



PEDA Training

2013/10

太引資訊 王雅莉

03-5726781 #105

yali_wang@tynesys.com



Outline

- } “ChartTool” Menu 繪圖功能選單
 - Piechart 圓餅圖
 - Barchart 長條圖
 - Pareto 柏拉圖
 - Histogram 直方圖
 - Scatter 散佈圖
 - Boxplot 盒鬚圖
- } “DataSet” Menu 資料整理選單
- } “Statistics” Menu 統計功能選單
 - Generic Report & Statistic 統計量報表
- } “WorkMap” Menu 智慧流程設計



“ChartTool” Menu 繪圖功能選單



Piechart 圓餅圖

Data File :
BloodType.csv

- } 圓餅圖，呈現每個項目對應數值佔數值總值的比例。
- } 元素：分群項目、分群個數

資料

分群項目 ↓ 分群個數 ↓

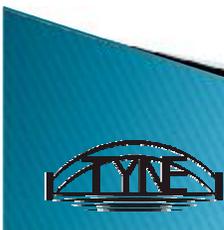
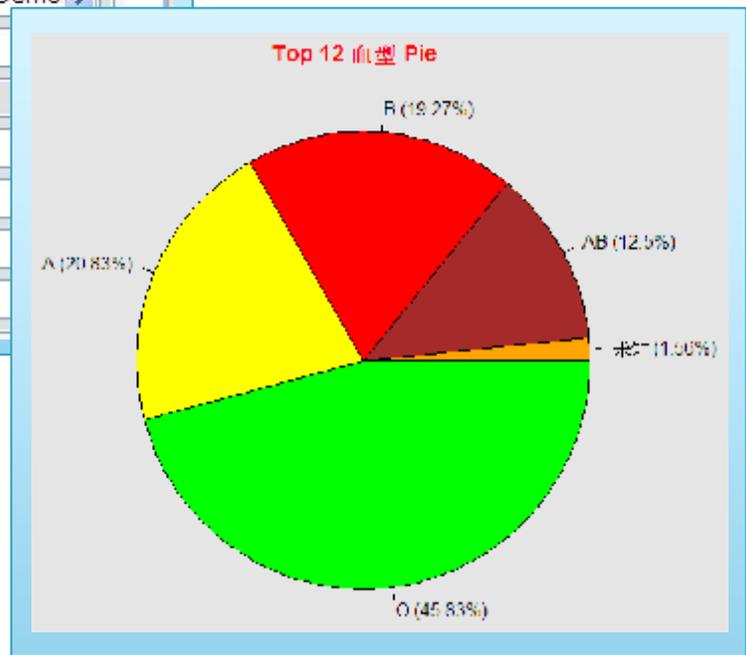
血型	次數
A	80
B	74
AB	48
O	176
未知	6

設定

分群項目 → Group By
分群個數 → NumVar

適用於"第二級資料"

輸出



Piechart--Count 圓餅圖

Data File :
New_99survey2.csv

- } 圓餅圖，呈現每個項目對應數值佔數值總值的比例。
- } 元素：分群項目(自動統計該項目共有幾筆資料)

資料

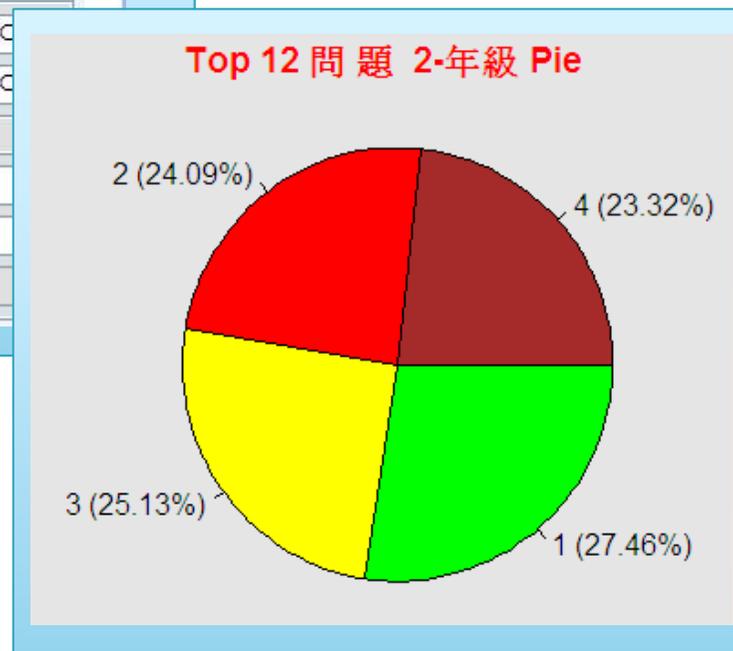
問題 2-年級
1
3
1
3
1
1
1
1
1
3
3
2
2
3

設定

分群項目 →

適用於"第一級資料"

輸出



Bar chart 長條圖。

Data File :
BloodType.csv

- } 長條圖，用以顯示或比較資料在不同 X 軸類別變數下的分布型態。
 - X 軸為類別變數，代表項目
 - Y 軸為對應的數值
- } 元素：分群項目、分群個數

適用於"第二級資料"

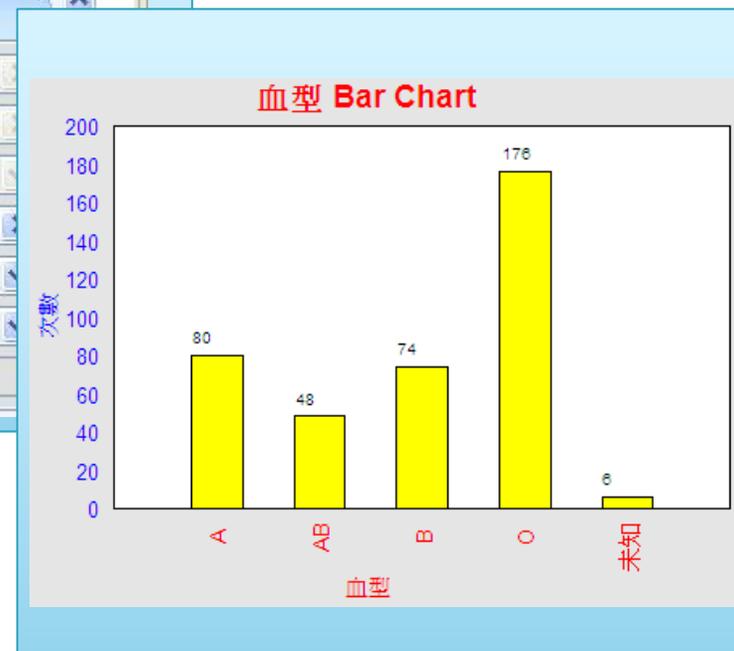
資料

分群項目	分群個數
血型	次數
A	80
B	74
AB	48
O	176
未知	6

設定



輸出



Barchart--Count 長條圖

Data File :
New_99survey2.csv

- } 長條圖，用以顯示或比較資料在不同 X 軸類別變數下的分布型態。
 - o X 軸為類別變數，代表項目
 - o Y 軸為對應的數值
- } 元素：分群項目(自動統計該項目共有幾筆資料)

資料

問題 2-年級
1
3
1
3
1
1
1
1
3
3
2
2
2

設定

分群項目 →

Criteria

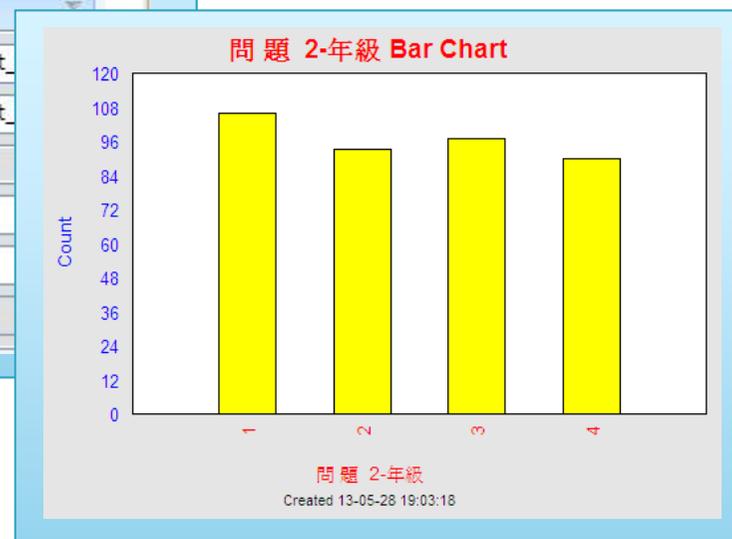
ChartTool: Bar Chart--Count

Main

- * Work Path: C:\PEDA\Stat
- * Data File: C:\PEDA\Stat
- Sheet Name:
- * X-Axis Var(s): 問題 2-年級
- MultiChart Var:
- * Job Name: barchart_c

適用於“第一級資料”

輸出



Pareto 柏拉圖

Data File :
BloodType.csv

- } 柏拉圖，可用以顯示或比較資料數值遞減排序之後在 X 軸類別變數的情況，藉以突顯與檢視較具影響程度的類別變數種類。
- } 元素：分群項目、分群個數

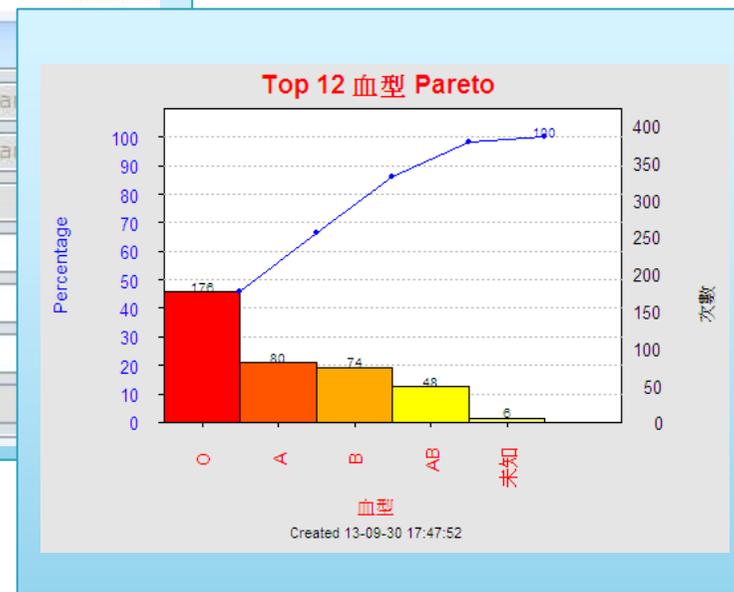
資料

分群項目	分群個數
血型	次數
A	80
B	74
AB	48
O	176
未知	6

設定

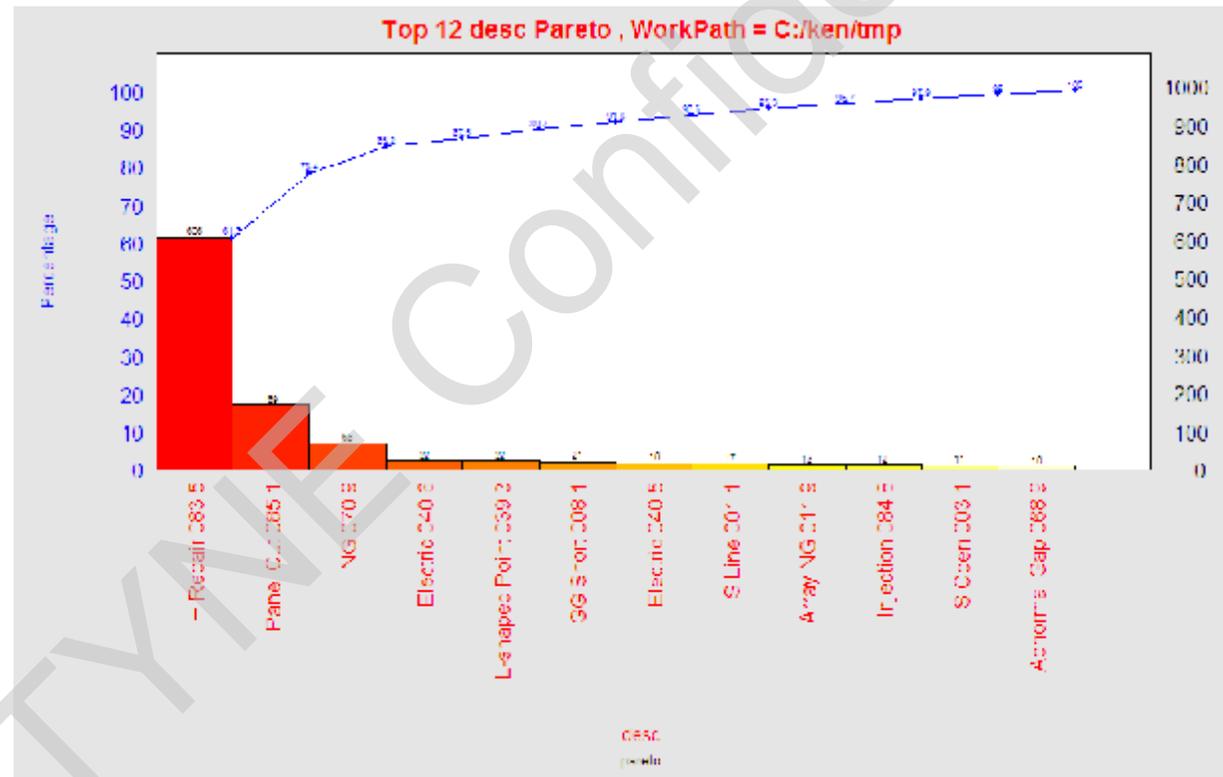
分群項目
分群個數

輸出



(補充) Pareto

- 柏拉圖，可用以顯示或比較資料數值遞減排序之後在 X 軸類別變數的情況，藉以突顯與檢視較具影響程度的類別變數種類。



(補充) Pareto (續)

Data

panel(cnt)	desc
606	H Repair 083 5
169	Panel Out 085 1
68	NG 070 9
22	Electric 040 6
22	L-shaped Point 039 3
21	GG Short 008 1
18	Electric 040 5
17	S Line 001 1
13	Array NG 011 9
12	Injection 084 8
11	S Open 003 1
10	Abnormal Gap 068 9

各項目之數值
遞減排序

繪製累積機率折線

計算累積機率值

→ $606/989=61.27\%$

→ $(606+169)/989=78.36\%$

→ $(606+169+68)/989=85.24\%$

·

·

·

·

·

→ $(606+169+\dots+11)/989=97.88\%$

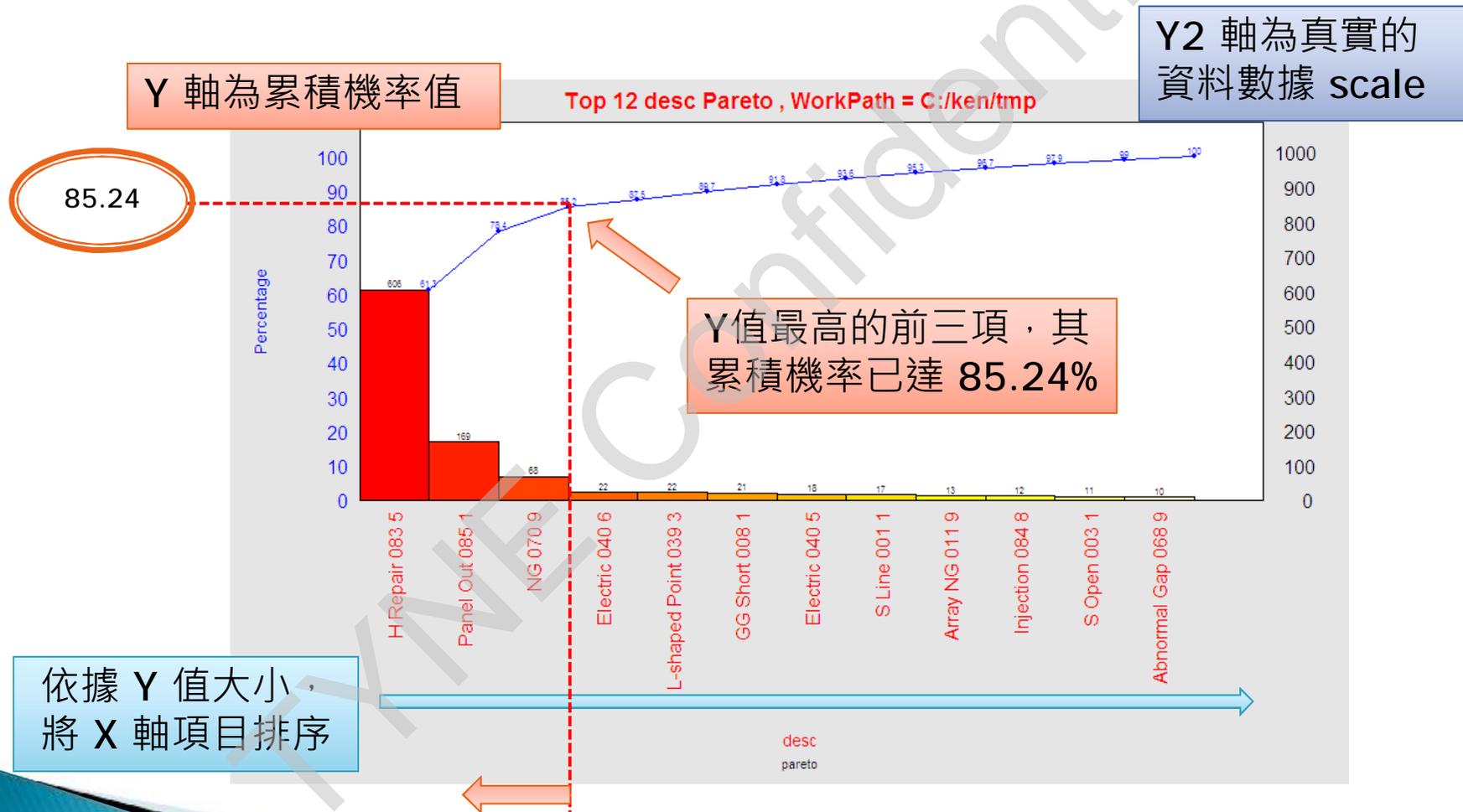
→ $(606+169+\dots+11+10)/989=100\%$

總計： $(606+169+\dots+10)=989$



(補充) Pareto (續)

} 找出影響產品品質的少數重要原因è 80/20法則



Histogram 直方圖

Data File :
StudHeight.csv

- } 直方圖，將連續型資料的數值範圍（X軸）分成相同組距（寬度）的組，自動計算出該組的計數或百分比，為該組 Y 軸之高度，以呈現資料數據的分布情況。
- } 元素：數值欄位

資料

性別	身高	班級
男	161.7	B班
男	169.8	B班
男	177.1	B班
男	163.4	B班
男	172.1	B班
女	156.7	A班
女	156.2	A班
女	159	A班
女	153.9	A班
女	152.2	A班
女	166	A班

設定

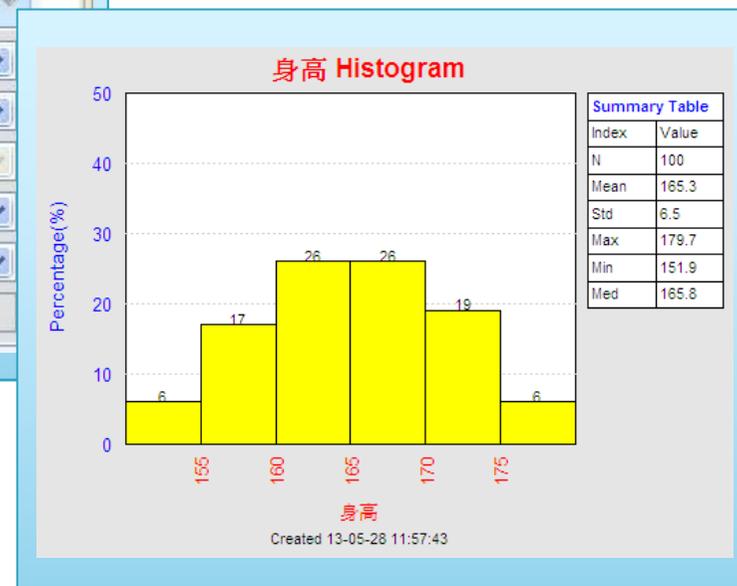
ChartTool: Histogram Chart

Main

- * Work Path: C:\PEDA\Stat_C
- * Data File: C:\PEDA\Stat_C
- Sheet Name: [dropdown]
- * X-Axis Var: 身高
- MultiChart Var: [dropdown]
- * Job Name: histogram

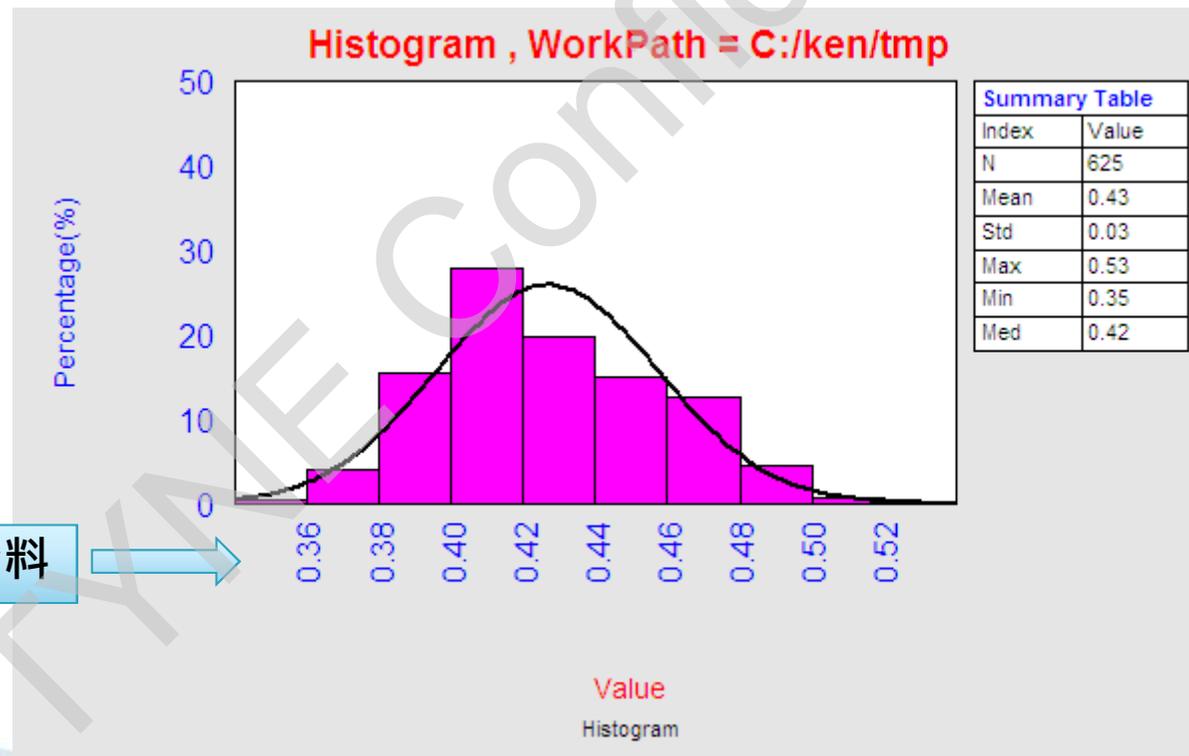
數值欄位 →

輸出



(補充) Histogram

- 直方圖，將連續型資料的數值範圍（X軸）分成相同組距（寬度）的組，自動計算出該組的計數或百分比，為該組 Y 軸之高度，以呈現資料數據的分布情況。



X 軸是連續型資料



(補充) Histogram

- 直方圖，將連續型資料的數值範圍（X軸）分成相同組距（寬度）的組，自動計算出該組的計數或百分比，為該組 Y 軸之高度，以呈現資料數據的分布情況。

Data

value
0.420748
0.412461
0.385155
0.441979
0.398104
0.422967
0.399386
0.433637
0.419183
0.417802
0.465462
0.43028
0.401197
0.371064
0.409968
0.418721
0.41994

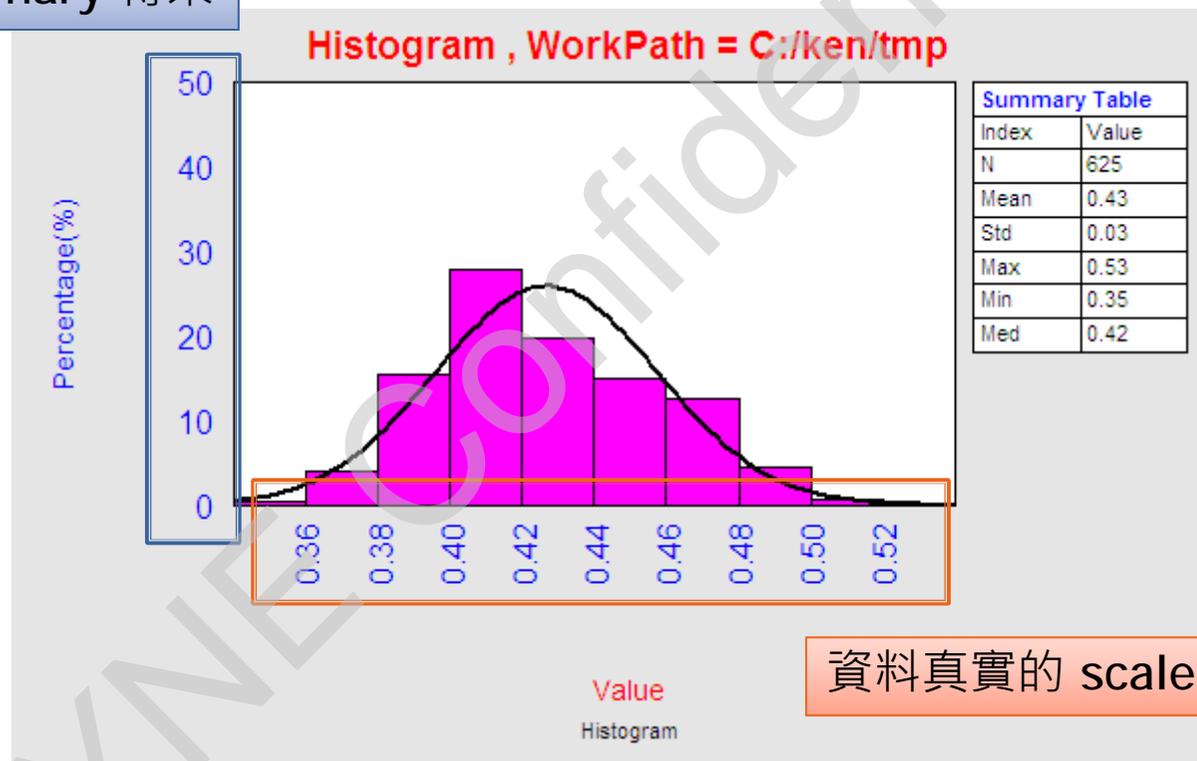
依據數值範圍分組，
計算每組內共有幾筆資料，
以及共有幾 % 的
資料數值落於該組距內

Range	Freq
0.34~0.36	3 (0.48%)
0.36~0.38	25 (4%)
0.38~0.40	96 (15.36%)
0.40~0.42	174 (27.84%)
0.42~0.44	123 (19.68%)
0.44~0.46	93 (14.88%)
0.46~0.48	78 (12.48%)
0.48~0.50	28 (4.48%)
0.50~0.52	4 (0.64%)
0.52~0.54	1 (0.16%)

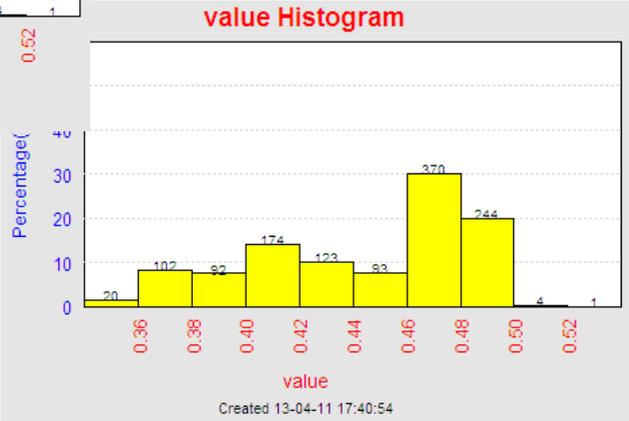
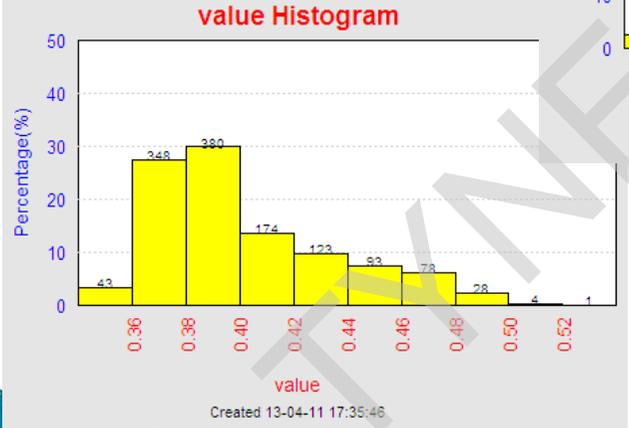
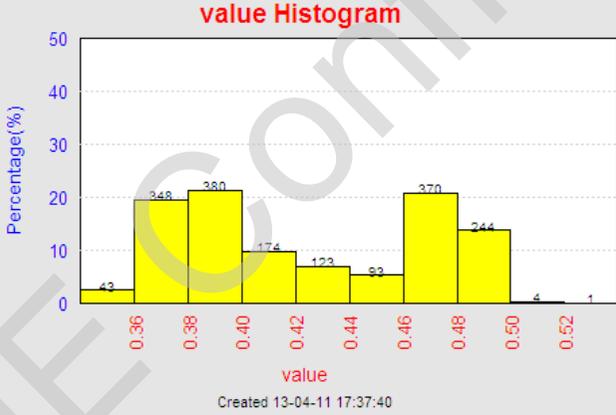
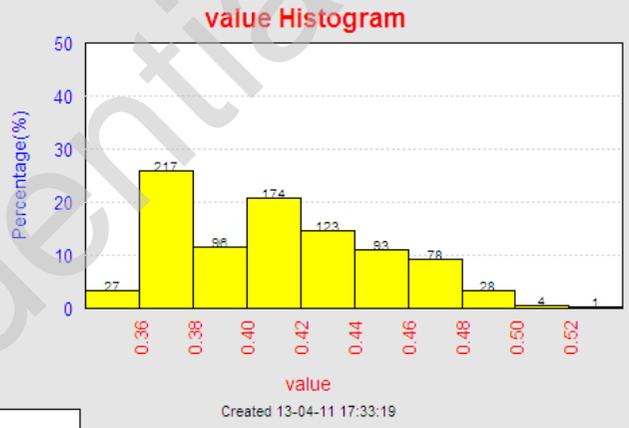
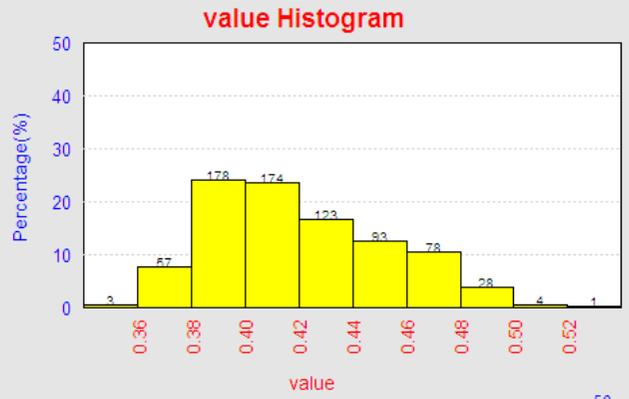
數值介於 0.46~0.48 的，共有 78 筆
(佔全部筆數的 12.48%)

(補充) Histogram

從資料 summary 得來



Histogram



Scatter 散佈圖

Data File :
StudHeight.csv

- 散佈圖，用以了解或呈現 X、Y 兩參數之間的資料關係 (relationship)。
- 元素：X 軸變數、Y 軸變數

資料

身高	體重
171.6	56.1
168.1	51.6
172	62.4
169	71
167.7	61.2
169.2	62.5
177	59.9
168.1	69.6
174.9	58.2
164.8	67.2
173.7	74
172.7	71.5
169.6	55.5
170.6	54.7
176.8	69.5

設定

ChartTool: X Y Scatter-XY Values

Main

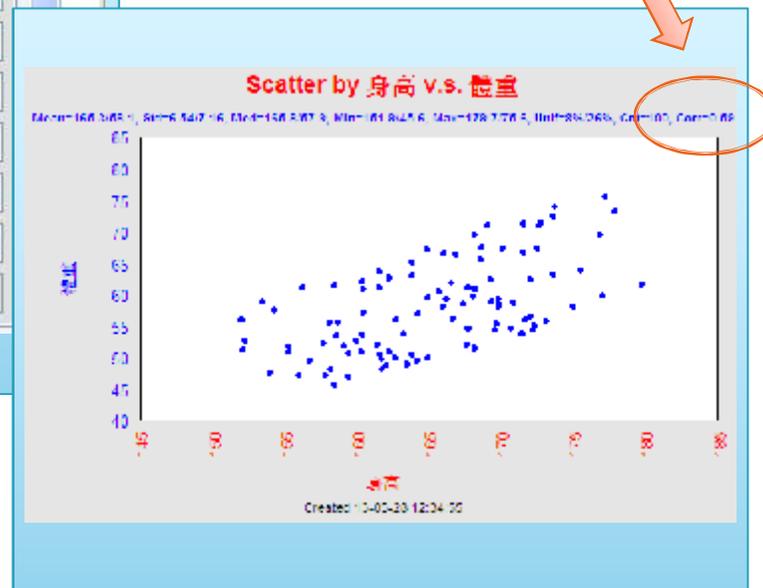
- * Work Path: C:\PEDA\Sti>
- * Data File: C:\PEDA\Sti>
- Sheet Name: [dropdown]
- * X-Axis Var: 身高
- * Y-Axis Var: 體重
- MultiChart Var: [dropdown]
- * Job Name: scatter

X 軸變數 →

Y 軸變數 →

相關係數

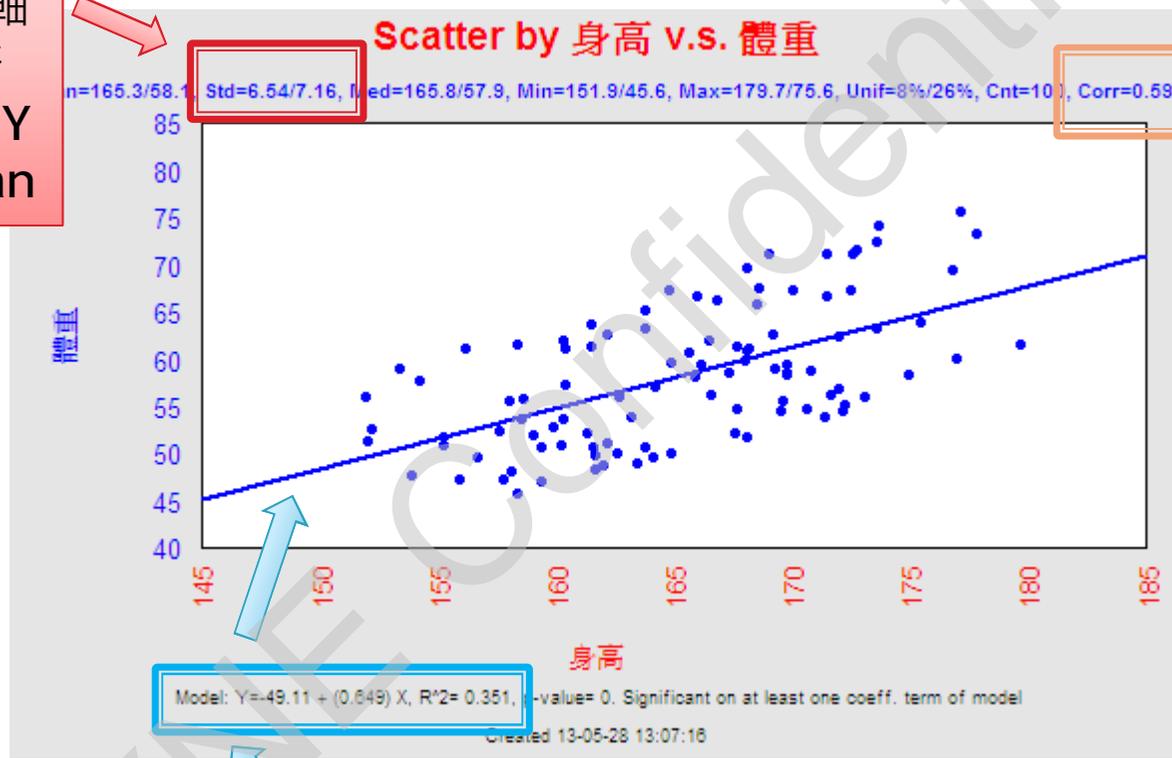
輸出



Scatter 散佈圖

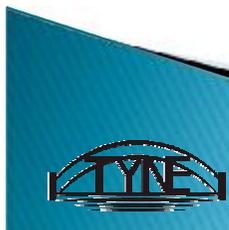
X、Y兩參數的相關係數

因 X 軸 & Y 軸
都是數值，所
以各自算 X、Y
兩參數的mean



Model 用來描述 X 變數與 Y 變數的關係式

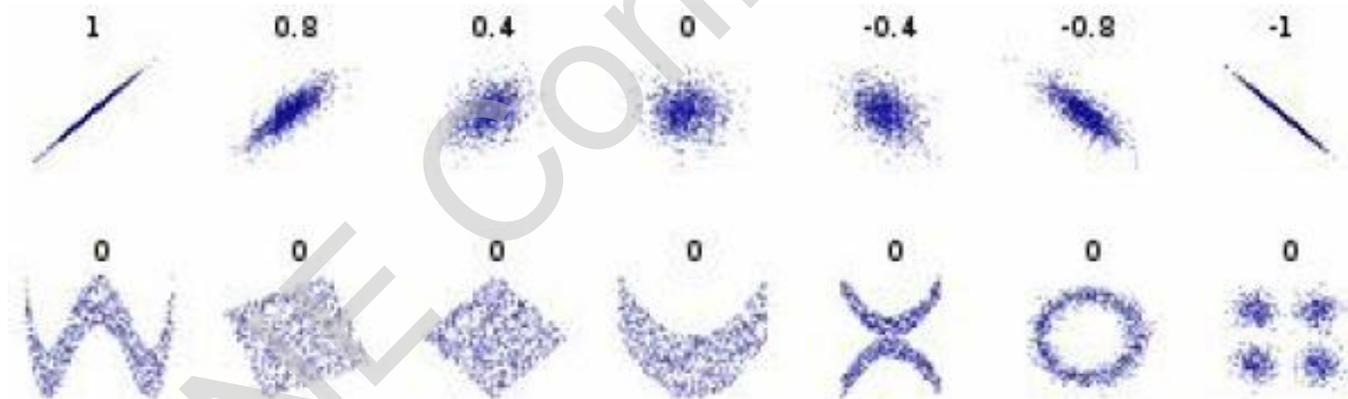
Further Options	
Group Var(s)	>
Fit Curve	1:Linear



(補充) Scatter

} 相關係數(Correlation coefficient):

- 兩組資料的**線性**相關程度, 介於-1 ~ 1 之間
- 相關係數絕對值達小於0.3 時, 為低度相關; 介於0.3~0.7 時, 為中度相關; 達0.7~0.8 時, 為高度相關; 若達0.8 以上時, 為非常高度相關



$$\text{公式: } r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}}$$

x_i 為第 i 個 x 的數值

y_i 為第 i 個 y 的數值

\bar{x} 為 x 變數的平均數

\bar{y} 為 y 變數的平均數

Boxplot 盒鬚圖

Data File :
StudHeight.csv

- } 盒鬚圖(Box-and-whisker-plot)，用箱子的概念呈現資料的集中及離散程度，可描述數值資料的分布狀況。
- } 元素：分群項目、數值欄位

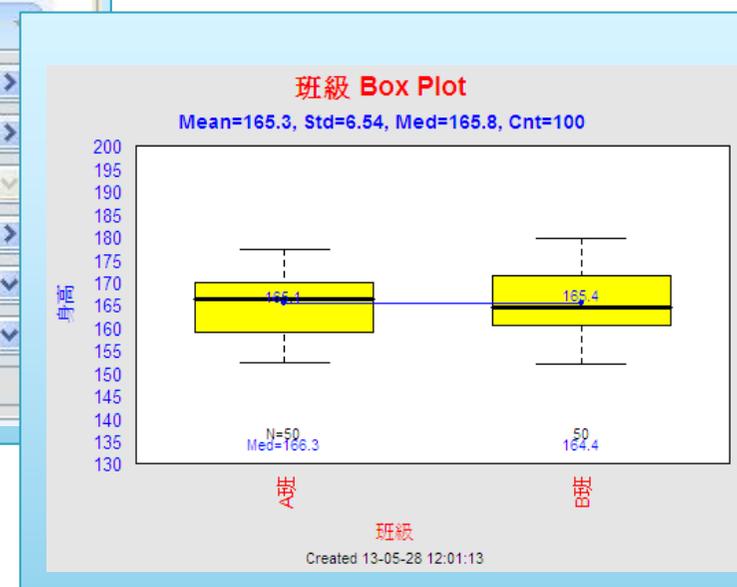
資料

性別	身高	班級
男	161.7	B班
男	169.8	B班
男	177.1	B班
男	163.4	B班
男	172.1	B班
女	156.7	A班
女	156.2	A班
女	159	A班
女	153.9	A班
女	152.2	A班
女	166	A班

設定



輸出



(補充) Boxplot

- 盒鬚圖(Box-and-whisker-plot)，用箱子的概念呈現資料的集中及離散程度，可描述數值資料的分布狀況。

Raw Data

	B	E
1	ecp	value
2	PA04-01	0.420748
3	PA04-01	0.441979
4	PA04-01	0.399386
240	PA04-01	0.475662
241	PA04-01	0.485806
242	PA04-02	0.403512
243	PA04-02	0.4092
244	PA04-02	0.437654
264	PA04-02	0.408596
265	PA04-02	0.394679
266	PA04-03	0.414114
267	PA04-03	0.404838
268	PA04-03	0.438964
288	PA04-03	0.427687
289	PA04-03	0.399651
290	PA04-04	0.396751
291	PA04-04	0.394147
292	PA04-04	0.404147
335	PA04-04	0.419738
336	PA04-04	0.420177
337	PA04-04	0.400988

計算 [PA04-01] 群的
P50, Q1, Q3, IQR

計算 [PA04-02] 群的
P50, Q1, Q3, IQR

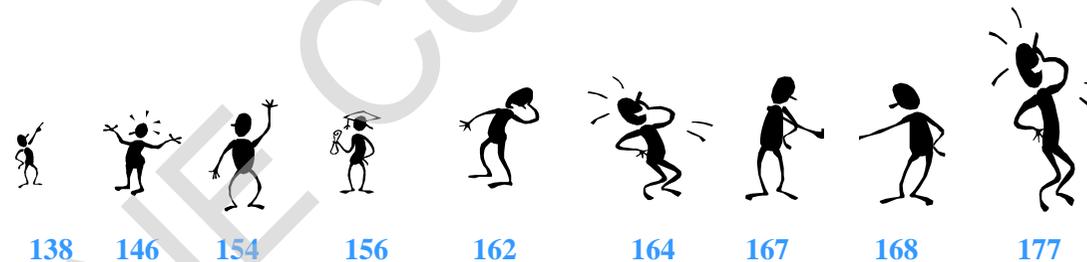
計算 [PA04-03] 群的
P50, Q1, Q3, IQR

計算 [PA04-04] 群的
P50, Q1, Q3, IQR

(補充)敘述統計量

} Percentile 百分位數

- 即為所有觀測值所形成之次數分配的詳細分割，把觀測值從小到大排列之後，我們要找一個數，使得有 **x%** 觀測值比它小，而 **(100-x)%** 比它大。這就是此組觀測值的第x百分位數。
- P50=中位數 (median)
- P25=Q1，第一四分位數(first quartile)
- P75=Q3，第三四分位數(third quartile)



à $P_x =$ 第 $\frac{x}{100} \times (n+1)$ 個位置的值

à $P_{25} =$ 第 $\frac{25}{100} \times (9+1)$ 個位置的值 = 第 2.5 個位置的值 = $\frac{(146+154)}{2} = 150$

(補充)敘述統計量

} IQR 四分位距 (Inner Quartile Range)

- 為樣本的第75百分位數(Q3)與第25百分位數(Q1)的差，意即 $P75 - P25 = Q3 - Q1$ 。
- 可顯示或衡量樣本中間百分之五十的資料範圍、變動幅度或分散狀況。



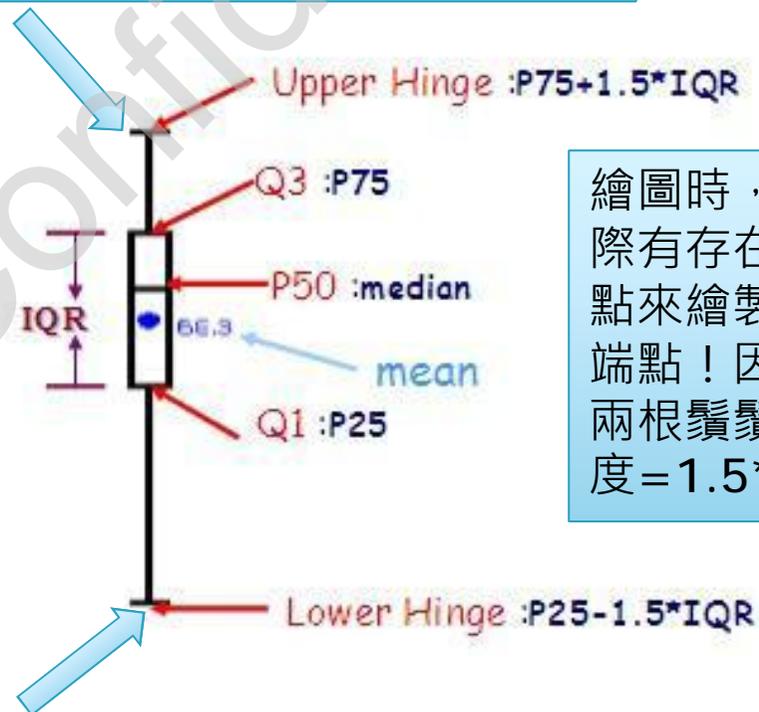
(補充) Boxplot (續)

} 盒鬚圖(Box-and-whisker-plot)

- Q1,P50,Q3 構成箱子(box)的部分
- Lower Hinge到Q1 和 Q3到Upper Hinge 構成鬚(whisker)的部分。

IQR
= Interquartile range
= Q3 - Q1
有 50% 的資料點落在 Box 內！

該點數值介於 $P75 \sim (P75 + 1.5 * IQR)$ 內
之數值最大的資料點



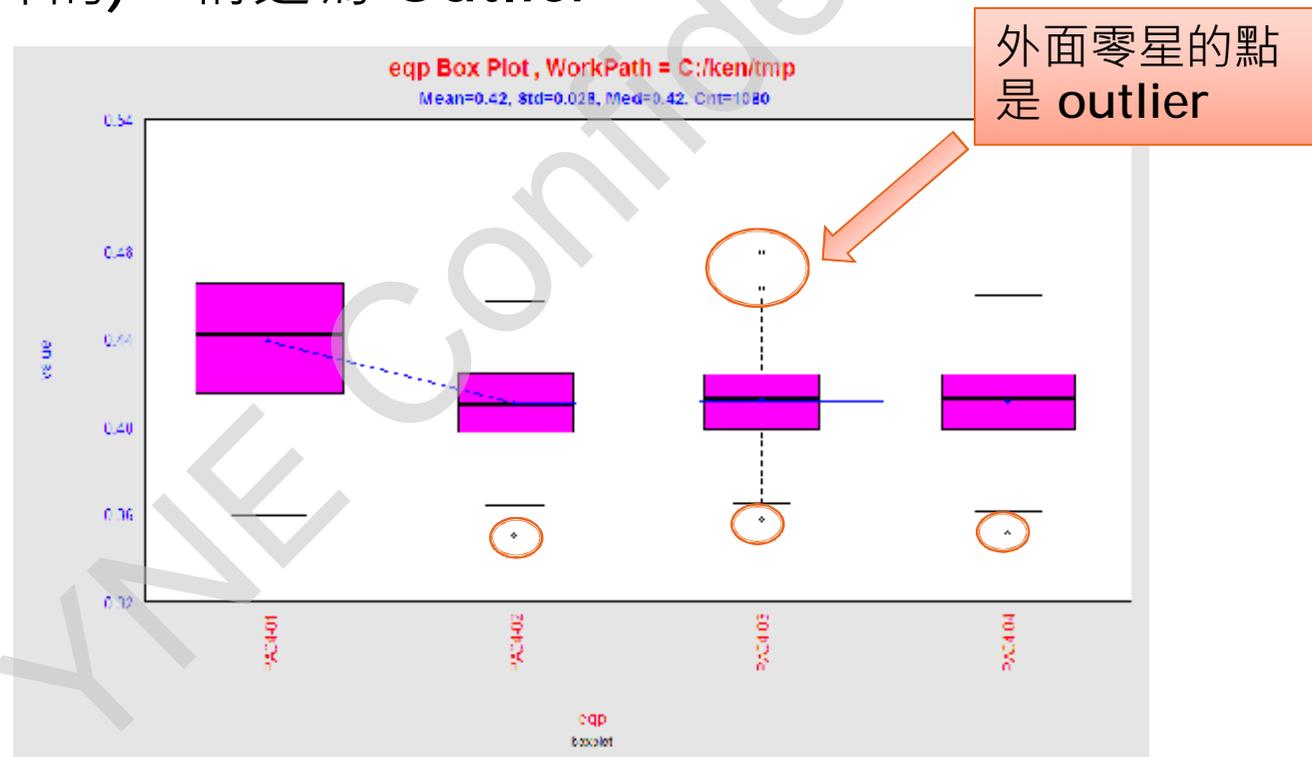
繪圖時，會以實際有存在的資料點來繪製鬚鬚的端點！因此上下兩根鬚鬚並非長度 = $1.5 * IQR$

該點數值介於 $(P25 - 1.5 * IQR) \sim P25$ 內
之數值最小的資料點

(補充) Boxplot

} **Outlier** 離群值

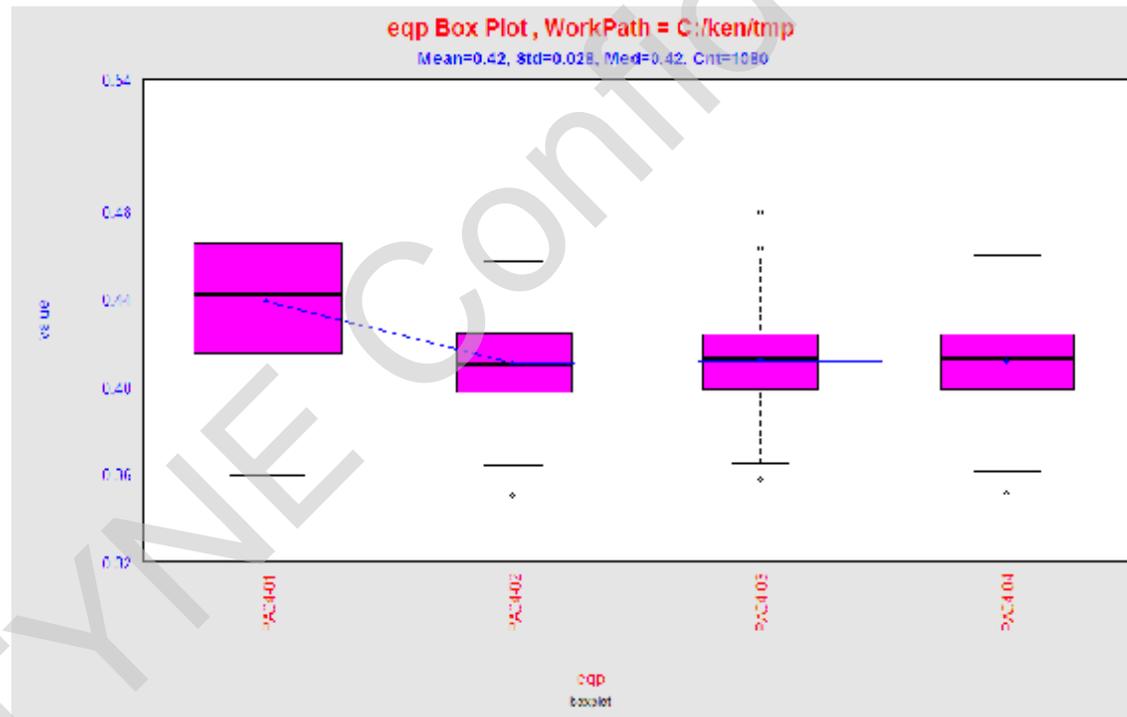
- 大於 Upper Hinge 或小於 Lower Hinge 的點(大於 $1.5 \times$ IQR 以外的), 稱之為 **Outlier**



(補充) Boxplot

} [練習 - Boxplot]

- 哪個機台的變異程度較大？
- 哪個機台的 median 較大？



練習題

8. 調查本校40位同學每日的上網時間 (小時), 資料如下。繪製 直方圖

0.9	1.0	1.0	1.0	1.0	1.1	1.2	1.2	1.3	1.4
1.4	1.4	1.4	1.4	1.5	1.5	1.6	1.8	1.8	1.8
1.9	2.0	2.0	2.0	2.0	2.0	2.1	2.1	2.3	2.4
2.5	2.6	2.6	2.7	2.7	2.7	2.8	3.1	3.3	3.4

9. 下表係本校某問卷調查各學院學院和各年級的受訪人數。

學院	人數	年級	人數
管理	112	一	100
工	87	二	98
商	86	三	108
電資	99	四	78
合計	384	合計	384

(a) 請繪製各學院受訪學生人數的長條圖與圓形圖。

(b) 請繪製各年級受訪學生人數的長條圖與圓形圖。

練習題

例 3.43.

下表示 8 對父子的身高資料，請計算他們的相關係數。

表 3.4: 8 位父親與他們的小孩身高

編號	父親身高 (X)	小孩身高 (Y)
1	163.9	161.3
2	174.4	175.6
3	171.4	169.9
4	169.7	171.7
5	174.7	174.2
6	179.0	185.3
7	170.2	168.1
8	176.6	171.5



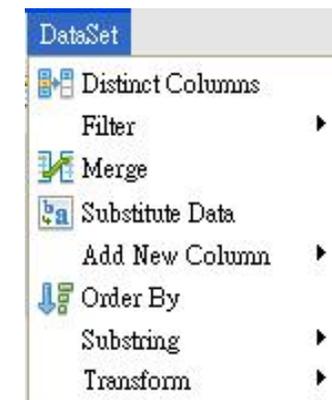
“DataSet” Menu 資料整理選單



資料整理函式

} “DataSet” Menu

- Distinct Columns : 異值篩選 (欄位內容值取唯一)
- Filter : 資料篩選
- Merge : 兩檔案合併
- Substitute Data : 資料替換
- Add New Column
 - Grouping : 資料分群
 - Expression : 公式計算
 - Paste : 欄位合併
- Order By : 排序
- Substring : 擷取字串 (Ex:取第4碼 ~ 第7碼)
- Transform
 - Horizon2Vertical : 橫擺資料轉直擺資料
 - Vertical2Horizon : 直擺資料轉橫擺資料



Horizon2Vertical

Data File :
FoodCust_Month(V).csv

} 橫擺資料轉直擺資料

資料

抱怨項目	一月	二月	三月	四月
少放香料	20	5	7	5
少放贈品	30	2	10	11
包裝不良	12	15	3	14
地址錯誤	3	2	1	4
冷掉了	10	15	8	7
其他	3	5	3	2
訂單錯誤	2	5	7	3
烤焦	40	5	20	20
數量錯誤	5	20	4	2
遲到	15	30	5	3

設定

欲轉成之
欄位名稱

欲轉橫擺之
內容值

資料列之ID

DataSet : Vertical2Horizon

Main

* Work Path C:\Documer >

* Data File C:\Documer >

Sheet Name

* Param Column 月份

* Value Column 次數

Group by 抱怨項目

* Job Name V2H

輸出

	抱怨項目	月份	次數
1	冷掉了	一月	10
2	遲到	一月	15
3	數量錯誤	一月	5
4	少放香料	一月	20
5	烤焦	一月	40
6	地址錯誤	一月	3
7	包裝不良	一月	12
8	訂單錯誤	一月	2
9	少放贈品	一月	30
10	其他	一月	3
11	冷掉了	二月	15
12	遲到	二月	30
13	數量錯誤	二月	20
14	少放香料	二月	5
15	烤焦	二月	5
16	地址錯誤	二月	2
17	包裝不良	二月	15
18	訂單錯誤	二月	5
19	少放贈品	二月	2
20	其他	二月	5



“Statistics” Menu

統計功能選單



Statistic 統計量報表

Data File :
StudHeight.csv

} Statistics → Generic Report → Statistic

資料

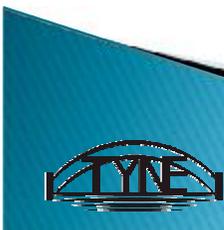
身高	體重	班級
168.1	60.9	A班
173.6	63.2	A班
162.2	62.6	A班
171.5	71.2	B班
172.2	55	B班
171.5	66.7	B班
171.4	53.8	B班
169.5	54.5	B班
163.8	63.1	B班
179.7	61.6	B班
173.1	55.9	B班
172.5	67.3	B班

設定

- N (Count) : 樣本數
- Mean : 樣本平均數
- P50 (Median) : 中位數
- P25 (Q1) : 第一四分位數
- P75 (Q3) : 第三四分位數
- Std : 樣本標準差
- Range : 全距
- IQR : 四分位距 (Q3-Q1)
- CV : 變異係數
- Skewness : 動差偏態係數

輸出

	Response	班級	N	Mean	P50	Std
1	身高	A班	50	165.1	166.3	6.6
2	身高	B班	50	165.4	164.4	6.53
3	體重	A班	50	58	58	6.8
4	體重	B班	50	58	56	7.6



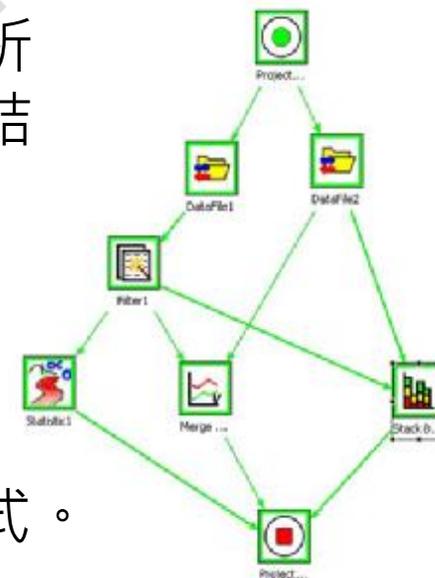
“WorkMap” 智慧流程設計



PEDA – WorkMap (簡介)

} 分析流程模版化

- 能一次進行多項資料處理、繪製多張不同的分析圖表或是使用相同的來源資料得到多種的分析結果
- 將一道道繁複的分析程序建立成模版
- 便於執行日常重複性工作
- 模版仍可做局部修正、重置或多工使用
- 期待分析方法經過多次修正後，可形成決策模式。
- 方便公司內部人員相互交流，多方驗證
- 將分析流程傳承給新進工程師，達到經驗傳承與知識累積的目的



PEDA – WorkMap (名詞)

} Model

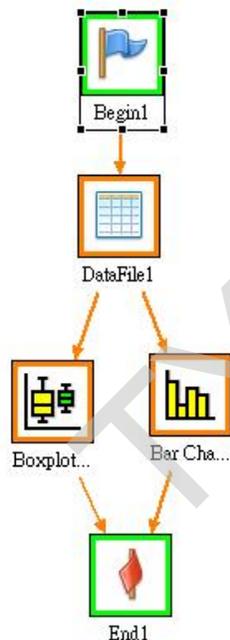
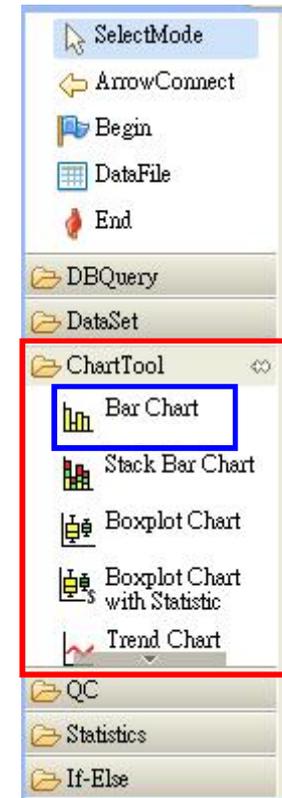
- 我們稱單一個分析步驟為“Model”
- 可想像為之前教的 Function, 譬如 Boxplot, Scatter, Histogram, ...

} Model Group

- 為 Model 的分類
- 可想像為之前教的 Menu, 譬如 DataSet, ChartTool, QC, Statistics, ...

Model

Model Group

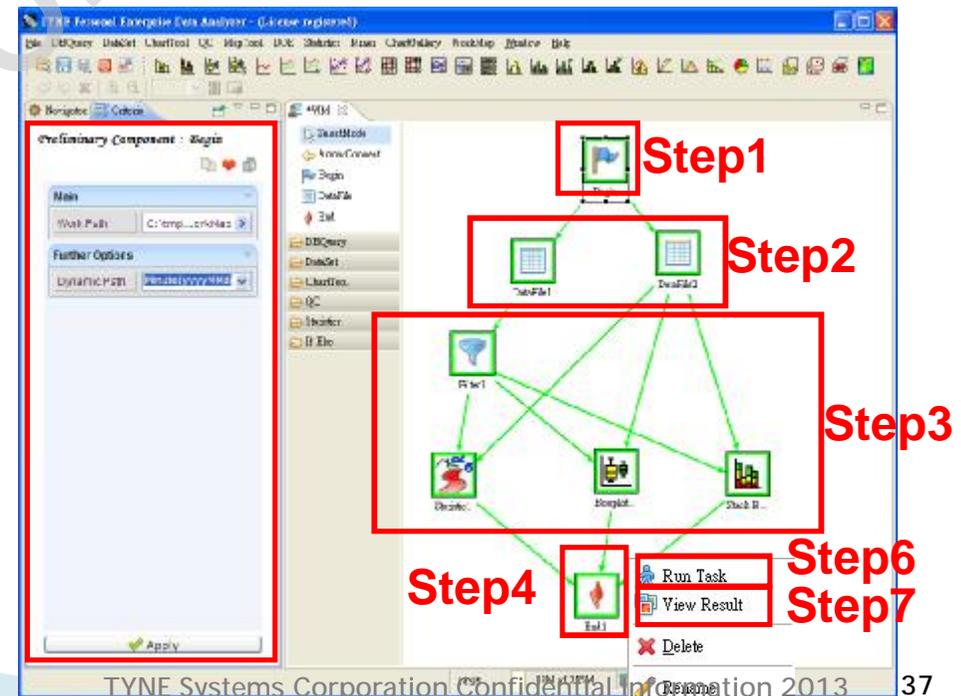


} Arrow

- 用 Arrow (箭頭) 來連接 Model, 以代表分析流程的順序關係

PEDA – WorkMap (步驟)

- } 步驟1. 起始 Model : 'Begin' model (分析流程的開頭)
- } 步驟2. 次要 Model : 資料來源檔 (DataFile or DataSQL Model)
- } 步驟3. 建立欲分析步驟的 Model (如資料篩選 Filter, 畫柏拉圖 Pareto, 做檢定 One-Sample T-Test, ...), 及建立 Model 間連線關係
- } 步驟4. 結尾 Model : 'ProjectEnd' model
- } 步驟5. 設定 Model 各自的 Criteria
- } 步驟6. 執行 : Run Task
- } 步驟7. 觀看結果 : View Result



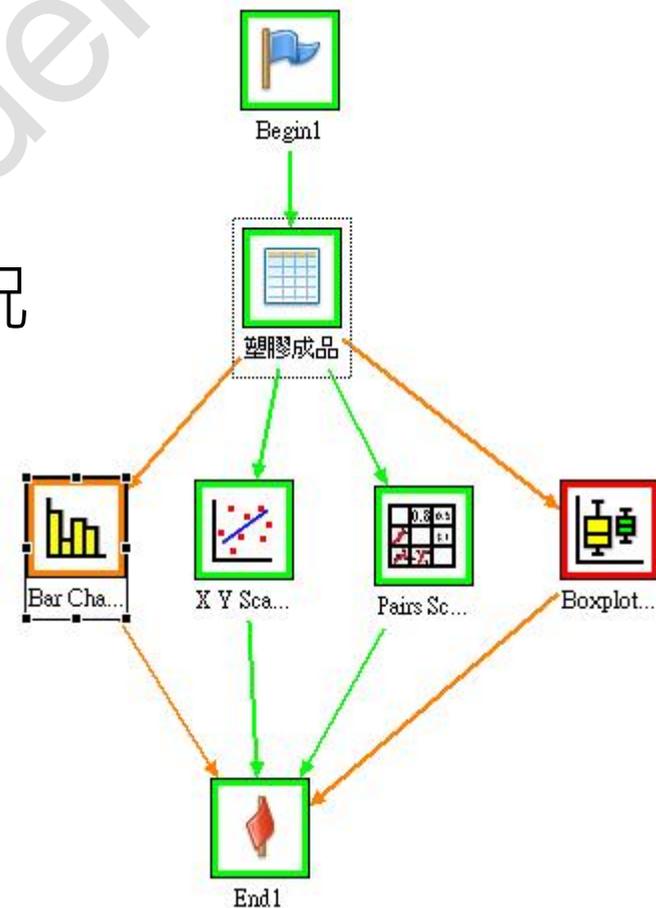
PEDA – WorkMap (顏色)

} Model 的顏色代表 Criteria 設定狀況

- 橘色：Criteria 尚未設定完成
- 綠色：Criteria 已設定完成
- 紅色：Criteria 設定有錯誤

} Arrow 的顏色代表分析執行狀況

- 橘色：尚未執行
- 綠色：執行完成
- 紅色：執行有錯誤



PEDA – WorkMap Model

必要！

Begin



必要！

資料來源檔
(DataFile or DataSQL Model)



資料處理
(Model Group : DataSet or If-Else, eg. Filter, Grouping, Expression, H2V, V2H,)



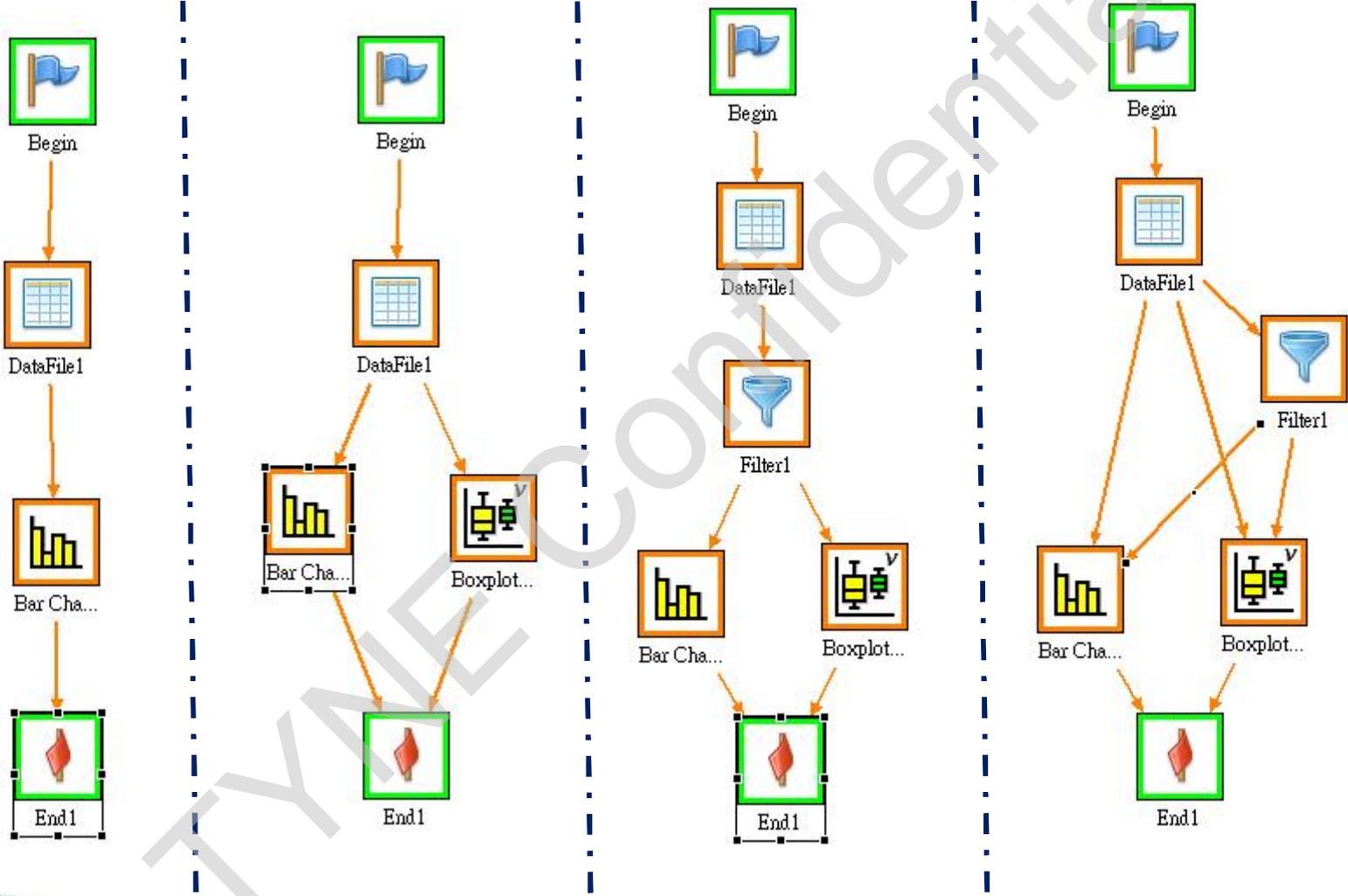
畫圖 or 統計分析
(Model Group : ChartTool, QC, or Statistics, eg. Boxplot, Control Chart, Hypothesis Test, ...)



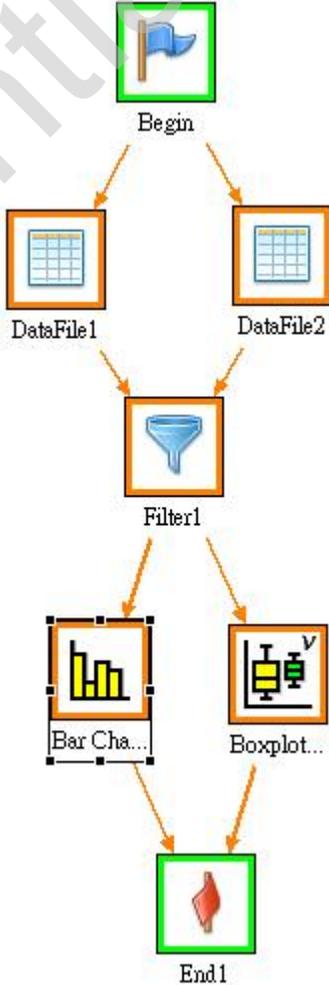
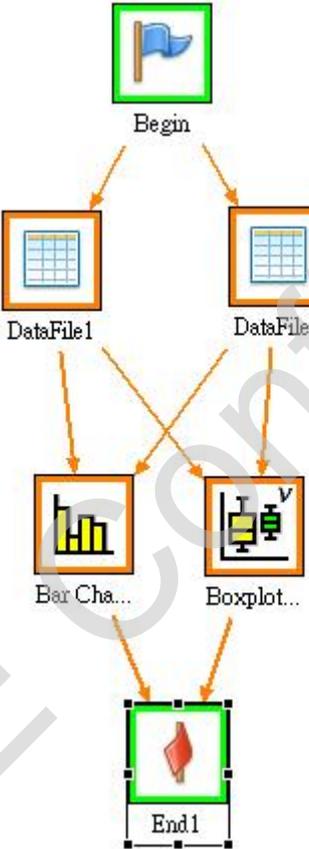
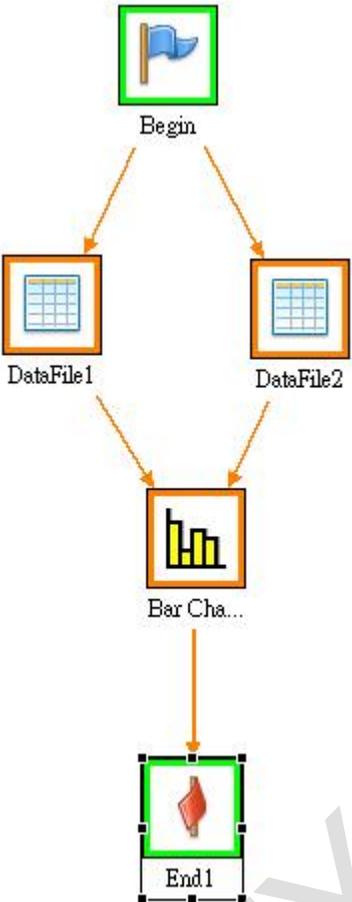
End



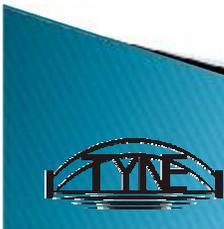
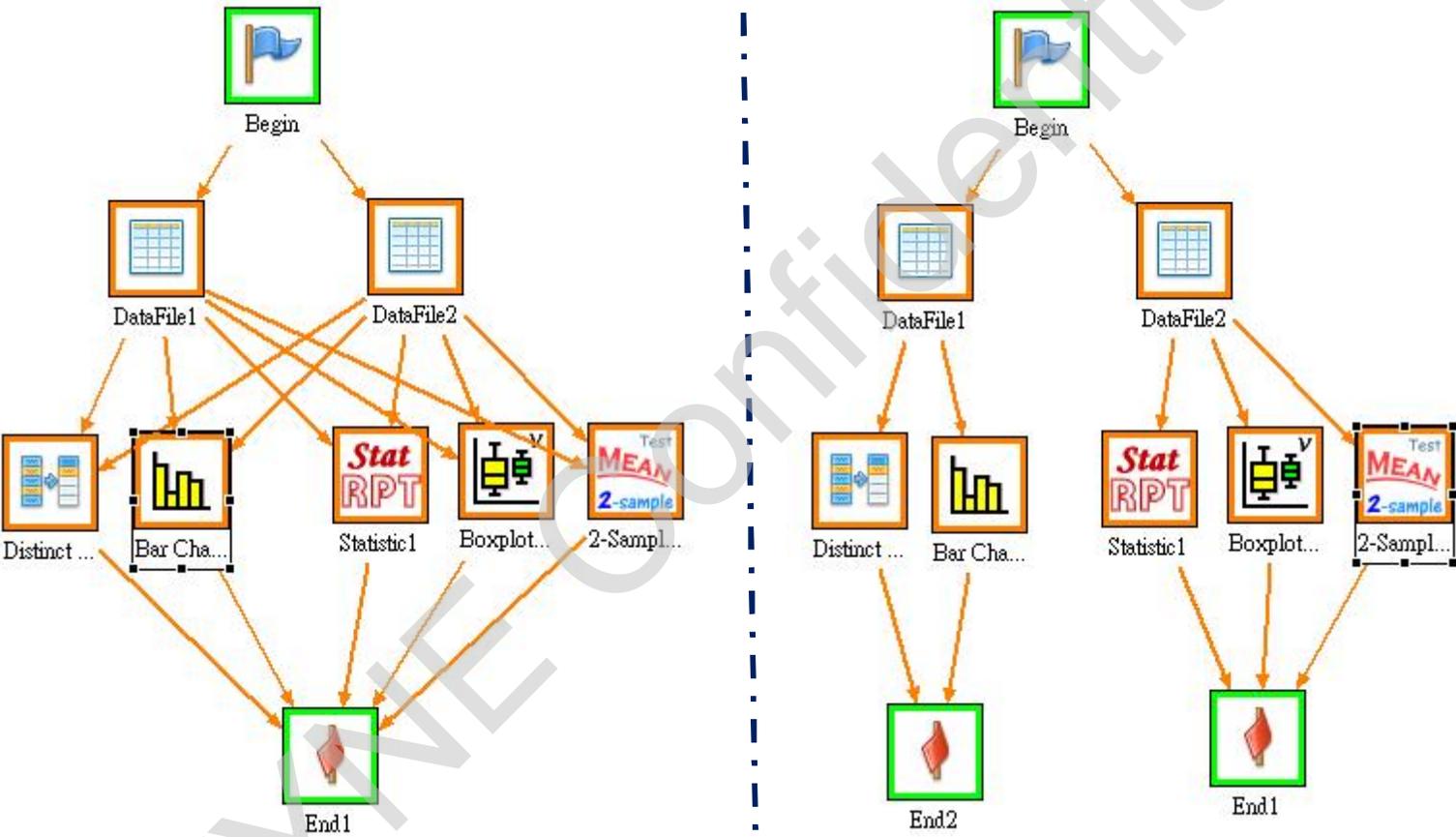
PEDA – WorkMap (範例)



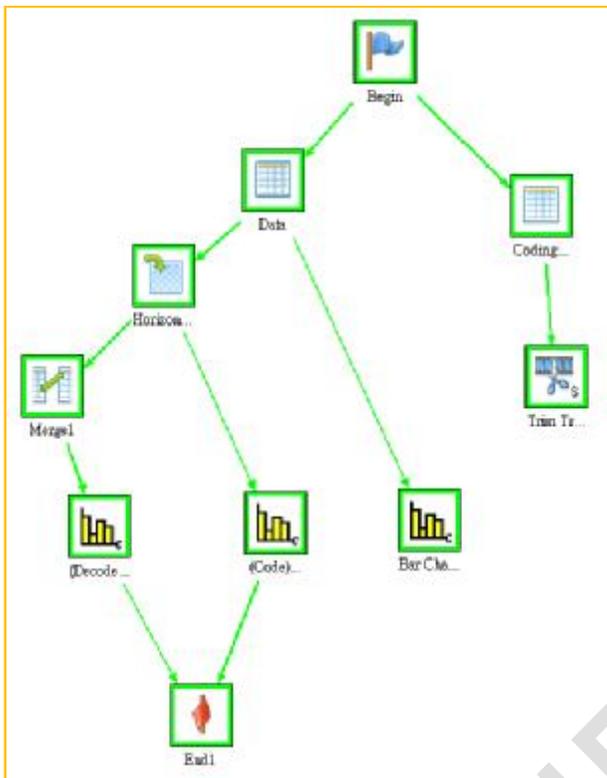
PEDA - WorkMap (範例)



PEDA – WorkMap (範例)

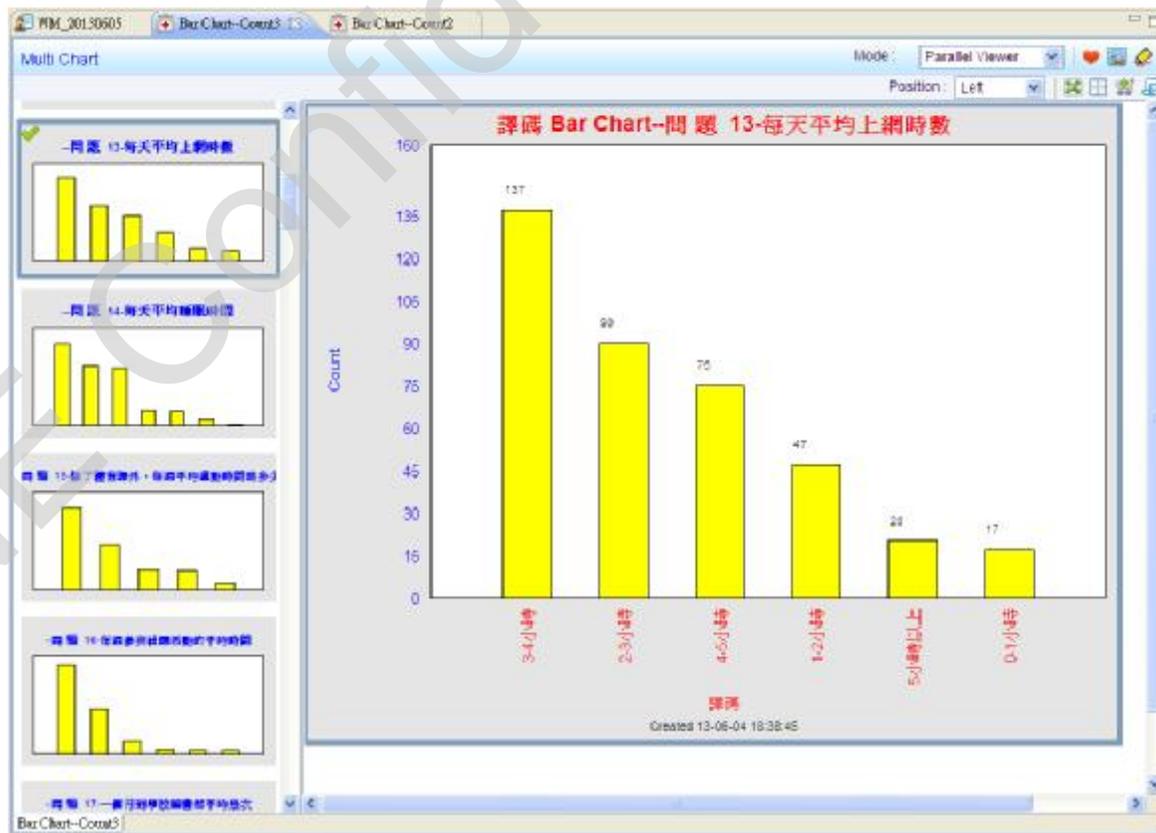


PEDA – WorkMap (範例)



Data File :
New_99survey2.csv

Coding_Book.xlsx



~ End ~

Thanks for your listening.

TYNE Confidential

