

# 國貿系商用統計學 A 組

期末考試題及作答-訂正版本

貿二 A 05154152 徐晰或

內容訂正的部分加上黃色標記

Q1

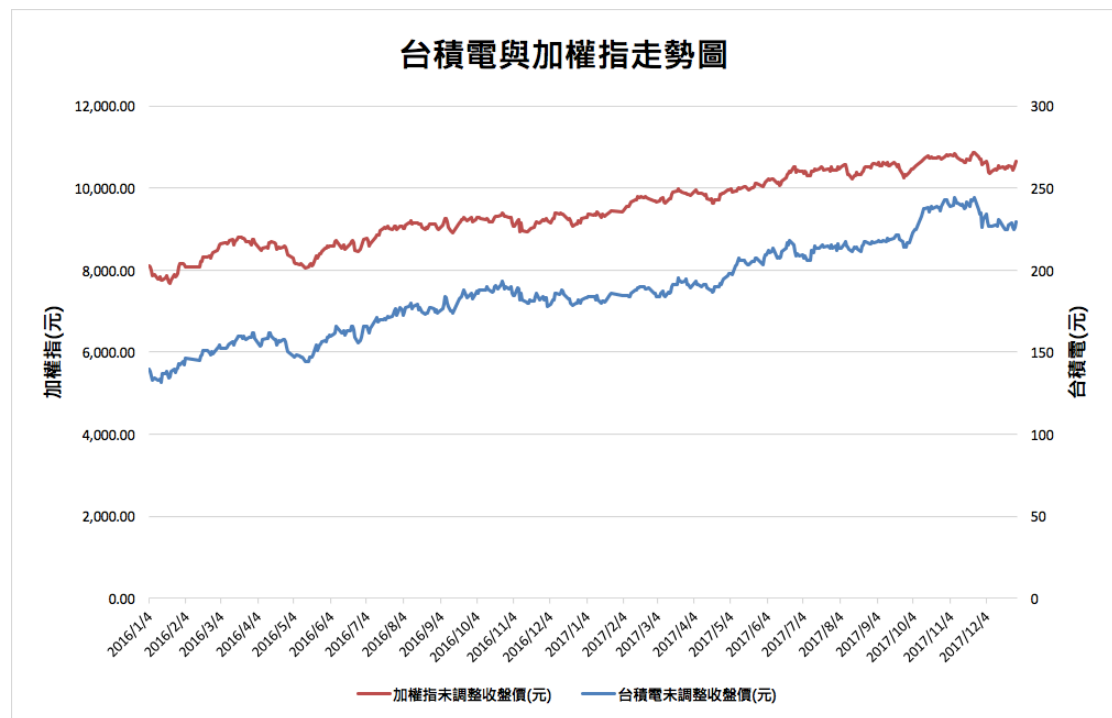
請至TEJ下在臺灣加權股價指數與台積電未調整的股價資料，時間是2016年1月1日至2017年12月31日。

- 1.繪製台灣加權股價指數與台積電的股價走勢圖
- 2.計算日報酬率，上網搜尋計算公式
- 3.繪製台灣加權股價指數與台積電的日報酬率資料散布圖，Y軸為台積電、X軸為加權股價指數
- 4.上網搜尋市場風險或市場特徵線，進行台積電對加權指數的迴歸分析。上述須截圖至Word
- 5.根據統計分析做出你的結論。

作答：

### 1.台積電與加權指股價走勢圖(雙軸)

此處將兩個圖訂正為合併成一個圖並且家軸上名稱加上



雙座標參考：<http://blog.xuite.net/wendyhg40/90833952/40513504>-如何利用 Excel 製作雙座標圖表

2.關於日報酬率的計算，老師上課時便提過使用 LN 計算，這裡補充在網路上找  
的資料，股票報酬率公式，參考內容如下。

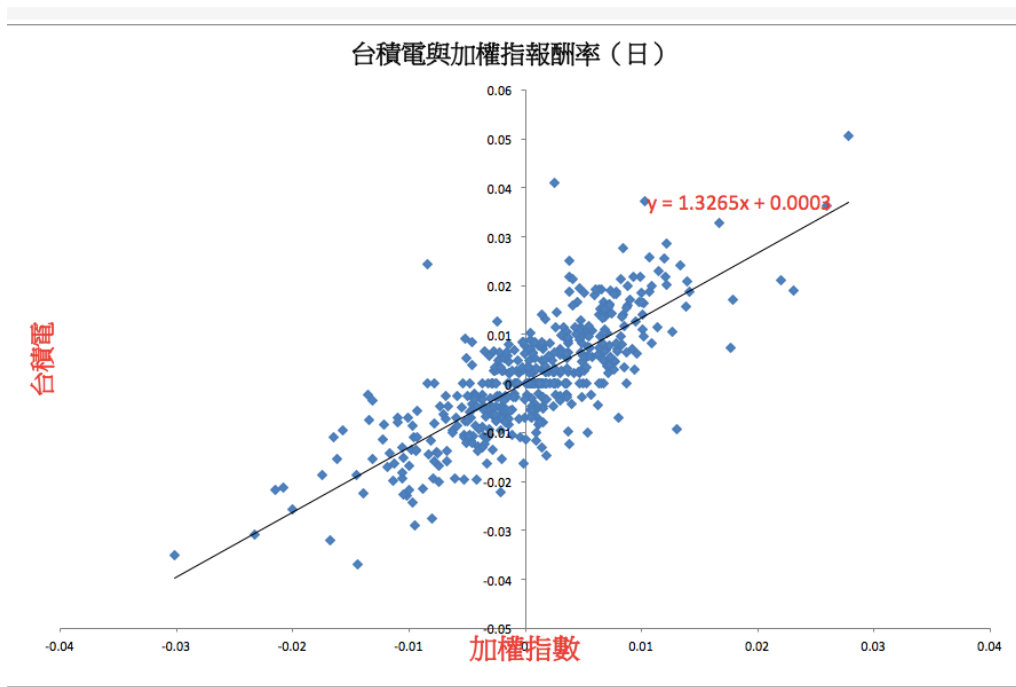
Unlike bonds, stocks don't pay owners a predetermined interest rate. However, the price of many stocks increases over time. To calculate log return, you must first find the initial value of the stock and the current value of the stock. In a spreadsheet, enter the formula "**=LN(current price/original price)**." For example, if you purchased a stock for \$25 a share that is currently \$50 a share, you would enter, "**=LN(50/25)**." The resulting figure is the continuously compounded rate of return for the stock for that time period.

擷取自：<https://www.sapling.com/6470290/calculate-log-return>

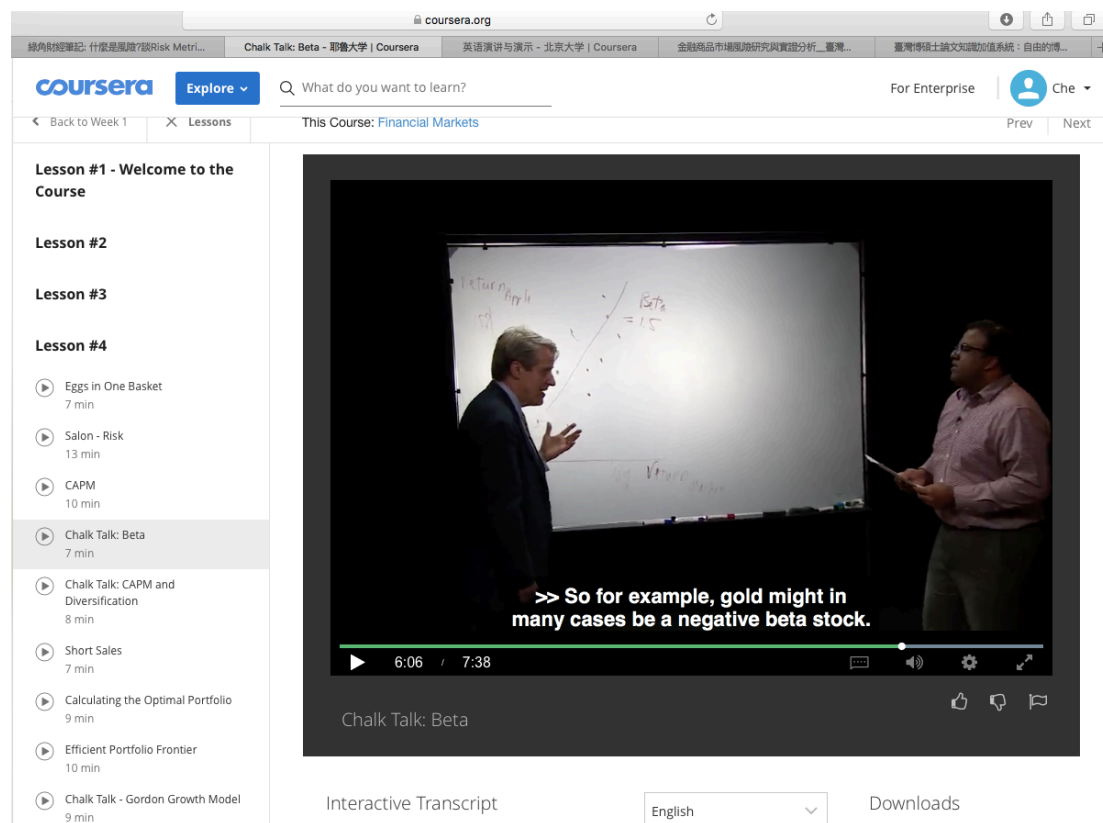
3.依照上述公式，報酬率計算為  $\ln(\text{今日股價}/\text{昨日股價})$ ，以下為 excel 操作截

fx =LN(M3/M4)							
	H	I	J	K	L	M	N
	2330 台積電				Y9999 加權指數		
	年月日	未調整收盤價(元)	報酬率		年月日	未調整收盤價(元)	報酬率
	2017/12/29	229.5	0.01537		2017/12/29	10,413.36	0.00709
	2017/12/28	226	0.00443		2017/12/28	10,567.64	0.00769
	2017/12/27	225	-0.0044		2017/12/27	10,486.67	0.00619
	2017/12/26	226	-0.011		2017/12/26	10,421.91	-0.0096
	2017/12/25	228.5	0.00439		2017/12/25	10,522.49	-0.0014
	2017/12/22	227.5	0.01105		2017/12/22	10,537.27	0.00459
	2017/12/21	225	0.00222		2017/12/21	10,488.97	-0.0015
	2017/12/20	224.5	-0.0022		2017/12/20	10,504.52	0.00355
	2017/12/19	225	-0.0088		2017/12/19	10,467.34	-0.0037
	2017/12/18	227	-0.0131		2017/12/18	10,506.52	0.00144
	2017/12/15	230	-0.0022		2017/12/15	10,491.44	-0.0044
	2017/12/14	230.5	0.0153		2017/12/14	10,538.01	0.00641
	2017/12/13	227	-0.0022		2017/12/13	10,470.70	0.00262
	2017/12/12	227.5	0		2017/12/12	10,443.28	-0.0029
	2017/12/11	227.5	0.0022		2017/12/11	10,473.09	0.00714
	2017/12/8	227	0.00221		2017/12/8	10,398.62	0.00413
	2017/12/7	226.5	-0.0022		2017/12/7	10,355.76	-0.0037
	2017/12/6	227	-0.011		2017/12/6	10,393.92	-0.0165
	2017/12/5	229.5	-0.0194		2017/12/5	10,566.85	-0.0079
	2017/12/4	234	0.0129		2017/12/4	10,651.11	0.00478
	2017/12/1	231	0.02188		2017/12/1	10,600.37	0.00377
	2017/11/30	226	-0.0369		2017/11/30	10,560.44	-0.0144
	2017/11/29	234.5	0.00213		2017/11/29	10,713.55	0.00061
	2017/11/28	234	-0.0127		2017/11/28	10,707.07	-0.0041
	2017/11/27	237	-0.0291		2017/11/27	10,750.93	-0.0095
	2017/11/24	244	0.00617		2017/11/24	10,854.09	-4E-05
	2017/11/23	242.5	0.00206		2017/11/23	10,854.57	0.00295

用日報酬率做的散佈圖如下截圖，Y 軸為台積電，X 軸為加權指數



關於市場風險以及市場特徵線，之前在做平台作業時上網學習過，參考 Robert Shiller 在 Coursera 上 Financial Market 課程中介紹。



The screenshot shows a Coursera video player for the course 'Financial Markets'. The video is titled 'Chalk Talk: Beta' and is 7:38 long. The video content shows two men standing in front of a whiteboard. The whiteboard has a scatter plot with a regression line and the text 'Beta = 1.5'. The video player interface includes a progress bar at 6:06 / 7:38, a subtitle 'Chalk Talk: Beta', and options for 'Interactive Transcript', 'English', and 'Downloads'. The video player also shows a list of lessons on the left side of the screen.

Beta 值可以用來評估一個投資組合和市場整體的相對起伏程度，舉例來說，一支基金的 Beta 值是 0.9，代表市場上升 1% 時，他會上升 0.9%，而市場下降 1% 時，他會下降 0.9%。Robert Shiller 在白板上畫出蘋果公司報酬以及市場報酬的散佈圖來做介紹，透過這個散佈圖，可以畫出一條回歸線（也就是市場特徵線），迴歸線的斜率就是 Beta，然而，散佈圖上的點，不是都在線上，而是跟現有距離，這是來自 idiosyncratic risk(非系統風險)又稱或可分散風險，是指只

對某個行業或個別公司的產生影響的風險，它是指由公司個別的因素引起，而只對個別或少數的收益產生影響。

參考：<http://si.secda.info/scu05154152/?p=1403>

5.對其做迴歸的結果，圖如上方散佈圖，摘要如下。

台積電對加權指數回歸								
Regression Statistics								
Multiple R		0.7861						
R Square		0.6179						
Adjusted R Square		0.6171						
Standard Error		0.0072						
Observations		489						
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	0.0406	0.0406	787.6080	0.0000			
Residual	487	0.0251	0.0001					
Total	488	0.0657						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	0.0003	0.0003	0.8664	0.3867	-0.0004	0.0009	-0.0004	0.00092
報酬率	1.3265	0.0473	28.0644	0.0000	1.2337	1.4194	1.2337	1.41941

5.解釋如下：

回歸方程式為  $\hat{y} = 1.3265x + 0.0003$

摘要的 F 值 看出其是否有解釋力，R 平方 則看其能夠解釋多少。

R 平方 = 0.6179 也就是說 y(台積電)的變異中有 61% 可以由 x(加權指)解釋。

此外，可以看到 F 檢定  $p < 0.05$ ，達顯著水準，此回歸有解釋力，Beta 值

1.3265， $p < 0.05$  也達顯著水準。

## Q2

1.根據說明，將原始資料簡化成五個測量變數: 樂趣、回應、互動、信任、忠誠度
2.簡化的方法是計算其平均數，例如樂趣是a1至a7，則計算這7題的平均數，命名為團購樂趣，注意有遺失值
3.進行上述5個變項的敘述統計量，並加以說明
4.進行上述5個變項的相關分析與散佈圖，並加以分析
5.研究者想知道，消費者的團購樂趣、主購者的回應、主購者的互動、對主購者的信任是否會影響到團購者的忠誠度，請進行適當的統計分析
6.檢視一下殘差分析
7.做出結論
8.上網找相關研究論文，截圖

作答：

此處對於置換的平均數有稍微訂正，上次直接去掉小數點了。

1.使用排序功能，找出所有遺漏值，並標記上顏色，以免忘記置換。

2個遺漏 平均為	1個遺漏 平均為	1個遺漏 平均為	1個遺漏 平均為	1個遺漏 平均為	2個遺漏 平均為	1個遺漏 平均為		1個遺漏 平均為	1個遺漏 平均為
3.549738	3.989556	3.206266	3.198433	4.240209	4.413613	4.331593		4.138381	4.093995

id	a1	a2	a3	a4	a5	a6	a7	b1	b2	b3	b4	b5	c1	c2	c3	c4	c5	c6	c7	e1	e2	e3	e4	
1	2	3	2	4	7	2	2	4	5	5	5	6	4	4	2	1	5	4	4	3	2	4	4	3
2	2	1	2	2	4	4	4	4	5	5	5	6	4	4	4	4	4	4	4	5	4	4	5	5
3	4	4	4	4	4	4	4	4	4	5	4	5	4	4	4	4	4	4	4	4	4	4	5	5
4	2	2	3	2	2	2	2	5	5	4	4	4	2	2	2	2	4	3	1	1	2	2	2	2
5	4	7	5	5	4	6	7	6	5	6	6	6	6	4	5	2	7	5	4	5	6	5	7	6
6	2	4	4	2	3	1	2	4	4	3	4	2	1	2	3	4	2	1	1	4	4	5	4	4
7	4	6	5	5	2	3	4	4	6	6	4	4	6	4	4	5	4	1	4	5	4	6	4	4
8	3	2	5	1	5	2	2	3	3	3	2	4	1	6	4	4	2	3	4	4	5	5	4	4
9	4	4	5	7	4	3.206266	4	2	3	3	5	4	3	4	3	4	4	3	1	4	4	4	5	4
10	4	4	3	3	4	2	5	4	4	5	4	4	5	4	5	3	3	5	5	4	2	4	3	3
11	3	3	4	4	4	2	3	5	5	3	4	3	2	2	5	5	4	4	5	4	4	6	2	2
12	3	7	4	2	4	5	3	4	4	5	3	2	2	3	2	6	5	5	4	4	2	2	1	1
13	2	5	5	2	4	4	2	4.240209	2	5	5	4	5	4	3	2	1	4	4	4	5	4	5	4
14	4	2	2	4	4	4	1	2	4	5	2	2	4	5	6	6	4	6	5	6	4	3	4	4
15	4	4	6	1	3.989556	3	6	4	4.413613	4	3	5	3	4	5	4	4	3	2	5	2	3	2	2
16	4	4	3	5	5	4	3	4	5	4	4	4.138381	4	4	4	5	5	5	4	6	6	6	5	5
17	4	4	4	3	3	3	3	4	5	4	5	4	3	3	4	4	4	5	5	4	4	5	5	5
18	5	4	4	5	2	4	2	6	6	6	6	6	6	5	4	2	4	5	6	2	3	4	4	4
19	4	5	4	3.549738	6	3	1	4	4	4	4	4	4	4	5	3	4	4	5	4	4	5	3	3
20	4	5	4	5	4	4	4	4	5	6	4	4	4	4.093995	5	5	6	4.183246	4	5	7	6	5	6
21	5	6	4	4	4	4	4	4	4	4	4	4	4	5	4	4	5	4	3	4	4	4	4	4
22	3	6	3	3	6	3	6	5	5	6	5	3	4	4	4	6	4	4	6	4	4	4	5	6
23	3	4	4	3.549738	4	3	5	5	5	5	6	4	2	4	4	3	5	4	5	3	4	5	4	4
24	2	4	2	6	4	2	3.198433	6	6	4	6	4	6	6	6	4	6	4.183246	6	4	4.18799	6	4.472585	6
25	2	3	3	2	4	3	2	5	5	4.331593	4	3	6	6	6	6	6	6	6	6	4	4	5	5
26	4	4	5	3	5	1	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	6	5	4
27	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	1	4	4	5	5	3	4	4	4
28	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	1	1
29	4	6	6	6	4	3	6	5	6	6	6	6	5	5	4	4	5	4	6	6	5	4	5	6
30	2	4	3	4	2	1	1	4	3	4	4	3	4	2	3	4	3	5	5	4	3	4	2	2
31	4	5	2	1	2	1	2	1	1	1	2	1	1	1	1	1	1	1	2	3	4	3	4	2
32	4	5	5	6	5	4	4	5	5	6	6	6	6	7	5	6	1	3	7	5	6	5	6	6

使用“平均數”來把遺漏值補上。

2.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	a1	a2	a3	a4	a5	a6	a7	團購樂趣	b1	b2	b3	b4	b5	回應	c1	c2	c3	c4	c5
2	2	3	2	4	7	2	2	3.14286	4	4	5	5	4	4.4	4	2	1	5	
3	2	1	2	2	4	4	4	2.71429	5	5	5	6	4	5	4	4	4	4	
4	4	4	4	4	4	4	4	4	4	5	4	5	4	4.4	4	4	4	4	
5	2	2	3	2	2	2	2	2.14286	5	5	4	4	4	4.4	2	2	2	4	
6	5	7	5	5	4	6	7	5.57143	6	5	6	6	6	5.8	4	5	2	7	
7	2	4	4	2	3	1	2	2.57143	4	4	3	4	2	3.4	1	2	3	4	
8	4	6	5	5	2	3	4	4.14286	4	6	6	4	4	4.8	6	4	4	5	
9	3	2	5	1	5	2	2	2.85714	3	3	3	2	4	3	1	6	4	4	
10	4	4	5	7	4	4	4	4.45804	2	3	3	5	4	3.4	3	4	3	4	
11	4	4	3	3	4	2	5	3.57143	4	4	5	4	4	4.2	5	4	5	3	
12	3	3	4	4	4	2	3	3.28571	5	5	3	4	3	4	2	2	5	5	
13	3	7	4	2	4	5	3	4	4	4	5	3	2	3.6	2	3	2	6	
14	2	5	5	2	4	4	2	3.42857	4.24021	2	5	5	4	4.04804	5	4	3	2	
15	4	2	2	4	4	4	1	3	2	4	5	2	2	3	4	5	6	6	
16	4	4	6	1	3.98956	3	6	3.99851	4	4.41361	4	3	5	4.08272	3	4	5	4	
17	4	4	3	5	5	4	3	4	4	5	4	4	4.13838	4.22768	4	4	4	5	
18	4	4	4	3	3	3	3	3.42857	4	5	4	5	4	4.4	3	3	4	4	
19	5	4	4	5	2	4	2	3.71429	6	6	6	6	6	6	5	4	2	4	
20	4	5	4	3.54974	6	3	1	3.79282	4	4	4	4	4	4	4	4	5	3	
21	4	5	4	5	4	4	4	4.28571	5	6	4	4	4	4.6	4.09399	5	5	6	4.18
22	5	6	4	4	4	4	4	4.42857	4	4	4	4	4	4	5	4	4	5	
23	3	6	3	3	6	3	6	4.28571	5	5	6	5	3	4.8	4	4	4	6	
24	3	4	4	3.54974	4	3	5	3.79282	5	5	5	6	4	5	2	4	4	3	
25	2	4	2	6	4	2	3.19843	3.31406	6	6	4	6	4	5.2	6	6	4	6	4.18
26	2	3	3	2	4	3	2	2.71429	5	5	4.33159	4	3	4.26632	6	6	6	6	
27	4	4	5	3	5	1	3	3.57143	4	4	4	4	4	4	4	4	4	4	
28	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	1	4	
29	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	
30	4	6	6	6	4	3	6	5	5	6	6	6	5	5.6	5	4	4	5	
31	2	4	3	4	2	1	1	2.42857	4	3	4	4	3	3.6	4	2	3	4	
32	4	5	2	1	2	1	2	2.42857	1	1	1	2	1	1.2	1	1	1	1	
33	4	5	5	6	5	4	4	4.71429	4	5	5	6	6	5.2	6	7	5	6	
34	3	4	3	3	4	3	4	3.42857	4	3	4	5	5	4.2	5	5	4	5	
35	4	7	7	6	6	6	6	6	6	6	6	6	6	6	7	6	6	7	

a1-a7 → 計算平均數，命名為新變數“團購樂趣”

b1-b5 → 計算平均數，命名為新變數“回應”

c1-c7 → 計算平均數，命名為新變數“互動”

e1-e6 → 計算平均數，命名為新變數“信任”

f1-f5 → 計算平均數，命名為新變數“忠誠度”



### 3. 跑出描述統計量

團購樂趣		回應		互動		信任		忠誠度	
平均數	3.630392	平均數	4.3044466	平均數	4.0388618	平均數	4.332752	平均數	4.4765625
標準誤	0.0477714	標準誤	0.0468767	標準誤	0.0455428	標準誤	0.0462178	標準誤	0.0531552
中間值	3.7142857	中間值	4.2	中間值	4	中間值	4.3333333	中間值	4.5
眾數	4	眾數	4	眾數	4	眾數	4	眾數	5
標準差	0.936125	標準差	0.9185921	標準差	0.892453	標準差	0.9056809	標準差	1.0416259
變異數	0.8763301	變異數	0.8438114	變異數	0.7964724	變異數	0.8202579	變異數	1.0849845
峰度	0.0793444	峰度	1.3046166	峰度	1.4601089	峰度	1.4046438	峰度	0.3327027
偏態	-0.261433	偏態	-0.053537	偏態	0.1186284	偏態	-0.048716	偏態	0.0524423
範圍	5.1428571	範圍	5.8	範圍	6	範圍	6	範圍	6
最小值	1	最小值	1.2	最小值	1	最小值	1	最小值	1
最大值	6.1428571	最大值	7	最大值	7	最大值	7	最大值	7
總和	1394.0705	總和	1652.9075	總和	1550.9229	總和	1663.7768	總和	1719
個數	384	個數	384	個數	384	個數	384	個數	384

修正過後重新跑一次敘述統計量，結果與上次差不多

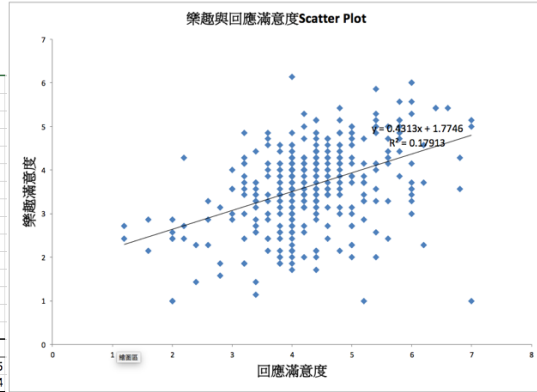
說明如下：

藍色部分為集中量數，粉紅色為變異量數。

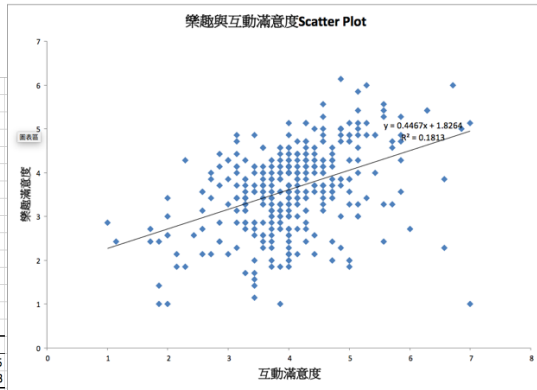
由集中量數來看，團購樂趣的滿意度相較其他變項來得低，而忠誠度的滿意度則相對較高(不論是在容易受極端值影響的平均數、出現最多次的眾數、或是位於中間的中位數都可以看得出來)。變異量數來看，忠誠度的變異數稍微較其他大一些。而從偏態係數 skewness 來看，團購樂趣、回應、信任滿意度屬於負偏(左偏)，互動與忠誠度的滿意度則為正偏(右偏)→絕對值不到 0.5 都是非常輕微的。從峰度係數 kurtosis 來看，都是屬於大於 0 的高峽峰，當中回應、互動、信任滿意度的峰度係數都大於 1→屬於較嚴重的峰度。

4.相關分析與散布圖及其分析如下：

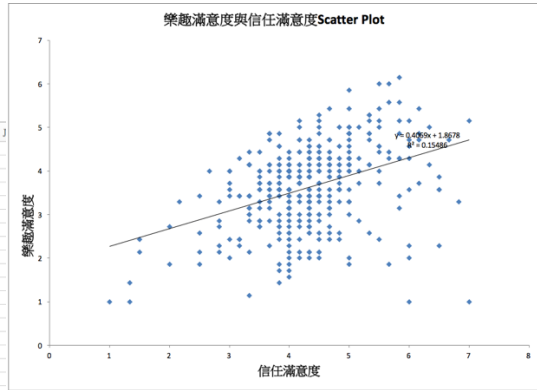
樂趣滿意度與回應滿意度								
<b>Regression Statistics</b>								
Multiple R	0.4232							
R Square	0.1791							
Adjusted R Square	0.1770							
Standard Error	0.8493							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	60.1258	60.1258	83.3593	0.0000			
Residual	382	275.5306	0.7213					
Total	383	335.6564						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.7746	0.2079	8.5376	0.0000	1.3659	2.1832	1.3659	2.18325
回應	0.4313	0.0472	9.1301	0.0000	0.3384	0.5241	0.3384	0.52414



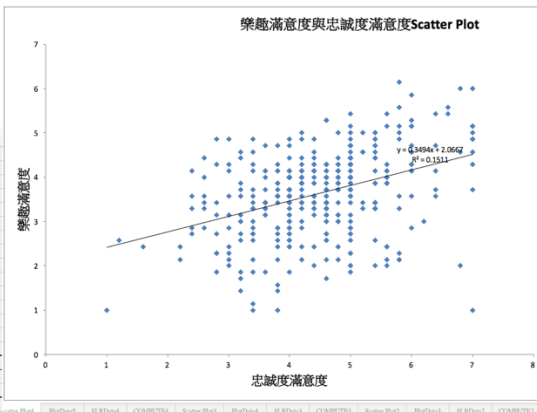
樂趣滿意度與互動滿意度								
<b>Regression Statistics</b>								
Multiple R	0.4258							
R Square	0.1813							
Adjusted R Square	0.1792							
Standard Error	0.8482							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	60.8547	60.8547	84.5938	0.0000			
Residual	382	274.8017	0.7194					
Total	383	335.6564						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.8264	0.2009	9.0919	0.0000	1.4314	2.2214	1.4314	2.22135
互動	0.4467	0.0486	9.1975	0.0000	0.3512	0.5422	0.3512	0.54223



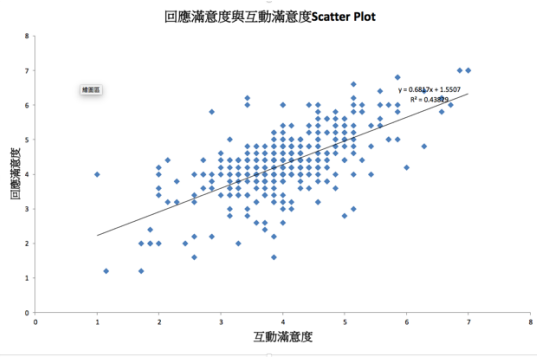
樂趣滿意度與信任滿意度								
<b>Regression Statistics</b>								
Multiple R	0.3935							
R Square	0.1549							
Adjusted R Square	0.1527							
Standard Error	0.8617							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	51.9811	51.9811	69.9983	0.0000			
Residual	382	283.6753	0.7426					
Total	383	335.6564						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.8678	0.2152	8.6781	0.0000	1.4446	2.2910	1.4446	2.29101
信任	0.4069	0.0486	8.3665	0.0000	0.3113	0.5025	0.3113	0.50249



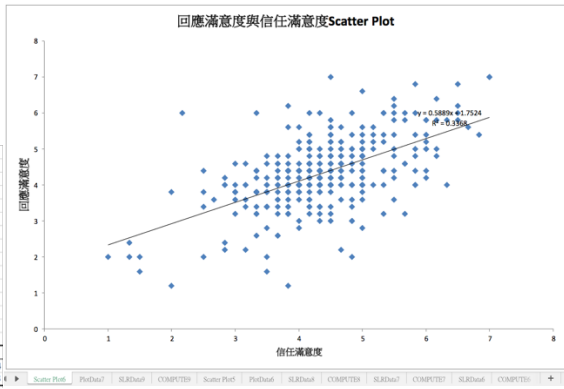
樂趣滿意度與忠誠度滿意度								
<b>Regression Statistics</b>								
Multiple R	0.3887							
R Square	0.1511							
Adjusted R Square	0.1489							
Standard Error	0.8637							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	50.7181	50.7181	67.9947	0.0000			
Residual	382	284.9383	0.7459					
Total	383	335.6564						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	2.0667	0.1947	10.6138	0.0000	1.6838	2.4495	1.6838	2.44951
忠誠度	0.3494	0.0424	8.2459	0.0000	0.2661	0.4327	0.2661	0.43266



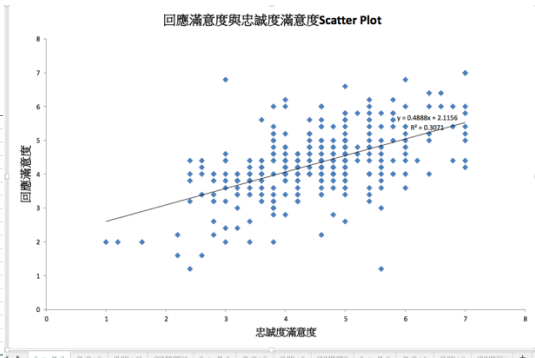
回應滿意度與互動滿意度								
<b>Regression Statistics</b>								
Multiple R	0.6620							
R Square	0.4383							
Adjusted R Square	0.4368							
Standard Error	0.6895							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	141.6885	141.6885	298.0674	0.0000			
Residual	382	181.5864	0.4754					
Total	383	323.2749						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.5507	0.1633	9.4961	0.0000	1.2296	1.8717	1.2296	1.8717
互動	0.6817	0.0395	17.2646	0.0000	0.6040	0.7593	0.6040	0.75928



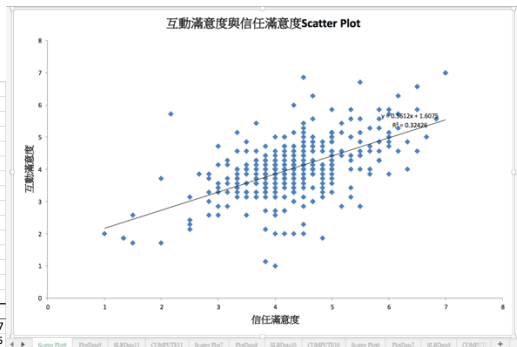
回應滿意度與信任滿意度								
<b>Regression Statistics</b>								
Multiple R	0.5803							
R Square	0.3368							
Adjusted R Square	0.3351							
Standard Error	0.7492							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	108.8805	108.8805	193.9994	0.0000			
Residual	382	214.3944	0.5612					
Total	383	323.2749						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.7524	0.1871	9.3657	0.0000	1.3845	2.1203	1.3845	2.12034
信任	0.5889	0.0423	13.9284	0.0000	0.5057	0.6720	0.5057	0.67198



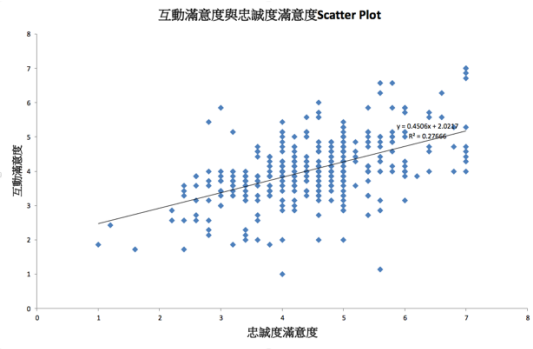
回應滿意度與忠誠度滿意度								
<b>Regression Statistics</b>								
Multiple R	0.5542							
R Square	0.3071							
Adjusted R Square	0.3053							
Standard Error	0.7658							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	99.2778	99.2778	169.3064	0.0000			
Residual	382	223.9971	0.5864					
Total	383	323.2749						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	2.1156	0.1726	12.2542	0.0000	1.7761	2.4550	1.7761	2.45503
忠誠度	0.4888	0.0376	13.0118	0.0000	0.4149	0.5626	0.4149	0.56264



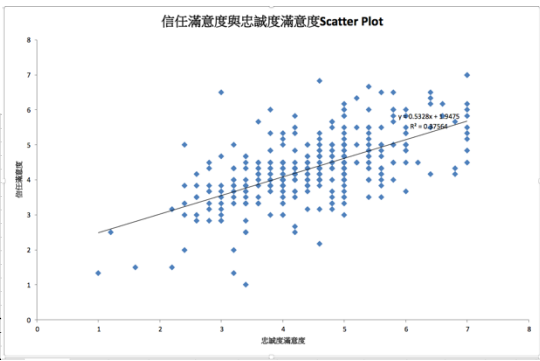
互動滿意度與信任滿意度								
<b>Regression Statistics</b>								
Multiple R	0.5694							
R Square	0.3243							
Adjusted R Square	0.3225							
Standard Error	0.7345							
Observations	384							
<b>ANOVA</b>								
	df	SS	MS	F	Significance F			
Regression	1	98.8779	98.8779	183.3049	0.0000			
Residual	382	206.0575	0.5394					
Total	383	304.9354						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.6075	0.1834	8.7631	0.0000	1.2468	1.9682	1.2468	1.96817
信任	0.5612	0.0414	13.5390	0.0000	0.4797	0.6427	0.4797	0.64265



互動滿意度與忠誠度滿意度								
Regression Statistics								
Multiple R	0.5260							
R Square	0.2767							
Adjusted R Square	0.2748							
Standard Error	0.7599							
Observations	384							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	84.3620	84.3620	146.1023	0.0000			
Residual	382	220.5734	0.5774					
Total	383	304.9354						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	2.0217	0.1713	11.8009	0.0000	1.6848	2.3585	1.6848	2.3585
忠誠度	0.4506	0.0373	12.0873	0.0000	0.3773	0.5239	0.3773	0.52386



信任滿意度與忠誠度滿意度								
Regression Statistics								
Multiple R	0.6129							
R Square	0.3756							
Adjusted R Square	0.3740							
Standard Error	0.7164							
Observations	384							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	1	117.9512	117.9512	229.8275	0.0000			
Residual	382	196.0485	0.5132					
Total	383	313.9997						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95%	Upper 95%
Intercept	1.9475	0.1615	12.0579	0.0000	1.5299	2.2650	1.5299	2.26505
忠誠度	0.5328	0.0351	15.1601	0.0000	0.4637	0.6019	0.4637	0.60187



因為圖片較多，依照上面結果以矩陣整理各個變相之間的相關

以下為 R 平方的值，回歸 F 檢定 p 都小於 0.05，達顯著。

	樂趣	回應	互動	信任	忠誠
樂趣		0.1791	0.1813	0.1549	0.1511
回應	0.1791		0.4383	0.3368	0.3071
互動	0.1813	0.4383		0.3243	0.2767
信任	0.1549	0.3368	0.3243		0.3756
忠誠	0.1511	0.3071	0.2767	0.3756	

互動與回應的相關相對其他的相關較高，樂趣與忠誠相對其他相關最低。

5.

其他變項對忠誠度-多元迴歸								
<b>Regression Statistics</b>								
Multiple R	0.6737							
R Square	0.4539							
Adjusted R Square	0.4481							
Standard Error	0.7738							
Observations	384							
<b>ANOVA</b>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	4	188.6188	47.1547	78.7538	0.0000			
Residual	379	226.9303	0.5988					
Total	383	415.5491						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.5670	0.2267	2.5013	0.0128	0.1213	1.0126	0.1213	1.0126
樂趣	0.1064	0.0482	2.2059	0.0280	0.0116	0.2013	0.0116	0.2013
回應	0.2321	0.0616	3.7707	0.0002	0.1111	0.3531	0.1111	0.3531
互動	0.1551	0.0629	2.4661	0.0141	0.0314	0.2788	0.0314	0.2788
信任	0.4380	0.0569	7.7009	0.0000	0.3262	0.5499	0.3262	0.5499

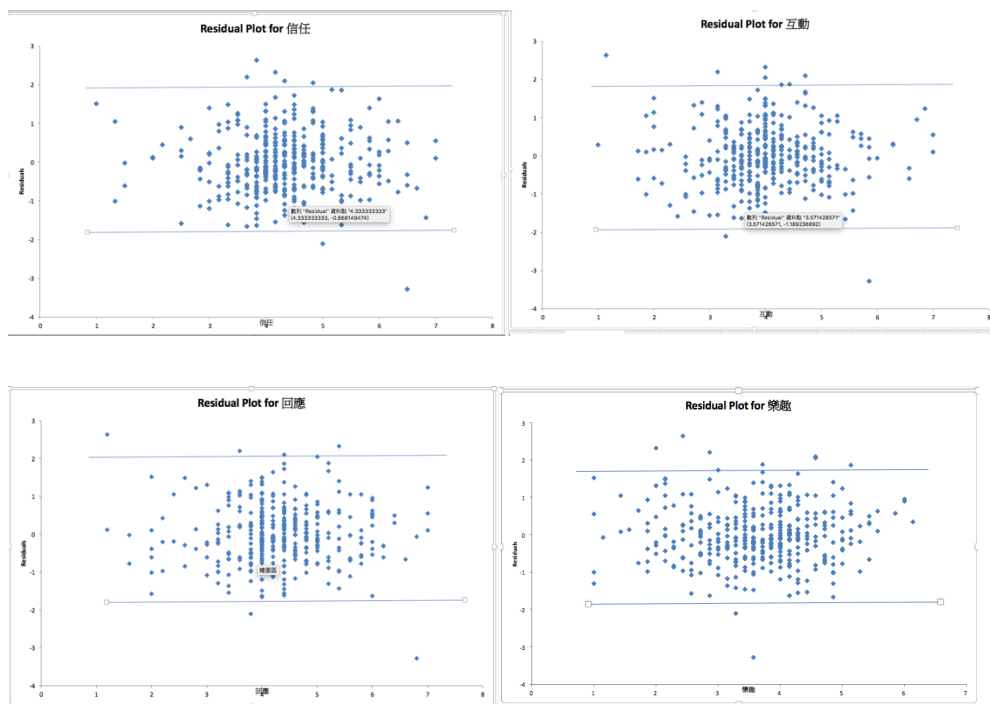
迴歸方程式:  $\hat{y} = 0.5670 + 0.1064X_{\text{樂趣}} + 0.2321X_{\text{回應}} + 0.1551X_{\text{互動}} + 0.4380X_{\text{信任}}$

此處重新做一次回歸，因為老師課堂指出有不顯著的結果，不過在跑了一次還

是沒發現不顯著的項目。

摘要輸出								
<b>迴歸統計</b>								
R 的倍數	0.67337322							
R 平方	0.4534315							
調整的 R 平	0.44766297							
標準誤	0.77412991							
觀察值個數	384							
<b>ANOVA</b>								
	<i>自由度</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>顯著值</i>			
迴歸	4	188.423034	47.1057586	78.60430022	1.6703E-48			
殘差	379	227.126028	0.59927712					
總和	383	415.549063						
	<i>係數</i>	<i>標準誤</i>	<i>t 統計</i>	<i>P-值</i>	<i>下限 95%</i>	<i>上限 95%</i>	<i>下限 95.0%</i>	<i>上限 95.0%</i>
截距	0.56947526	0.22674373	2.51153697	0.012435711	0.123642	1.01530853	0.123642	1.01530853
團購樂趣	0.1064174	0.04826339	2.20493039	0.028059018	0.01151986	0.20131495	0.01151986	0.20131495
回應	0.23124986	0.06161201	3.75332416	0.000201853	0.11010567	0.35239404	0.11010567	0.35239404
互動	0.15526662	0.06295806	2.46619113	0.014097695	0.03147577	0.27905747	0.03147577	0.27905747
信任	0.43811584	0.05689516	7.70040608	1.19E-13	0.32624613	0.54998555	0.32624613	0.54998555

## 6. 殘差分析



殘差是觀察值與預測值的差。用殘差來檢視回歸分析的假設。

### 變異數同質性

變異數同質性是在說，不同 X 給定的 Y，要有相同的變異數。

殘差圖落在以 0 為中心線的水平帶內，且正、負並沒有任何規則的傾向，則表示

變異數具有同質性。

### 獨立性

當誤差項無自我相關(獨立性)時，殘差圖將呈現無規則性。

圖形呈現水平的隨機分配，符合變異數同質性與獨立性

此外，因為殘差圖落在以 O 為中心線的水平帶內，且正、負都沒有任何規則性的傾向，則表示迴歸函數為線性模型。

7.

$\hat{y} = 0.5670 + 0.1064X_1 + 0.2321X_2 + 0.1551X_3 + 0.4380X_4$  此方程式有有解釋力，

並且可發現，在樂趣、回應、互動、信任會影響忠誠度，當中，信任的影響是最大的。

Q3

1. Cancer為1為得到肝癌0為正常
2. age為年齡
3. Coffee為一天喝幾杯黑咖啡
4. 執行logistic迴歸分析，解釋變項為年齡與咖啡
6. 標題設為: 咖啡與肝癌關係
7. 執行時會出現說coffee要不要.....，請按 否

問題一: 請問資料中得到肝癌比例是多少?喝咖啡的次數分配表為何?平均數?  
 問題二: 根據課本，請解釋報表中的咖啡迴歸係數或轉換後係數的意義?  
 問題三: 繪製Y為肝癌預測機率與X軸年齡的散布圖，這就是邏輯斯曲線。  
 問題四: 上網搜尋與本問題相關的新聞報導截圖。

### 咖啡與肝癌關係

Predictor	Coefficients	SE Coef	Z	p-Value
Intercept	-9.8741	1.5190	-6.5002	0.0000
年齡	0.1958362	0.03	6.531276027	0.0000
咖啡	-0.9484897	0.2342	-4.04997783	0.0001
Deviance	170.0232243			

1.

此樣本得到肝癌的比例是 0.56 (56%)，喝咖啡的次數分配如下

	A	B	C	D
1	Frequency Distribution for Coffee			
2	for Coffee			
3	咖啡日用量	Frequency	Percentage	Cumulative Pctage.
4		0	82	41.00%
5		1	66	33.00%
6		2	36	18.00%
7		3	10	5.00%
8		4	3	1.50%
9		5	3	1.50%

平均數為 0.975



## 2. 根據課本

$$\ln(\text{estimated odds ratio}) = b_0 + b_1X_1 + b_2X_2$$

$$\ln(\text{estimated odds ratio}) = -9.8741 + 0.19584X_1 - 0.948494X_2$$

也就是  $\ln(\text{estimated odds ratio}) = -9.8741 + 0.19584 \text{ 年齡} + -0.948494 \text{ 咖啡日用量}$

$y=0$  正常  $y=1$  得到  $X_1 \rightarrow$  年齡  $X_2 \rightarrow$  咖啡日用量

以下是對於數據的解釋訂正

Logistic 迴歸分析結果的解讀

logistic regression 結果的係數

說明：探討年齡與咖啡用量與癌症的關係，自變數分別是咖啡用量與年齡。果年齡與性別的係數分別是 0.195 與 -0.948，若直接從係數值來看，我們應該說癌症率與年齡呈正相關，年紀愈大，癌症的機率愈大；癌症機率與咖啡日用量呈負相關。

轉換成 odds ratio

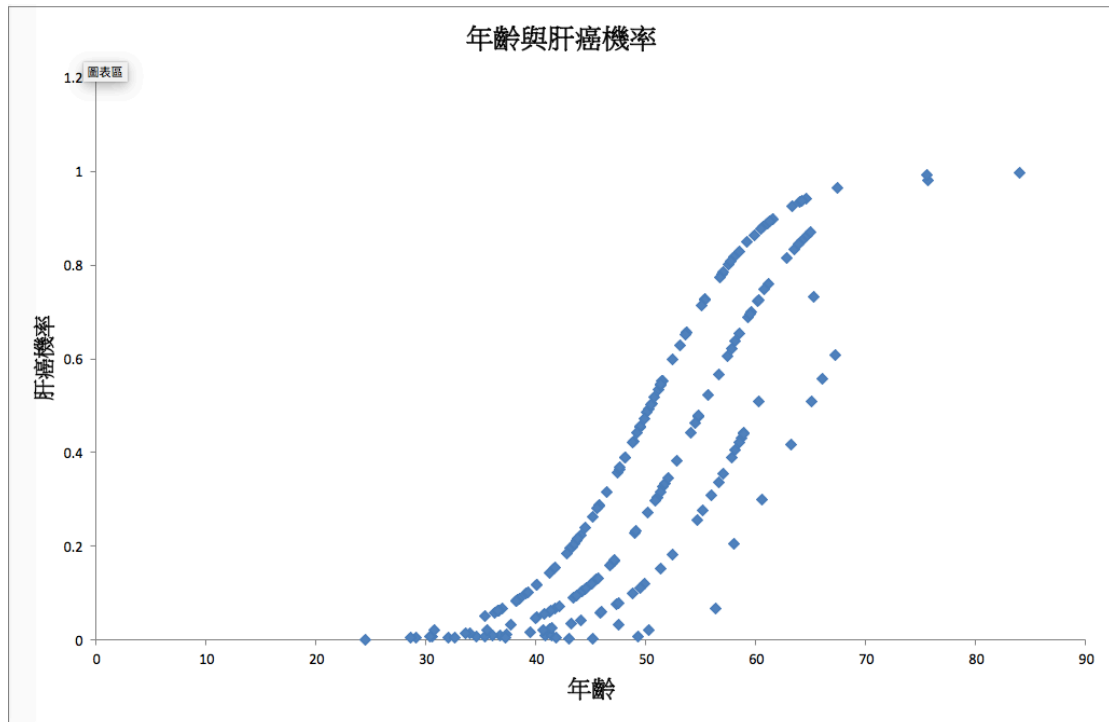
如果將係數轉換成 odds ratio，年齡與咖啡日用量的 odds ratio 分別 1.216320113 與 0.387360304 ( $\text{odds ratio} = \exp(\text{係數值})$ )。

解釋：

年齡每增加 1 歲，癌症風險為 1.21 倍

每日咖啡用量多一杯，癌症風險為 0.387 倍

### 3.散佈圖(羅吉斯曲線)



### 4.相關報導

#### 一天1杯黑咖啡，肝癌風險降42%！肝臟最喜歡的10種食物

非讀BOOK | 撰文者 | 克里斯汀·柯爾派翠克, 易普欣·漢諾納 | 2017-08-15 | 瀏覽數: 18684

+A -A 讚 0 分享 G+ 0回應



2005年，日本的研究人員在分析了兩場前瞻性世代研究（prospective cohort study）的成果後發現，咖啡的攝取量和肝癌的得病率呈反比：與不喝咖啡的人相比，每天喝一到兩杯咖啡者，其得到肝癌的風險降低了42%。

由於咖啡中含有成千上萬種物質，所以我們不可能直接點名是當中的哪些成分貢獻了這些護肝功效，因為很可能必須要有多種物質相輔相成（例如當中的強效抗氧化劑和抗癌物質相互搭配），方能產生這股守護肝臟的力量；不過最近有研究發現，或許是咖啡中的綠原酸成分所致。不論是飲用黑咖啡或是拿鐵都可以獲得這方面的好處，唯一需要注意的一點，就是沖泡咖啡時最好要有過濾這道手續，因為這個動作可以濾除咖啡中大部分的咖啡固醇（cafestol）和咖啡白脂（kahweol）；這是兩種會讓血中膽固醇含量上升的油性物質。

資料來源 <https://health.businessweekly.com.tw/AArticle.aspx?id=ARTL000095352>

## Q4

問題一: 請估計各教育程度使用Apply PAY系統意願的平均分數。  
 問題二: 分別以女性、國中為參考選項進行虛擬編碼，進行多元迴歸分析，以使用意願為依變項，性別與教育程度為解釋變項。  
 問題三: 男性是否使用意願高於女性? 研究所學歷使用意願是否高於國中?

先求算個教育程度的樣本平均

樣本平均	
國中	2.44928
高中職	3.74684
大專	4.02151
研究所	5.08475

各個教育程度的母體平均數估計(95 信心下)

將性別與教育程度以女性跟國中為參考點做 Dummy 如下

dummy2	D1	D2	D3	D4
國中2	0	0	0	0
女性1	0	0	0	0
男性2	1	0	0	0
大專4	0	1	0	0
研究所5	0	0	1	0
國小1	0	0	0	1

將當中各個教育程度依照 dummy 重新編碼

資料編碼列		P	Q	R	S
I	J	D1	D2	D3	D4
性別	1 女性	0	0	0	0
	2 男性	1	0	0	0
教育程度	1 國小	0	0	0	0
	2 國中	1	0	0	0
	3 高中職	0	0	0	0
	4 大專	0	0	1	0
	5 研究所	0	0	0	0
問題一: 請估計各教育程度使用Apply PAY系統意願的平均分數。		0	0	1	0
問題二: 分別以女性、國中為參考選項進行虛擬編碼，進行多元迴歸分析，以使用意願為依變項，性別與教育程度為解釋變項。		1	0	0	0
問題三: 男性是否使用意願高於女性? 研究所學歷使用意願是否高於國中?		0	0	0	0
		0	0	1	0
		1	0	0	0
		0	0	0	0
		1	0	0	0
樣本平均		1	0	0	0
國中	1.55072	1	0	0	0
高中職	1.63291	0	1	0	0
大專	1.56989	0	1	0	0
研究所	1.81356	0	1	0	0
		0	0	1	0

跑多元迴歸分析如下

A	B	C	D	E	F	G	H	I
<b>Regression Analysis</b>								
<b>Regression Statistics</b>								
Multiple R	0.1383							
R Square	0.0191							
Adjusted R Square	0.0125							
Standard Error	1.3137							
Observations	300							
<b>ANOVA</b>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	2	9.9959	4.9980	2.8958	0.0568			
Residual	297	512.6007	1.7259					
Total	299	522.5967						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	3.5586	0.1247	28.5380	0.0000	3.3132	3.8040	3.3132	3.8040
D1	0.3779	0.1571	2.4058	0.0167	0.0688	0.6871	0.0688	0.6871
D2	0.0000	0.0000	#DIV/0!	#DIV/0!	0.0000	0.0000	0.0000	0.0000

此回歸沒有通過 F 考核，依照性別做的迴歸不具解釋力。

<b>教育程度-使用意願多元迴歸</b>								
<b>Regression Statistics</b>								
Multiple R	0.6606							
R Square	0.4364							
Adjusted R Square	0.4287							
Standard Error	0.9992							
Observations	300							
<b>ANOVA</b>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	4	228.0549	57.0137	57.1024	0.0000			
Residual	295	294.5418	0.9984					
Total	299	522.5967						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	2.4493	0.1203	20.3610	0.0000	2.2125	2.6860	2.2125	2.6860
D1高中	1.2976	0.1646	7.8808	0.0000	0.9735	1.6216	0.9735	1.6216
D2大專	1.5722	0.1588	9.9029	0.0000	1.2598	1.8847	1.2598	1.8847
D3研究所	2.6355	0.1772	14.8745	0.0000	2.2868	2.9842	2.2868	2.9842
D4國小	0.0000	0.0000	#DIV/0!	#DIV/0!	0.0000	0.0000	0.0000	0.0000

依照教育程度跑的迴歸，當中的截距便是參考點國中的平均，而且從 D3 得，

研究所意與國中差距 2.6355，所以有高於國中，達顯著。

訂正在下一頁

訂正如下並且做性別與教育程度一起回歸

新的編碼，沒有國小學程度故刪除國小

id	D男	D高中	D大專	D研究	使用意願
1	1	0	1	0	2
2	0	0	0	1	2
4	1	0	1	0	3
6	0	1	0	0	4
8	0	0	1	0	5
11	1	0	1	0	4
14	1	1	0	0	2
24	0	0	0	0	2
25	0	0	1	0	3
28	1	0	0	1	1

跑迴歸結果如下

迴歸統計								
R 的倍數	0.27575832							
R 平方	0.07604265							
調整的 R 平方	0.06351442							
標準誤	1.27937621							
觀察值個數	300							
ANOVA								
	自由度	SS	MS	F	顯著值			
迴歸	4	39.7396369	9.93490922	6.06970188	0.00010486			
殘差	295	482.85703	1.63680349					
總和	299	522.596667						
	係數	標準誤	t 統計	P-值	下限 95%	上限 95%	下限 95.0%	上限 95.0%
截距	4.33588965	0.14103939	30.7424015	5.4869E-94	4.05831875	4.61346055	4.05831875	4.61346055
D男	-0.2228876	0.15813022	-1.4095196	0.15973491	-0.5340939	0.08831866	-0.5340939	0.08831866
D高中	-0.682053	0.21068719	-3.2372779	0.00134427	-1.0966934	-0.2674126	-1.0966934	-0.2674126
D大專	-0.4309309	0.19506609	-2.2091532	0.02793228	-0.8148284	-0.0470334	-0.8148284	-0.0470334
D研究	-0.8007711	0.23139968	-3.4605542	0.00061867	-1.2561745	-0.3453678	-1.2561745	-0.3453678

方程式是： $\hat{y} = 4.33 - 0.222D_{男} - 0.682D_{高} - 0.4309D_{大專} - 0.8007D_{研究所}$

男生是否顯著比女生高？答案為否，p 並沒有達到顯著！

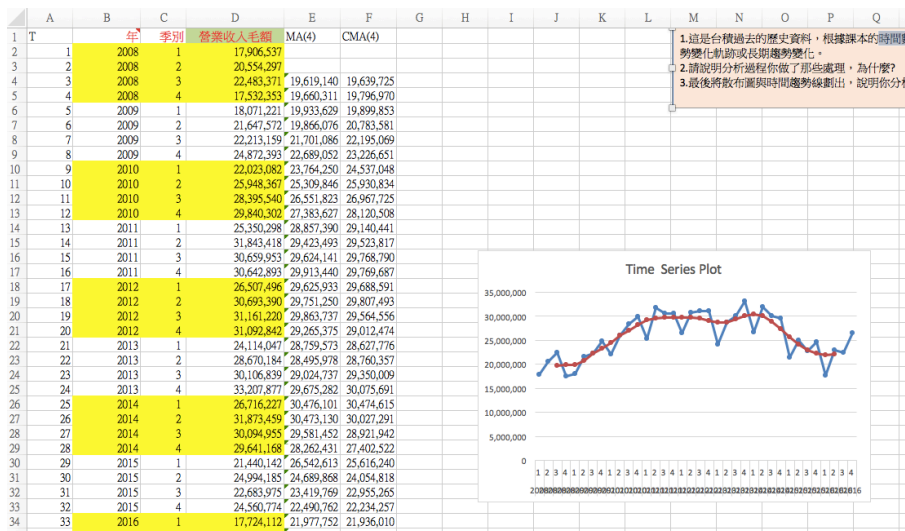
研究所是否高於國中？答案為是，p 達到顯著！

Q5

- 1.這是台積過去的歷史資料，根據課本的時間數列迴歸，請進行台積這三年營收毛額的時間趨勢變化軌跡或長期趨勢變化。
- 2.請說明分析過程你做了那些處理，為什麼?
- 3.最後將散布圖與時間趨勢線劃出，說明你分析的結論。

作答：

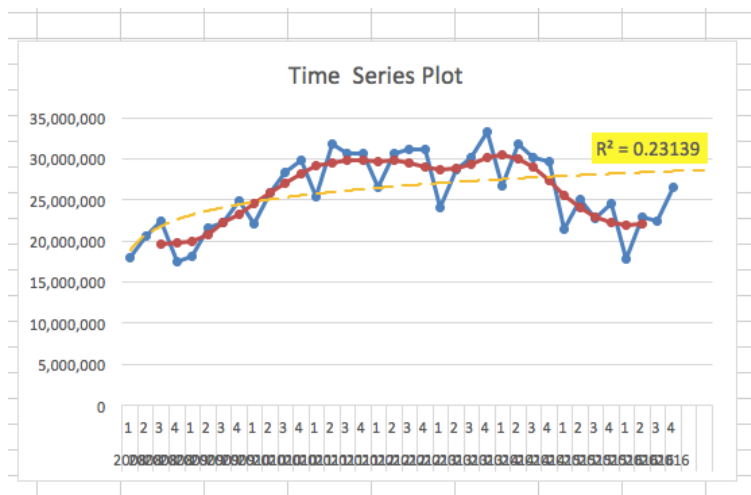
三年來營收毛額趨勢變化軌跡如下



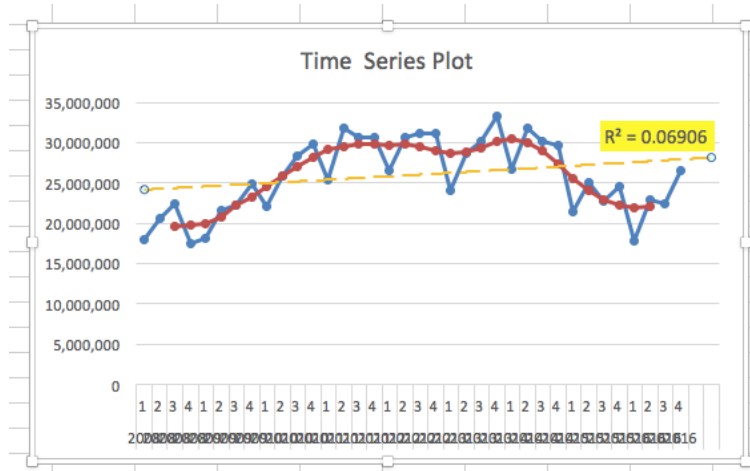
為了可以更看出隨著時間，銷售額變化的趨勢，這裡使用四個季度的移動平均

法將其平滑化，畫出散佈圖，看其趨勢。

如果要做回歸方程式，這裡嘗試使用對數以及線性，效果都不佳



使用對數的趨勢線，解釋力不佳。



使用線性的趨勢線，解釋力更差。

訂正如下，上次沒將其依照 0-35 排序，這裡更正，並做平均移動

