

# Chapter 2

## Graphic and Tabular Descriptive Techniques

# Agenda

- 助教介紹與課程規定
- 本課程回顧
- 不同資料型態
  - 下列這是什麼資料型態？
  - 請找出圖表錯誤的地方
- 各種圖形怎麼畫？
  - 畫圖步驟
  - Bar chart and Histogram有何差異？
  - Stem & Leaf 如何解讀？
  - 樞紐分析表
- 練習題

# 助教介紹

- 助教: 王珊慧
- 身分: 博士班
- 聯絡方式: [Shanhueiwan@gmail.com](mailto:Shanhueiwan@gmail.com)
- 工作: 實習課、出作業、複習~~



# 課程規劃

日期	課程	章節	作業	批改助教	實習講演
Sep-22	Introduction to Statistics	C1			X
Sep-29	Graphical and Tabular Descriptive Statistics	C2	1	陳昱君	
10-Jun	Numerical Descriptive Techniques	C4	2	張雅惠	
Oct-13	Probability	C6	3	陳昱君	<ol style="list-style-type: none"> <li>1. 認識SPSS</li> <li>2. SPSS資料輸入及讀取其他格式資料檔。</li> <li>3. 使用SPSS做出長條圖、圓餅圖、直方圖、折線圖與散佈圖</li> </ol>
Oct-20	Random Variables and Discrete Probability Distr	C7	4	張雅惠	<ol style="list-style-type: none"> <li>1. 使用Excel與SPSS產生各種敘述性統計量</li> <li>2. 使用Excel與 SPSS畫出box plot</li> <li>3. 使用EXCEL查詢機率與建立二項分配與Possion分配</li> </ol>
Oct-27	Continuous Probability Distributions	C8	5	陳昱君	使用EXCEL查詢常態分配，t分配、卡方分配與F分配值
3-Nov	Sampling Distributions	C9	6	張雅惠	
10-Nov	Review		計畫	張雅惠	X
17-Nov	Mid-term Exam				X
24-Nov	Statistical Inference: Estimation	C10	7	陳昱君	
1-Dec	Statistical Inference: Hypothesis Testing	C11	no		
8-Dec	Statistical Inference: Hypothesis Testing	C11	8	張雅惠	<ol style="list-style-type: none"> <li>1. 使用Excel與SPSS進行單一母體z-test</li> <li>2. 使用Excel與 SPSS產生單一母體信賴區間</li> </ol>
15-Dec	Statistical Inference: Inference about a Populatio	C12	no		
22-Dec	Statistical Inference: Inference about a Populatio	C12	9	陳昱君	<ol style="list-style-type: none"> <li>1. 使用Excel與SPSS進行兩母體t檢定，成對母體t檢定</li> <li>2. 使用Excel與SPSS進行母體比例t檢定，母體變異數檢定</li> </ol>
29-Dec	Guest speech/Review				
5-Jan	Final Exam				X

# 一、本課程回顧

---

9/22	Introduction to Statistics	Ch 1
9/29	Graphical and Tabular Descriptive Statistics	Ch 2
10/6	Numerical Descriptive Techniques	Ch 4
10/13	Probability	Ch 6
10/20	Random Variables and Discrete Probability Distributions	Ch 7
10/27	Continuous Probability Distributions	Ch 8
11/3	Sampling Distributions	Ch 9
11/10	Review	
11/17	Mid-term Exam	
11/24	Statistical Inference: Estimation	Ch 10
12/1	Statistical Inference: Hypothesis Testing	Ch 11
12/8	Statistical Inference: Hypothesis Testing	Ch 11
12/15	Statistical Inference: Inference about a Population	Ch 12
12/22	Statistical Inference: Inference about a Population	Ch 12
12/29	Guest speech/Review	
1/5	Final Exam	

# 一、本課程回顧

- 今天行銷經理問你，Pockmon Go的下載率（熱門度）能否超過三個月？
- 請問你要如何去驗證？

## 本課程回顧:

統計是一門用科學化的方式，處理不確定性的問題

- **Stage 1: 確定假設 (C10, C11)**
  - Pockmon Go 下載率會超過三個月 (C10)
  - $H_0: u < 3 \text{ month}$
  - $H_1: u > 3 \text{ month}$
- **Stage 2: 確定母體/樣本的分配 (決定用 Z、t、F)**
  - 抽取樣本 (發問卷) ~ 你要抽多少個才合理? (C9)
  - 抽完要如何驗證? (你覺得他是什麼分配)? (C7, C8)
- **Stage 3: 選擇critical point ~ (查表, 需要信心水準)**
- **Stage 4: 拒絕 or 不拒絕 hypothesis**

# 一、本課程回顧

9/22	Introduction to Statistics	Ch 1
9/29	Graphical and Tabular Descriptive Statistics	Ch 2
10/6	Numerical Descriptive Techniques	Ch 4
10/13	Probability	Ch 6
10/20	Random Variables and Discrete Probability Distributions	Ch 7
10/27	Continuous Probability Distributions	Ch 8
11/3	Sampling Distributions	Ch 9
11/10	Review	
11/17	Mid-term Exam	
11/24	Statistical Inference: Estimation	Ch 10
12/1	Statistical Inference: Hypothesis Testing	Ch 11
12/8	Statistical Inference: Hypothesis Testing	Ch 11
12/15	Statistical Inference: Inference about a Population	Ch 12
12/22	Statistical Inference: Inference about a Population	Ch 12
12/29	Guest speech/Review	
1/5	Final Exam	

一個母體

2/20	Inference about Two Populations	Ch 13
2/27	Inference about Two Populations	Ch 13
3/6	Analysis of Variance (ANOVA)	Ch 14.1, 3
3/13	Analysis of Variance (ANOVA)	Ch 14.4, 5
3/20	Analysis of Variance (ANOVA) review	Ch 14.2
3/27	1 <sup>st</sup> Midterm Exam	
4/10	Chi-square Test	Ch 15
4/17	Nonparametric Methods	Ch 19
4/24	Simple Linear Regression	Ch 16
5/1	Simple Linear Regression	Ch 16
5/8	Multiple Regression	Ch 17
5/15	2 <sup>nd</sup> Midterm exam	
5/22	Multiple Regression	Ch 17
5/29	Regression Analysis: model building	Ch 18
6/5	Time Series Analysis and Forecasting	Ch 20
6/12	Time Series Analysis and Forecasting	Ch 20
6/19	Final Exam	

二個母體

三個母體

小樣本

## 二、資料種類(1)

- 資料的型態有三種
  - 名目資料(Nominal data)
    - 資料的值是不同的類別
  - 順序資料(Ordinal data)
    - 看起來像名目資料，但是它們的數值是有順序的
  - 區間資料(Interval data)
    - 是真實的數字，諸如身高、體重、所得和距離之類
    - 李克特五點 量表(Likert scale)
    - 長度0公分就是沒有長度，時間0秒是「沒有時間」，但是溫度攝氏0度只是水的冰點，不是「沒有溫度」
- 屬質(Qualitative) 和屬量 (Quantitative) 資料
- 橫斷面資料(Cross-Sectional)：同一時間點，不同人的比較。
  - EX:2016年，7-11與全家的營收比較
- 時間序列資料(Time-Series Data)：同一個人，不同時間的比較
  - EX: 7-11 營收是否有成長？(2015-2016年)

## 二、下列是什麼資料型態？

### 陳林滿天下

附表:台閩地區前二十大姓氏統計

單位：萬人

姓氏	陳	林	黃	張	李	王	吳	劉	蔡	楊
排名	1	2	3	4	5	6	7	8	9	10
人數	250.9	187.1	135.8	119.2	116.0	92.9	91.4	71.6	65.9	59.8
姓氏	許	鄭	謝	洪	郭	邱	曾	廖	賴	徐
排名	11	12	13	14	15	16	17	18	19	20
人數	52.6	42.6	39.8	33.9	33.8	33.1	32.5	30.5	29.8	28.5

備註：九十三年一月底臺閩地區總人口數約為 2,260 萬人。

- 1) Nominal (名目尺度)
- 2) Ordinal Data (順序尺度)
- 3) Interval (區間尺度)

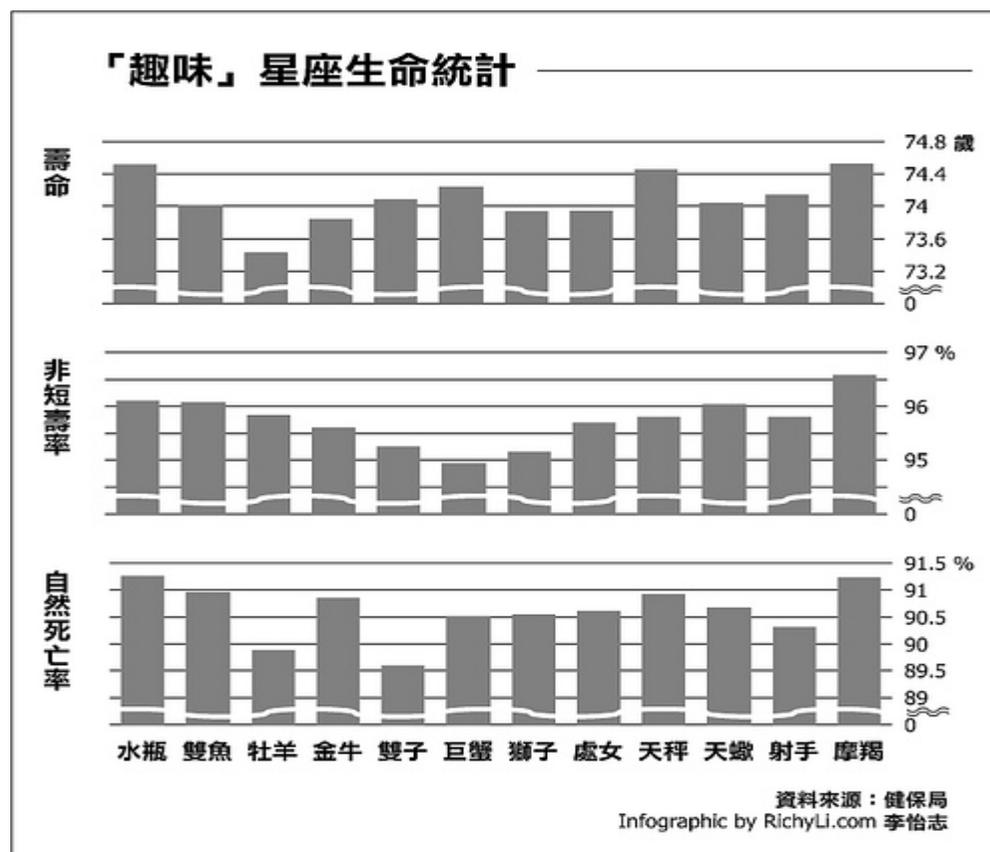
答：Nominal data

哪裡有問題？如何解釋？

## 二、下列是什麼資料型態？

### 星座統計

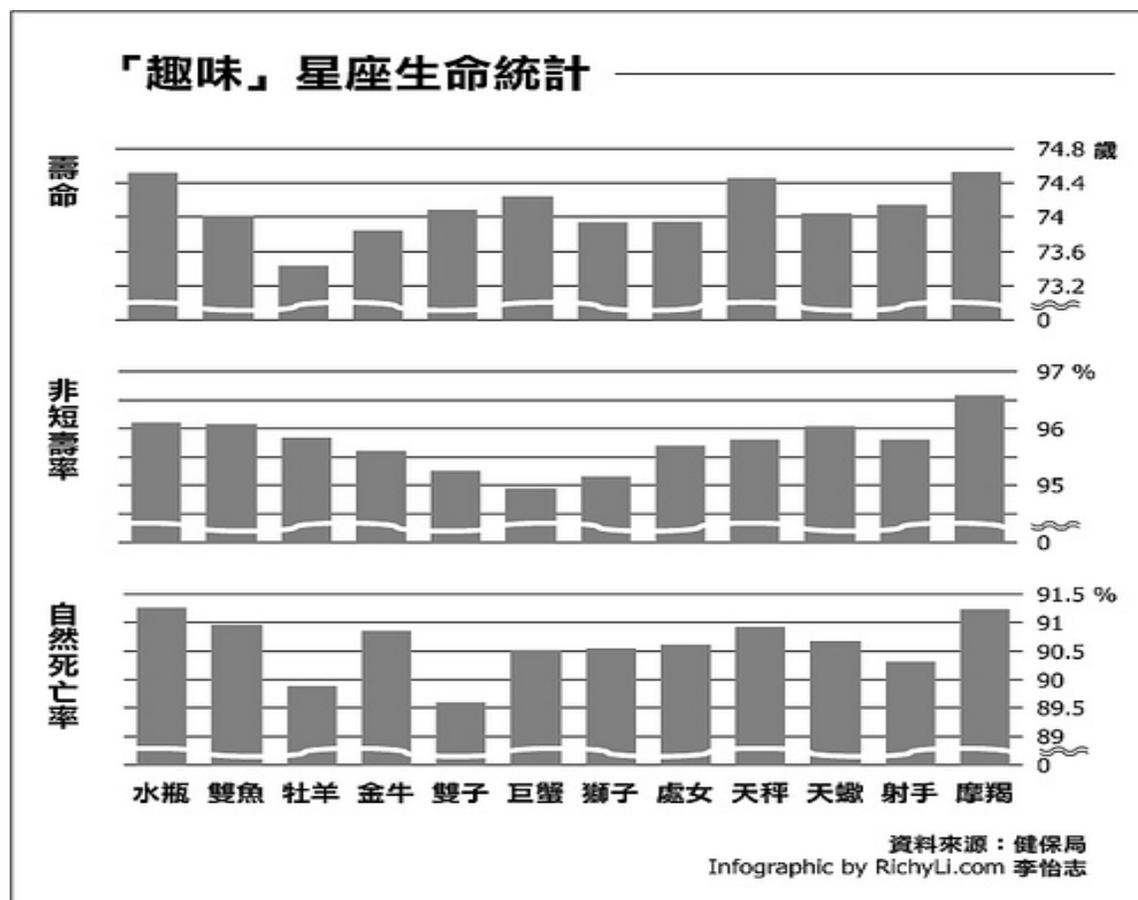
- 健保局統計室提供的2006年台灣地區死亡總人數共十三萬五千多人資料，以西洋十二星座日期進行分析，發現一些有趣的統計。
- 你們如何解釋？
  - 統計發現，平均死亡年齡最年長的前三個星座分別是魔羯座、水瓶座、天秤座（七十四點四六歲）
  - 大致上有一個趨勢，就是春、夏生的人比秋、冬生的人，更要好好珍惜生命



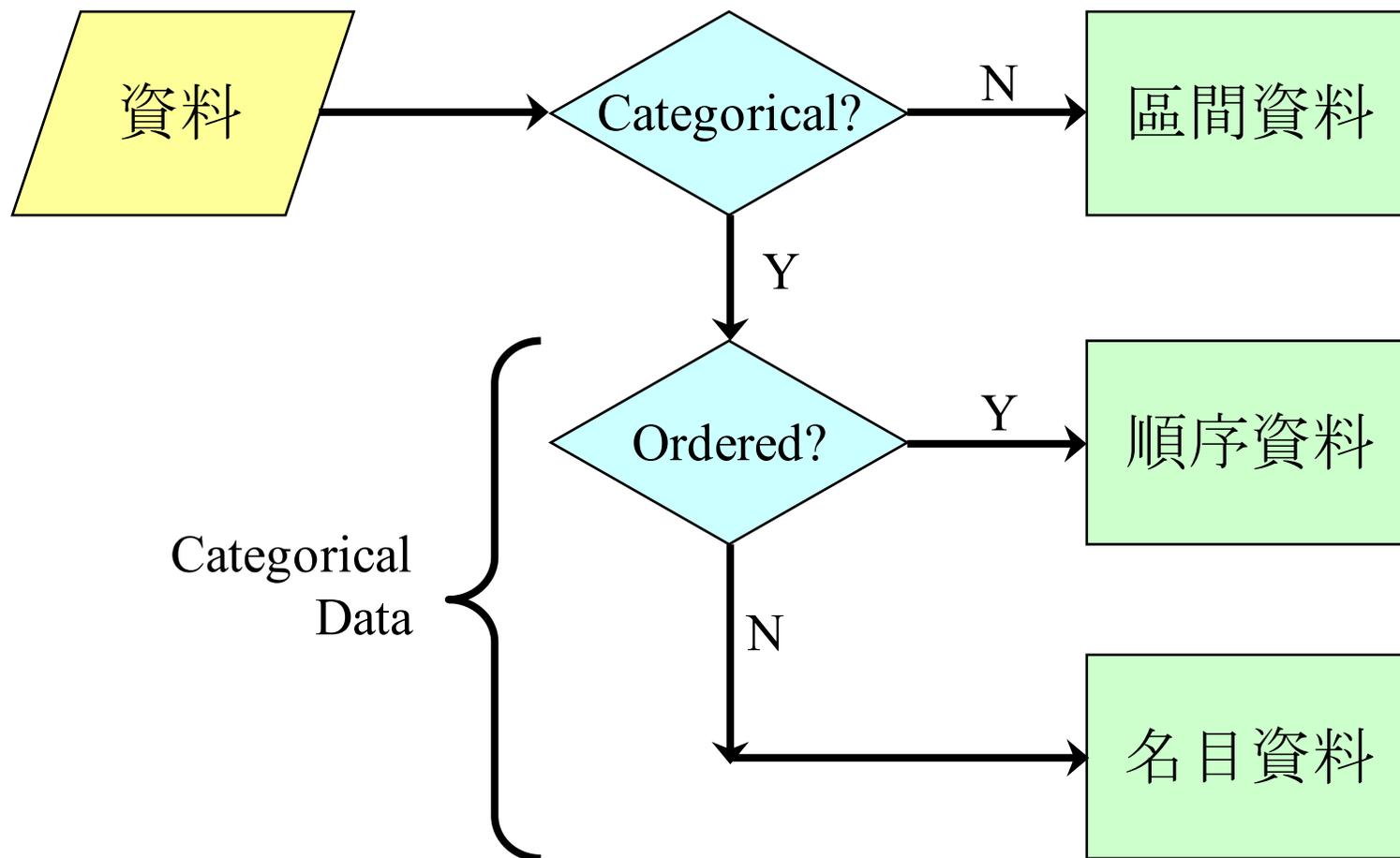
# 下列是什麼資料型態？

## 星座統計

- 健保局統計室提供的2006年台灣地區死亡總人數共十三萬五千多人資料，以西洋十二星座日期進行分析，發現一些有趣的統計。
- 這是什麼資料型態？
  - 1) Nominal (名目尺度)
  - 2) Ordinal Data (順序尺度)
  - 3) Interval (區間尺度)
- 這是什麼圖型？
  - 1) Bar chart (長條圖)
  - 2) Histogram (直方圖)
- 這是屬於下列哪種？
  - 1) Frequency
  - 2) Relative frequency



## 二、各種資料 & 資訊...



### 三、各圖形怎麼畫-長條圖與圓型圖

- 長條圖和圓型圖(Bar chart and pie chart)
  - 先選取你要跑出資料(整理過資料)→插入→選取直條圖(Bar) or 圓形圖(Pie)，圖型便會自動跑出
  - 如要修改圖表標題，先點選圖表→點選Excel上方圖表工具之版面配置→可以點選圖表標題、坐標軸標題、資料標籤等進行修改

- 若只有原始資料，可以使用Excel函數的countif計算出各種分類下的次數

- 在任何一格鍵入=countif(Input range, Criteria)
- Input range 為所要輸入的範圍，例 B1:B286
- Criteria 為計算的準則，例鍵入1即可算出資料為1的次數
- 如果是要算有幾個 A +呢？
- Countif(Input range, “A+”)

編號	個數
1	70
2	89
3	22
4	30

編號	個數
A +	70
A	89
A -	22
B+	30

# 三、各圖形怎麼畫--直方圖

編號	數值
1	70.3
2	70.5
3	80
4	82
5	82.4
6	77
7	78.5

- 直方圖(Histogram)
  - 要先決定好分幾組，並輸入每組的上界於Excel
    - **Sturges's rule:  $1+3.3*\log(n)$**
    - 依照課本表3.2選取或方便性選取
  - 點選Excel上方資料→資料分析→直方圖→將輸入範圍與組界範圍選取好
    - 標記表示資料是否有選取到標題，如有選到必須打勾
    - 記得勾選圖表輸出
  - 直方圖出現後進行下列步驟修改
    - 將組界為其它且頻率為0那一系列內容刪除
    - 點選圖表上直條圖並按滑鼠右鍵→選取資料數列格式→將列別間距改為無間距
  - 每組是大於下界，小於等於上界(第一組除外)

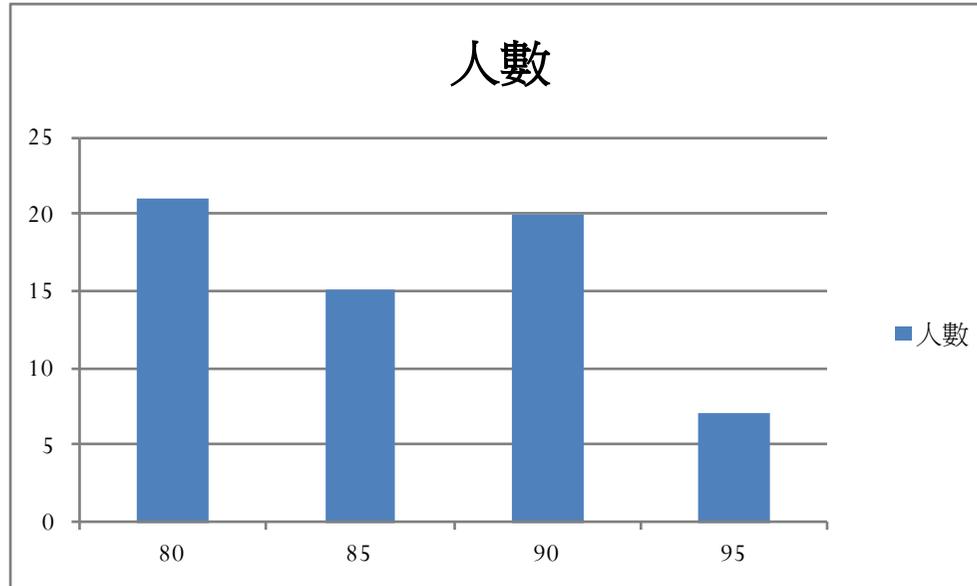
## 三、各圖形怎麼畫--累積機率圖

- 累積機率圖(Ogive)
  - 點選Excel 上方資料→資料分析→直方圖→選取輸入範圍與組界範圍，不要選標記(組界範圍第一個要多加一個第一組的下界)(勾選圖表輸出與累積百分率)
  - 圖型出現後進行下列步驟修改
    - 將組界為其它且頻率為0那一系列內容刪除
    - 將出現之直方圖刪除(點選任一矩形，並按滑鼠右鍵刪除)
    - 改變刻度大小(如果直軸最大值不為1):點選直軸→並按滑鼠右鍵點選坐標軸格式→將坐標軸選項中的最大值改為1
    - 點選橫軸→並按滑鼠右鍵點選坐標軸格式→將座標軸位置中刻度與刻度之間相距(W)改成刻度上(K)
  - 或是用累積相對機率畫圖。
    - 插入→直線圖

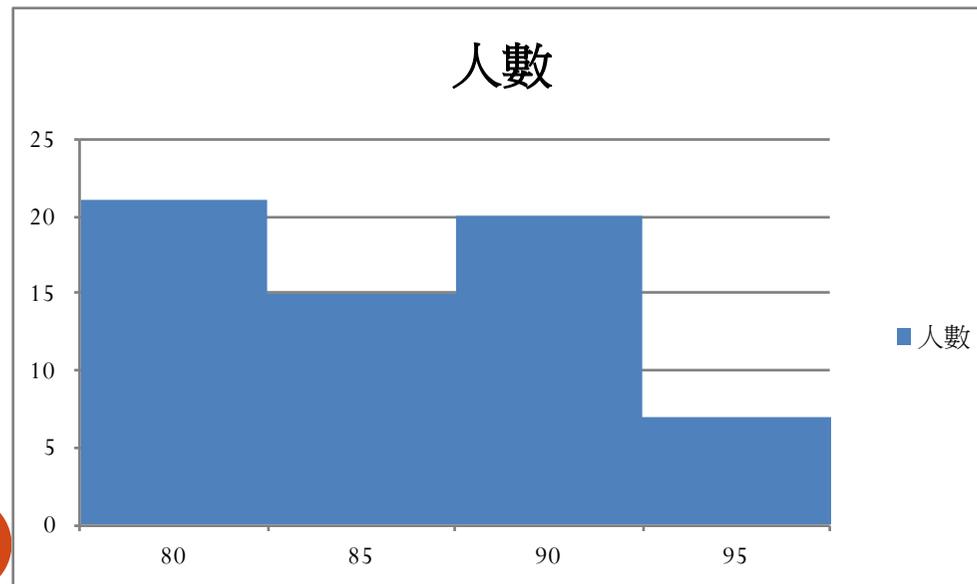
## 三、各圖形怎麼畫--莖葉圖

- 莖葉圖(Stem-and-Leaf display)
  - 增益集 → Data Analysis Plus → Stem-and-leaf- Display
    - Input range 要輸入資料範圍
    - Increment 表示 difference between stems
    - 資料如有選取標題則要勾選 Labels
- 折線圖(Line chart)
  - 時間序列資料
  - 選取資料 → 插入 → 折線圖
- 散佈圖(Scatter diagram)
  - 探討兩變數之間的關係
  - 選取資料 → 插入 → 散佈圖

### 三、Bar chart and Histogram有何差異？



分數	人數	比例
80 -84	21	33.333
85-89	15	23.810
90-94	20	31.746
95-100	7	11.111
Toatl	63	100.000



- 1) 這兩張圖分別是？
- 2) 哪一張圖才是對的？

# 三、莖葉圖如何解讀？

- 統計第一次小考成績如下，請問

1. 請問4, 2表示什麼？
2. 請問考0分有幾個？
2. 請問考30分有幾個？

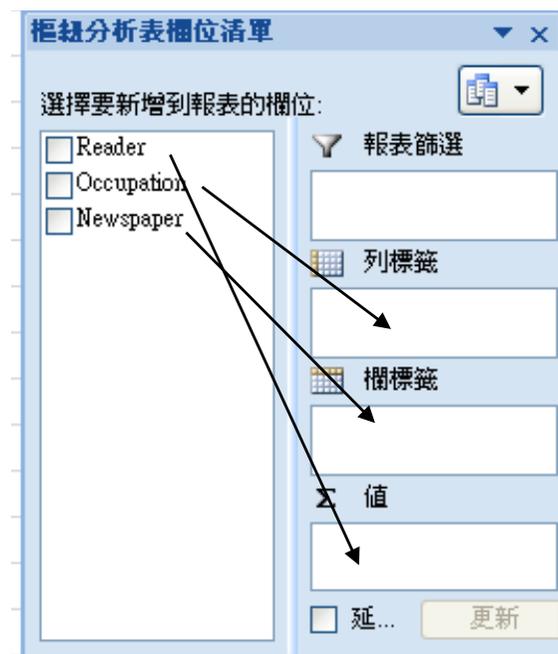
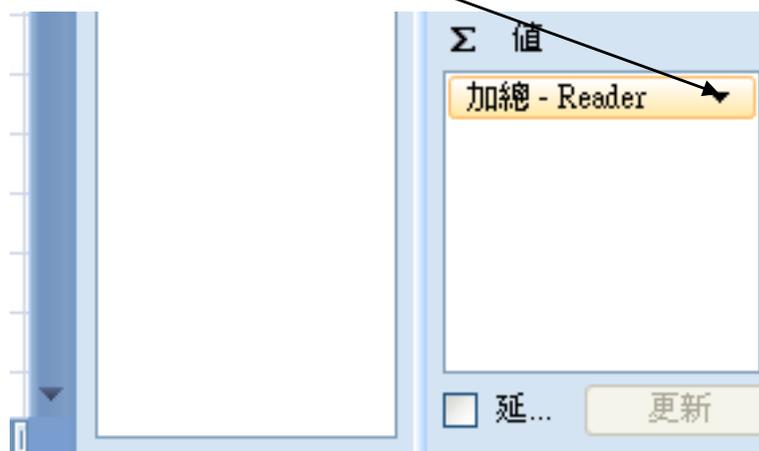
## Stem Leaf

0	000000000111112222223333345555556666666778888999999
1	00001111233333334455555667889999
2	000011112344666778999
3	001335589
<b>4</b>	<b>1</b> 24445589
5	33566
6	3458
7	022224556789
8	334457889999
9	00112222233344555999
10	001344446699
11	124557889

Thus, we still have access to our original data point's **value!**

## 四、樞紐分析表(Pivot Table)

- 為例樞紐分析表(以 Example 2.4 )
  - 插入→樞紐分析表→選取好你要的資料範圍→按確定
  - 右邊會出現樞紐分析表欄位清單→可用拖曳方式將想要的資料拉到列與欄，可得到cross-classification table
  - 在此處更改直欄位設定，改成項目個數（要用count人數）



2. The Nielsen Home Technology Report provided information about home technology and its usage. The following data are the hours of personal computer usage during one week for a sample of 50 people.

4.1	1.5	10.4	5.9	3.4	5.7	1.6	6.1	3.0	3.7
3.1	4.8	2.0	14.8	5.4	4.2	3.9	4.1	11.1	3.5
4.1	4.1	8.8	5.6	4.3	3.3	7.1	10.3	6.2	7.6
10.8	2.8	9.5	12.9	12.1	0.7	4.0	9.2	4.4	5.7
7.2	6.1	5.7	5.9	4.7	3.9	3.7	3.1	6.1	3.1

Summarize the data by constructing the following:

- A frequency distribution (use a class width of three hours)
- A relative frequency distribution
- A histogram
- An ogive
- Comment in what the data indicate about personal computer usage at home.

# 第一步：先key資料

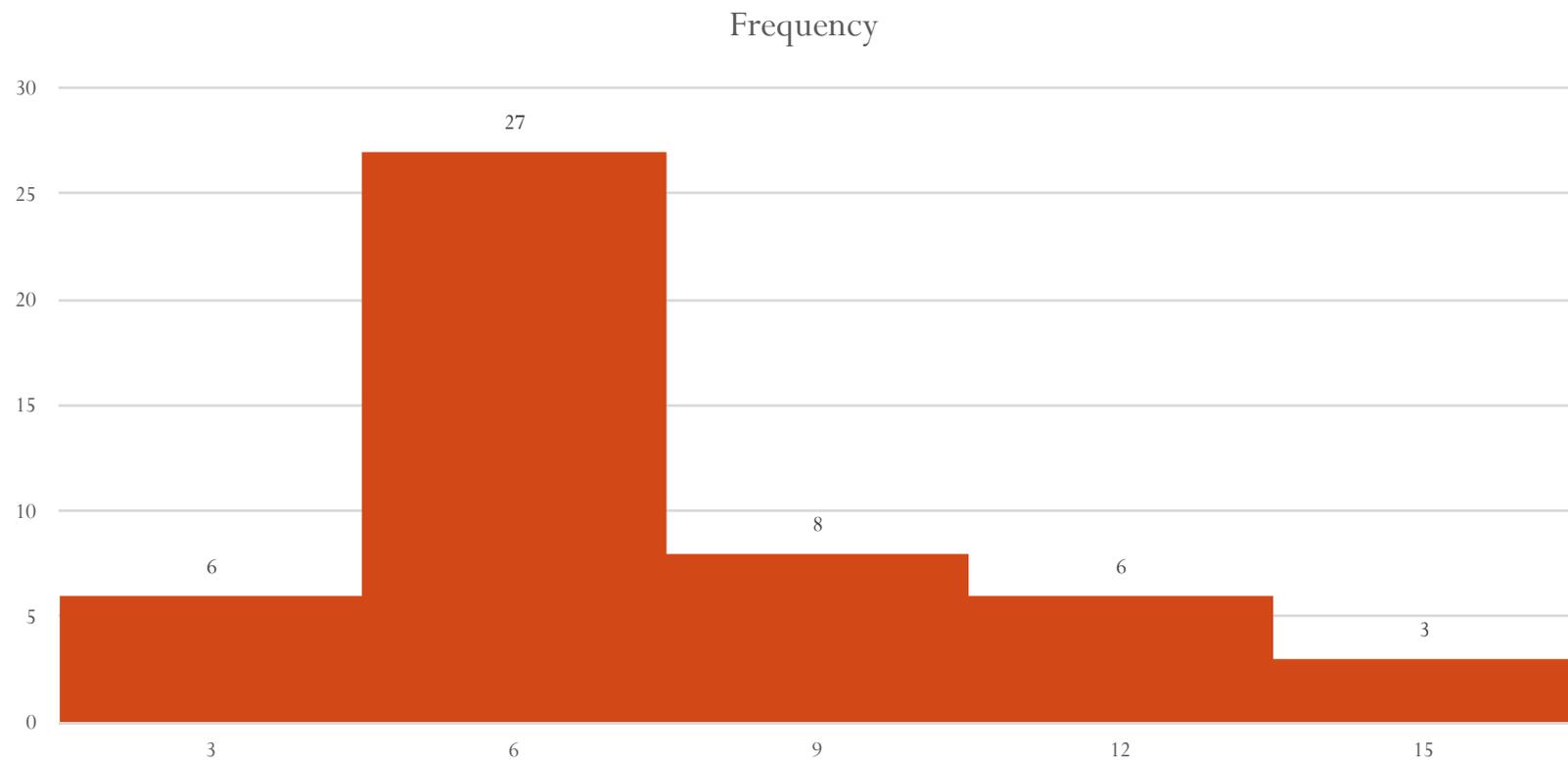
- (a) A frequency distribution (use a class width of three hours)
- (b) A relative frequency distribution
  - 排序（由小到大）
  - 依序分類

題目問的是distribution,  
所以要呈現右表

Hours	Frequency	Relative frequency
0~3	6	0.12
3~6	27	0.54
6~9	8	0.16
10~12	6	0.12
12~15	3	0.06
Toatl	50	1

- C) A histogram
- 因為題目已經提示用3hour為一組，無需再取組距
- 如果題目沒給提示，可以用下列公式
- Sturges's rule:  $1+3.3*\log(n)=1+3.3*\log(50)=6.60$
- (最大-最小) / 幾組 =  $(14.8-0.7)/7=2.01$  (取2個為一組)
- 點選Excel上方資料→資料分析→直方圖→將輸入範圍與組界範圍選取好
  - 標記表示資料是否有選取到標題，如有選到必須打勾
  - 記得勾選圖表輸出
- 直方圖出現後進行下列步驟修改
  - 將組界為其它且頻率為0那一列內容刪除
  - 點選圖表上直條圖並按滑鼠右鍵→選取資料數列格式→將列別間距改為無間距

# 記得histogram圖要連起來



- D) An ogive
- 用剛剛寫a) b)兩題的資料

	(a)	(b)	(c)
Hours	Frequenc y	Relative frequency	accumulate relative frequency
0~3	6	0.12	0.12
3~6	27	0.54	0.66
6~9	8	0.16	0.82
0~12	6	0.12	0.94
12~15	3	0.06	1
Toatl	50	1	

$$0.66=0.54+0.12$$

$$0.82=0.66+0.16$$

- D) An ogive
- 用剛剛寫a) b)兩題的資料,算出累積的相對次數(b)
- 選取資料→插入→折線圖(line Chart)

