

題號	題目	答案項目	線上問卷(A)			email 問卷回覆(B)			總回覆(A+B)		
			次數	比率	人數	次數	比率	人數	次數	比率	人數
15	(15) Sample Power	1 有興趣	14	88%		5	83%		19	86%	
6	(6) 統計圖表的製作	1 有興趣	17	85%		4	67%		21	81%	
9	(9) 醫學統計中尋找切點的方法---ROC curve	1 有興趣	15	83%		4	67%		19	79%	
11	(11) Survival analysis I--Log rank test vs survival plot	1 有興趣	14	82%		4	67%		18	78%	
12	(12) Survival analysis II--Cox model	1 有興趣	14	82%		4	67%		18	78%	
5	(5) 連續樣本檢定及線性關係分析	1 有興趣	15	75%		5	83%		20	77%	
13	(13) GEE之原理及範例分析	1 有興趣	13	76%		4	67%		17	74%	



# 統計圖表的製作

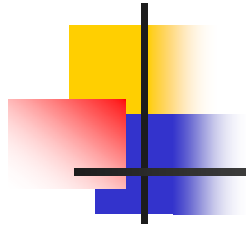
---



# 變項的類型

---

- 分類資料 (discrete variable)
  - 類別尺度 (nominal scale)
    - 例如：疾病、性別等
  - 序位尺度 (ordinal scale)
    - 除了分類以外，有大小程度順序之關係
    - 例如：職稱、學歷等
- 連續性資料 (continuous scale)
  - 等距尺度 (interval scale)
    - 除了分類、順序以外，代表相對距離，但沒有絕對的0
    - 例如：溫度等
  - 等比尺度 (ratio scale)
    - 使用標準化的單位，有絕對的0
    - 例如：身高、體重、年齡等



# 分類資料的描述

比例或百分比 (ratio or %)

# 分類資料的描述

新聞首頁 政治 財經 社會 地方 影

圖片集錦 熱門新聞 心情新聞 新聞總覽 新聞專

## 時事民調

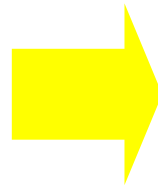
### 時事民調

中秋節連假四天，你想好要怎麼過了嗎？(可複選，至多選4項)

- 開烤肉趴
- 狂嗑月餅、文旦
- 返鄉陪家人
- 出遊
- 窩在家
- 仍要上班
- 其他
- 不知道

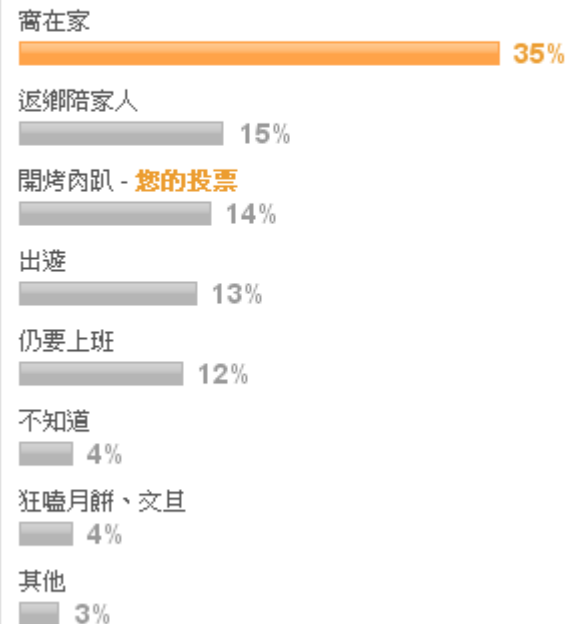
投票

投票結果



## 時事民調

中秋節連假四天，你想好要怎麼過了嗎？(可複選，至多選4項)

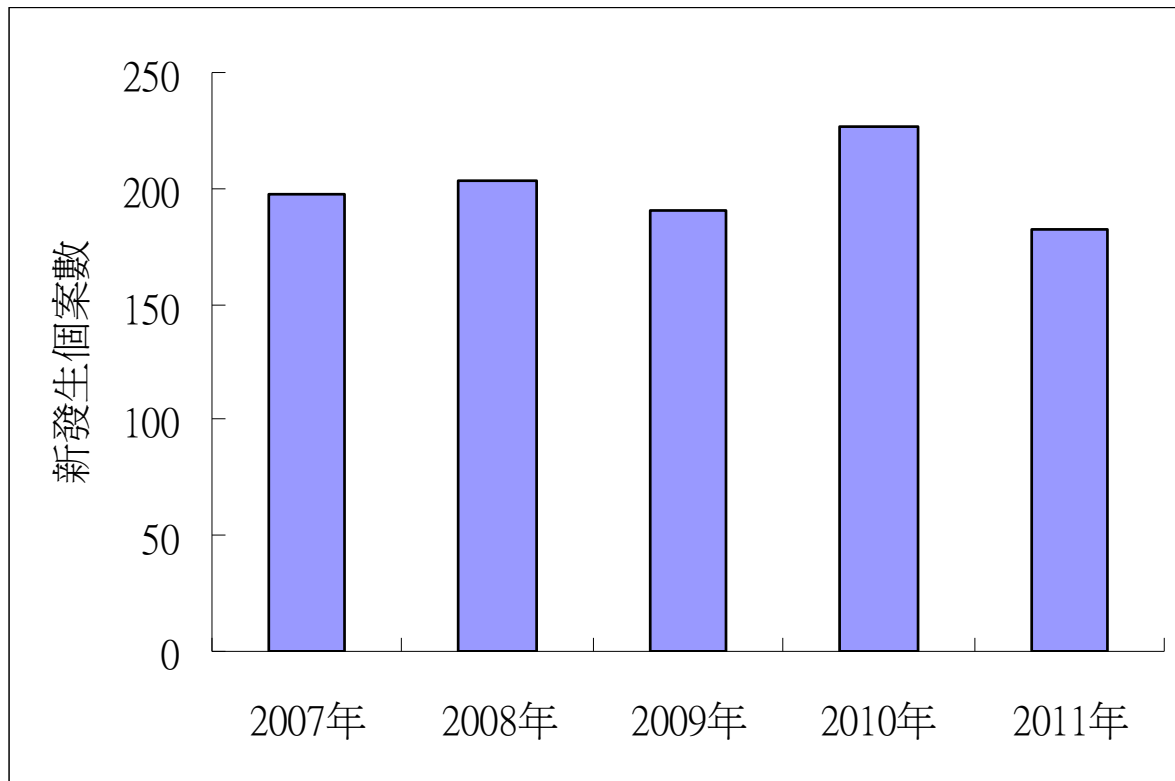


自2013/09/17起過去11小時總計2,638票

Yahoo新聞時事民調

# 分類資料的基本統計圖1

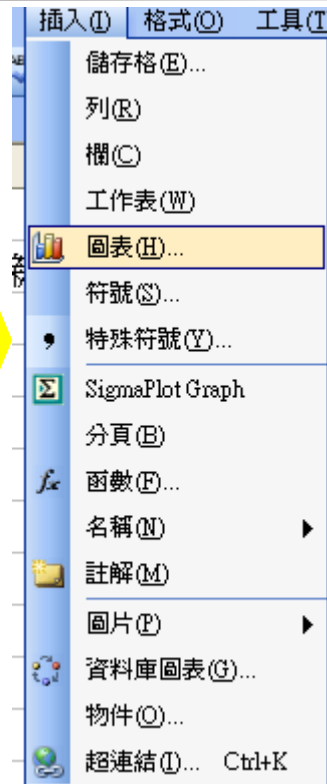
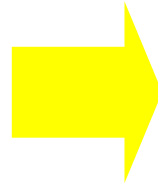
- 柱狀圖 (Bar charts)



# 柱狀圖繪製步驟1

## excel 2003 為例

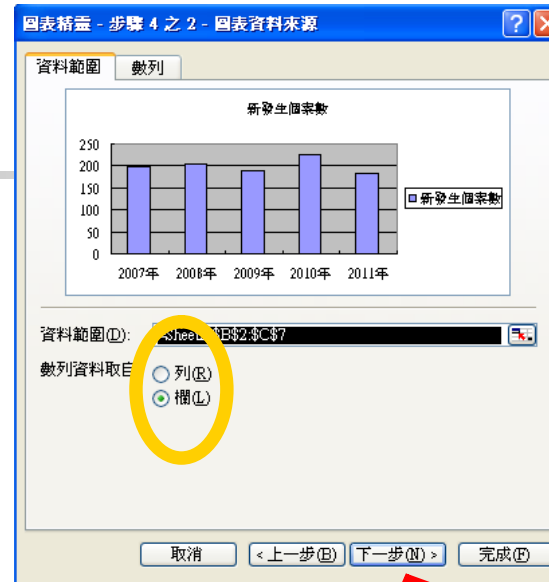
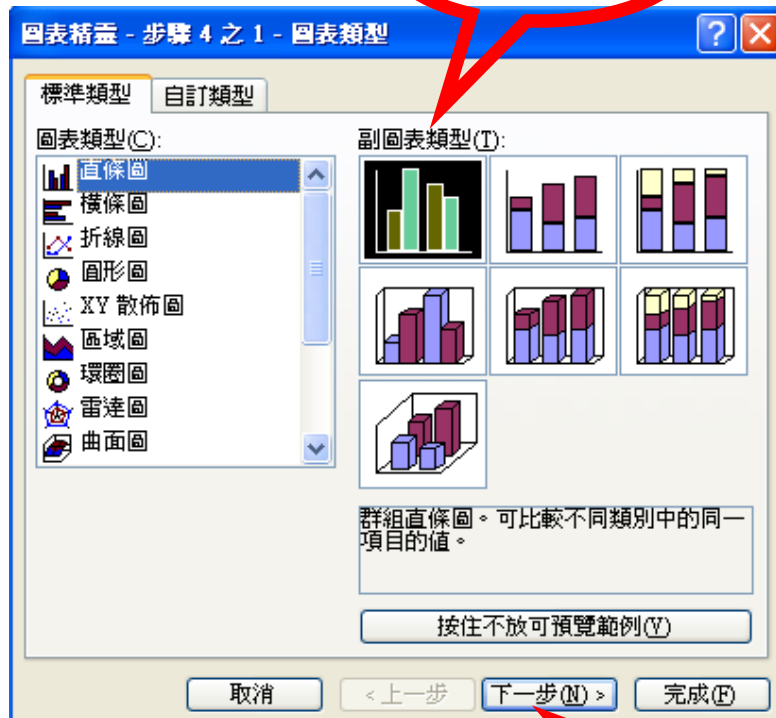
	A	B	C
1			
2			新發生個案數
3		2007年	197
4		2008年	203
5		2009年	190
6		20010年	227
7		2011年	182
8			



OR



# 柱狀圖繪製步驟2 excel 2003 為例



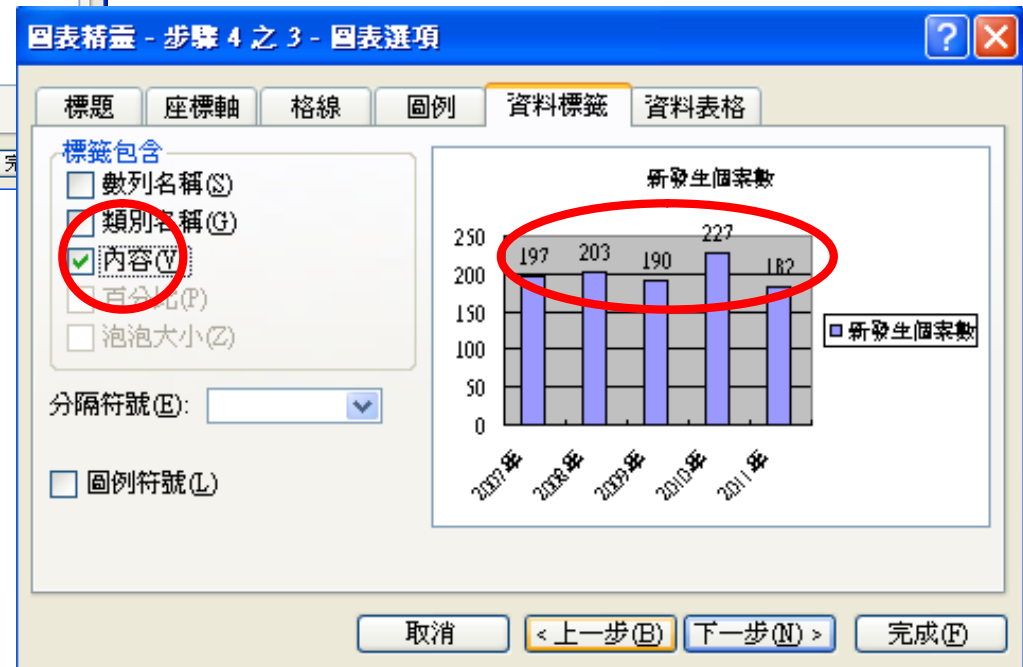
1

2

3



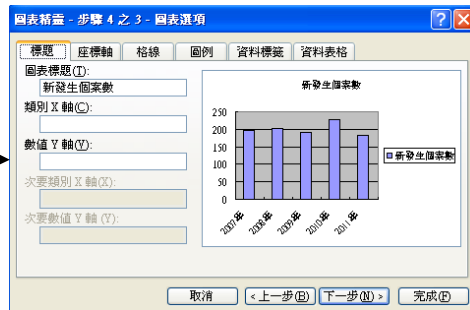
# 柱狀圖繪製步驟3 excel 2003 為例



# 柱狀圖繪製後，如何更改 excel 2003 為例

圖上按滑鼠右鍵

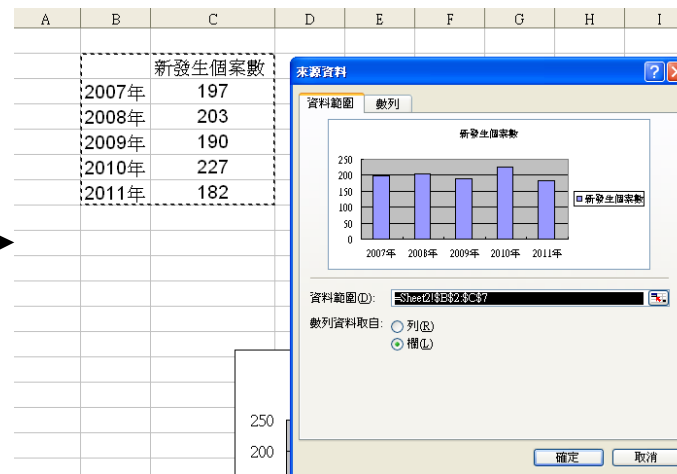
- 圖表區格式(O)...
- 圖表類型(Y)...
- 來源資料(S)...
- 圖表選項(I)...
- 圖表位置(L)...
- 檢視立體圖表(V)...
- 圖表視窗(W)
- 剪下(I)
- 複製(C)
- 貼上(P)
- 清除(A)
- 提到最上層(T)
- 移到最下層(K)
- 指定巨集(N)...



- 圖表區格式(O)...
- 圖表類型(Y)...
- 來源資料(S)...
- 圖表選項(I)...
- 圖表位置(L)...
- 檢視立體圖表(V)...

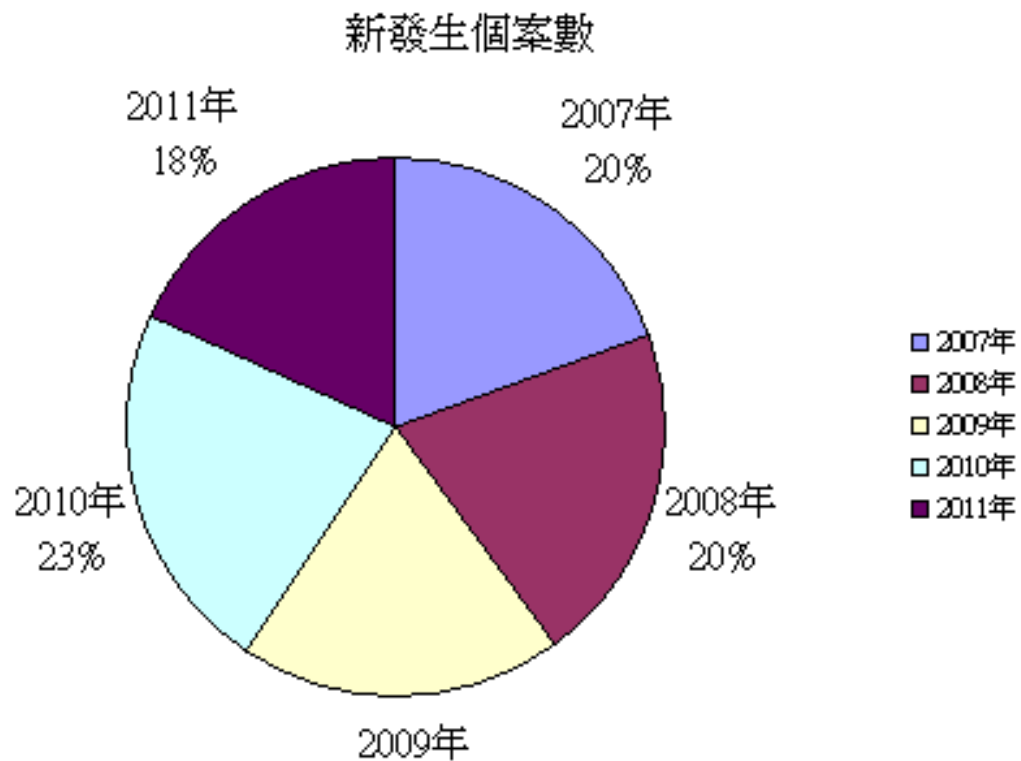


- 圖表區格式(O)...
- 圖表類型(Y)...
- 來源資料(S)...
- 圖表選項(I)...
- 圖表位置(L)...

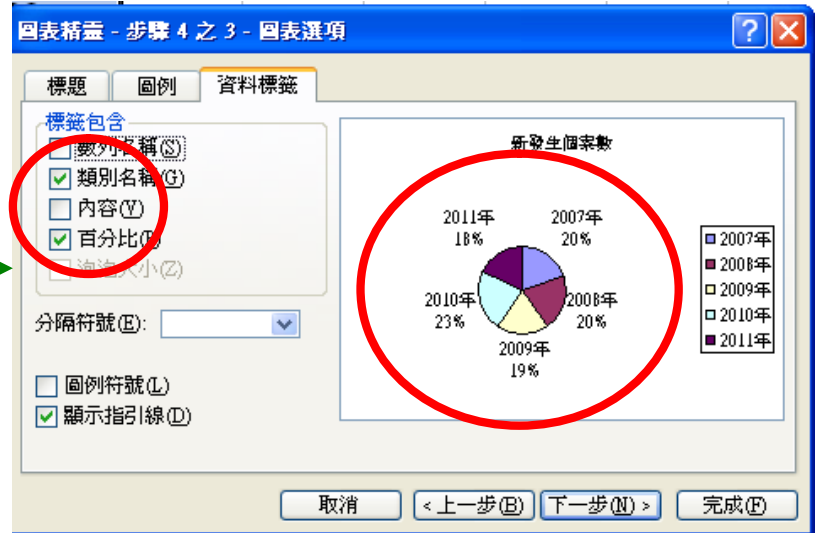
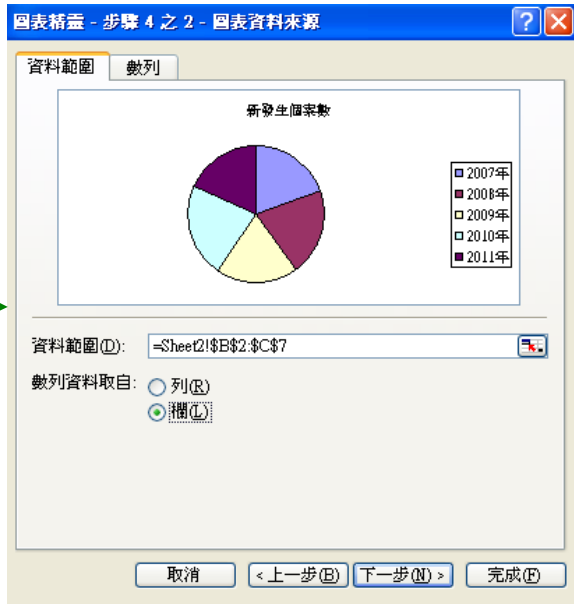


# 分類資料的基本統計圖2

- 派餅圖 (Pie charts)



# 派餅圖繪製步驟1 excel 2003為例



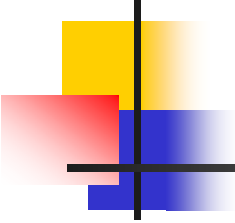


# 連續性資料的描述

---

集中量數

變異量數



# 集中量數

---

- 平均值 (mean,  $M$ )
  - 總和/個數
- 中位數 (median,  $Mdn$ )
  - 全部個體小到大排序後，最中間的一個
- 眾數 (mode,  $Mo$ )
  - 出現頻率最多的數值
- 平均值 = 中位數 = 眾數 → 常態分布

## 三種集中量數的比較

	眾數	中位數	平均數
類別尺度	○	×	×
序位尺度	○	△	×
連續尺度	○	○	○
優點	簡單，不受極端值影響	算簡單，不受極端值影響	最精密，考慮到每個觀察值
缺點	粗略，無法反應所有觀察值	無法反應所有觀察值	易受極端值影響



# 變異量數

- 組距 (range)
  - 最大值與最小值之差，觀察值的範圍
- 變異數 (variance, Var)
  - $\{\Sigma(\text{每個數值}-\text{平均數})^2\}/\text{個數}$
- 標準差 (standard deviation, SD)
  - $(\sqrt{\text{變異數}})$
- 標準誤 (standard error, SE)
  - 標準差 /  $(\sqrt{\text{個數}})$  (樣本平均值間的差別)
- 變異係數 (coefficient of variance)
  - 標準差 / 平均值 (主要比較不同變項或單位的變異程度)
- 四分位差 (inter-quartile range, IQR)
  - 觀察值由小到大排序後，高分組第25% (Q3) 的數值 - 低分組後25% (Q1) 的數值



# 連續資料的描述



Y!奇摩首頁 > 新聞首頁 > 焦點新聞

## 台平均薪資4萬5 亞洲四小龍墊底

華視 - 2013年9月10日 上午6:00

-字 +字

### 相關內容



台平均薪資4萬5 亞洲四小龍墊底

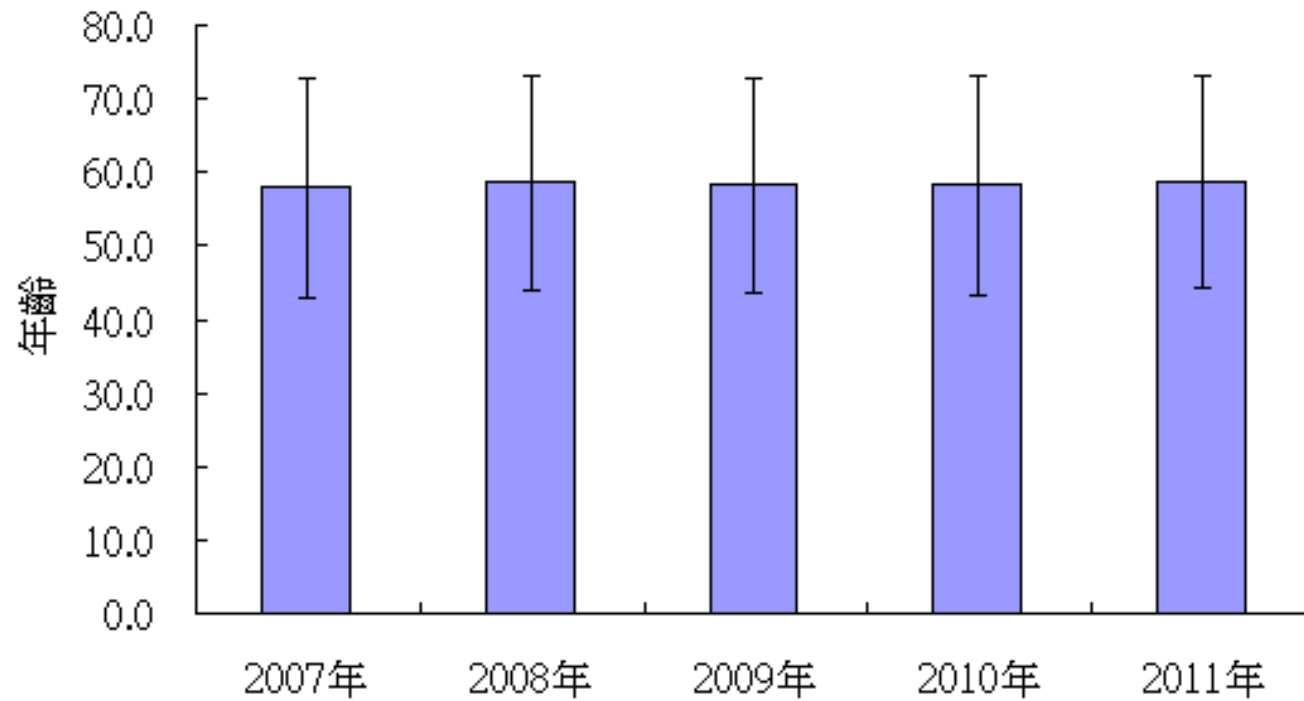
根據勞委會公布的最新國際勞動統計，你知道嗎，我國勞工去年平均月薪是亞洲四小龍最低，只有4萬5888元，相較於其他國家如新加坡有十萬元，韓國的七萬元，香港的六萬多元，反而是我們的家戶儲蓄率是偏高的，遠高於韓國跟日本，勞工團體認為，國內低薪化嚴重，對工作不安全感偏高，所以才會把錢都拿去存起來了。

台灣的小朋友懂得存錢，就連大人也愛跑銀行存錢，這是勞委會公佈的最新國際勞動統計發現，我國勞工雖然薪水賺得沒有比其他國家的人多，像去年平均月薪只有4萬5888元，新加坡就超過10萬，韓國七萬，香港六萬，是亞洲四小龍最低。

但是我們台灣存錢的儲蓄率呢，卻一點都不低，平均16.2%都拿去儲蓄了，韓國只有6.3%，日本只有6.4%，甚至也高過美國、德國等十多個國家，為什麼賺得少，存得會比較多，勞工團體說是因為大家都缺乏安全感，才會把錢存起來。

# 連續資料的基本統計圖1

- 柱狀圖 (Bar charts)

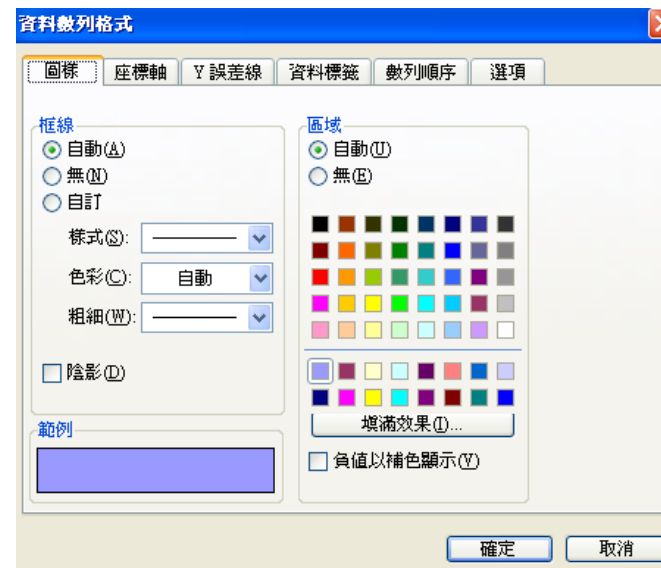
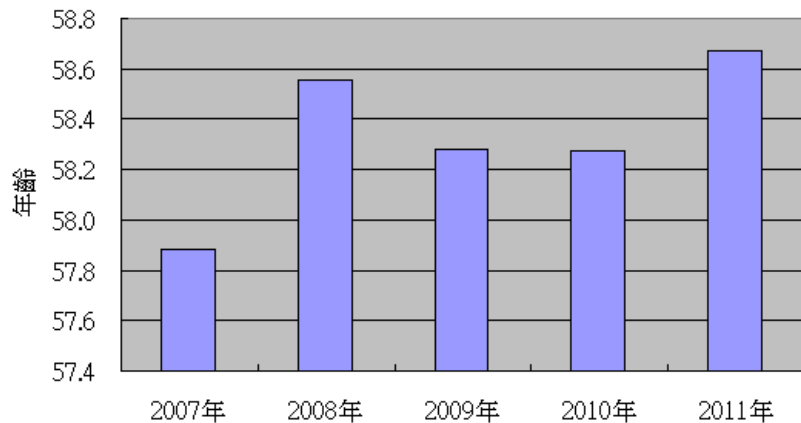
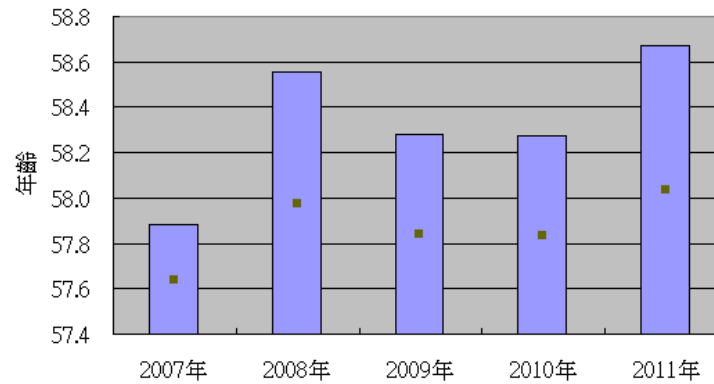


# 柱狀圖繪製步驟1

## excel 2003 為例

	A	B	C	D
1				
2				
3		年齡	mean	SD
4		2007年	57.9	15.0
5		2008年	58.6	14.7
6		2009年	58.3	14.7
7		2010年	58.3	15.0
8		2011年	58.7	14.5
9				

其餘同分類資料的柱狀圖  
步驟



在柱上雙擊滑鼠左鍵

# 柱狀圖繪製步驟2 excel 2003為例

**1**

資料數列格式

圖樣 座標軸 Y 誤差線 資料標籤 數列順序 選項

顯示(D)

兩者 正差 負差 無

誤差量

定值(D): 0.2

百分比(P): 5 %

標準差(S): 1

標準誤差(E)

自訂(C): +

確定 取消

**2**

資料數列格式 - 自訂 +

4

	A	B	C	D	E	F	G	H	I	J	K
1											
2		年齡	mean	SD							
3		2007年	57.9	15.0							
4		2008年	58.6	14.7							
5		2009年	58.3	14.7							
6		2010年	58.3	15.0							
7		2011年	58.7	14.5							
8											

資料數列格式 - 自訂 +

=Sheet2!\$D\$3:\$D\$7

**5**

**3**

資料數列格式

圖樣 座標軸 Y 誤差線 資料標籤 數列順序 選項

顯示(D)

兩者 正差 負差 無

誤差量

定值(D): 0.2

百分比(P): 5 %

標準差(S): 1

標準誤差(E)

自訂(C): +

確定 取消

**6**

資料數列格式 - 自訂 -

7

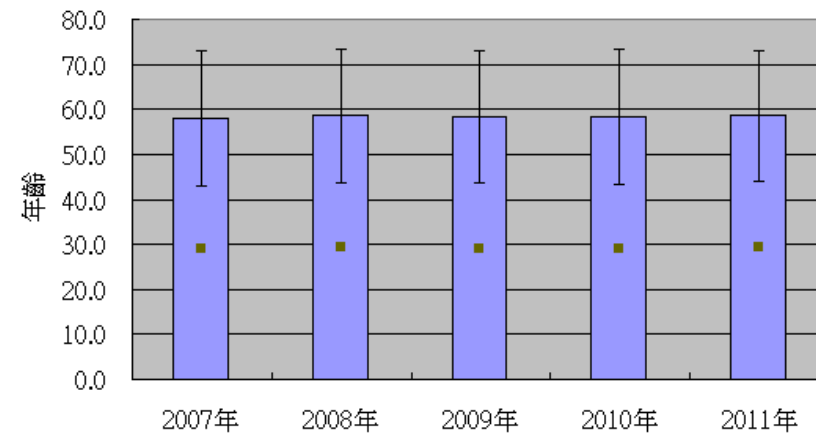
	A	B	C	D	E	F	G	H	I	J	K
1											
2		年齡	mean	SD							
3		2007年	57.9	15.0							
4		2008年	58.6	14.7							
5		2009年	58.3	14.7							
6		2010年	58.3	15.0							
7		2011年	58.7	14.5							
8											

資料數列格式 - 自訂 -

=Sheet2!\$D\$3:\$D\$7

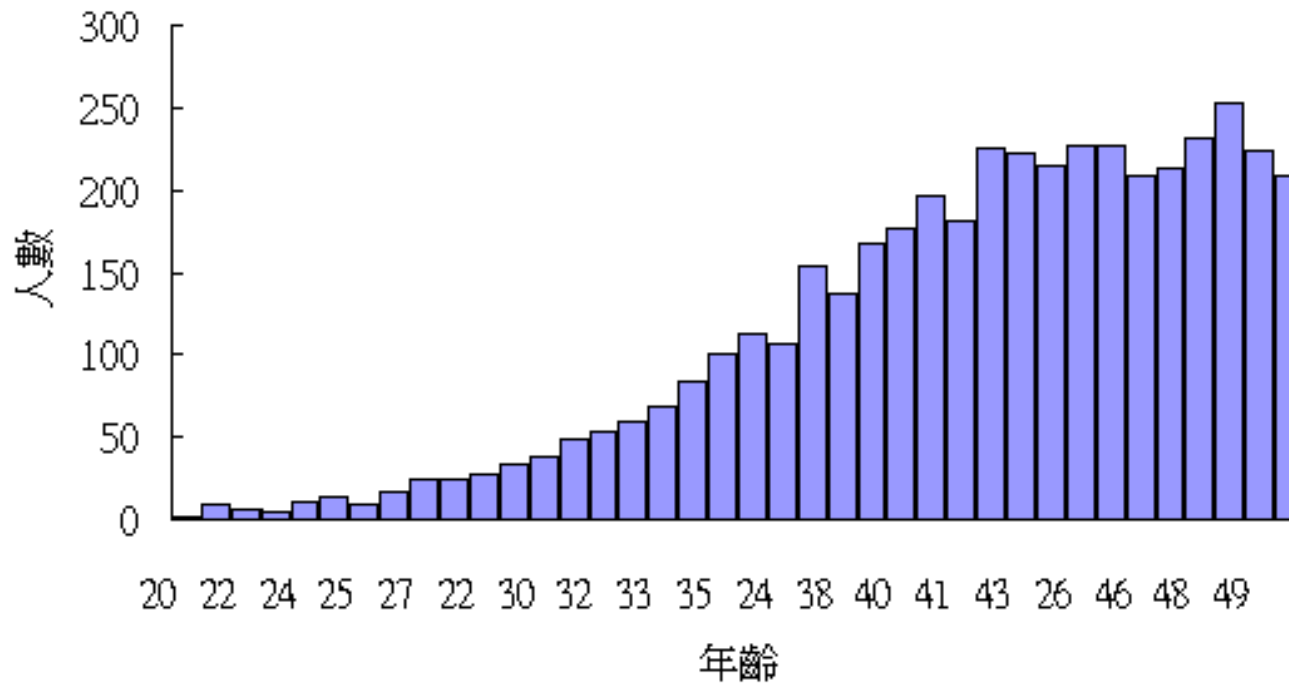
**8**

# 柱狀圖繪製步驟3 excel 2003為例



# 連續資料的基本統計圖2

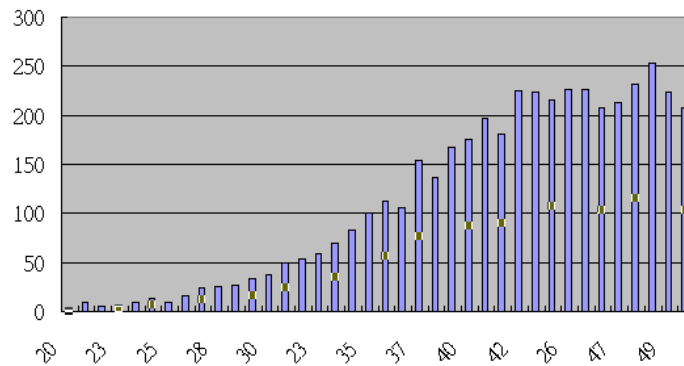
- 直方圖 (Histogram)



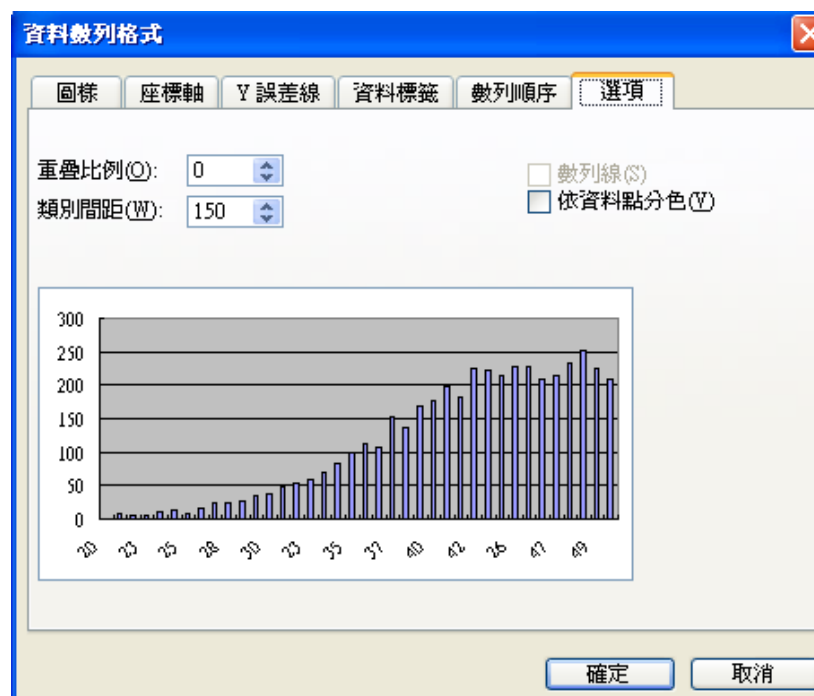
# 直方圖繪製步驟1 excel 2003為例

	A	B	C
1			
2		年齡	人數
3		20	1
4		21	9
5		22	6
6		23	5
7		24	10
8		21	14
9		25	9
10		26	16
11		27	24
12		28	25
13		22	27
14		29	34
15		30	38
16		31	49
17		32	53
18		23	59
19		33	69
20		34	83
21		35	100
22		36	113
23		24	106
24		37	154
25		38	137
26		39	168
27		40	176

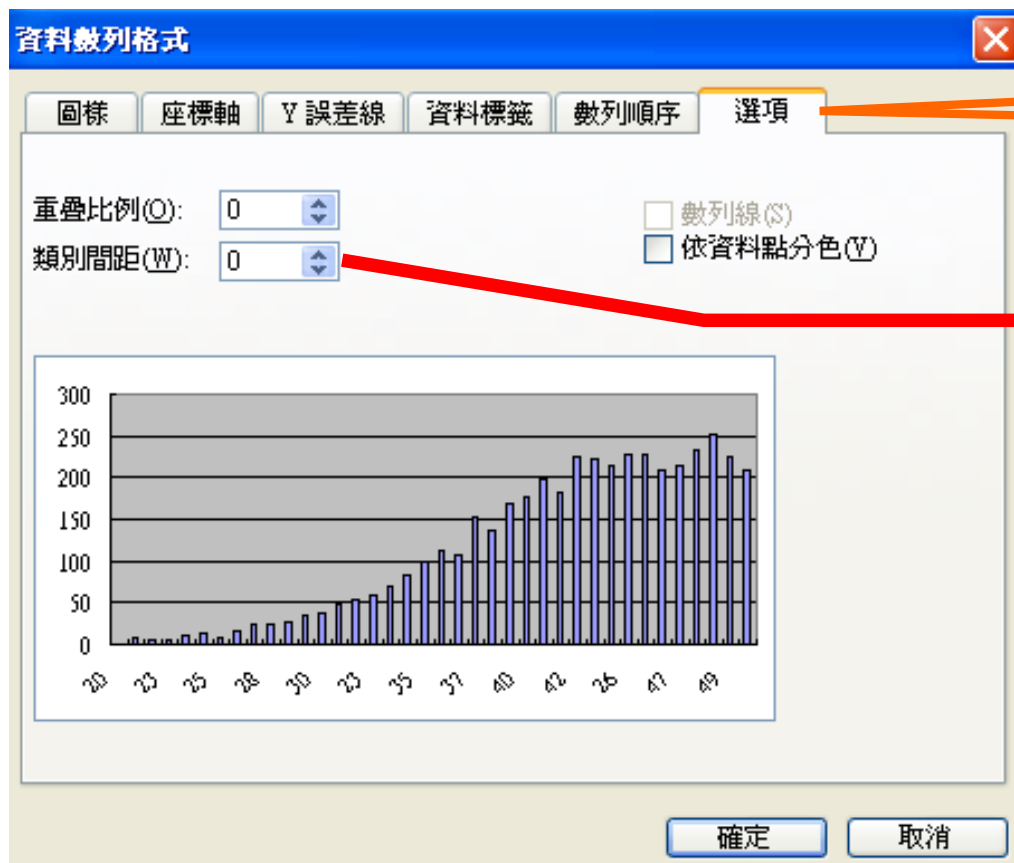
其餘同分類資料的柱狀圖步驟



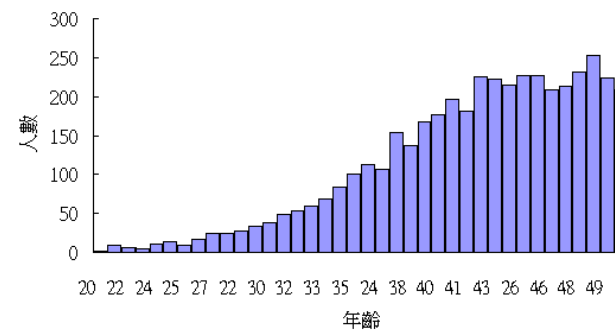
在柱上雙擊滑鼠左鍵



# 直方圖繪製步驟2 excel 2003為例



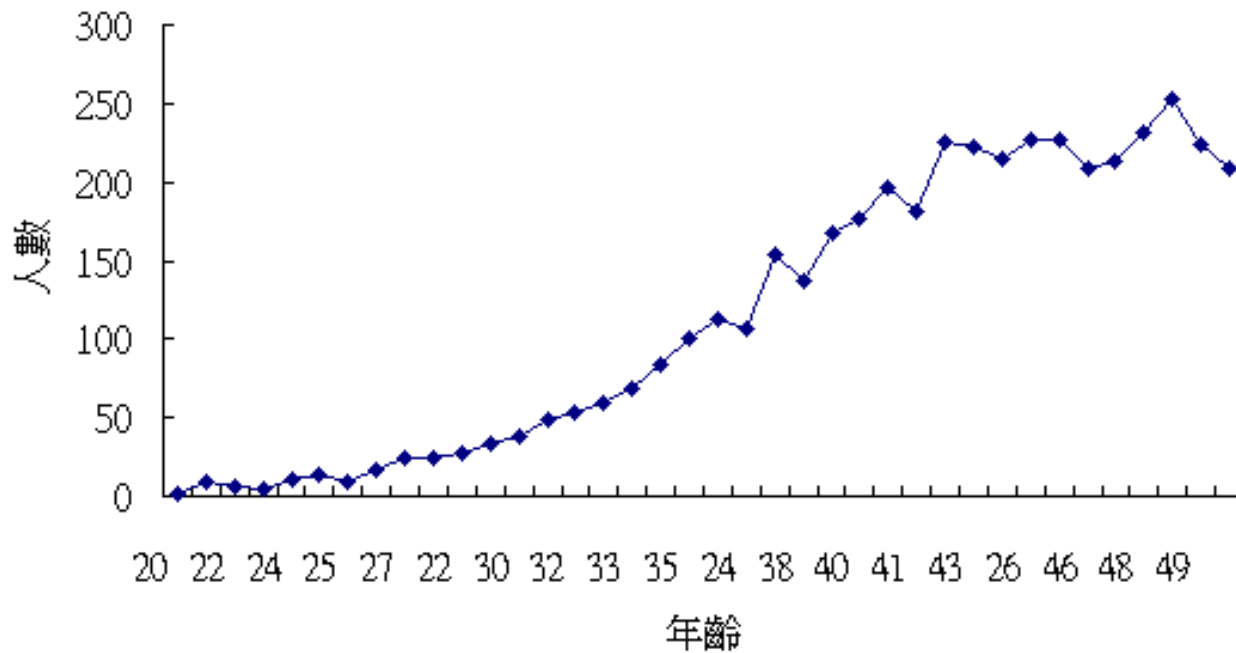
直方圖，X為連續變項，  
因此X間之間距=0



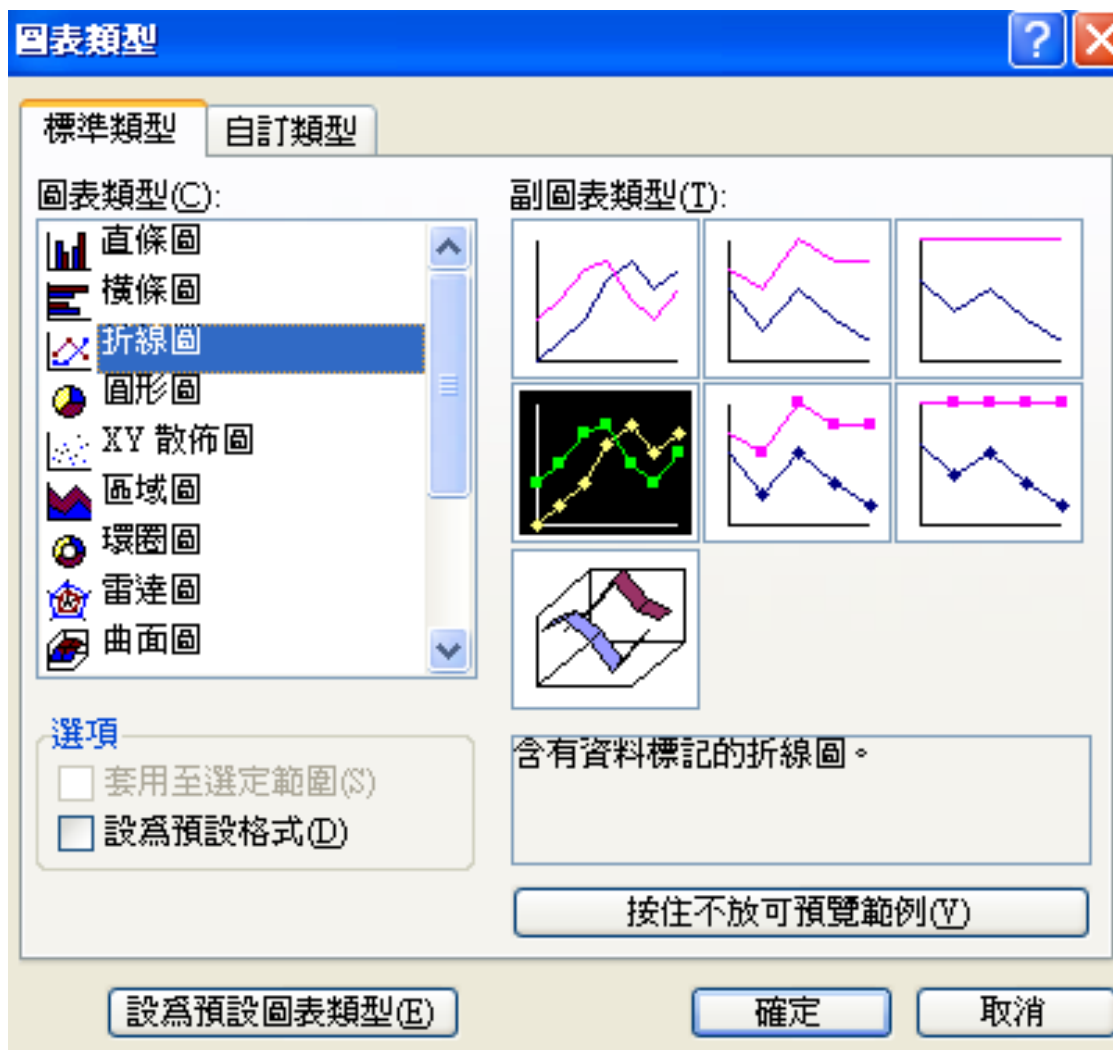


# 連續資料的基本統計圖3

## ■ 折線圖 (Line charts)

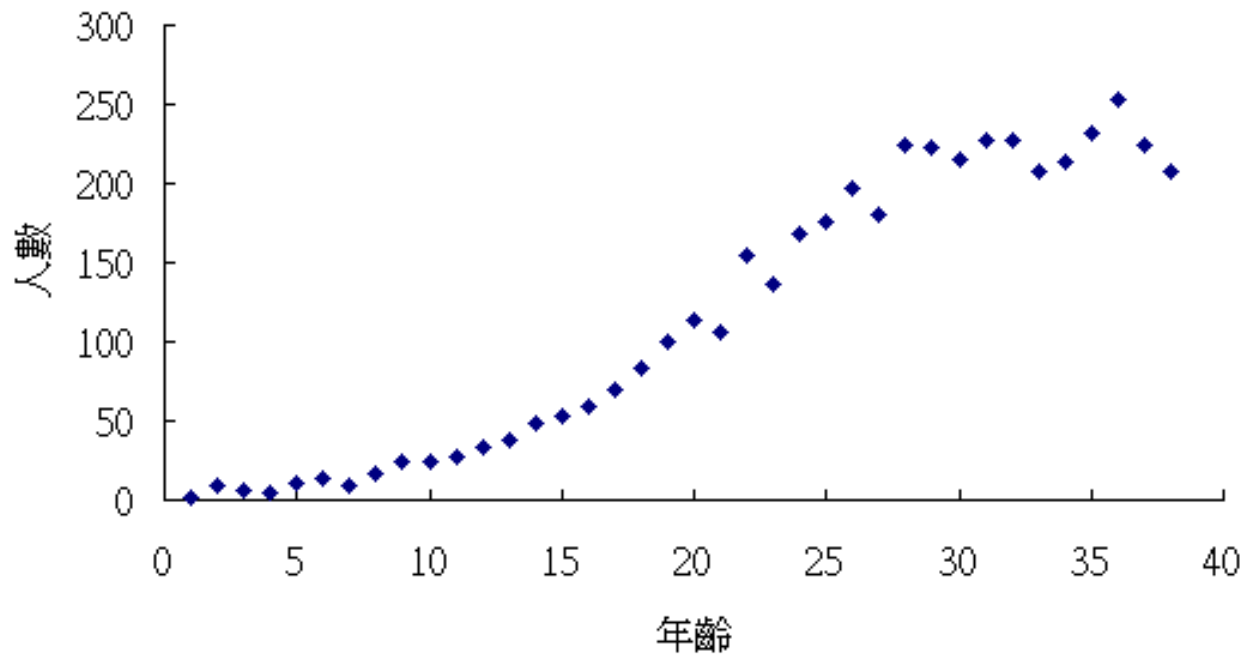


# 折線圖繪製步驟 excel 2003為例

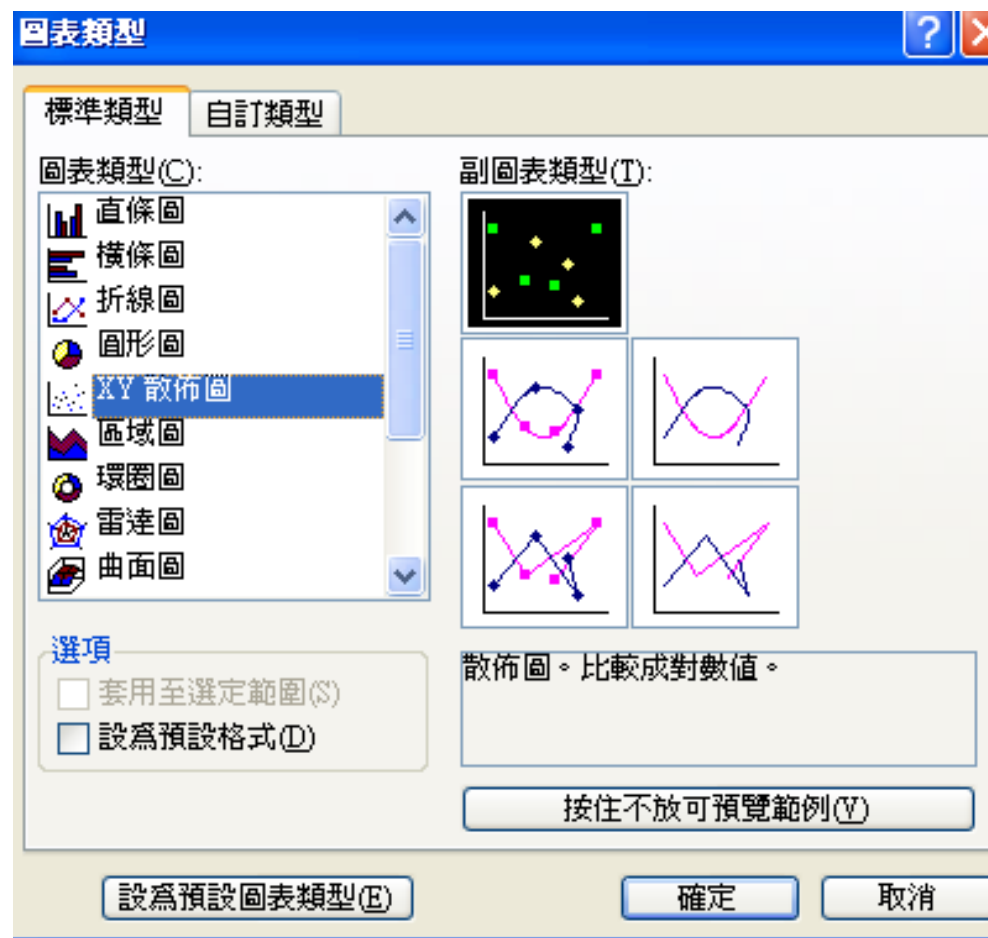


# 連續資料的基本統計圖4

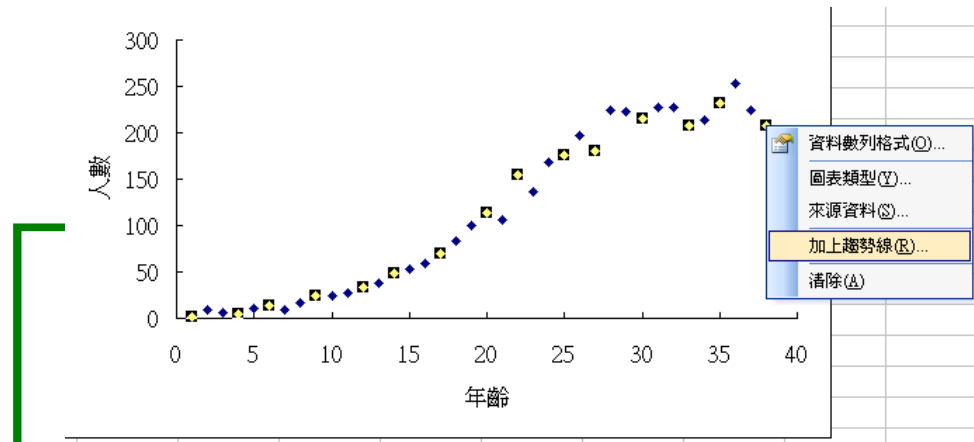
## ■ 散佈圖（Scatter/Dot plots）



# 散佈圖繪製步驟 excel 2003為例



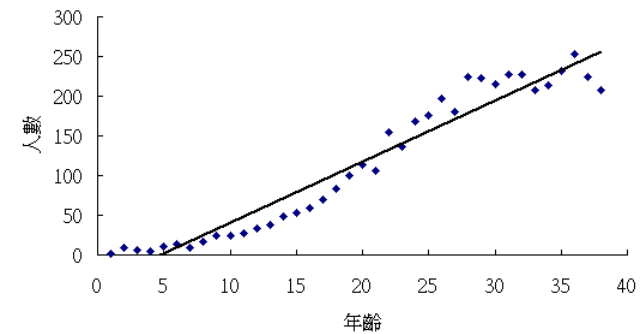
# 散佈圖加上趨勢線步驟 excel 2003 為例



在任一點上單擊滑鼠右鍵

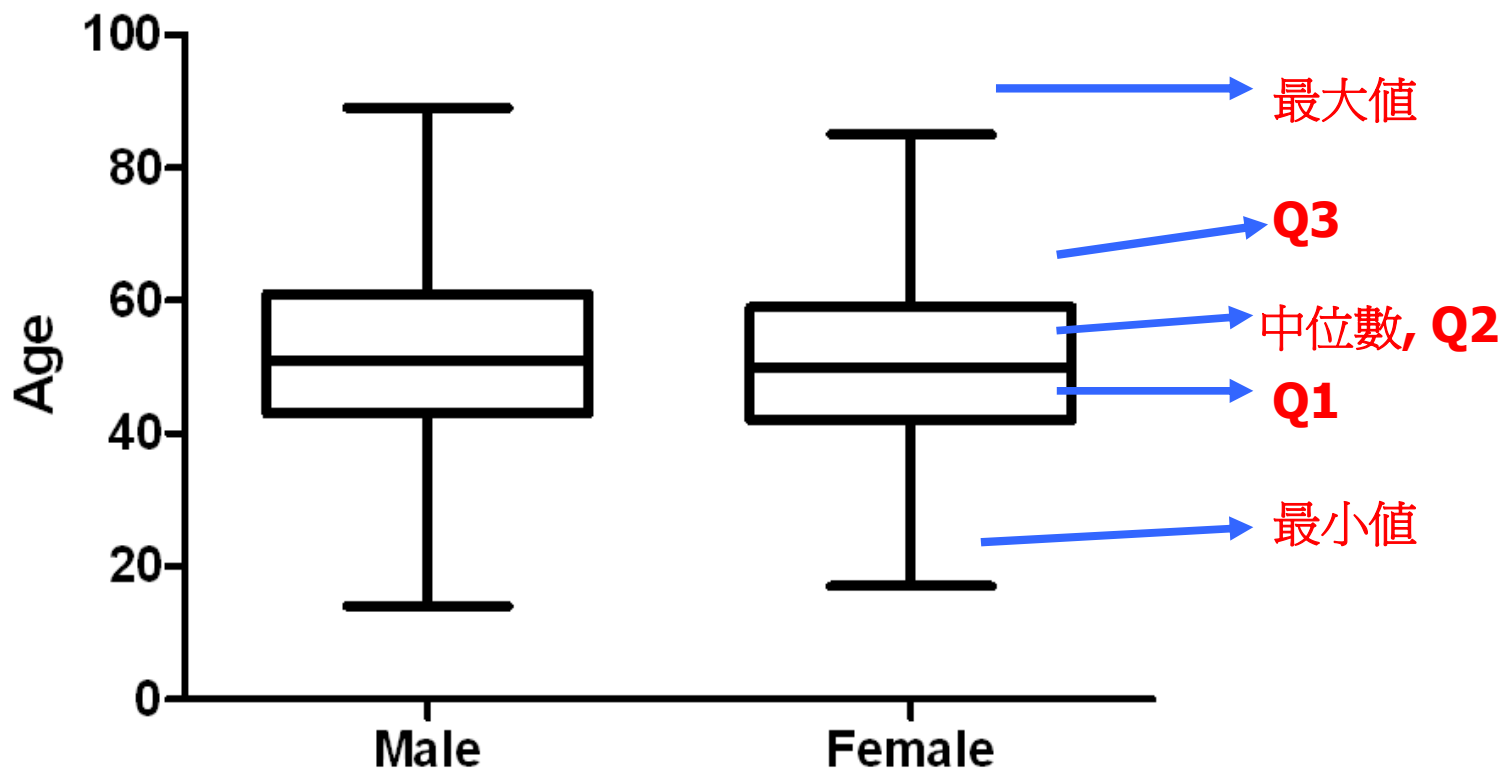


再點選適用的類型，一般以線性為主



# 連續資料的基本統計圖 (5)

## ■ 箱型圖 (Box plots)





# 連續資料的基本統計圖 (6)

- 莖葉圖 (Stem-and-Leaf Plot)  
(個位數)

Frequency	Stem &	Leaf
1.00	2 .	6
2.00	3 .	18
8.00	4 .	14457889
8.00	5 .	12255589
4.00	6 .	1229
1.00	7 .	3



# Receiver Operative Characteristic Curve (ROC)

---





# What is the difference between a screening and a diagnostic test?

---

- A screening test is not definitive
  - A personalized risk estimate
  - Increased risk
- A diagnostic test is definitive and indicates whether or not your baby has a certain disease or condition.

		Disease	
		D+	D-
Test	T+	A	D
	T-	B	C

- 敏感度(sensitivity, Sen.)
  - $P(T+|D+)=A/(A+B)$
- 特異性(specificity, Spe.)
  - $P(T-|D-)=C/(C+D)$
- 陽性預測值(Positive Predictive value, PPV)
  - $P(D+|T+)=A/(A+D)$
- 陰性預測值(Negative Predictive value, NPV)
  - $P(D-|T-)=C/(B+C)$

事前機率Prior probability,  $P(D+)$

事後機率Posterior probability,  $P(D+|T+)$

Sensitivity 95% Specificity.98%

N=1000000

Prevalence=0.04%

<u>Test result</u>	Present	Absent	Total
Positive	<b>380</b>	<b>19,992</b>	20,372
Negative	<b>20</b>	<b>979,608</b>	979,628
Total	400	999,600	1,000,000

$$\text{PPV} = 380 / 20372 = 0.019$$

$$\text{NPV} = 979608 / 979628 = 0.99998$$

Sensitivity 95% Specificity.98%

N=1000

Prevalence=10.0%

<u>Test result</u>	Present	Absent	Total
Positive	<b>95</b>	<b>18</b>	113
Negative	<b>5</b>	<b>882</b>	887
Total	100	900	1,000

$$\text{PPV} = 95/113 = 0.841$$

$$\text{NPV} = 882/887 = 0.994$$

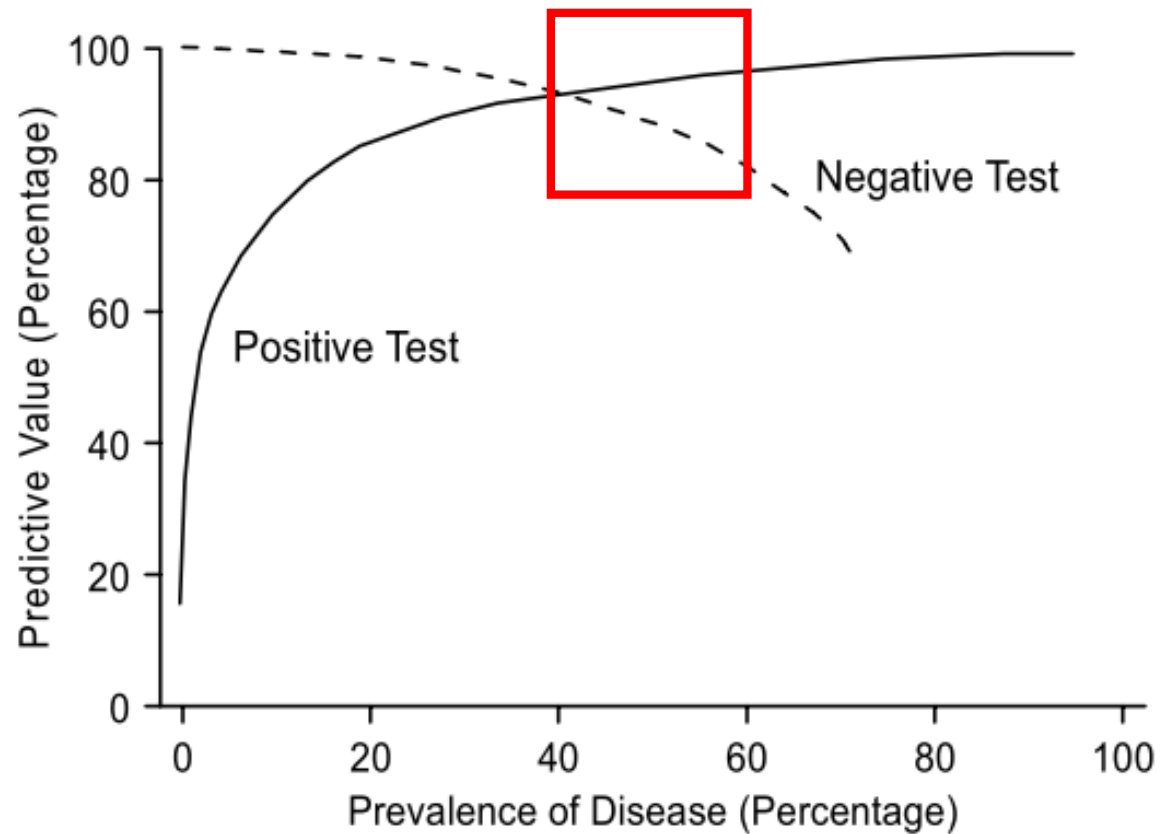


# Prevalence & Screening

(Sensitivity 99%, Specificity 95%)

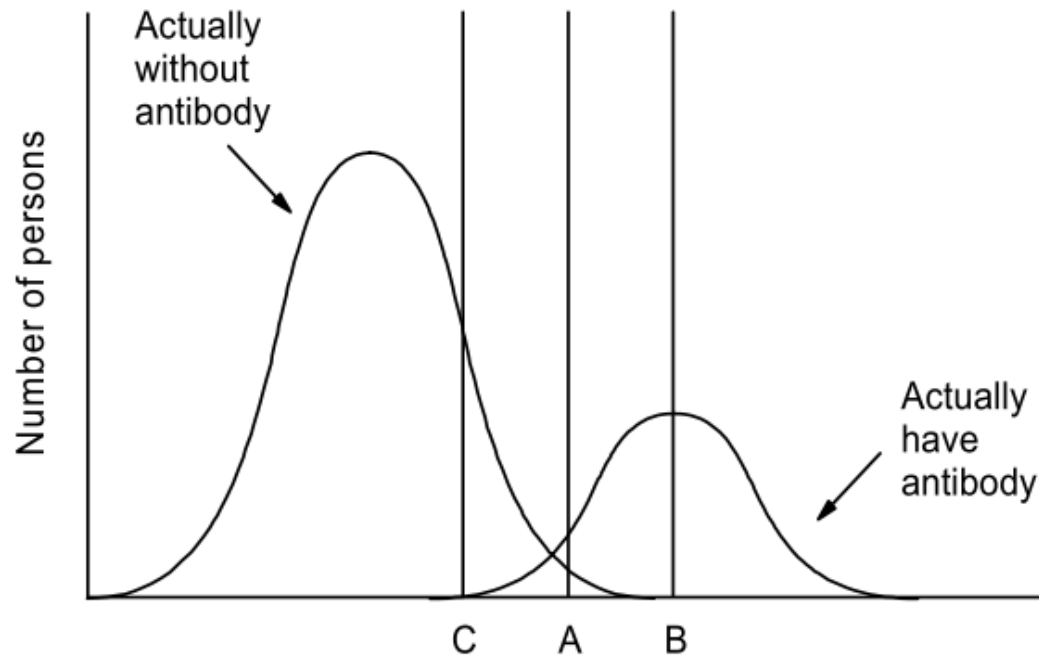
Prevalence	1%	5%
N	1000	1000
D+	10	50
D-	990	950
Test+	$9.9 + 49.5 = 60$	$49.5 + 47.5 = 97$
PPV	17%	51%
NPV	100%	100%

Relationship between prevalence of disease and predictive value,  
with sensitivity and specificity held constant at 95 percent  
(Adapted from Vecchio, 1966)

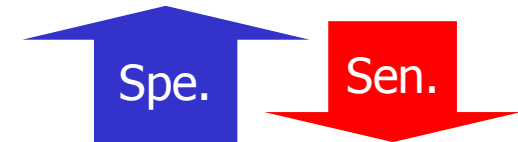


Screening test perform best when prevalence is between 40-60%

# Chang cut-off



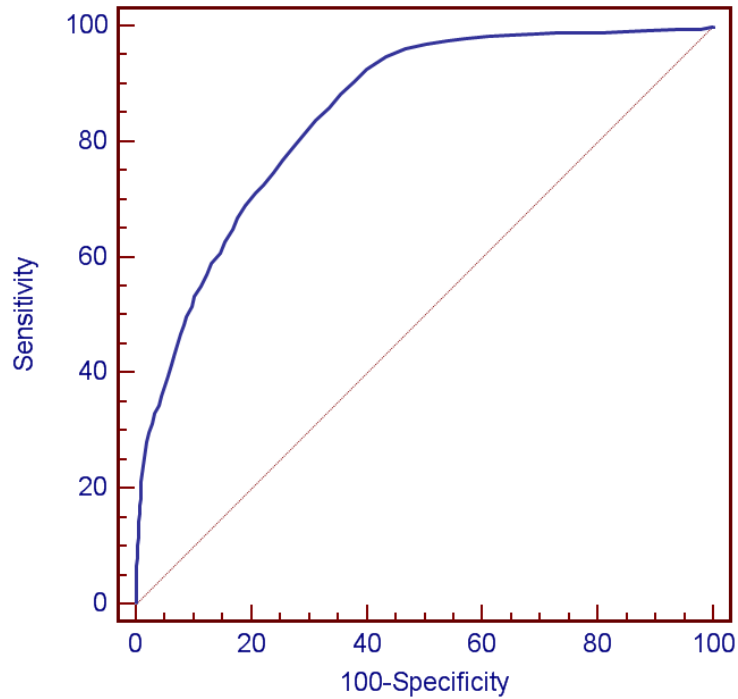
■ A -> B



■ A -> C



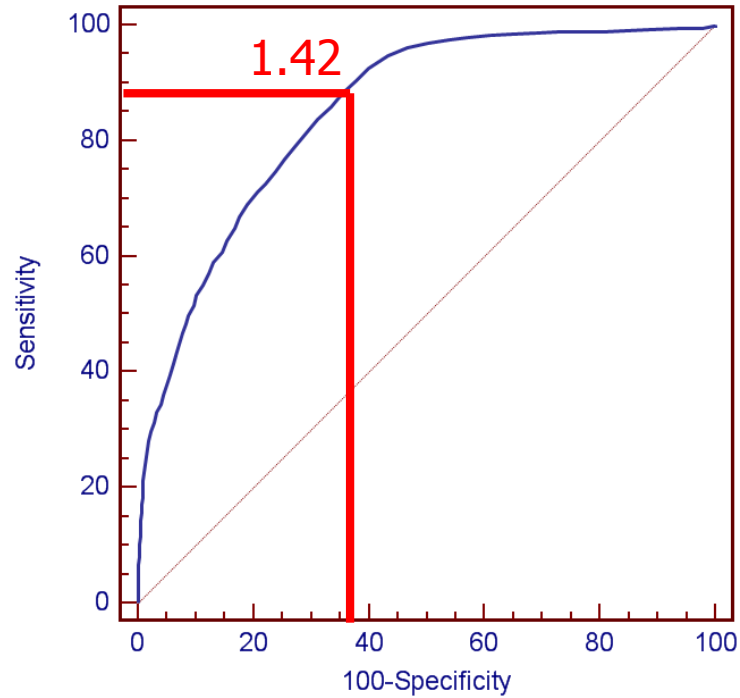
# ROC curve



Positive if Greater Than or Equal To <sup>a</sup>	Sensitivity	1 - Specificity
-7.2500	1.000	1.000
-6.1250	1.000	1.000
-5.8750	1.000	1.000
-5.7083	1.000	.999
-5.1250	1.000	.999
-4.4583	1.000	.998
-4.1667	1.000	.997
-3.5417	1.000	.997
-2.5417	1.000	.997
-1.0000	.999	.997
.0417	.996	.977
.0833	.995	.952
.1250	.994	.933
.1667	.994	.916
.2083	.992	.886
.2917	.991	.844
.3333	.990	.818
.3750	.990	.804
.4167	.990	.792
.4583	.990	.763
.5417	.989	.726
.5833	.988	.702
.6250	.987	.689
.6667	.986	.676
.7083	.985	.647
.7917	.983	.609



# ROC curve



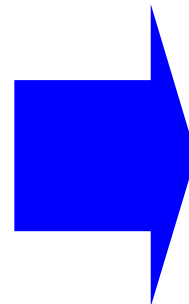
Area under curve (AUC)	0.852
95% C.I. of AUC	0.844~0.859
p-value	<0.001
Best cut-of	1.42
Sensitivity	88.3%
Specificity	64.5%
PPV	87.0%
NPV	67.2%

# 繪製ROC SPSS17.0為例

\*test.sav [DataSet3] - PASW Statistics Data Editor

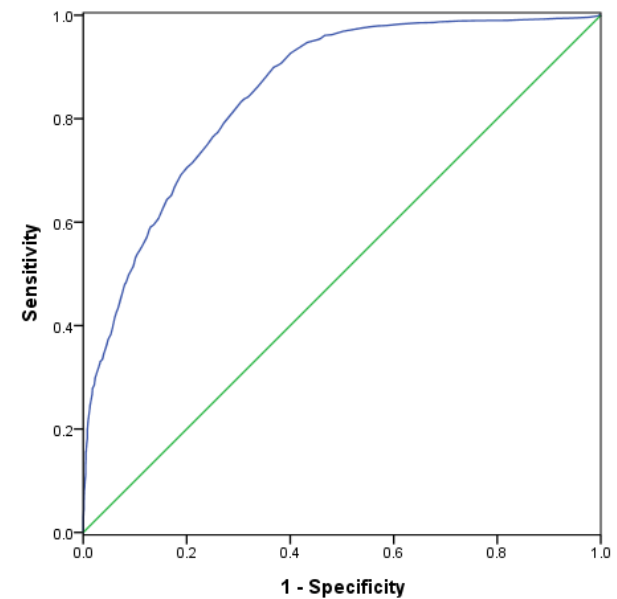
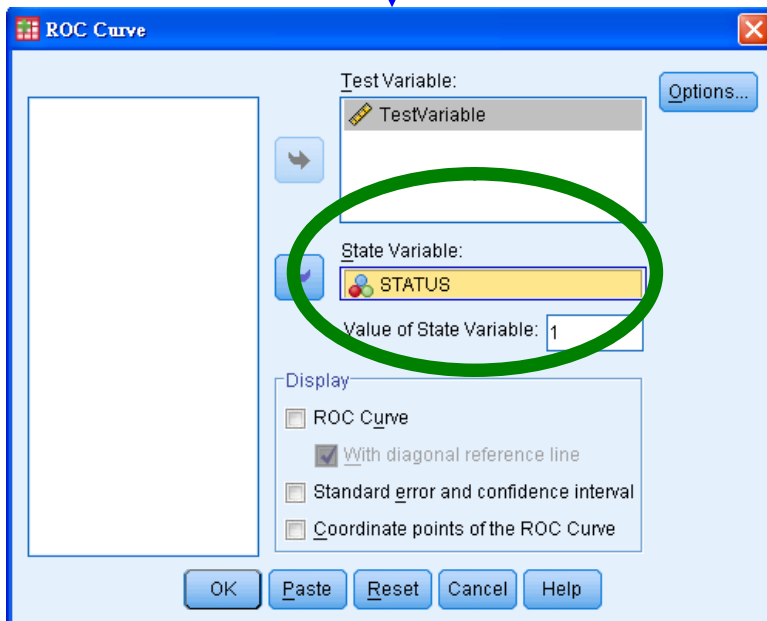
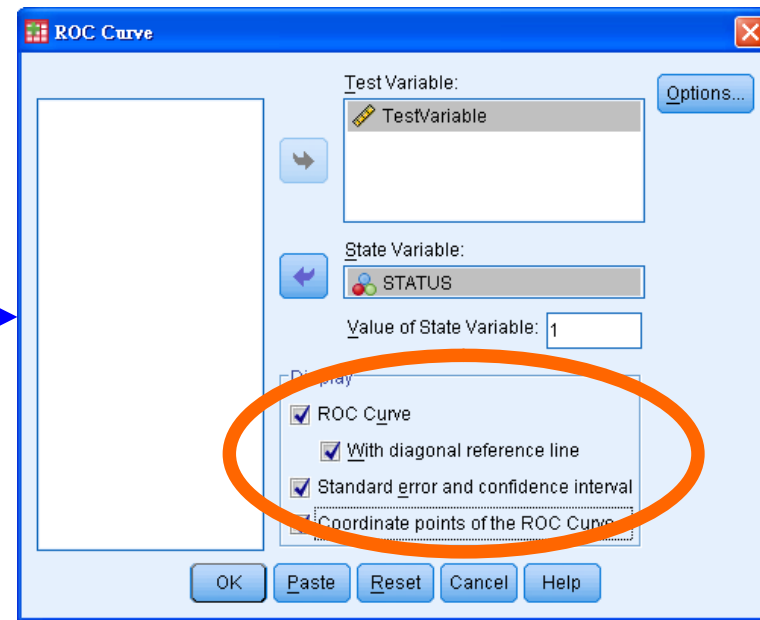
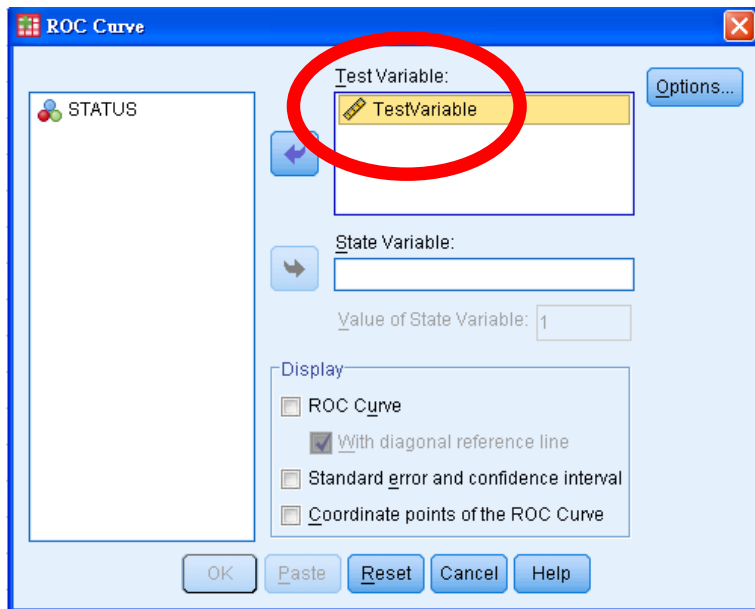
File Edit View Data Transform Analyze Direct Marketing Graphs

	STATUS	TestVariable	var	var	var
1	1	4.17			
2	0	1.75			
3	0	2.25			
4	1	5.08			
5	1	4.92			
6	1	6.17			
7	1	5.33			
8	0	1.00			
9	0	3.17			
10	1	4.83			
11	1	5.50			
12	1	4.83			
13	1	6.00			
14	1	5.58			
15	0	.08			
16	0	.25			
17	0	.08			
18	0	.50			



Analyze Direct Marketing Graphs

- Reports ▶
- Descriptive Statistics ▶
- Tables ▶
- Compare Means ▶
- General Linear Model ▶
- Generalized Linear Models ▶
- Mixed Models ▶
- Correlate ▶
- Regression ▶
- Loglinear ▶
- Neural Networks ▶
- Classify ▶
- Dimension Reduction ▶
- Scale ▶
- Nonparametric Tests ▶
- Forecasting ▶
- Survival ▶
- Multiple Response ▶
- Missing Value Analysis...
- Multiple Imputation ▶
- Complex Samples ▶
- Quality Control ▶
- ROC Curve...**



**Case Processing Summary**

STATUS	Valid N (listwise)
Positive <sup>a</sup>	6314
Negative	2342

1

3

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is 1.

**Area Under the Curve** 2

Test Result Variable(s): TestVariable

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.851	.005	.000	.842	.861

The test result variable(s): TestVariable has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Positive if Greater Than or Equal To <sup>a</sup>	Sensitivity	1 - Specificity
-7.2500	1.000	1.000
-6.1250	1.000	1.000
-5.8750	1.000	1.000
-5.7083	1.000	.999
-5.1250	1.000	.999
-4.4583	1.000	.998
-4.1667	1.000	.997
-3.5417	1.000	.997
-2.5417	1.000	.997
-1.0000	.999	.997
.0417	.996	.977
.0833	.995	.952
.1250	.994	.933
.1667	.994	.916
.2083	.992	.886
.2917	.991	.844
.3333	.990	.818
.3750	.990	.804
.4167	.990	.792
.4583	.990	.763
.5417	.989	.726
.5833	.988	.702
.6250	.987	.689
.6667	.986	.676
.7083	.985	.647
.7917	.983	.609