

基本統計介紹

統計概念

- 統計學係指蒐集、整理及分析統計資料，並由分析的結果做較大範圍的推論，使其在不確定性(uncertainty)的情況下，獲得普遍性結論的科學方法。
- 統計學主要分為敘述統計學與推論統計學

統計概念

- 敘述統計學—包括統計方法中的蒐集、整理及分析統計資料的部份，其目的在了解統計資料的特徵，以解決統計問題。而敘述統計學逕討論統計資料本身，並不分析結果。
- 推論統計學—探討如何在不確定性的情況下，利用敘述統計學的結果，對類似問題做較大範圍的推測，以獲得普遍性的結論。
- 實驗設計—對干擾變數的控制下，實驗者藉由操作自變數，以觀察其對應變數的影響，以了解兩變數之間的關聯。

範例

- 某研究針對某四製造商進行電池產品的調查，其資料如下：

製造商	售價(元/個)
A	15
B	15
C	12
D	10

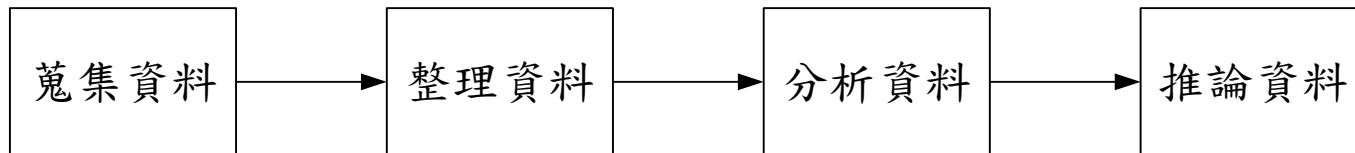
- 試就下列敘述說明，何者是敘述統計學?推論統計學?實驗設計?

範例

- (1) 四家製造商中，電池廠品每個平均售價13元($(15+15+12+10)/4=13$)
--- 敘述統計學
- (2) 台灣所有電池製造商，電池產品每個平均售價13元---推論統計學
- (3) 台灣有一半電池製造商，電池產品每個平均售價15元---推論統計學
- (4) 四家製造商中，電池產品每個售價低於15元者，有C及D兩家製造商
--- 敘述統計學
- (5) 研究者A製造商的電池產品做價格調整，分別每個以15、13、10元等價格出售，以觀察其對銷售量的影響 --- 實驗設計

統計方法

- 統計方法首先要對產生問題的統計資料加以“蒐集”，而初步資料的蒐集往往雜亂無章，必須透過“整理”，如此資料的特徵才得以顯現。
- 單只整理資料未必能完全符合研究者的需要，因此常必須進一步“分析”。
- 統計的目的在於推論工作，而推論是在不確定的情況下進行，以期望能獲得普遍性的結論，建立合理的原則與判斷。



兩個基本概念---母體與樣本

- 母體—對所欲推論對象的全體，進行量測或蒐集某些特性的紀錄之完整集合。母體所代表的是調查的目標，而資料蒐集過程的目的在於取得母體特徵的結論。
- 樣本—來自於母體的部分組合，也就是在調查過程中實際上量測或計數特性的集合。
- 母體與樣本的關係
 - (1)樣本來自於母體，具有母體的特徵。
 - (2)先確立母體才能確立樣本。
 - (3)由於時間與成本考量，大多數的情況下對樣本調查。
 - (4)由於樣本具有母體的特徵，因此可利用樣本的特徵推論母體的特徵。

統計圖表介紹

統計圖表

- 當統計人員得到一組數據後，第一步進行的工作就是繪圖展示，由顯現的圖型中可看出整體型態或趨勢。
- 如果所得數據很少的話，可將數據由大到小（或由小到大）順序排列，所得的結果為數列。

點圖

- 例: 某人因工程師量測九位工人的抓力，得到的量測值分別為

原始資料	128.3	106.5	93.9	116.6	152.4	125.0	132.1	105.8	136.7
由小到大 排序	93.3	105.8	106.5	116.6	125.0	128.3	132.1	136.7	152.4

單位: 磅

- 點圖可以繪製在一條直線上，在線上標寫容納所有該組數據的尺度。

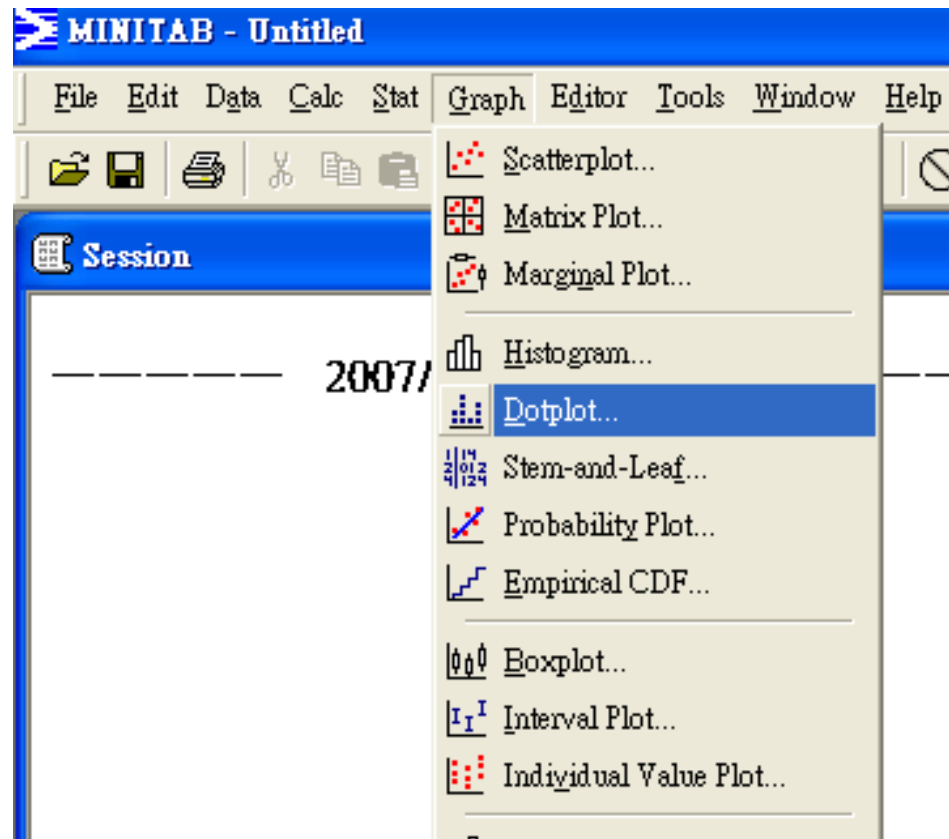
點圖

將九位工人抓力的
資料輸入此欄位

↓	C1	C2	C3
	工人抓力		
1	128.3		
2	106.5		
3	93.9		
4	116.6		
5	152.4		
6	125.0		
7	132.1		
8	105.8		
9	136.7		
10			

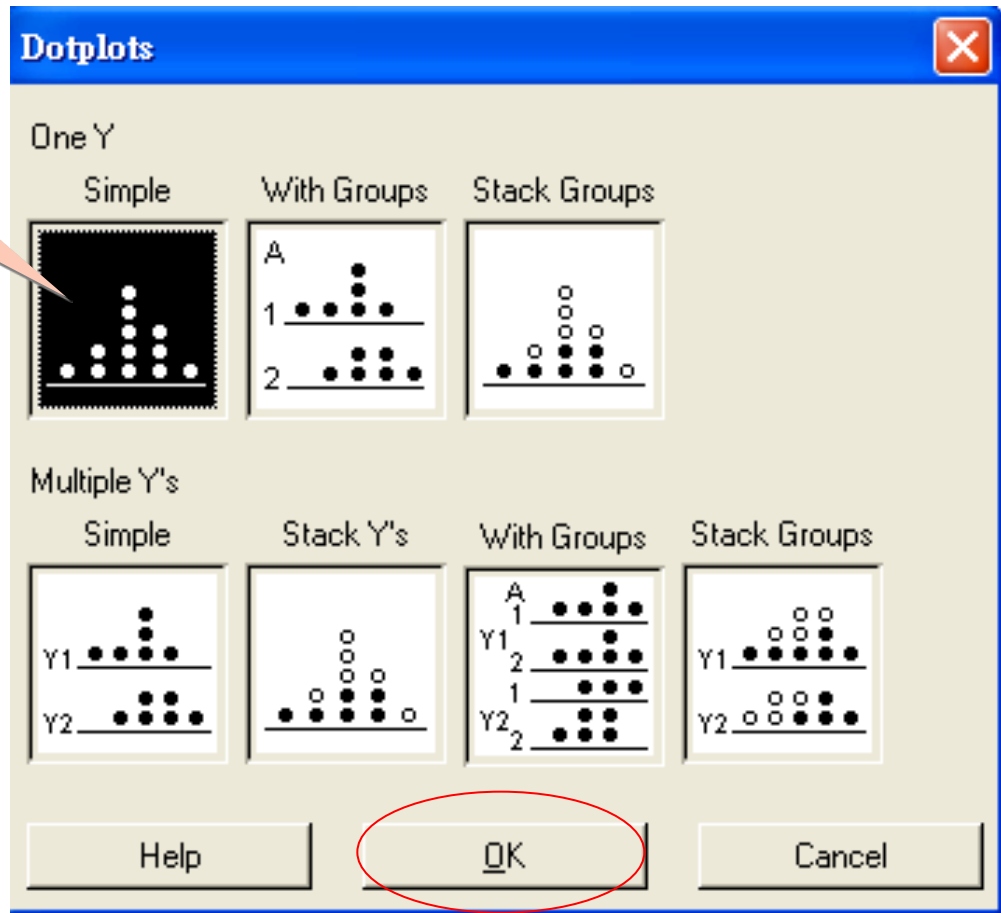
點圖

點選路徑 Graph > Dotplot

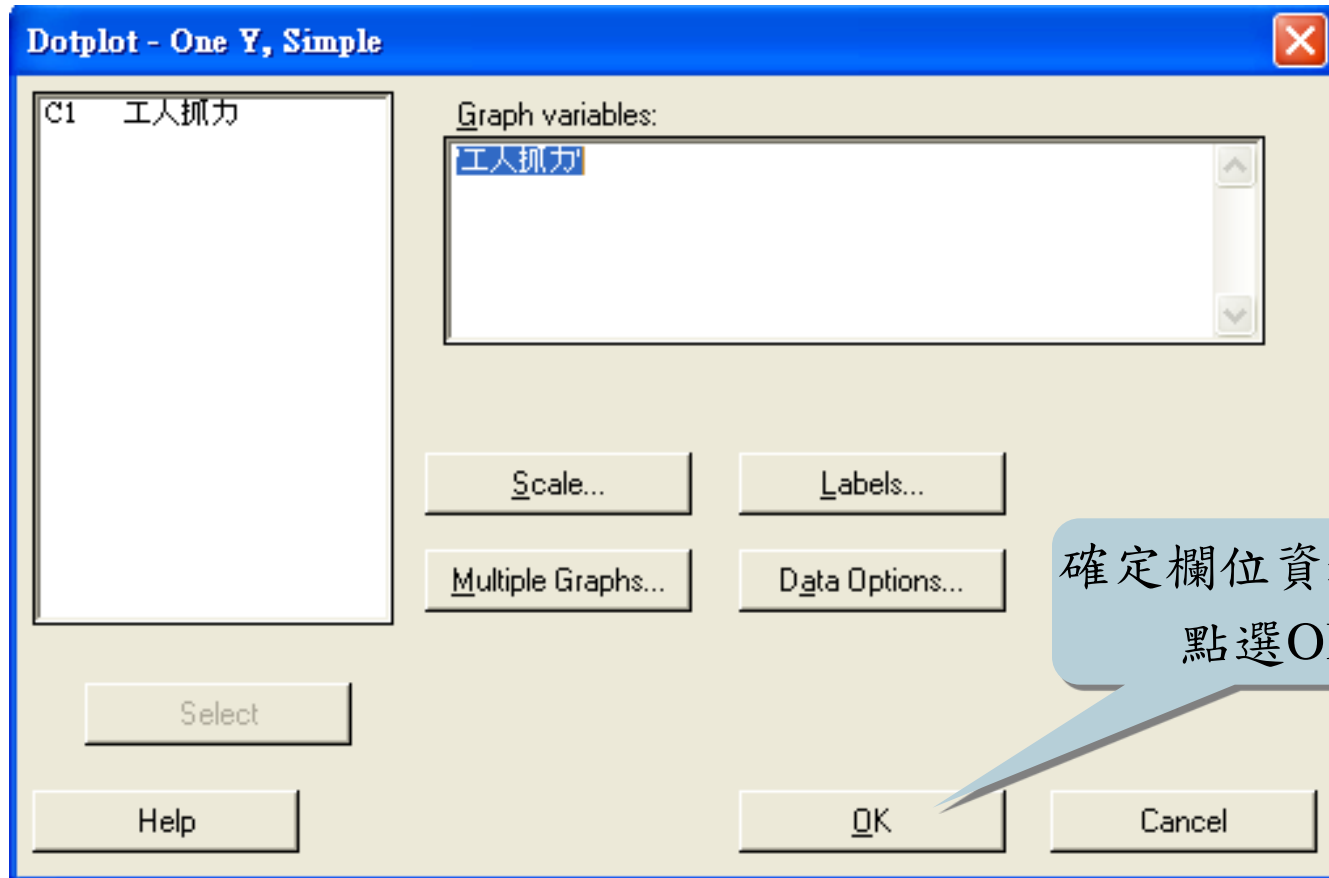


點圖

點選此功能後，
點選OK。

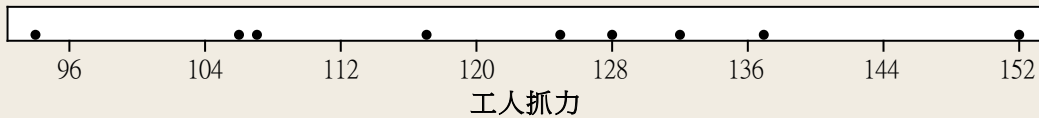


點圖



點圖

Dotplot of 工人抓力



此為點圖分佈
結果。

從此圖可初步看出工人抓力大多分布在100~140磅之間，而其中有兩點在96磅以下與150磅以上，可針對其個別狀況去探究其原因。

點圖

- 例: 若想了解兩組工人中各九個人的抓力，得到的量測值分別為：

A組	128.3	106.5	93.9	116.6	152.4	125.0	132.1	105.8	136.7
B組	91.3	102.7	110.5	106.6	120.0	118.3	112.1	133.7	142.4

單位: 磅

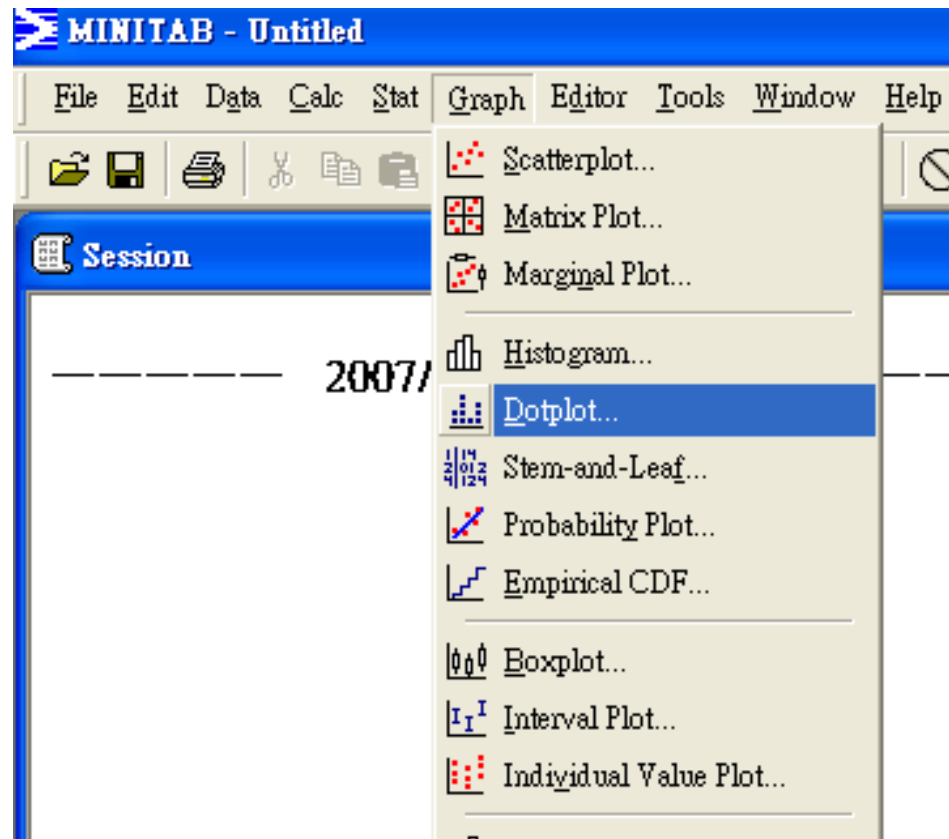
點圖

↓	C1-T	C2	C3
		工人抓力	
1	A	116.6	
2	A	152.4	
3	A	125.0	
4	A	132.1	
5	A	105.8	
6	B	136.7	
7	B	91.3	
8	B	102.7	
9	B	110.5	
10	B	106.6	
11	B	120.0	
12	B	118.3	
13	B	112.1	
14	B	133.7	
15	B	142.4	
16			

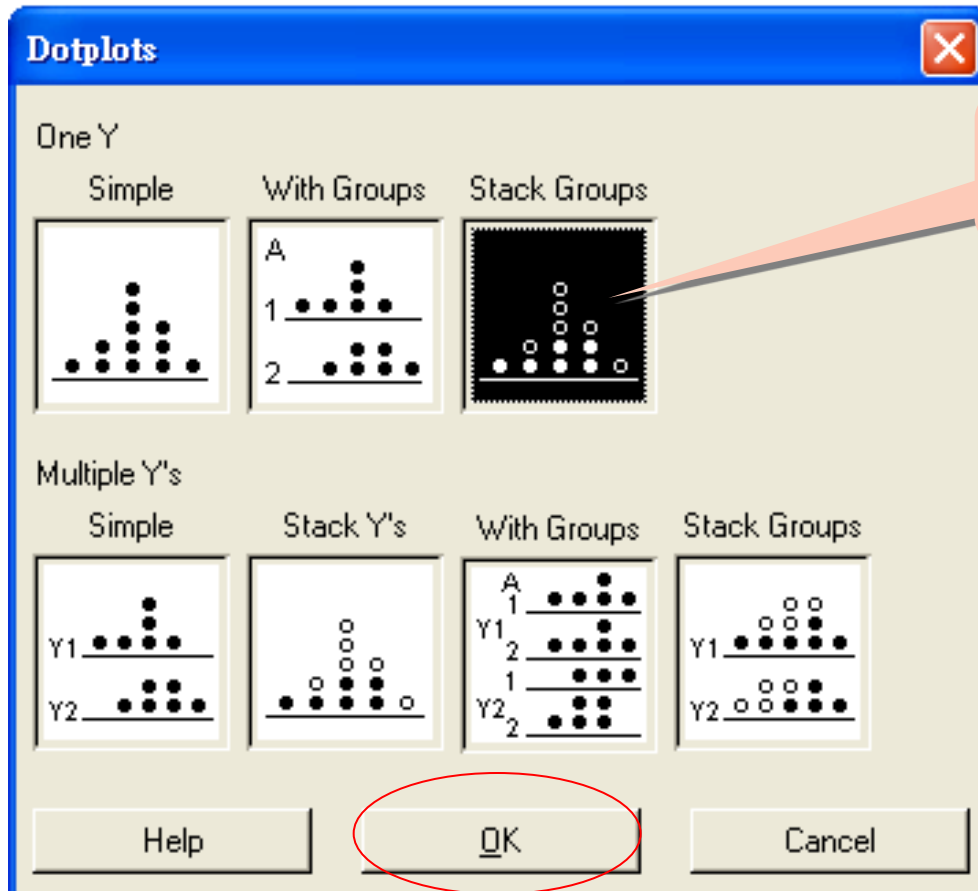
將A、B兩組各九位工人抓力的資料輸入欄位如右。

點圖

點選路徑 Graph > Dotplot



點圖



點選此功能後，
點選OK。

點圖

Dotplot of 工人抓力 vs C1



此為點圖
分佈結果。

從此圖可初步看出兩組工人抓力大多分布在100~140磅之間，而其中有兩點在96磅以下與150磅以上，可針對其個別狀況去探究其原因。

直方圖

- 係指依照次數分配之分組數量，以若干長條來表示分組數量之次數的多寡，長條圖適合於連續數量特性（或者資料數量大）的資料。
- 例:某工廠所生產的五十個零組件的厚度資料，其資料如下:

41	40	45	46	44	43	45	42	37	45
43	41	46	45	46	40	40	50	44	42
50	48	44	38	40	45	46	42	45	42
43	40	38	40	39	45	45	46	44	51
39	38	44	39	43	46	45	46	40	44

單位: mm

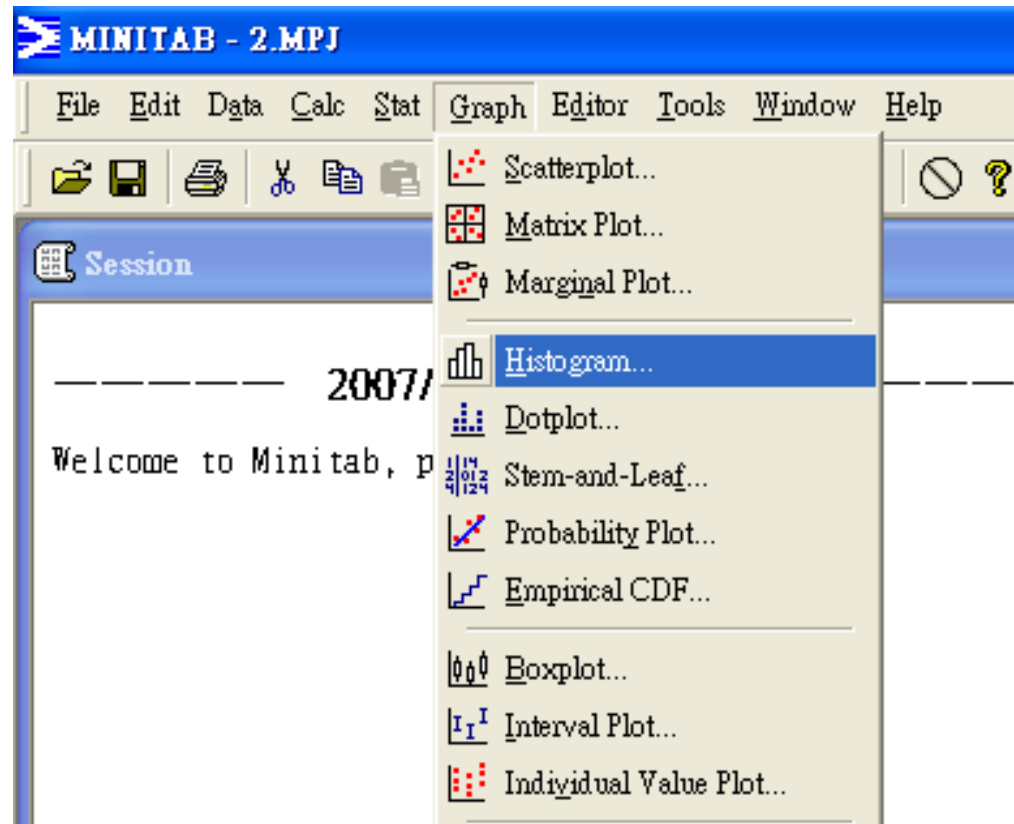
直方圖

將五十個零件厚度的資料輸入此欄位。

↓	C1	C2
	厚度	
1	41	
2	40	
3	45	
4	46	
5	44	
6	43	
7	45	
8	42	
9	37	
10	45	
11	43	
12	41	

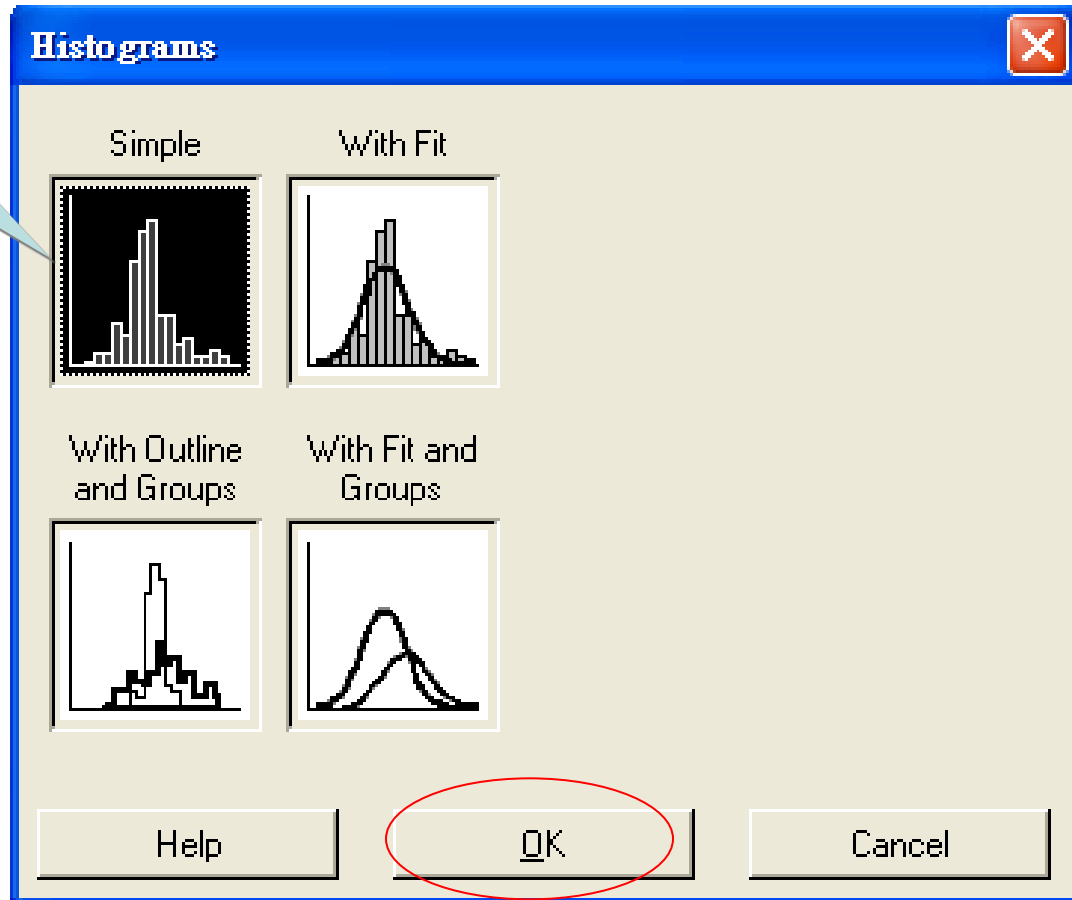
直方圖

點選路徑 Graph > Histogram

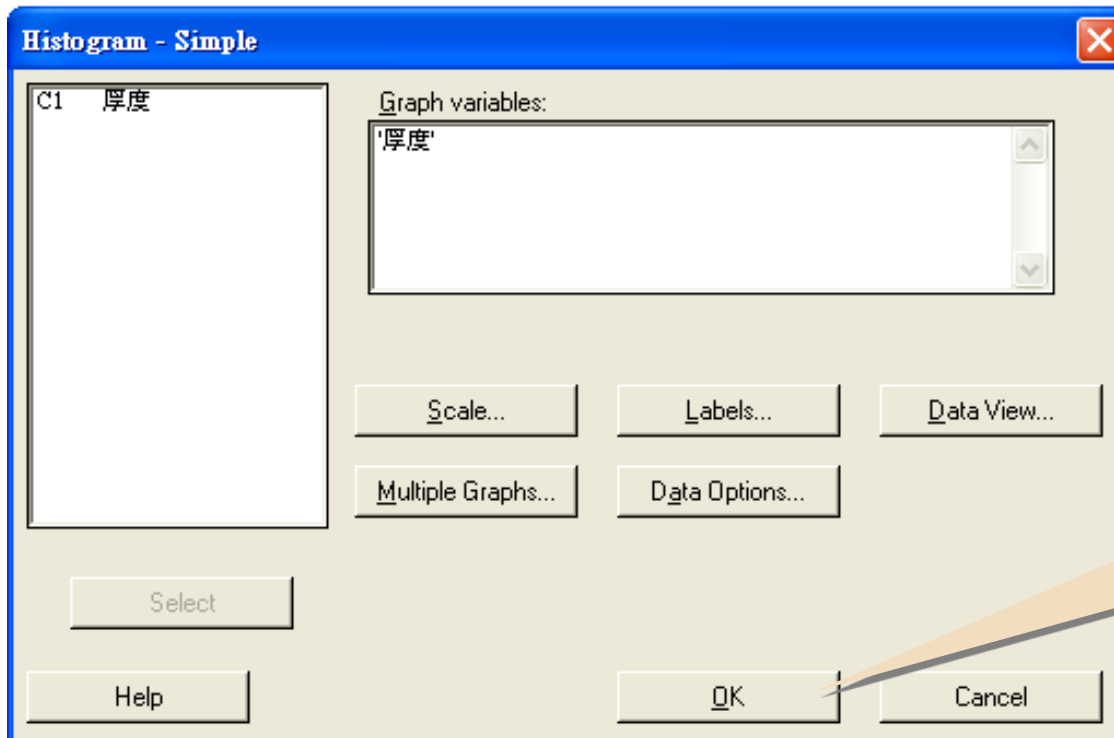


直方圖

點選此功能後，
點選OK。

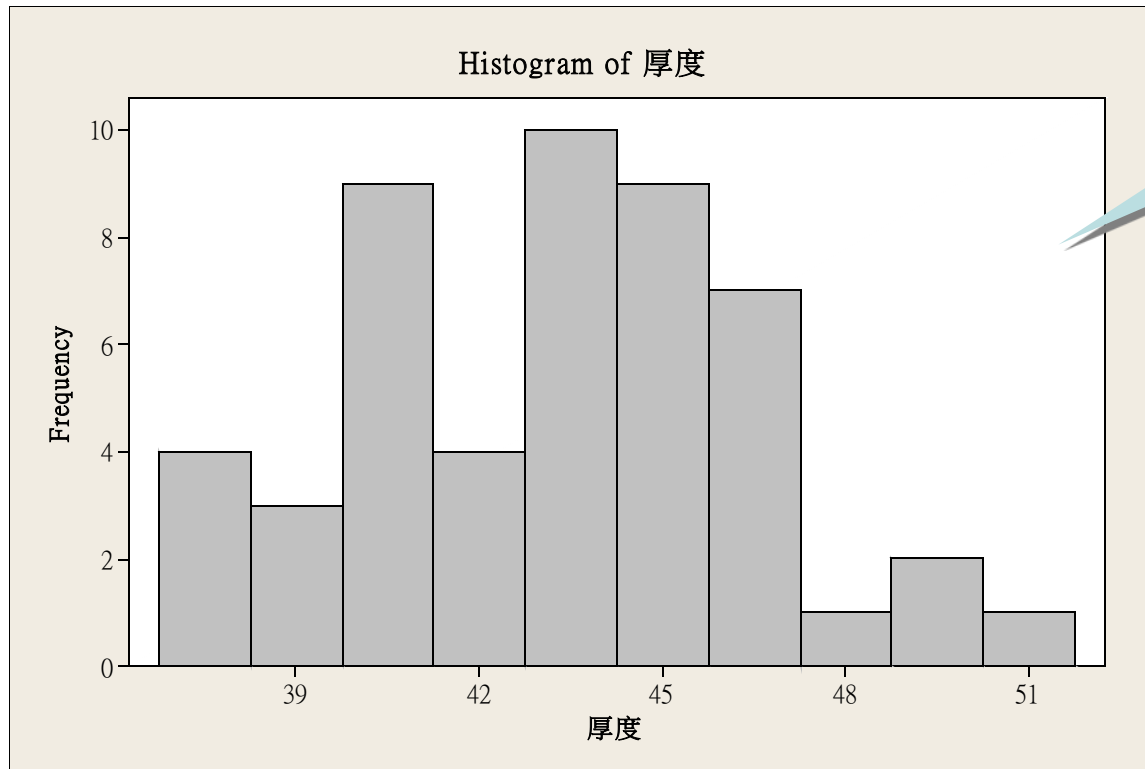


直方圖



確定欄位資料
後，點選OK。

直方圖

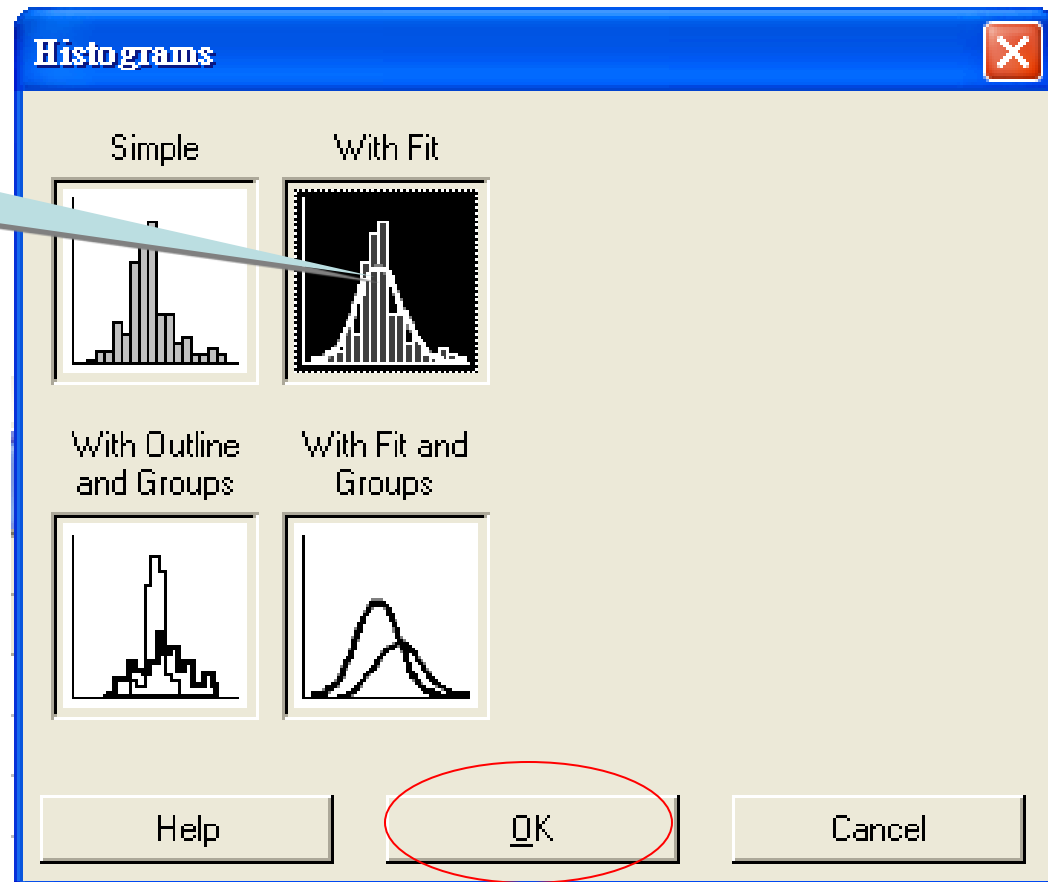


此為直方圖分佈結果。

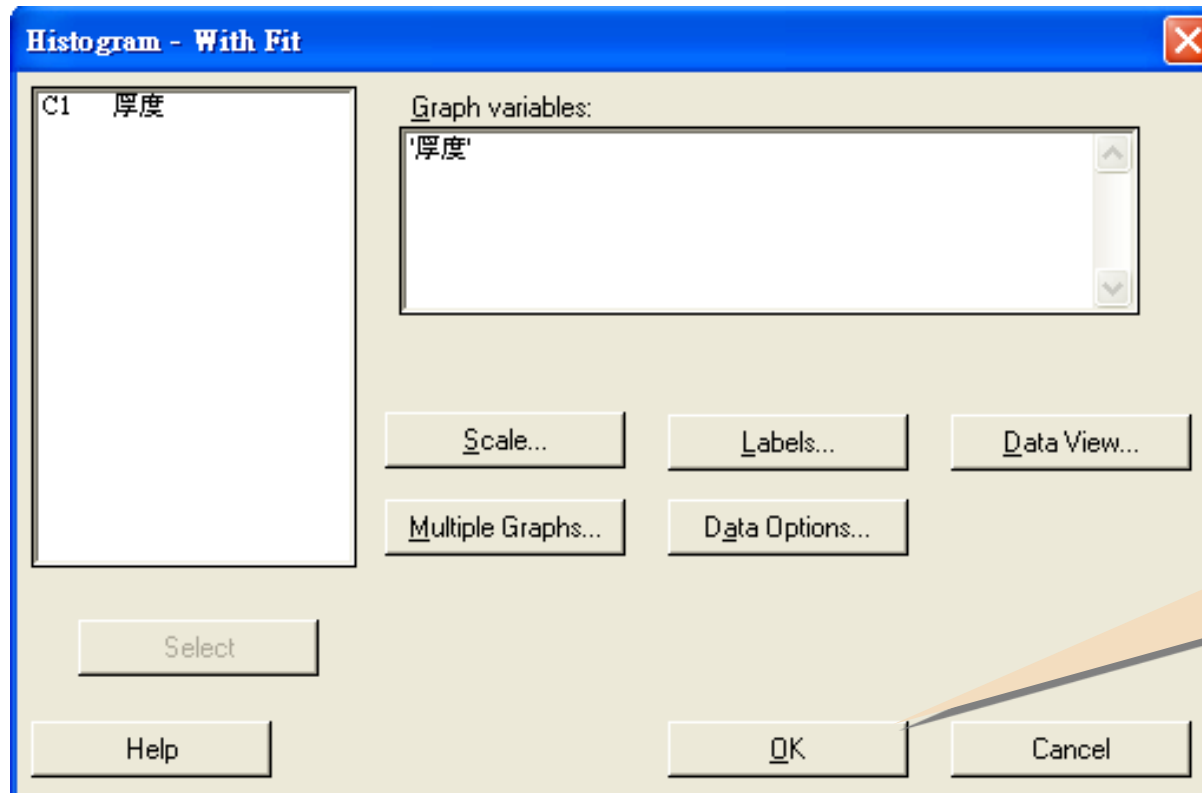
從此圖可初步看出零件厚度大多分布在40~47mm之間居多，而體重在50mm以上者只有少數。

直方圖

若點選此功能後，點選OK。

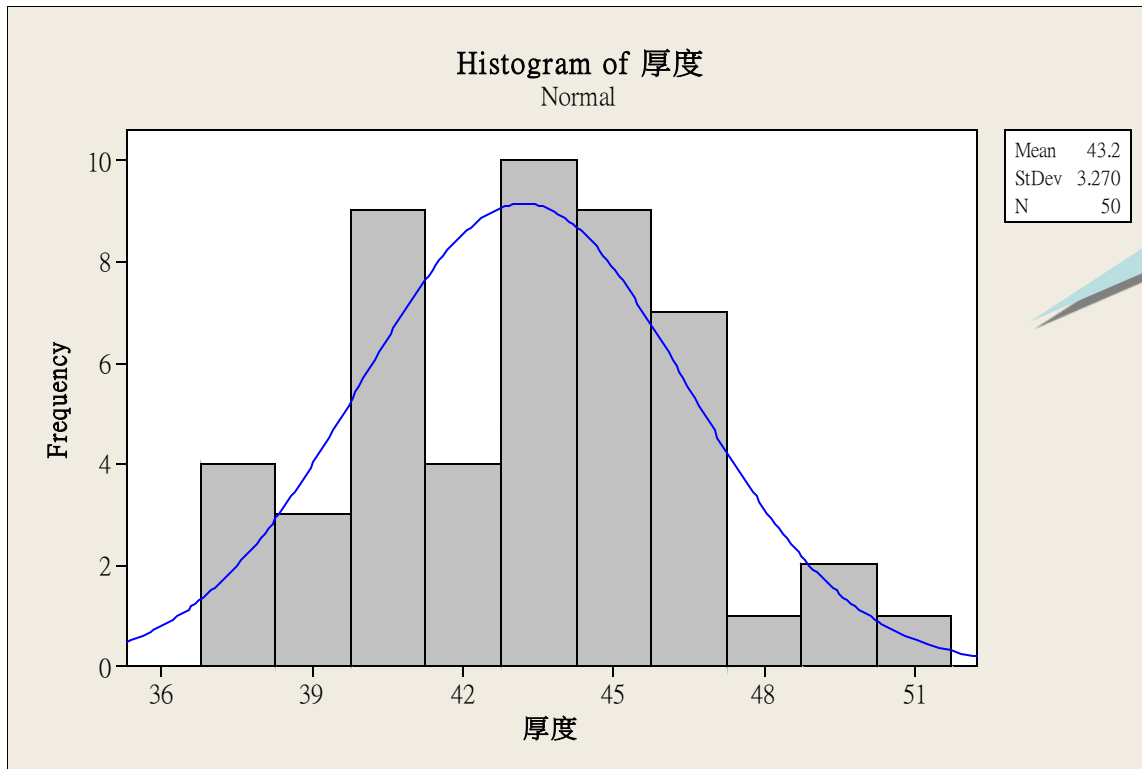


直方圖



確定欄位資料
後，點選OK。

直方圖



此為直方圖分佈結果。

從此圖可初步看出零件厚度大多分布在40~47mm之間居多，而體重在50mm以上者只有少數。有一條常態曲線可看出其直方圖符合常態的假設。

莖葉圖

- 結合數字與圖形的統計資料之表達圖示，其外觀與橫式直方圖類似，但比橫式直方圖能夠提供更多統計資料的特性。
- 例: 以下為某工廠抽樣70批產品中所發生的不良品個數，其資料如下

20	58	31	64	32	59	76	30	50	43
38	29	65	42	53	55	33	55	75	33
51	37	50	46	75	34	40	50	47	69
40	51	35	25	54	47	60	48	50	40
45	41	53	36	60	25	44	71	42	45
65	56	30	67	55	57	63	50	70	62
49	52	66	44	35	77	58	25	49	50

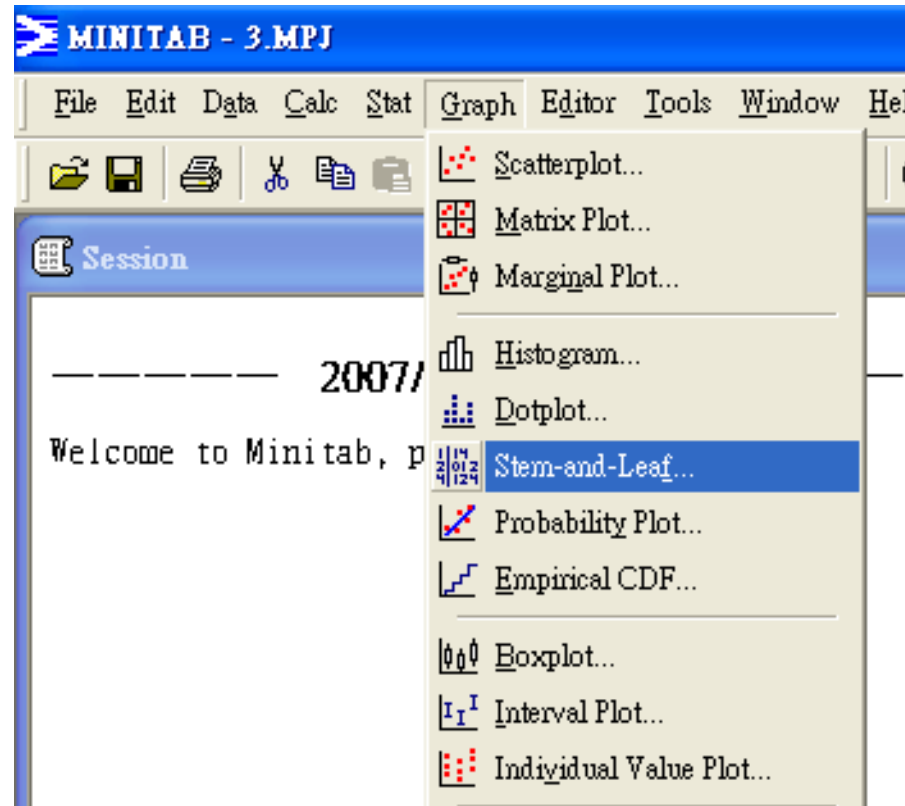
莖葉圖

將七十批產品中不良品個數的資料輸入此欄位。

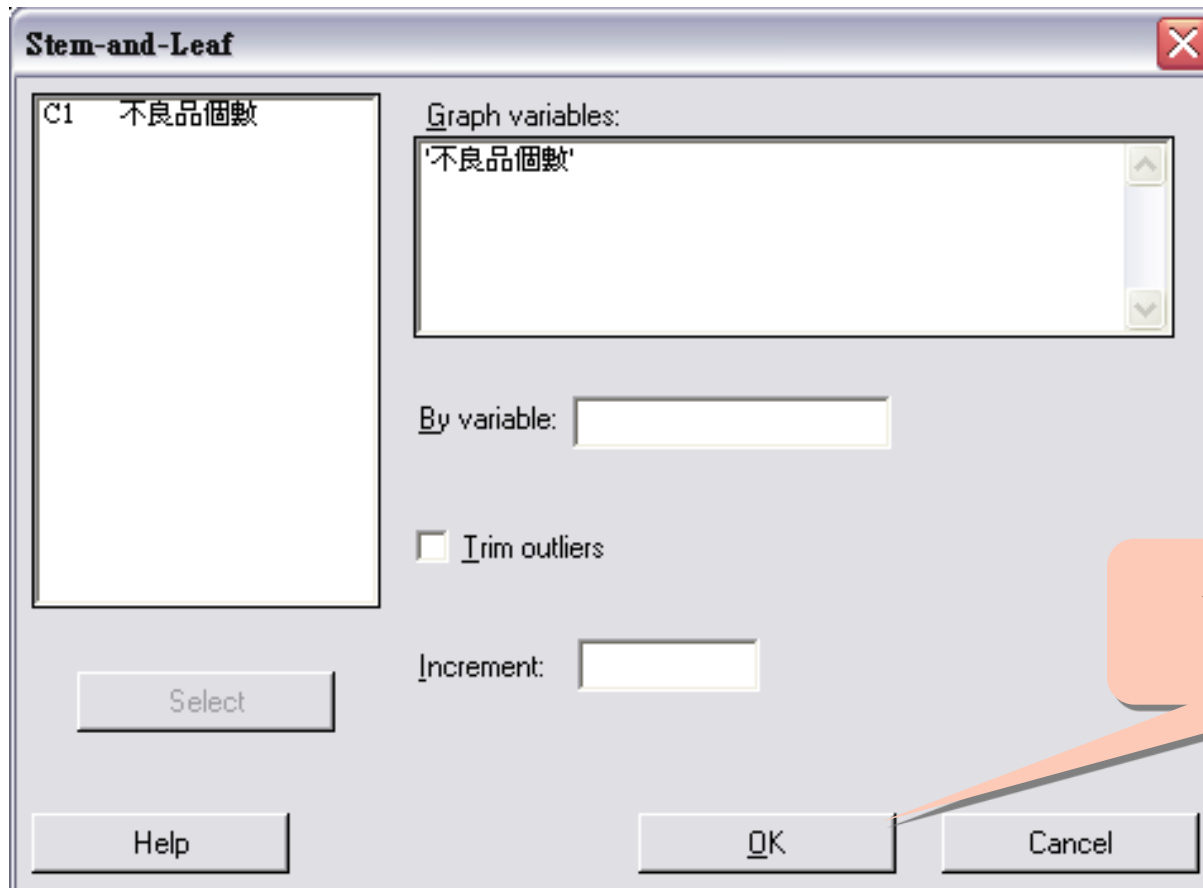
↓	C1	C2
	不良品個數	
1	20	
2	58	
3	31	
4	64	
5	32	
6	59	
7	76	
8	30	
9	50	
10	43	
11	38	
12	29	

莖葉圖

點選路徑 Graph > Stem-and-Leaf



莖葉圖



確定欄位資料
後，點選OK。

莖葉圖

從此圖可初步看出不良品個數大多分布在40~60個之間，而其中低於20個與超過70個的不良品極少。右圖中(12)所代表的是中位數所在位置(中位數為50)。

此為莖葉圖分佈結果。



盒鬚圖

- 又稱為箱型圖，在繪製上首先必須計算出全部資料的中位數、第一四分位數與第三四分位數，然後後出一盒狀長條。
- 例：某工廠製造銅絲所蒐集到的24條樣本資料如下：

0.76	0.74	0.45	0.80	0.95	0.84	0.82	0.78
0.82	0.89	0.75	0.81	0.85	0.75	0.89	0.76
0.89	0.99	0.71	0.77	0.55	0.85	0.77	0.87

銅絲直徑:公分

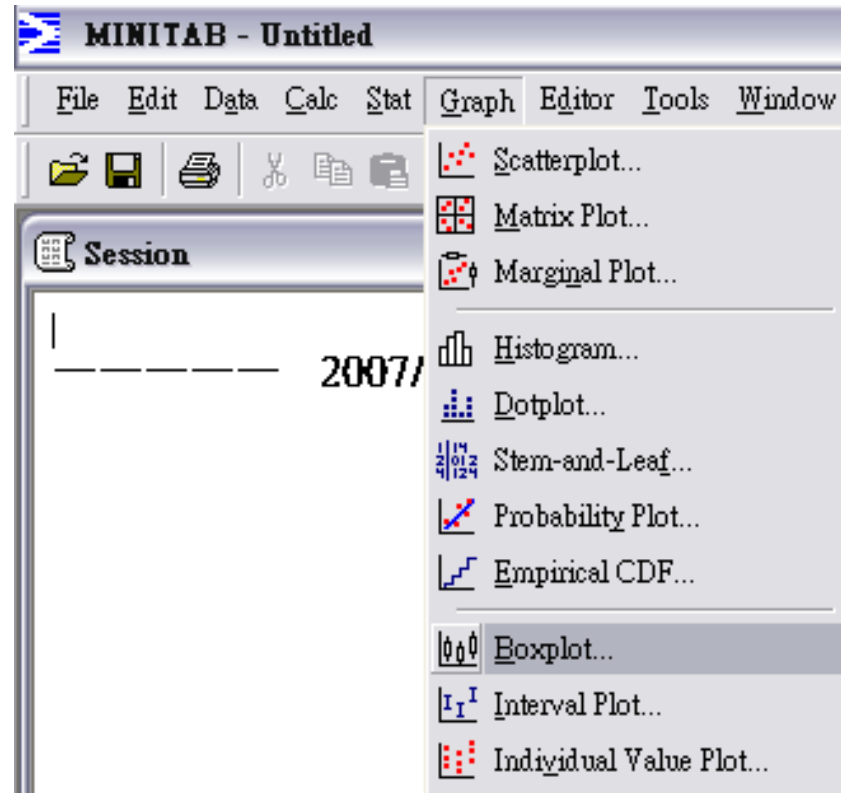
盒鬚圖

將二十四條銅絲直徑的資料輸入此欄位。

+	C1	C2	C3
	銅絲直徑		
1	0.76		
2	0.74		
3	0.45		
4	0.80		
5	0.95		
6	0.84		
7	0.82		
8	0.78		
9	0.82		
10	0.89		
11	0.75		
12	0.81		
13	0.85		

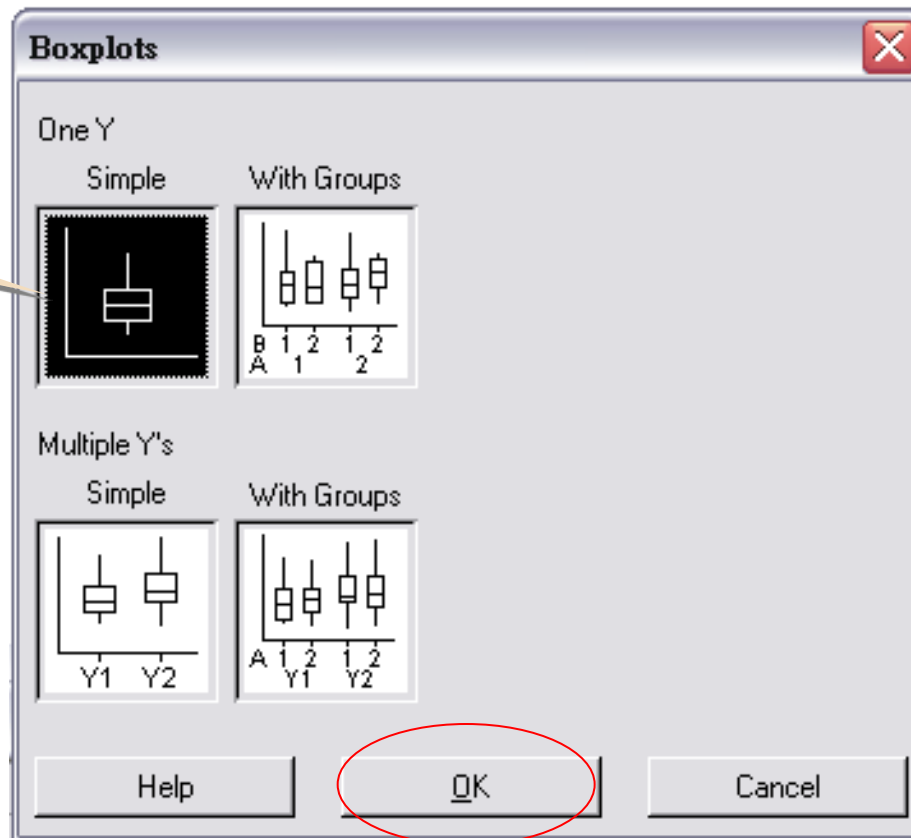
盒鬚圖

點選路徑 Graph > Boxplot

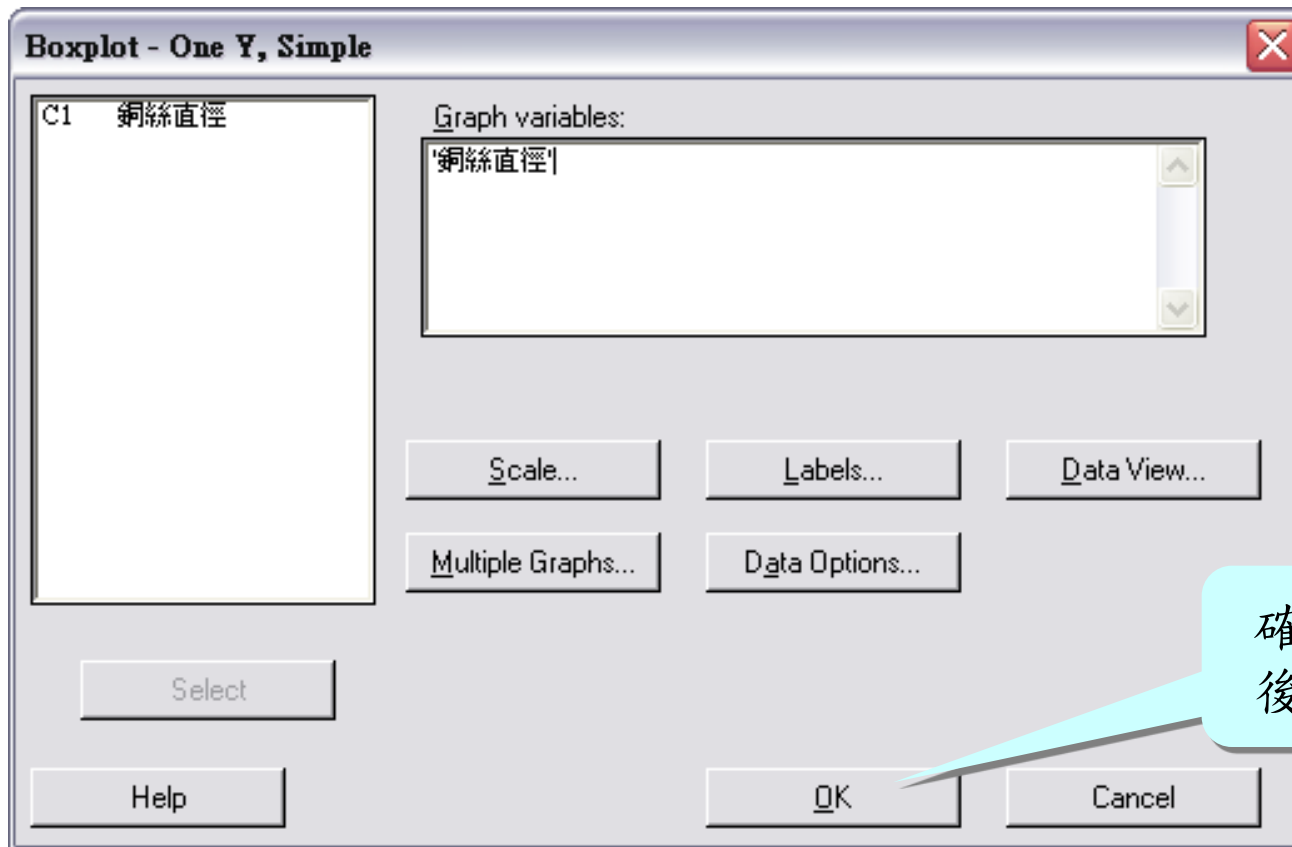


盒鬚圖

點選此功能後，
點選OK。



盒鬚圖

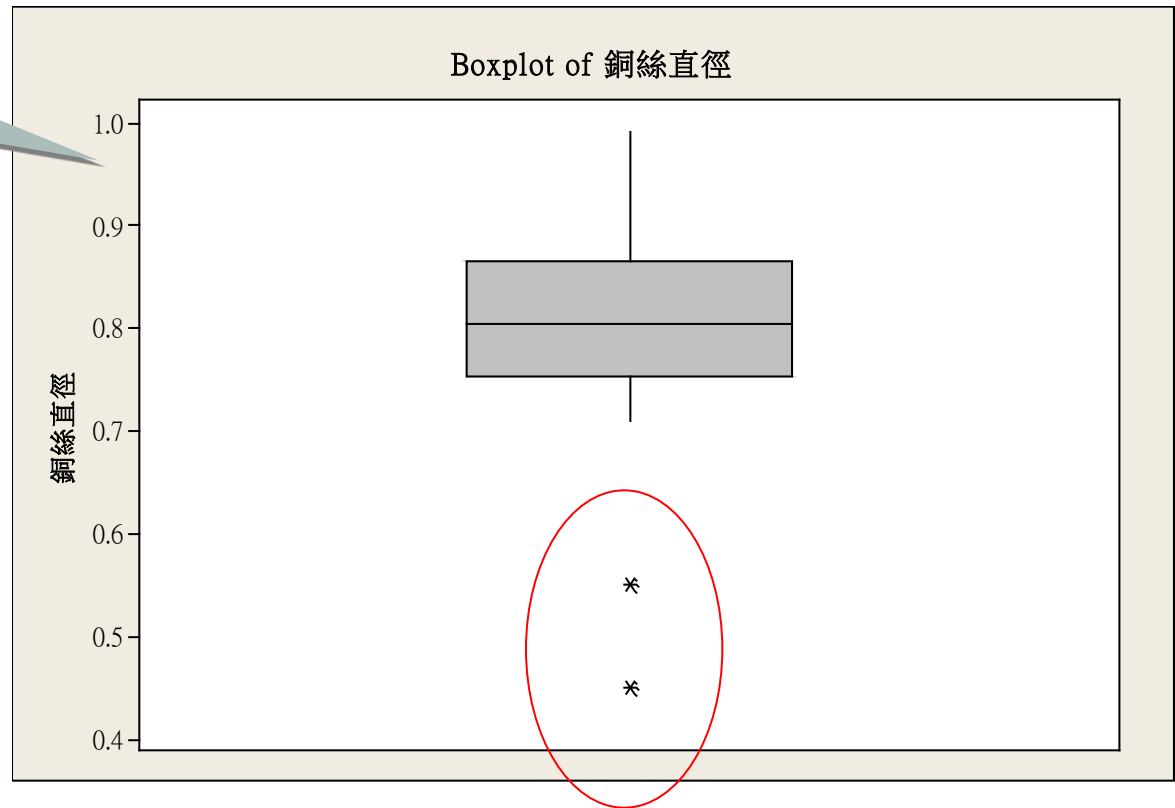


確定欄位資料
後，點選OK。

盒鬚圖

此為盒鬚圖的分佈結果。

- 從圖中可看出其分佈之情況，而紅色框起來的則為離群值(Outlier)，此為資料中有問題之數據，需進一步探討其發生的原因。



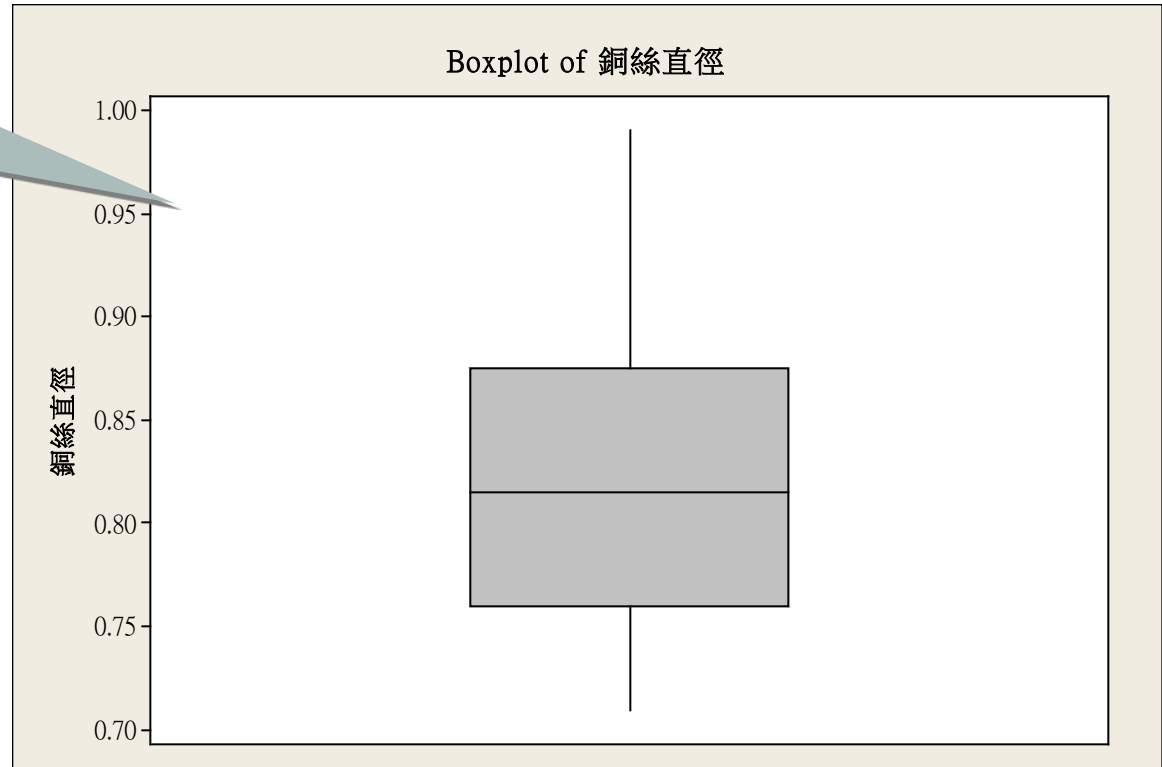
盒鬚圖

- 將離群值(0.45、0.55)去除後，重新製作盒鬚圖。

0.76	0.74	0.45	0.80	0.95	0.84	0.82	0.78
0.82	0.89	0.75	0.81	0.85	0.75	0.89	0.76
0.89	0.99	0.71	0.77	0.55	0.85	0.77	0.87

盒鬚圖

此為修正後盒鬚圖
的分佈結果。



數據特徵值的計算

集中趨勢量數

- 研究一組量測數據時，最重要的步驟之一為求出中心值的位置，常見的有樣本平均數、樣本中位數與樣本眾數。
- 樣本平均數: 平均數是代表一組數據的中心（重心）。
- 例: 一食品檢驗員檢驗七個某品牌鮪魚罐頭的隨機樣本，以測定外來不潔物質的百分比，其資料如下:

1.8	2.1	1.7	1.6	0.9	2.7	1.8
-----	-----	-----	-----	-----	-----	-----

不潔物質百分比: %

平均數

- 定義: 如果 x_1, x_2, \dots, x_n 代表大小為 n 的一個隨機樣本，則其養本平均值定義為

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- 此範例其計算公式如下:

$$\bar{x} = \frac{1.8 + 2.1 + 1.7 + 1.6 + 0.9 + 2.7 + 1.8}{7} = 1.8(\%)$$

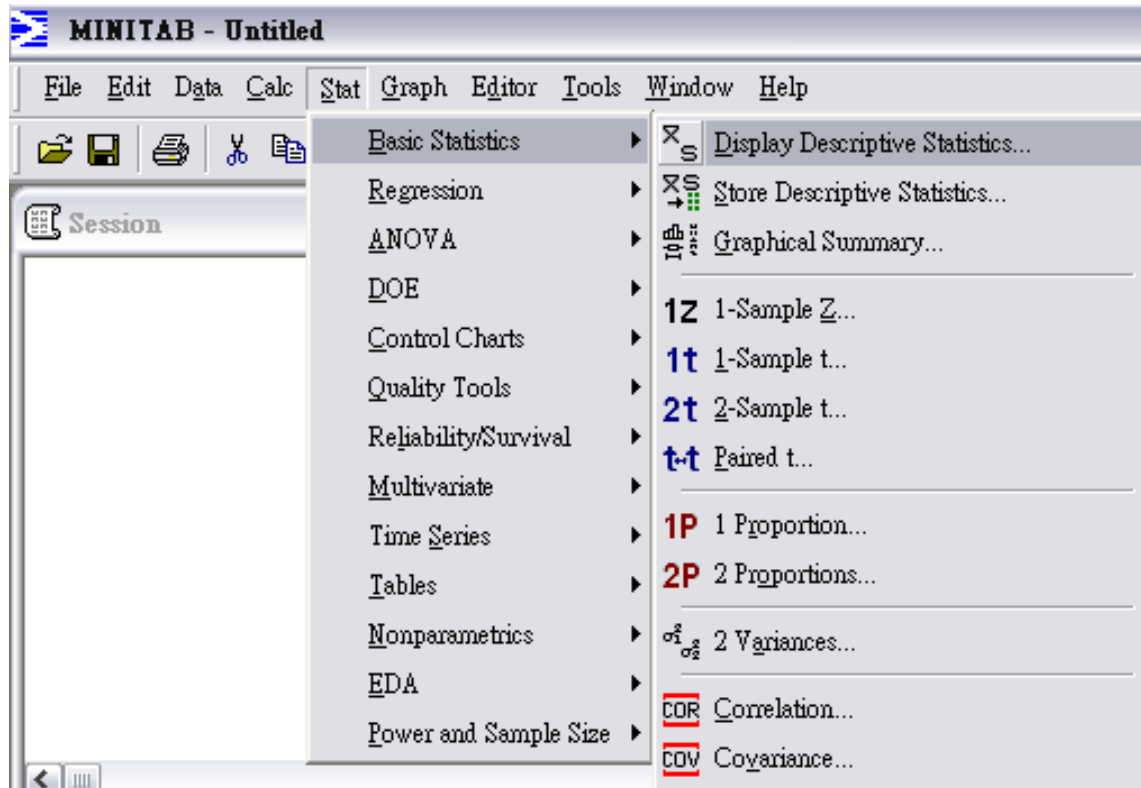
平均數

將七筆不潔物質百分比的資料輸入此欄位。

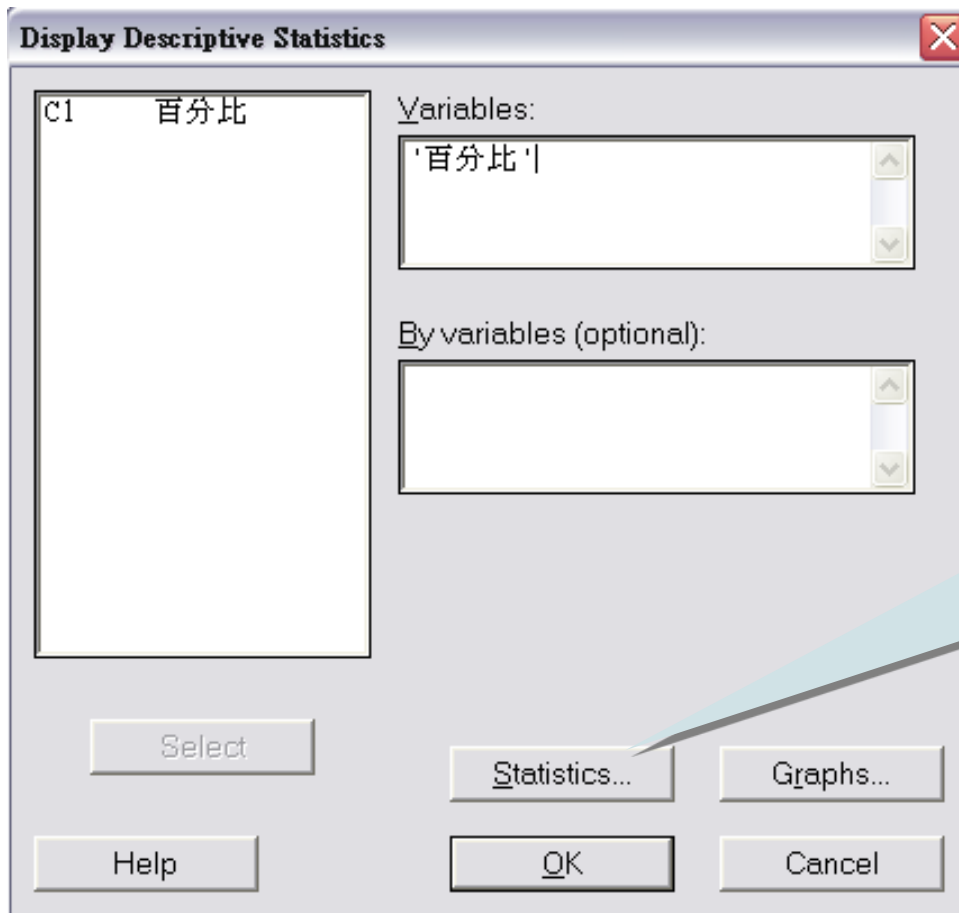
+	C1	C2
	百分比	
1	1.8	
2	2.1	
3	1.7	
4	1.6	
5	0.9	
6	2.7	
7	1.8	
8		

平均數

點選路徑 Stat > Basic Statistics > Display Descriptive Statistics



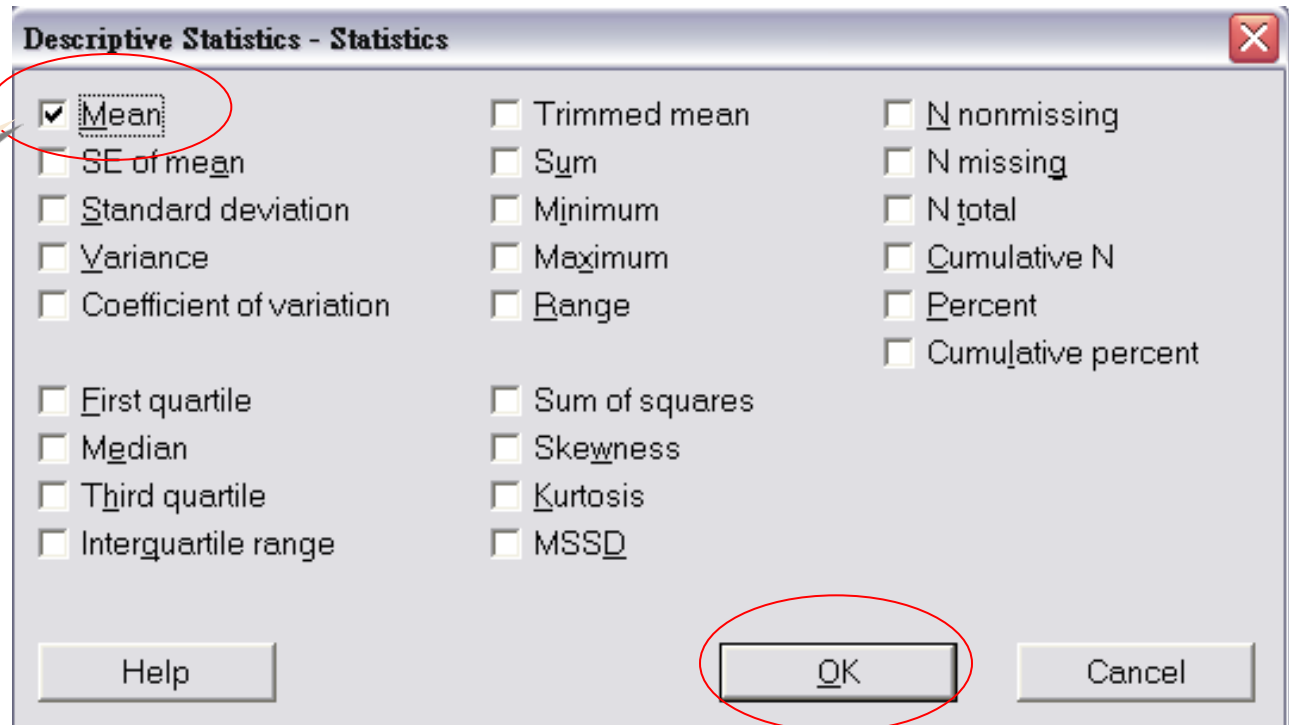
平均數



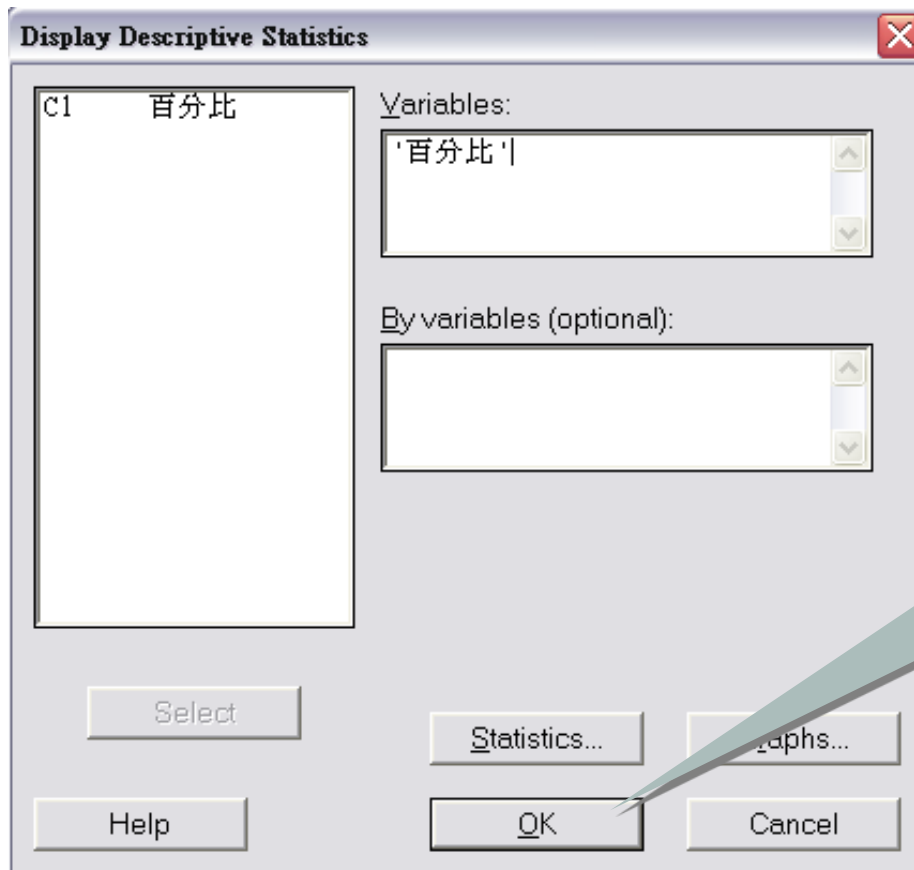
確定欄位資料後，點選此功能進入點選所要的統計量。

平均數

點選所需要的統計量後，按OK。



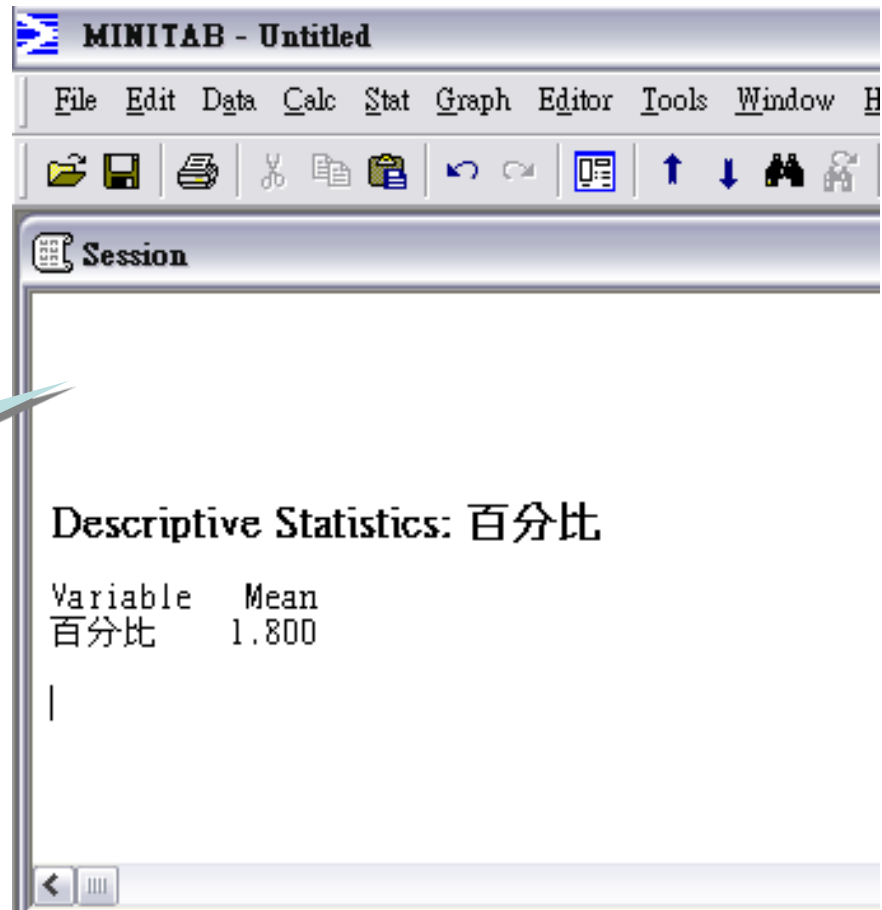
平均數



回到此畫面，確定欄位資料後，點選OK。

平均數

在此視窗內將顯示所需要統計量的計算結果。



MINITAB - Untitled

File Edit Data Calc Stat Graph Editor Tools Window H

Session

Descriptive Statistics: 百分比

Variable	Mean
百分比	1.800

中位數

- 定義: 如果 x_1, x_2, \dots, x_n 代表大小為 n 的一個隨機樣本，且依大小遞增的順序排列，則其樣本中位數為

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{如果 } n \text{ 是偶數} \\ \frac{x_{n/2} + x_{n/2+1}}{2} & \text{如果 } n \text{ 是奇數} \end{cases}$$

- 例: 某工廠製造香煙，其中六根香煙的隨機樣本中尼古丁含量資料如下:

2.3	2.7	2.5	2.9	3.1	1.9
-----	-----	-----	-----	-----	-----

中位數

- 如果將尼古丁含量以遞增的順序排列，則資料為

1.9	2.3	2.5	2.7	2.9	3.1
-----	-----	-----	-----	-----	-----

- 中位數為2.5與2.7的平均值，

$$\tilde{x} = \frac{2.5 + 2.7}{2} = 2.6$$

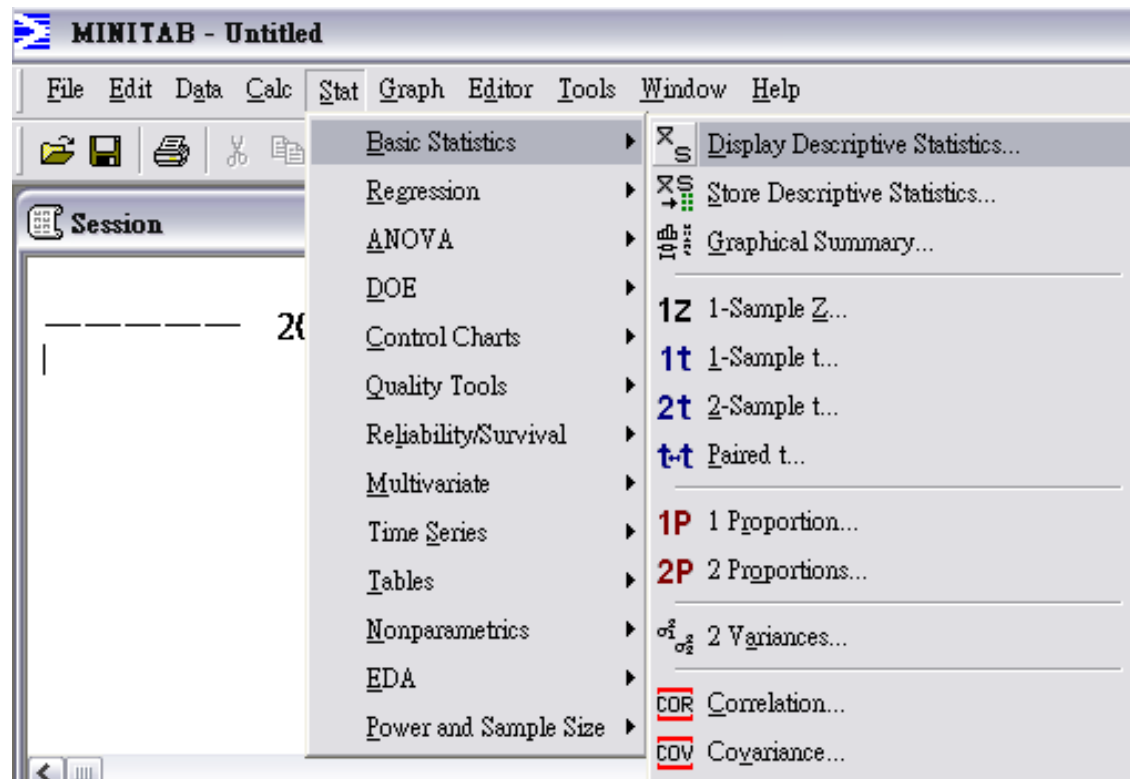
中位數

將六根香煙古丁含量的
資料輸入此欄位。

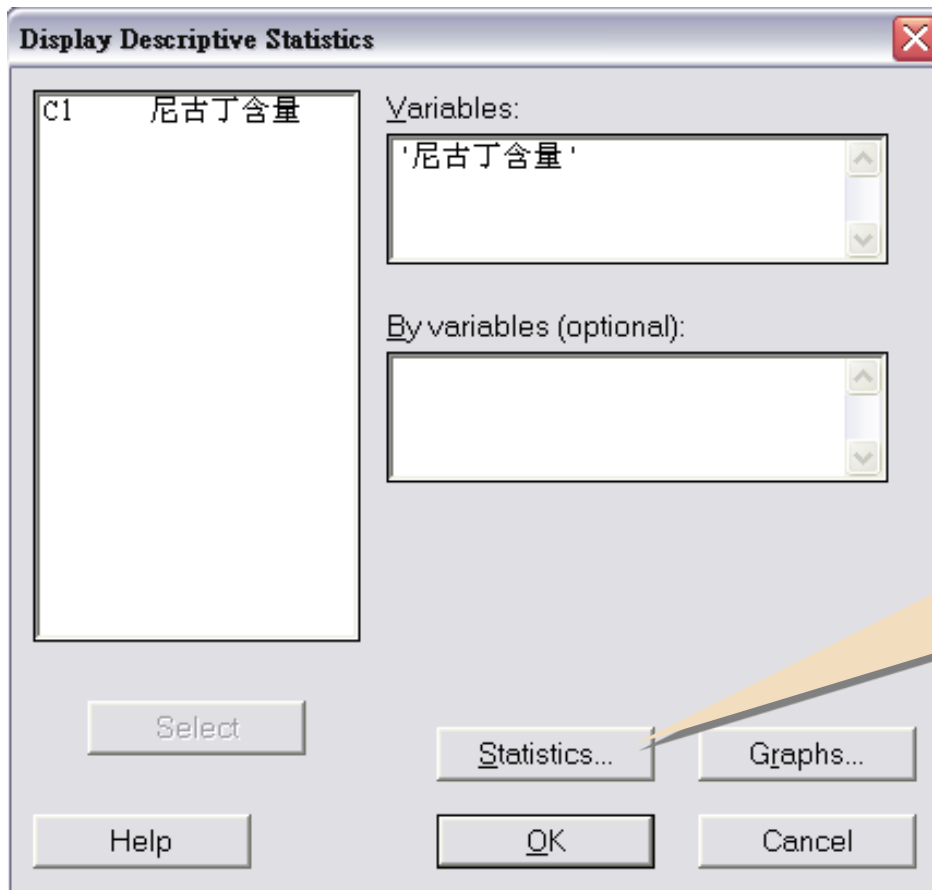
↓	C1	C2	C3
	尼古丁含量		
1	2.3		
2	2.7		
3	2.5		
4	2.9		
5	3.1		
6	1.9		
7			
8			

中位數

點選路徑 Stat > Basic Statistics > Display Descriptive Statistics



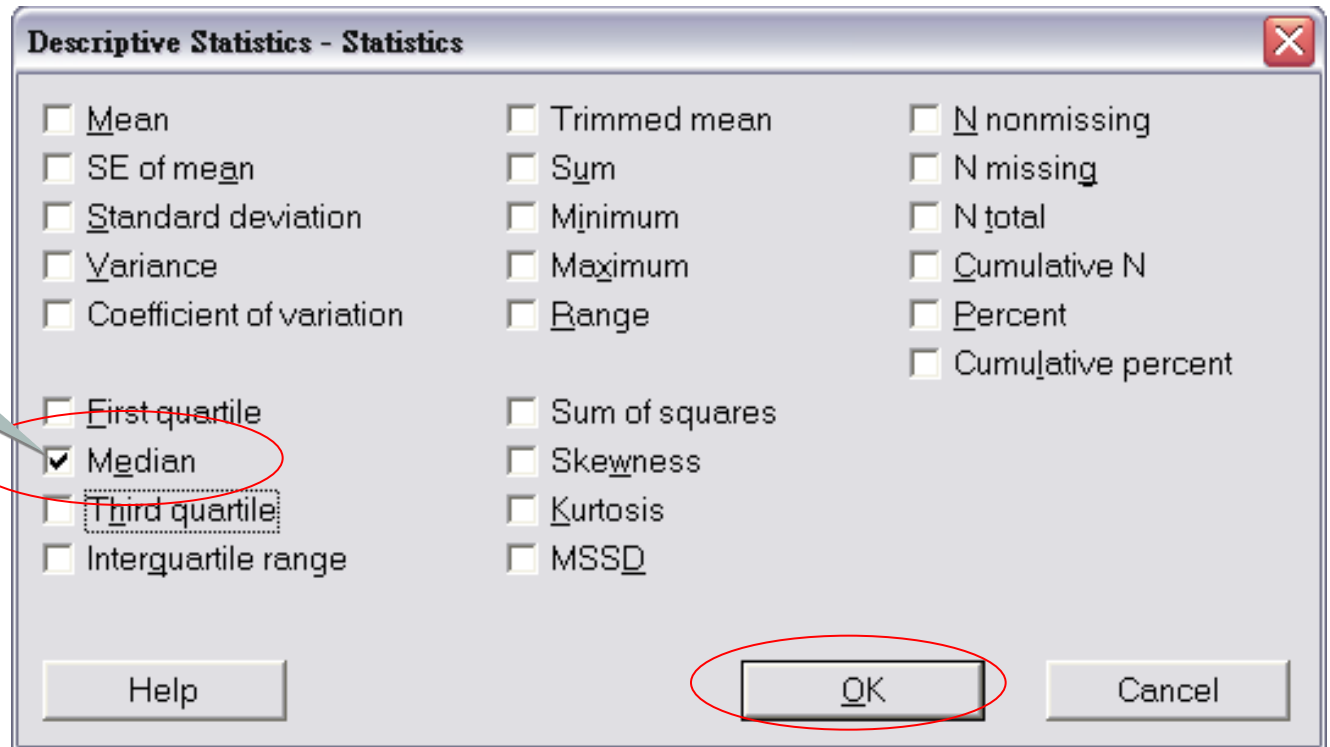
中位數



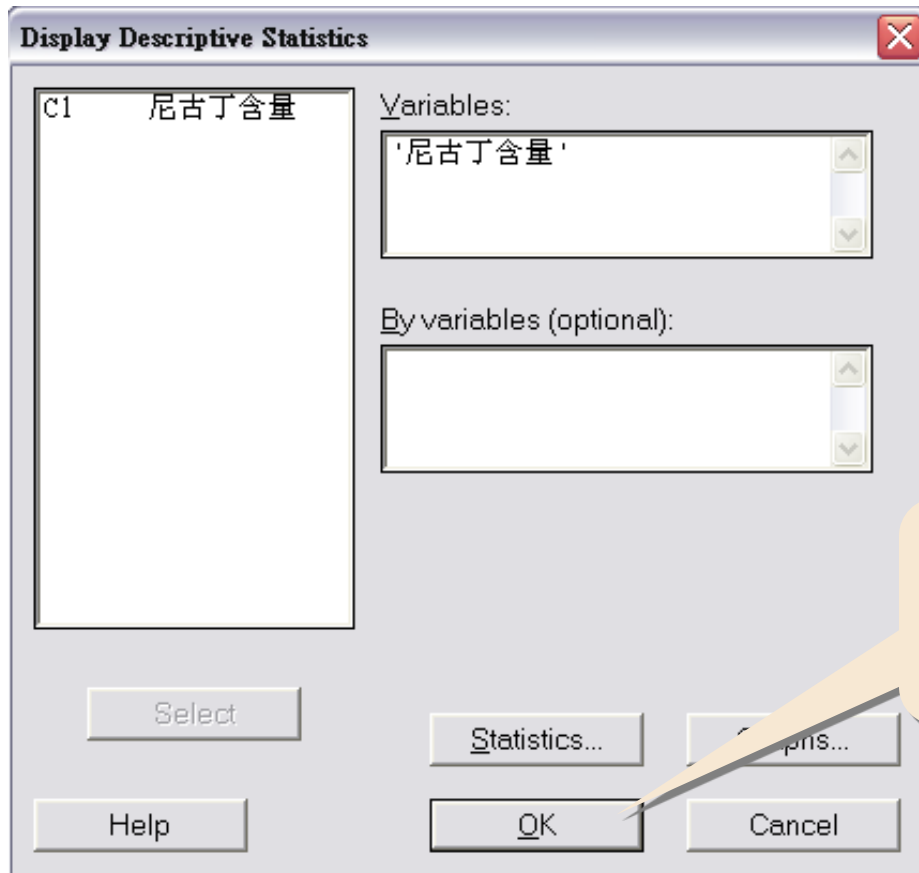
確定欄位資料後，點選此功能進入點選所要的統計量。

中位數

點選所需要的統計量後，按OK。



中位數



中位數



Session

----- 2007/2/19 21:28:28 -----

Results for: S.MTW

Descriptive Statistics: 尼古丁含量

Variable	Median
尼古丁含量	2.600

在此視窗內將顯示所需要統計量的計算結果。

眾數

- 定義: x_1, x_2, \dots, x_n (不需全部不同), 代表大小為 n 的隨機樣本, 則眾數 M 為最常發生或發生頻率最大的樣本值, 眾數可以不存在, 且當其存在時不需要是唯一的。
- 例: 1, 2, 3, 4, 5 以上五筆資料中, 無眾數存在。
- 例: 1, 2, 2, 2, 3, 4, 5, 5, 5 以上九筆資料中, 眾數為2和5。

眾數

- 例: 某碳酸飲料的飲用期限，隨機選出10罐樣本作測試，其資料如下:

108	124	123	124	124
134	139	123	159	163

飲料期限單位: 天

- 從以上資料可看出124天為此資料的眾數。

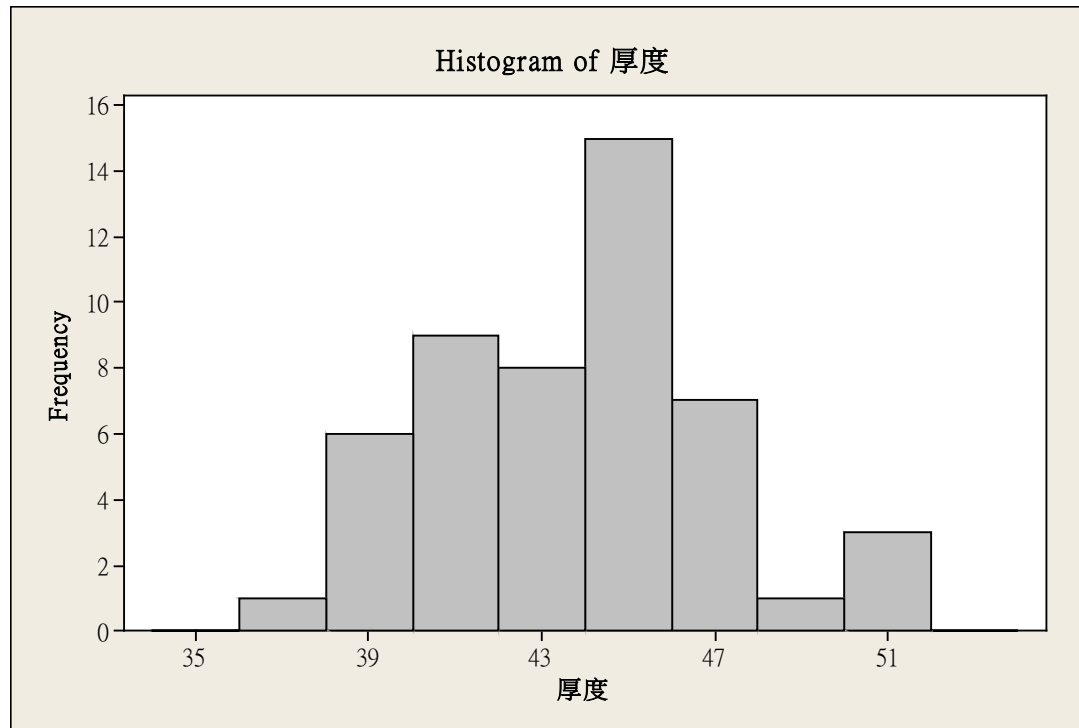
結論

	優點	缺點
平均數	1. 定義明確 2. 簡單易懂 3. 計算容易 4. 感應靈敏 5. 適合代數運算 6. 受抽樣方法變動的影響小	1. 易受極端值的影響 2. 有敞開組時無法求出
中位數	1. 定義明確 2. 簡單易懂 3. 計算容易 4. 受抽樣方法變動的影響小 5. 有敞開組時可以求出 6. 量化或質化資料均可求得	1. 感應較不靈敏 2. 不適合代數運算
眾數	1. 簡單易懂 2. 不受極端值影響 3. 有敞開組時可以求出 4. 適用質化的資料 5. 近似的眾數值易求	1. 感應較不靈敏 2. 不適合代數運算 3. 受組數與組距的影響過大 4. 當次數分配無顯著集中時， 沒有代表性。

➤ 註: 所謂敞開組指的是資料為一區間的數據，例如: 天氣溫度資料可能為 24~27°C，或資料量大時，先用組距做組別的分類。

結論

- 例: 此為敞開組的資料(直方圖), 可初步看出中位數在44~46這一組內, 而眾數也在44~46這一組內, 但無法得知所有原始數據的值, 所以平均數無法計算。



離中趨勢量數

- 表示數據組內各數散佈的情形，常見的有全距 (Range)，變異數 (Variance) 與標準差 (Standard deviation)。
- 全距: 係指一群資料中最大值與最小值之差的量數，其計算公式為:

$$R = X_{\max} - X_{\min} \quad \text{其中 } X_{\max} = \text{一群數值中之最大值}$$
$$X_{\min} = \text{一群數值中之最小值}$$

全距

➤ 例: 某檢驗員抽取5件產品，其重量資料如下:

60	70	52	68	82
----	----	----	----	----

單位: 公克

➤ 計算公式如下:

$$82 - 50 = 30(\text{公克})$$

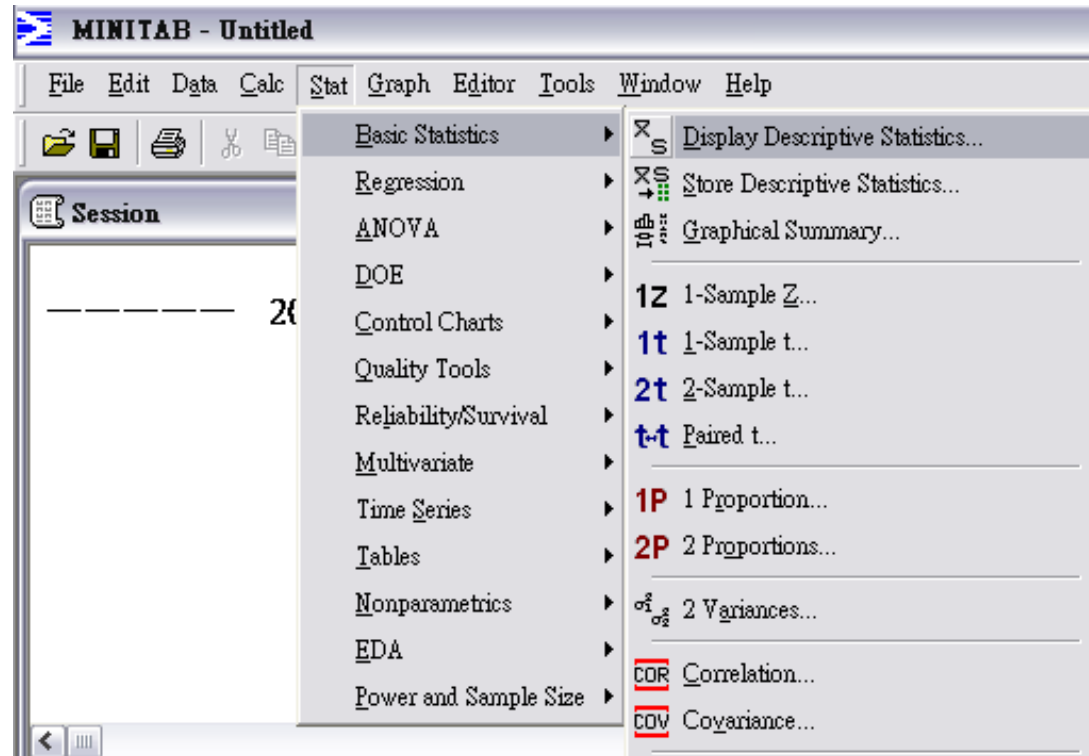
全距

將5件產品重量的資料輸入此欄位。

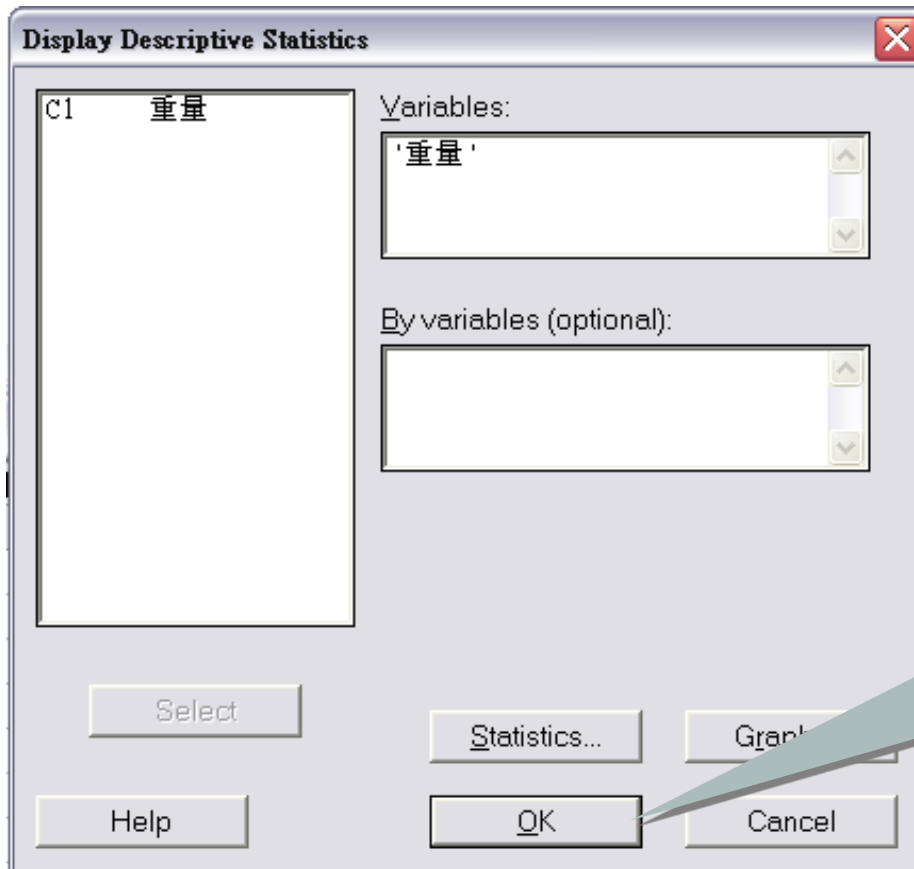
+	C1	C2
	重量	
1	60	
2	70	
3	52	
4	68	
5	82	
6		

全距

點選路徑 Stat > Basic Statistics > Display Descriptive Statistics



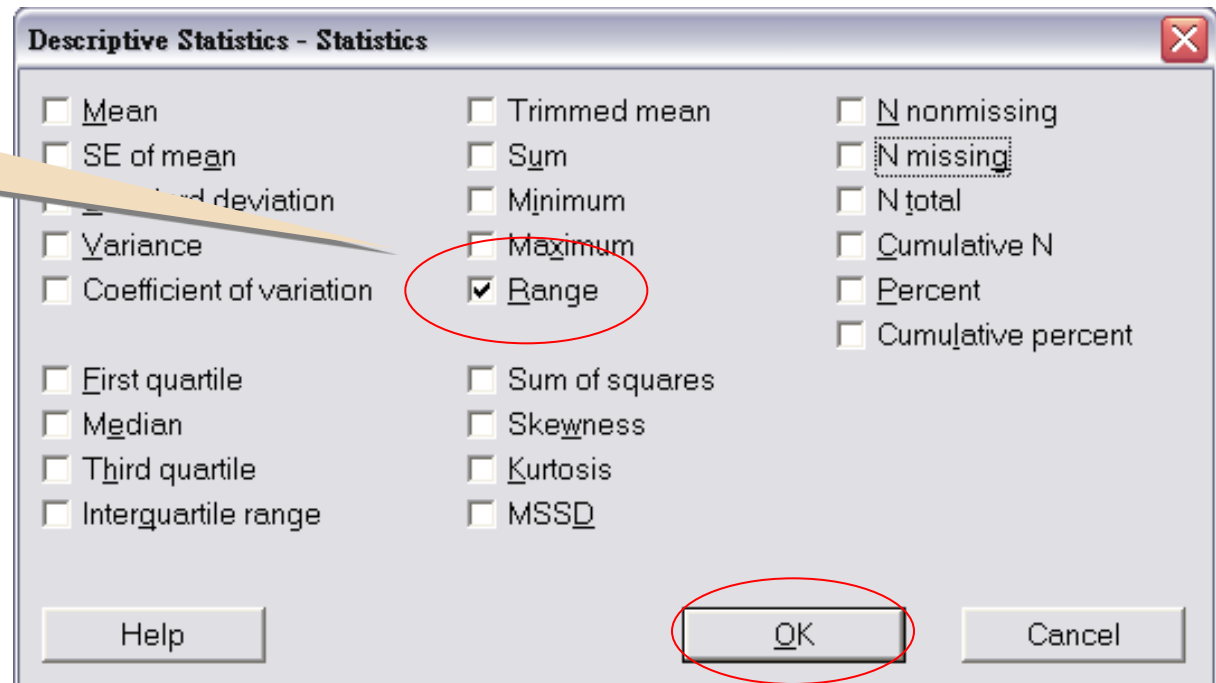
全距



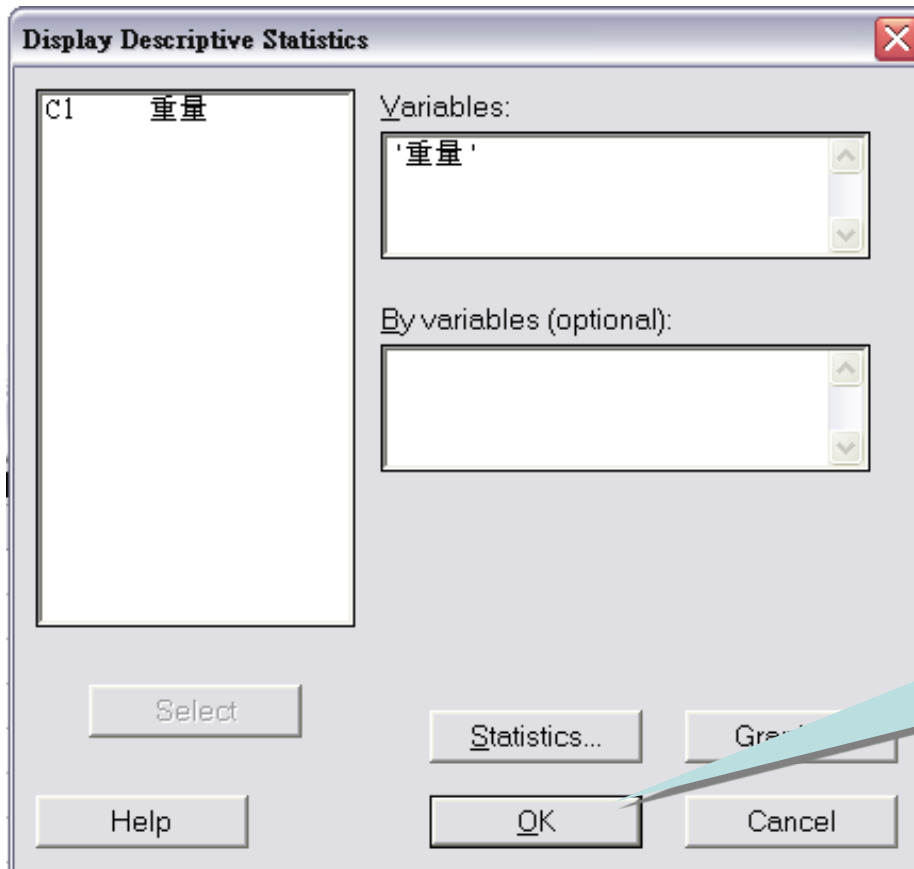
確定欄位資料後，點選此功能進入選擇所要的統計量。

全距

點選所需要的統計量後，按OK。



全距



回到此畫面，確定欄位資料後，點選OK。

全距



Session

Descriptive Statistics: 重量

Variable	Range
重量	30.00

在此視窗內將顯示所需要統計量的計算結果。

變異數

- 利用到全部資料值的離散量數，根據每一個資料值與平均數之差而求得。
- 定義：如果 x_1, x_2, \dots, x_n 代表大小為 n 的一個隨機樣本，則其變異數定義為：

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

變異數

➤ 例:有一檢驗員隨機抽取15件鎢絲熔點資料如下:

330	358	373	346	343
334	318	325	295	288
321	337	328	348	318

鎢絲熔點單位:攝氏(°C)

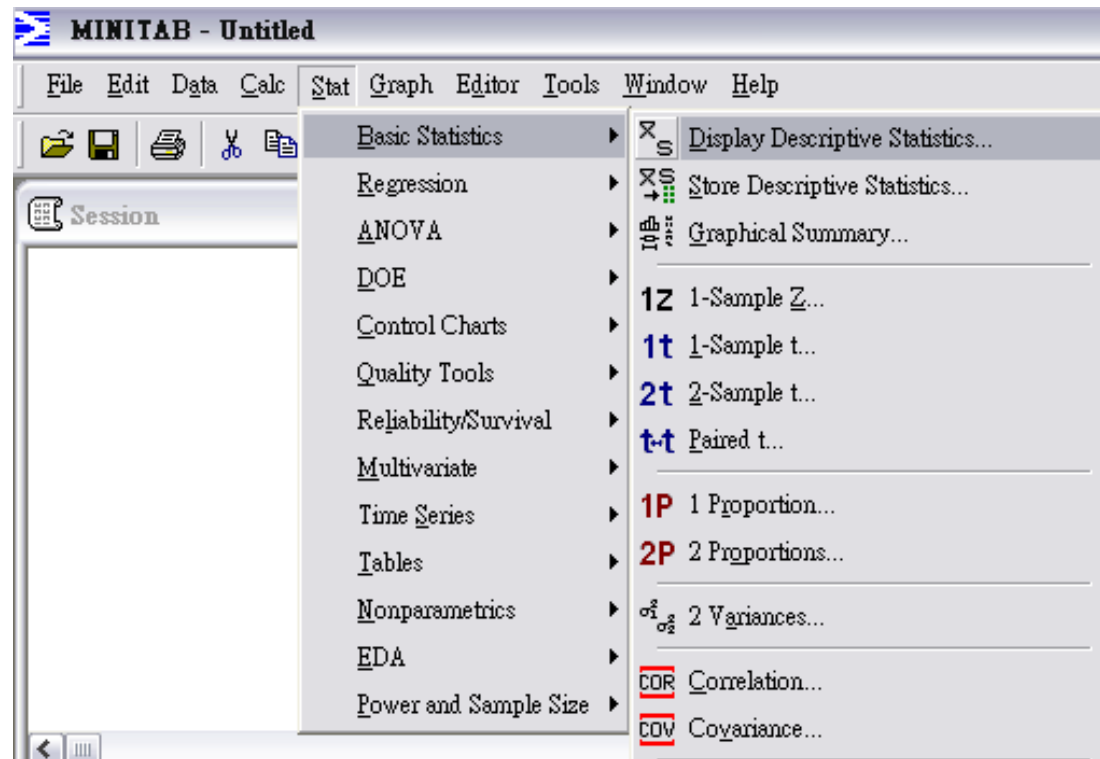
變異數

將十五筆鎢絲熔點的
資料輸入此欄位。

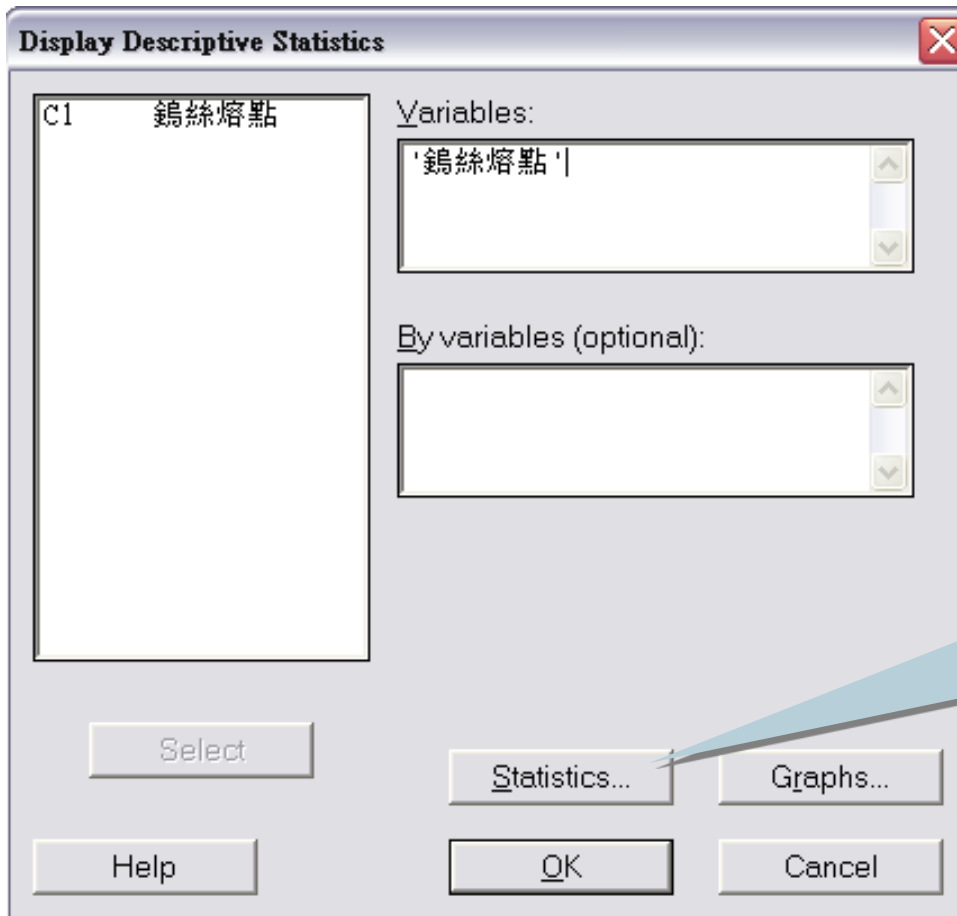
↓	C1	C2	C3
	鎢絲熔點		
1	330		
2	358		
3	373		
4	346		
5	343		
6	334		
7	318		
8	325		
9	295		
10	288		
11	321		
12	337		
13	328		
14	348		
15	318		
16			

變異數

點選路徑 Stat > Basic Statistics > Display Descriptive Statistics



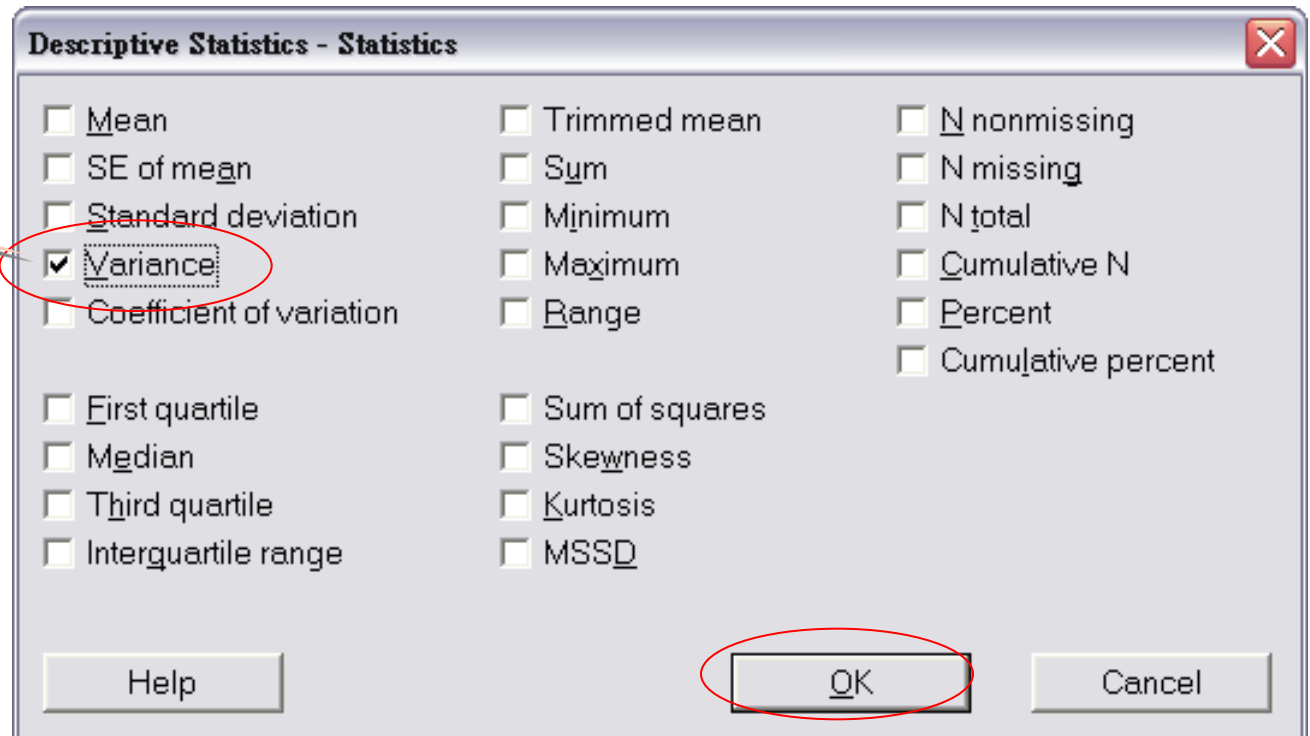
變異數



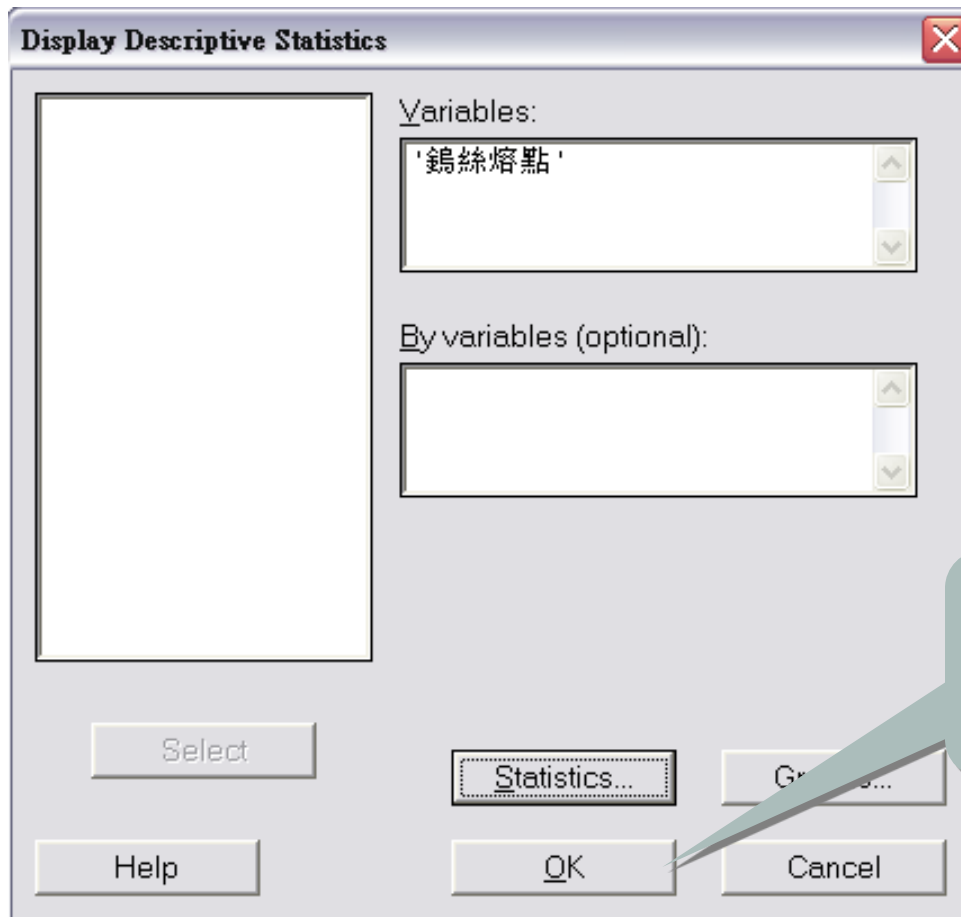
確定欄位資料後，點選此功能進入選擇所要的統計量。

變異數

點選所需要的統計量後，按OK。



變異數



回到此畫面，確定欄位資料後，點選OK

變異數

在此視窗內將顯示所需要統計量的計算結果。



Session

Results for: 7.MTW

Descriptive Statistics: 錫絲塔點

Variable	Variance
錫絲塔點	487.46

|

Navigation icons: back, list

標準差

- 定義：以 S 為代表符號的樣本標準差，此為樣本變異數的正平方根。

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

- 標準差因利用了全部資料，所以反應靈敏，並適合代數方法的計算。

標準差

- 例: 九位被實驗者，對某一刺激的反應時間的隨機樣本資料如下:

2.5	3.6	3.1	4.3	2.9	2.3	2.6	4.1	3.4
-----	-----	-----	-----	-----	-----	-----	-----	-----

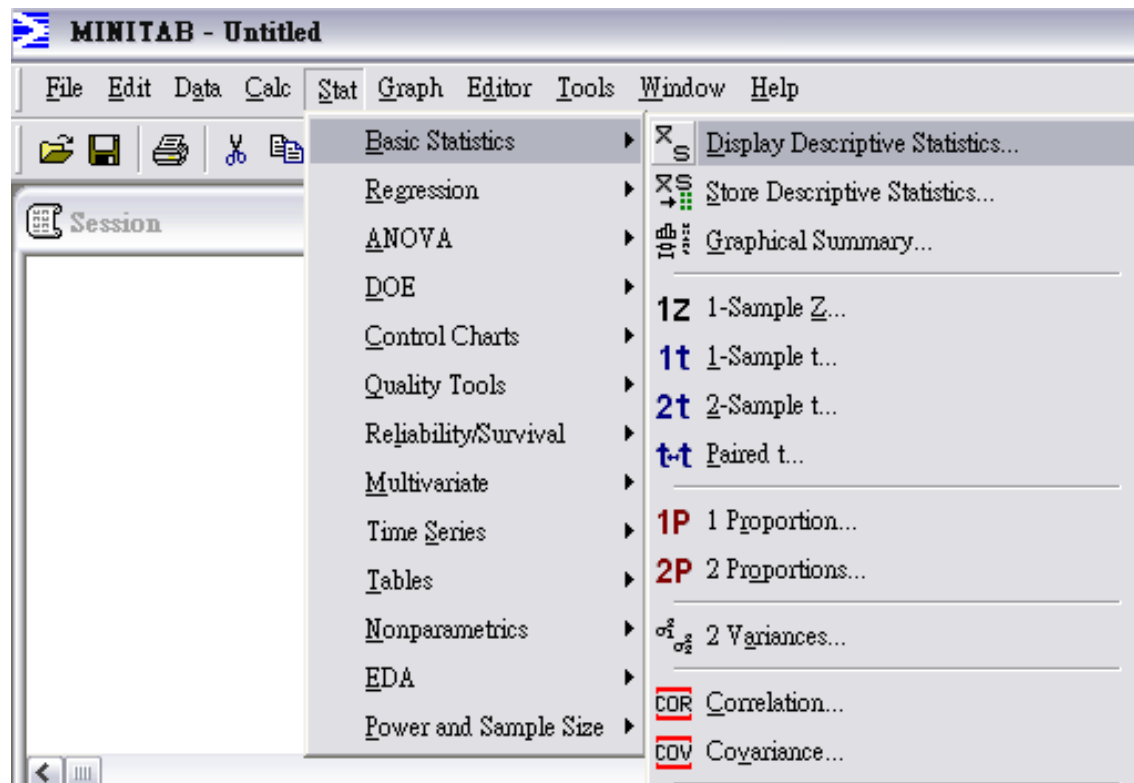
標準差

+	C1	C2	C3
	反應時間		
1	2.5		
2	3.6		
3	3.1		
4	4.3		
5	2.9		
6	2.3		
7	2.6		
8	4.1		
9	3.4		
10			

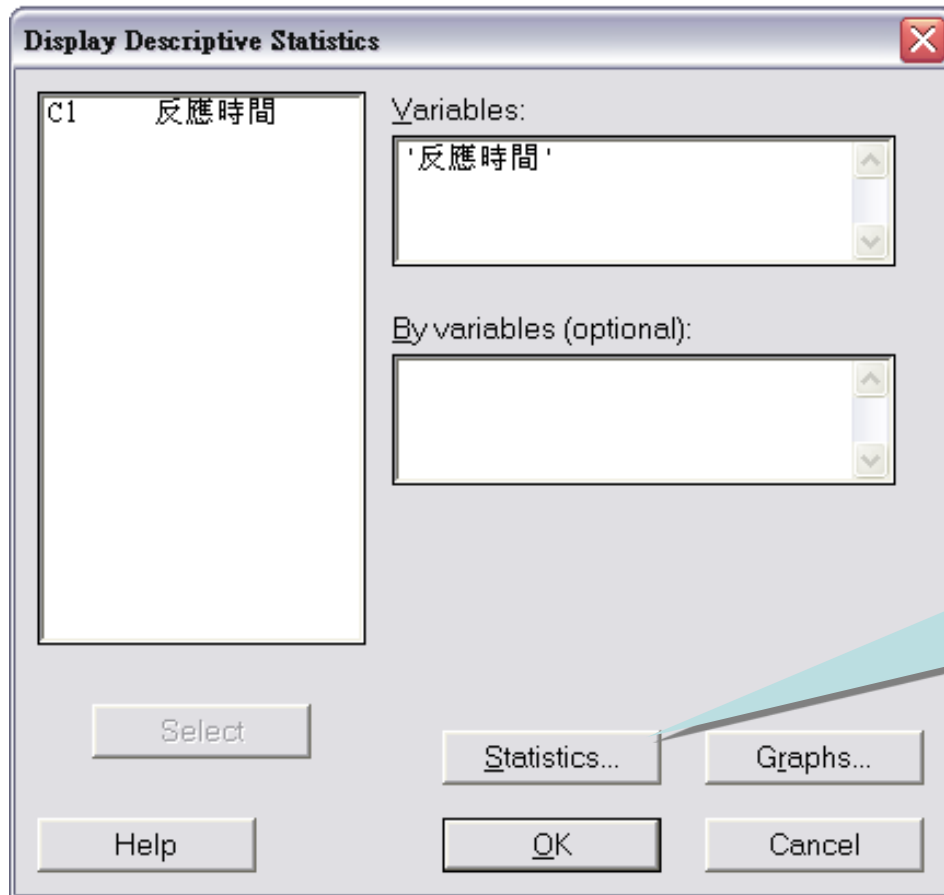
將九位被實驗者反應時間的資料輸入此欄位。

標準差

點選路徑 Stat > Basic Statistics > Display Descriptive Statistics



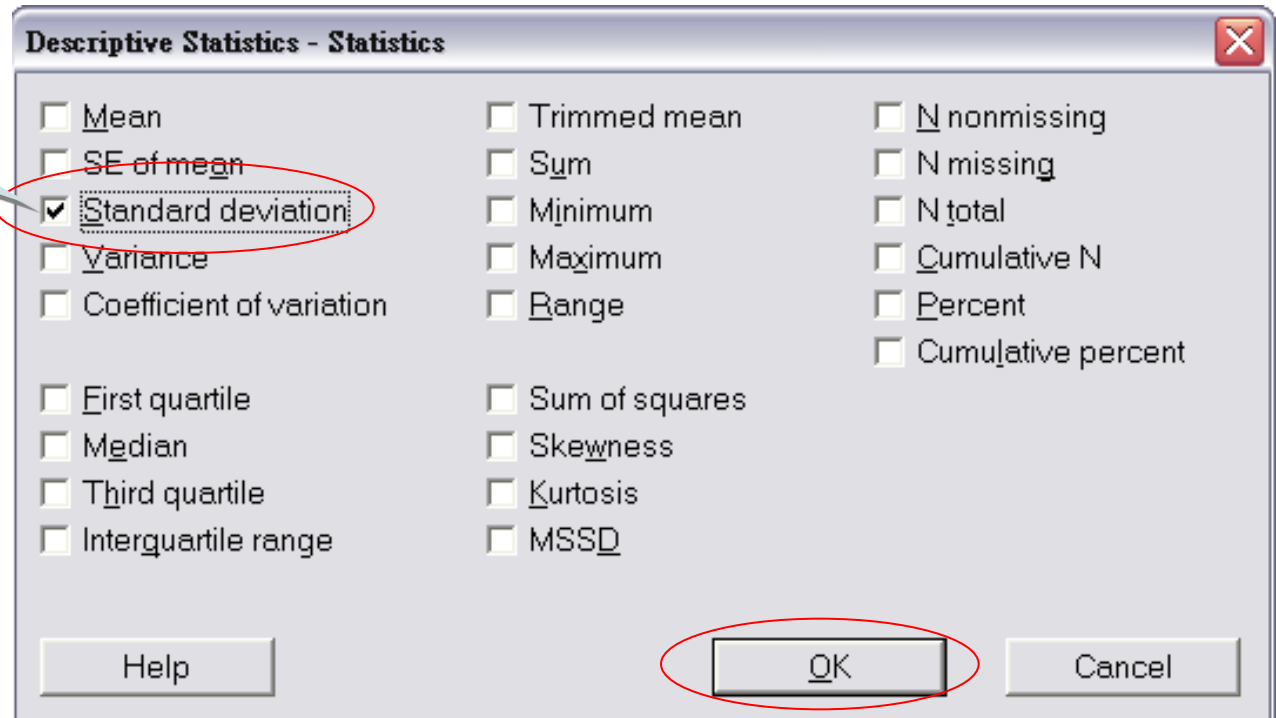
標準差



確定欄位資料後，點選此功能進入選擇所要的統計量。

標準差

點選所需要的統計量後，按OK。



標準差



在此視窗內將顯示所需要統計量的計算結果。

結論

	優點	缺點
全距	<ol style="list-style-type: none"> 1. 簡單易懂 2. 計算容易 	<ol style="list-style-type: none"> 1. 感應較不靈敏 2. 不適合代數運算 3. 受抽樣方法變動的影響大
變異數	<ol style="list-style-type: none"> 1. 定義明確 2. 感應較靈敏 3. 適合代數運算 4. 受抽樣方法變動的影響小 	<ol style="list-style-type: none"> 1. 有敞開組時，無法求出 2. 計算較複雜 3. 易受極端值影響
標準差	<ol style="list-style-type: none"> 1. 定義明確 2. 感應較靈敏 3. 適合代數運算 4. 受抽樣方法變動的影響小 	<ol style="list-style-type: none"> 1. 有敞開組時，無法求出 2. 計算較複雜 3. 易受極端值影響

註：全距可使用於管制圖中管制界限的制定。