0.5	Pr >=  M  1.0000
Signed Rank	S 9.5	Pr >=  S  0.8414

由此表可看出 p-value 的值皆明顯  $> 0.05$ ，所以接受  $H_0$ ，即殘差平均數為 0，符合基本假設  $E(\epsilon_i) = 0$ 。

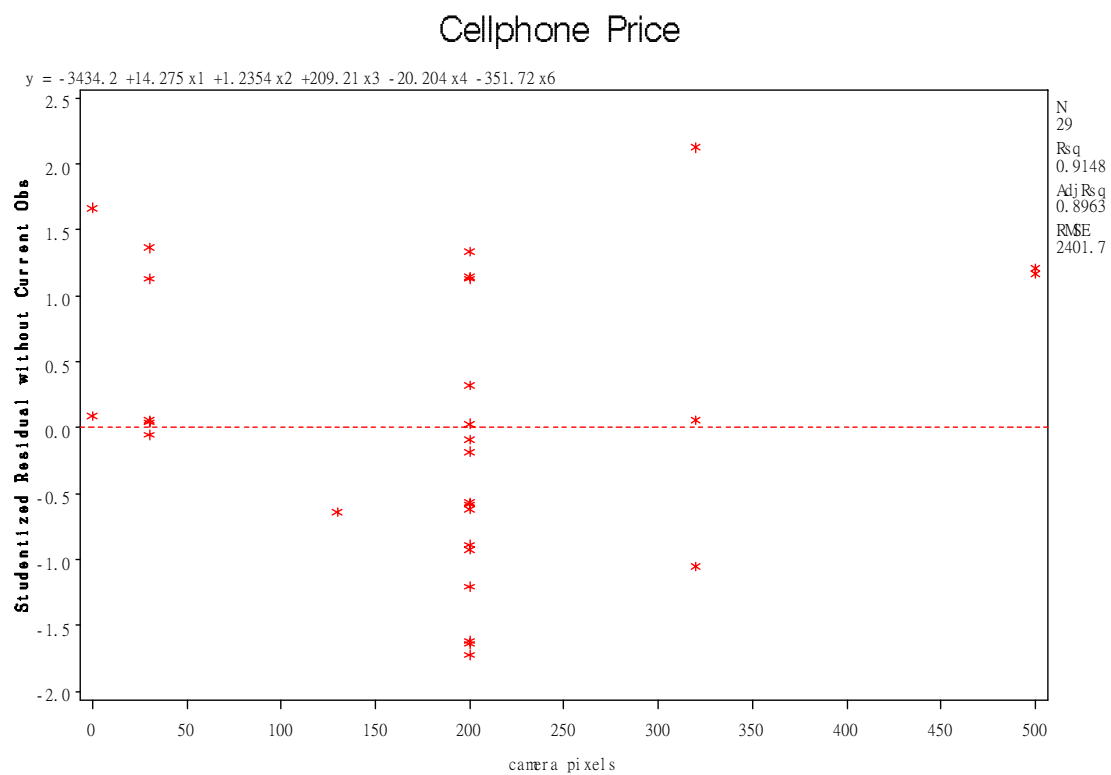
#### (二) 檢查殘差是否來自常態



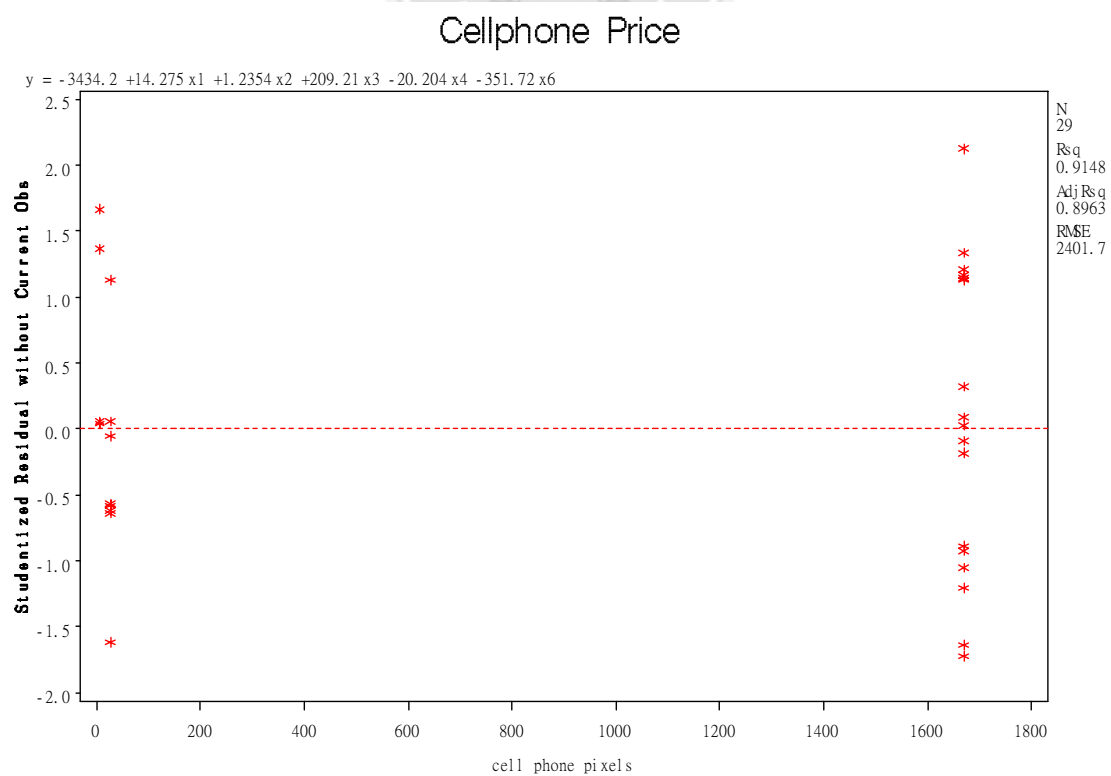
由以上兩個圖可看出，殘差來自常態。

#### 四、R-student 殘差圖

1. Y(手機價格)和X<sub>1</sub>手機內建相機畫素

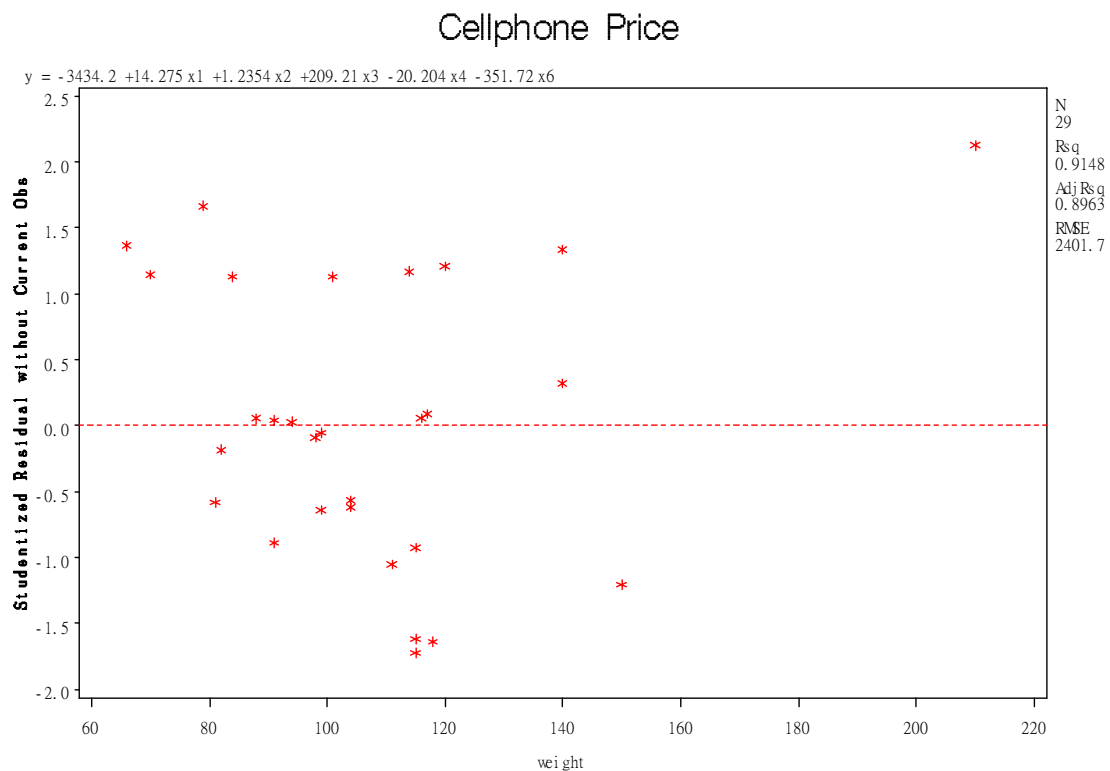


2. Y手機價格和X<sub>2</sub>手機介面畫素

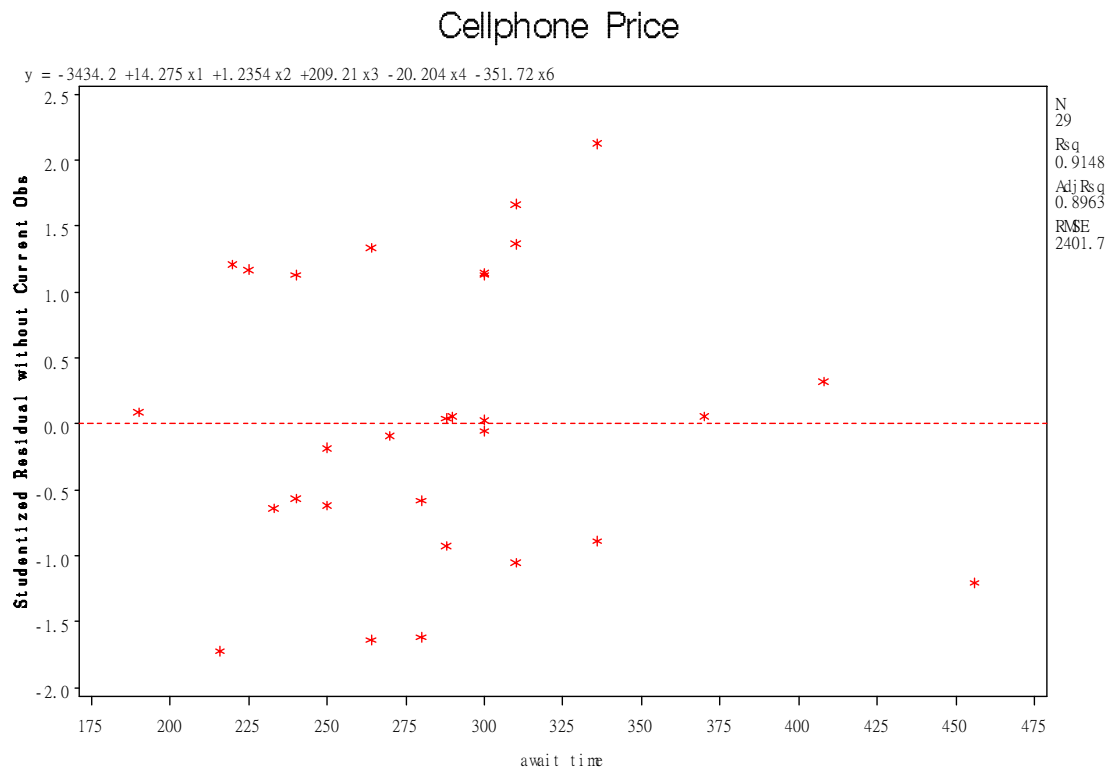


影響手機價格高低不同的因素

### 3. Y手機價格和X<sub>3</sub>手機重量



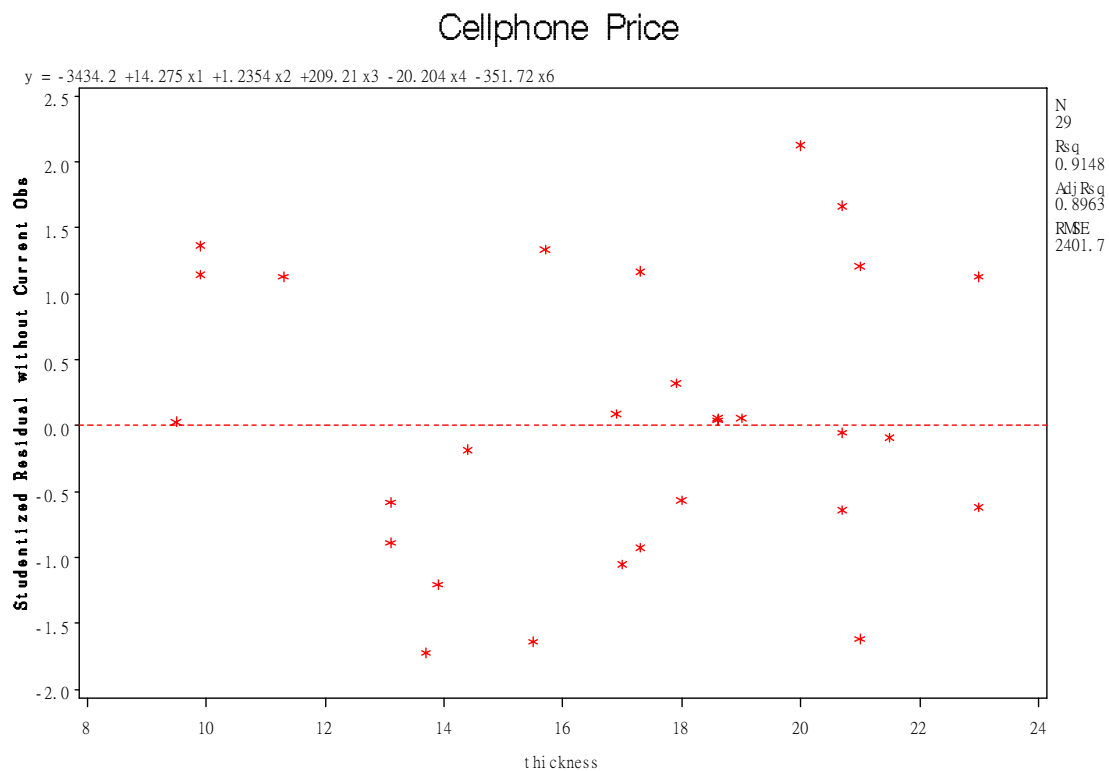
### 4. Y(手機價格)和X<sub>4</sub>待機時間



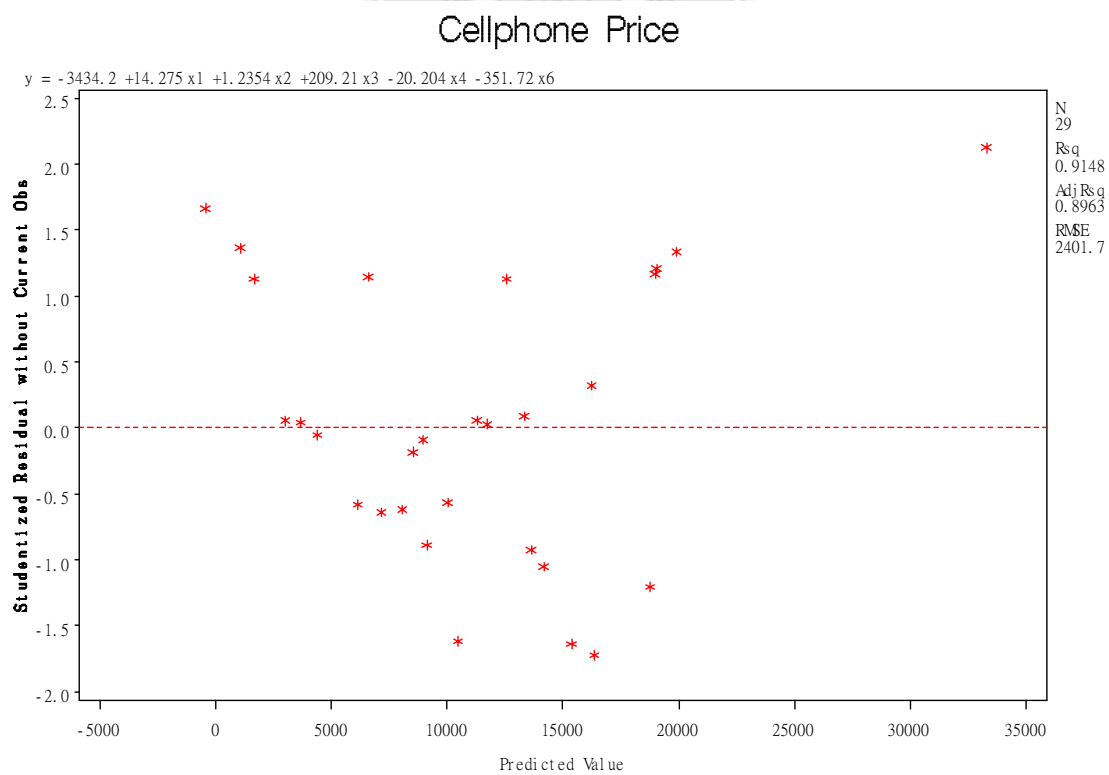


影響手機價格高低不同的因素

5. Y(手機價格)和X<sub>6</sub>手機厚度



6. 殘差預測圖



## 第四章 結論

在資料分析中，我們使用了五種方法來選取變數，分別是向後消去法、向前選取法、逐步選取法、 $R^2$ 、CP 選取法，經過這五種選取的方法我們選取了  $X_1$  (手機內建相機畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度) 為最佳模型的變數。因而我們可以大概推估此四項變數與手機價格有相關性，而手機價格與  $X_2$  (手機介面畫素)、 $X_5$  (手機通話時間) 較無相關性，其配適迴歸線為

$$\hat{y}_i = 2008.7715 + 12.8834x_1 + 269.2046x_3 - 39.5082x_4 - 604.9516x_6。$$

手機內建相機畫素每增加一萬畫素，手機價格則會增加 12.8834 元。

手機重量每增加 1 克，則手機價格則增加 269.2046 元，這是由於手機與其它電子產品結合，而增加了重量，例如，手機與筆電結合的 nokiaE60、手機與音樂播放器結合等。

待機時間每增加一小時，手機價格則會減少 39.5082 元，這是由於現在手機與其它電子產品結合，耗電量增加，使得待機時間較短，價格較高而較舊款的手機僅在乎通話功能，待機時間因而較長、價格較低，故手機價格與待機時間成負相關。

手機厚度每增加一公厘，手機價格則會減少 604.9516 元，現在人們選擇手機外型大多是偏向外觀薄型的手機，故手機越厚，較無人想購買，以較低價來銷售。

此外我們偵測異常點和影響點以免有異常點或影響點讓整個分析有誤，檢查的結果發現一個異常點需要被刪除，而刪除後，其調整後的判斷係數因而提高，也就是說各變數對價格的解釋能力提高，故將其異常點移除。

將異常點移除後，再作一次迴歸分析，選取的重要變數為  $X_1$  (手機內建相機畫素)、 $X_2$  (手機介面畫素)、 $X_3$  (手機重量)、 $X_4$  (待機時間)、 $X_6$  (手機厚度)，而手機價格與  $X_5$  (手機通話時間) 較無相關性，其配適迴歸線為

$\hat{y}_i = -3434.1506 + 14.2755x_1 + 1.2354x_2 + 209.2073x_3 - 20.2042x_4 - 351.7187x_6$ ，由此發現手機內建相機畫素、手機介面畫素、手機重量與手機價格呈現正相關，而待機時間與手機厚度與手機價格呈現負相關。

手機內建相機畫素每增加一萬畫素，手機價格則會增加 14.2755 元。

手機介面畫素每增加一萬畫素，手機價格則會增加 1.2354 元。

手機重量每增加 1 克，則手機價格則增加 209.2073 元，這是由於手機與其它電子產品結合，而增加了重量，例如，手機與筆電結合的 nokiaE60、手機與音樂播放器結合等。

待機時間每增加一小時，手機價格則會減少 20.2042 元，這是由於現在手機與其它電子產品結合，耗電量增加，使得待機時間較短，價格較高而較舊款的手機僅在乎通話功能，待機時間因而較長、價格較低，故手機價格與待機時間成負相關。

影響手機價格高低不同的因素

手機厚度每增加一公厘，手機價格則會減少 351.7187 元，現在人們選擇手機外型大多是偏向外觀薄型的手機，故手機越厚，較無人想購買，以較低價來銷售。

### 參考文獻

1. Kutner, MH, Nachtsheim, CJ, and Neter, J. ( 2004 4th ed.)  
“Applied Linear Regression Models”
2. 手機王 <http://www.sogi.com.tw/>
3. 中央大學-同步遠距教學網站  
<http://webclass.ncu.edu.tw/~tang0/>

