



25種統計方法秘技

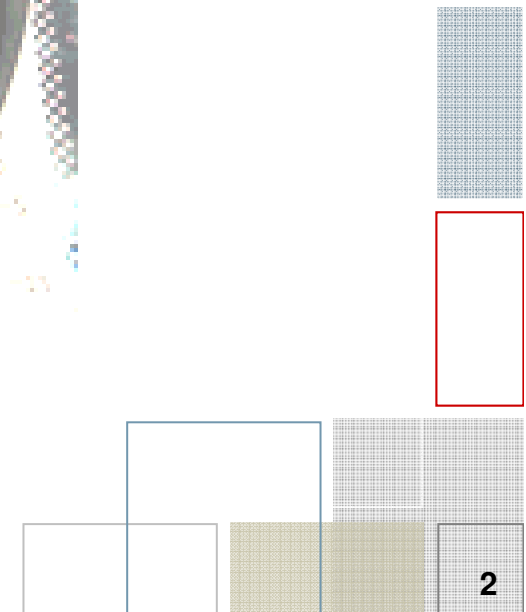
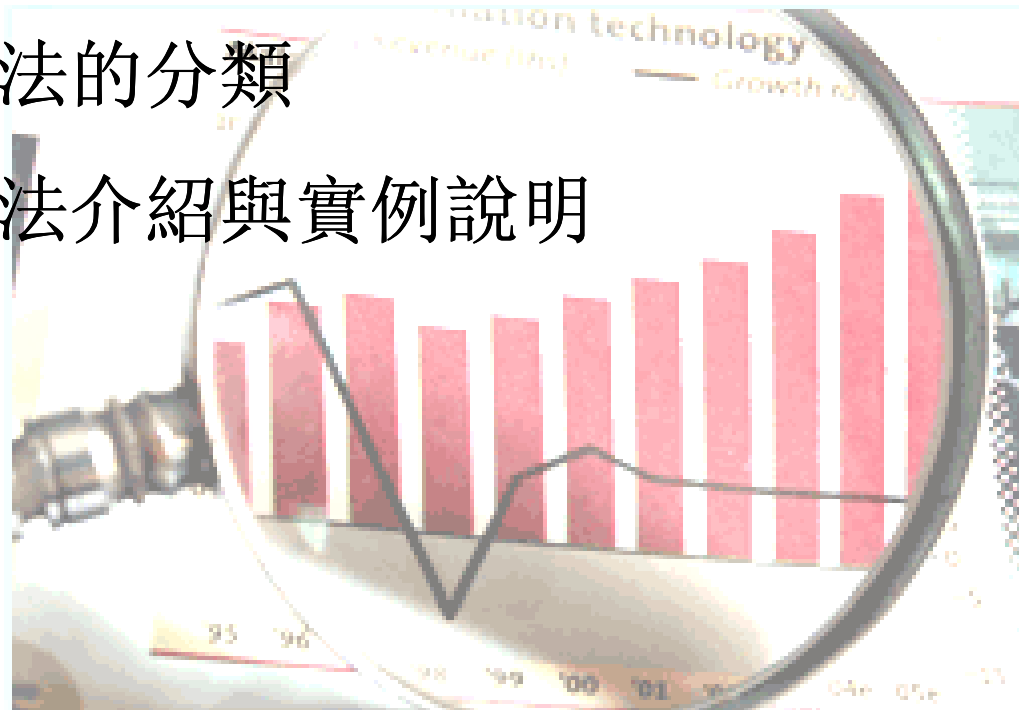
(I)

SPSS14.0中文版

SPSS Taiwan — 宏德國際

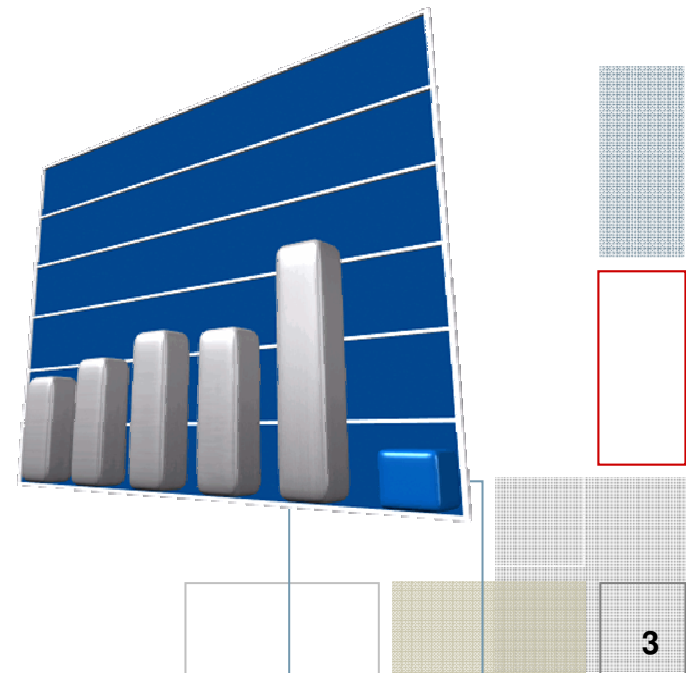
> Agenda

- 變數量尺種類介紹
- 統計方法的分類
- 統計方法介紹與實例說明



› 變數量尺種類(measurement scales)

- 分類型變數(因子)
 - 名義量尺(Nominal Scale)
 - 次序量尺(Ordinal Scale)
- 數值型變數(共變量)
 - 等距量尺(Interval Scale)
 - 等比量尺(Ratio Scale)



變數量尺種類

名義量尺(Nominal Scale)

由變項所擁有的某種特徵，分成幾個**互斥的類別**，類別與類別間並沒有必要的關聯

次序量尺(Ordinal Scale)

由變項特徵或屬性的大小順序加以排列而得。而在次序量尺上任取兩相鄰等級，其間的**間隔並不一定相等**

員工代碼： 001
性別：男、女
生日 2/3/52
教育程度(年)： 15
職別：職員、保全人員、管理人員
目前薪資： 57000
起薪： 27000
在本公司的年資(月)： 98
以前的資歷(月)： 144
種族區隔：有、無

變數量尺種類

員工代碼：_____ 001 _____
性別：男、女
生日 2/3/52
教育程度(年)：_____ 15 _____
職別：職員、保安人員、管理人員
目前薪資：_____ 57000 _____
起薪：_____ 27000 _____
在本公司的年資(月)：_____ 98 _____
以前的資歷(月)：_____ 144 _____
種族區隔：有、無

比例量尺(Ratio Scale)

量尺上任意兩個相鄰數字間的時間隔兩兩相等且還有一個真實的零點，由於比例量尺具有以上兩個特點，測量所得的數字可以用加減乘除等所有的算術運算來加以處理。

等距量尺(Interval Scale)

等距量尺上任兩個相鄰數字間的時間隔均相等。等距量尺並沒有真實的零點，因此用等距量尺測量兩個人或物之屬性，所得的兩個數字之間並沒有比例關係存在。

變數量尺種類

■ 類別資料的產生

■ 原發性類別資料(自然分類)

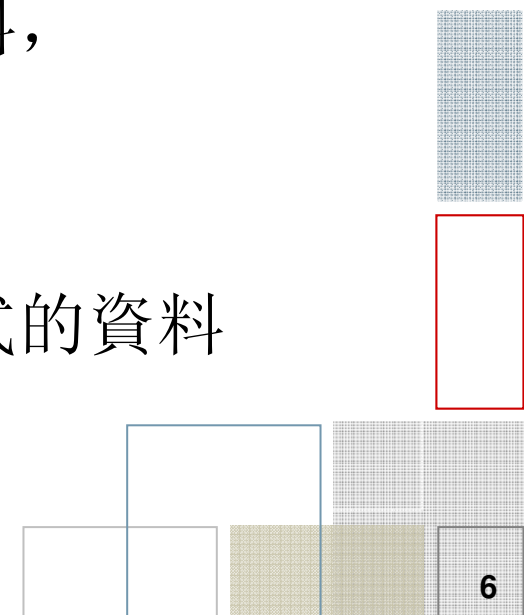
當被測定的變項的本質是名義性的屬性，
例如性別資料。

■ 操弄性類別資料

以人為操作的手段所獲致的分類性資料，
例如實驗操弄的分類結果。

■ 虛擬化類別資料

由其他類型的資料型態轉換成類別形式的資料，
例如由連續變項轉換來的類別變項。



直覺化變數類型圖示

測量水準	資料類型			
	數值	字串	日期	時間
尺度		無		
次序				
名義				

複選題分析表格

列(W):

欄(C):

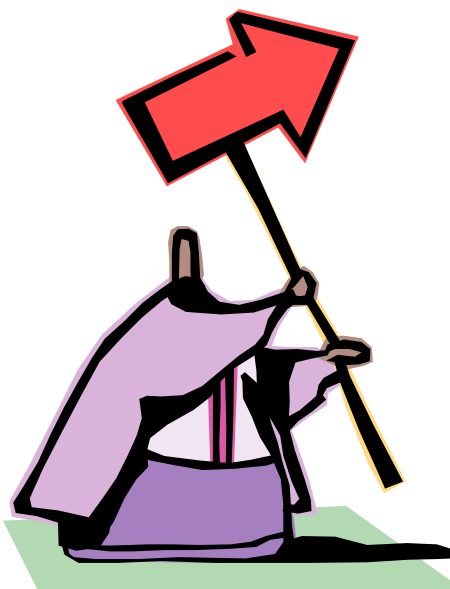
圖層(L):

巢狀:

- 列(C)
- 行(R)
- 圖層(A)

變數列表 (紅圈):

- 員工代碼 [id]
- 性別 [gender]
- 生日 [bdat]
- 教育程度 [E]
- 職別 [jobcat]
- 目前薪資 [salar]
- 起薪 [begin]
- score
- way

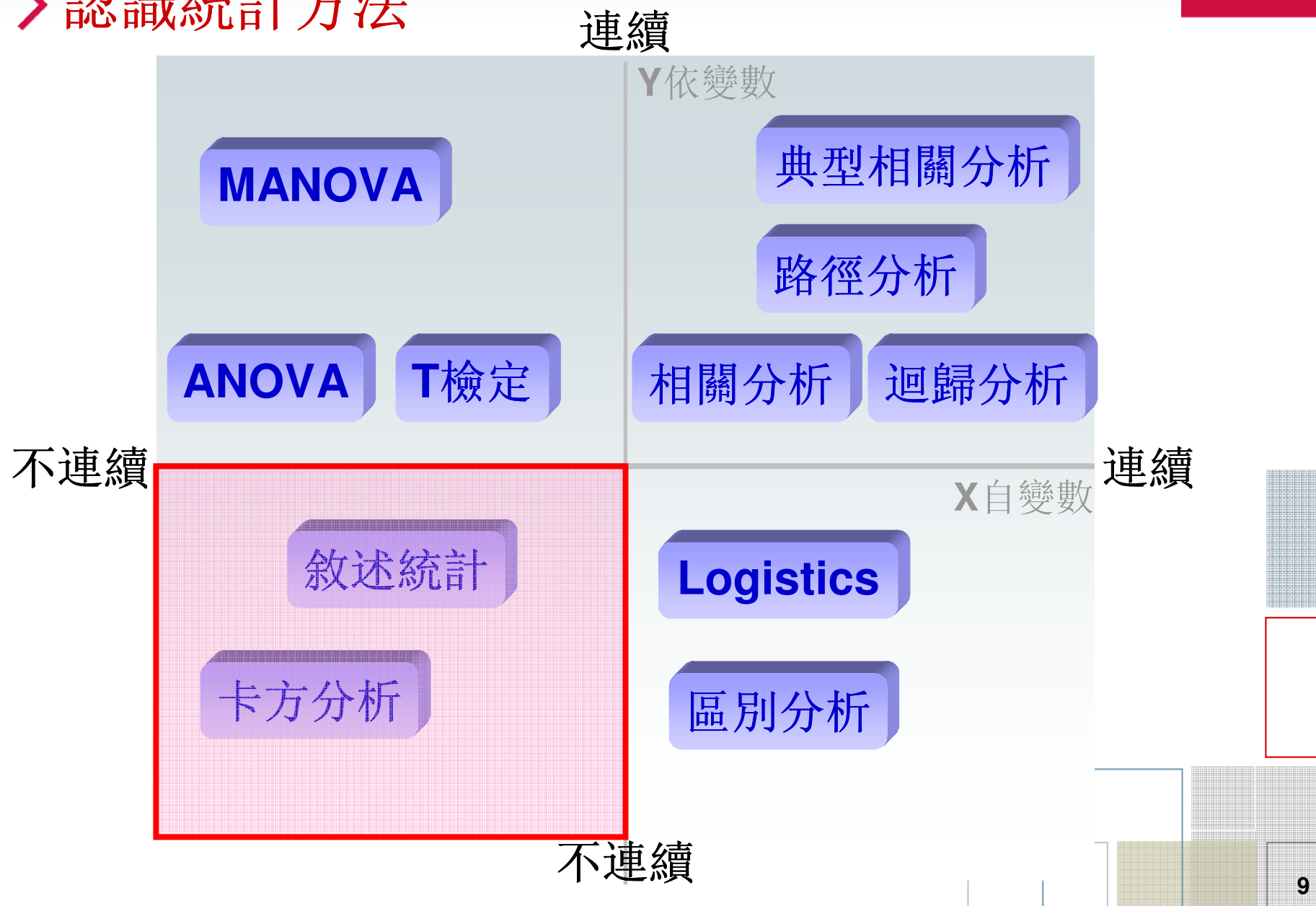


複選題集，多重類別	
複選題集，多重二分集	

統計方法的分類

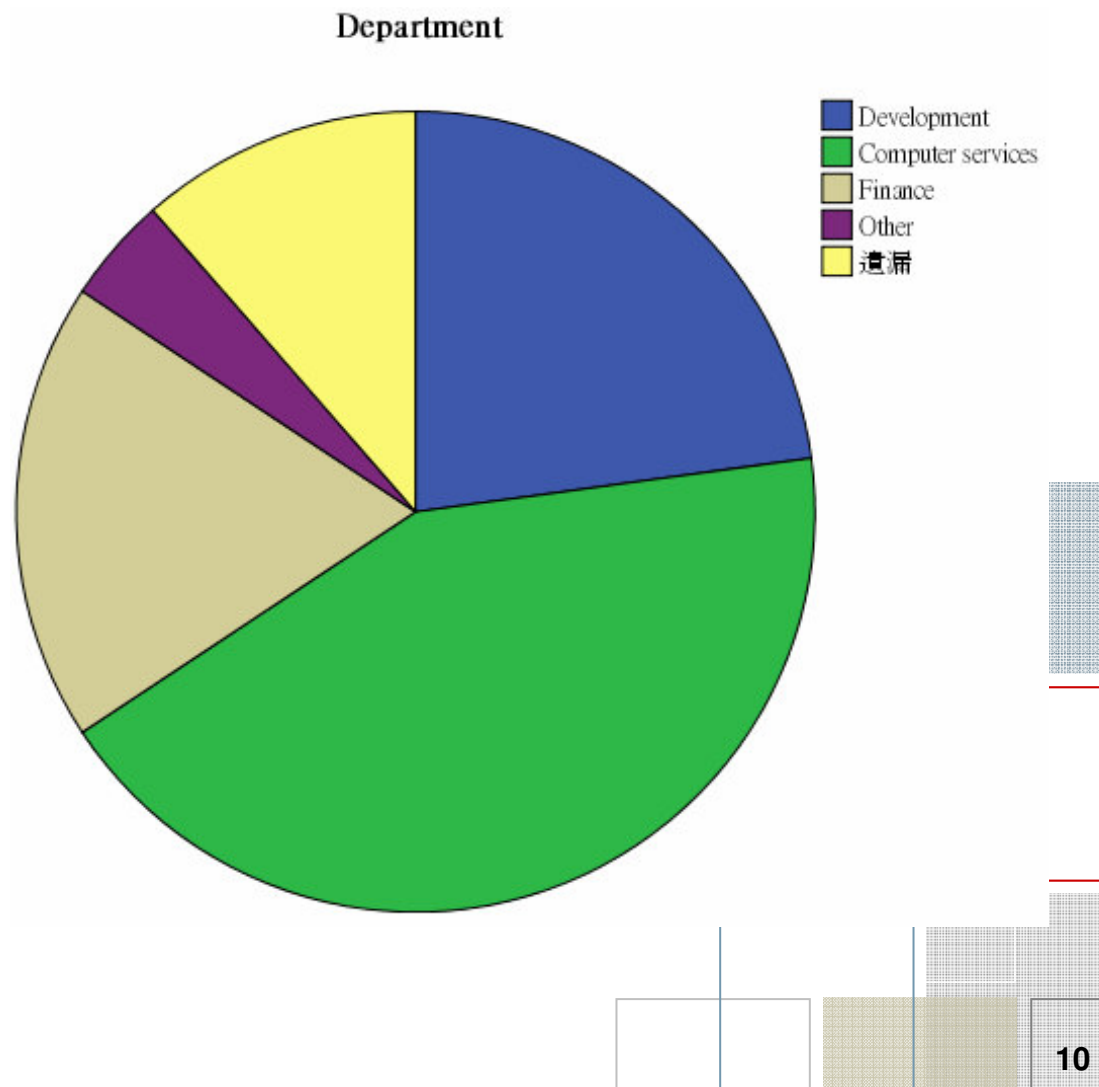
自變項 依變項		等距或等比(屬量)		類名、順序(屬質)	
		1個變數	2個變數及以上	1個變數	2個變數及以上
無分自變數依變數		常態分配檢定 描述統計 單樣本t檢定 時間序列	簡單迴歸 主成份分析 因素分析 集群分析	適合度卡方檢定	對數線性模式 卡方獨立性檢定
等距或等比(屬量)	1個變數	簡單迴歸 簡單相關 生存分析 時間序列— 轉換模式	多元相關 多元迴歸 多元複迴歸 生存分析	T檢定 單因子ANOVA 生存分析	多因子ANOVA 生存分析 Multiple-classification分析 變異數分析
	2個變數及以上	多元相關 典型相關	典型相關 因徑分析 結構模式 (PLS, Amos)	MANOVA Hotelling-T ² Profile analysis	MANOVA
類名或順序(屬質)	1個變數	區別分析 Logistic probit 雙樣本t檢定	區別分析 Logistic probit	卡方檢定 間斷型的區別分析	對數線性模式 Logistic迴歸 間斷型的區別分析 聯合分析
	2個變數及以上	區別分析	區別分析	對數線性模式 間斷型的區別分析	對數線性模式 Latent Class

> 認識統計方法



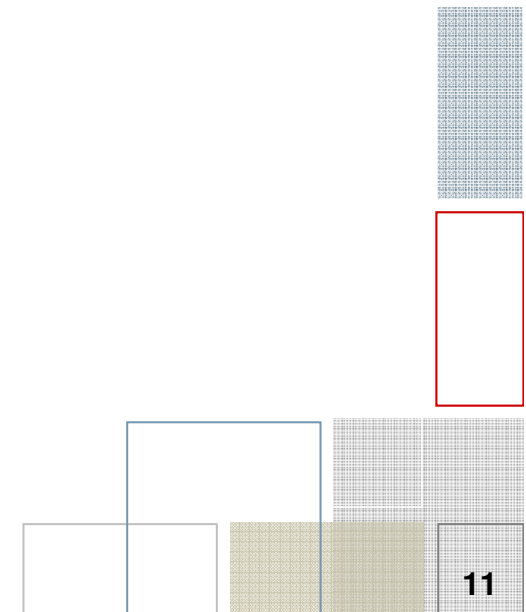
> 敘述統計

- 次數分配表
- 交叉分析
- 多重交叉分析
- 複選題分析



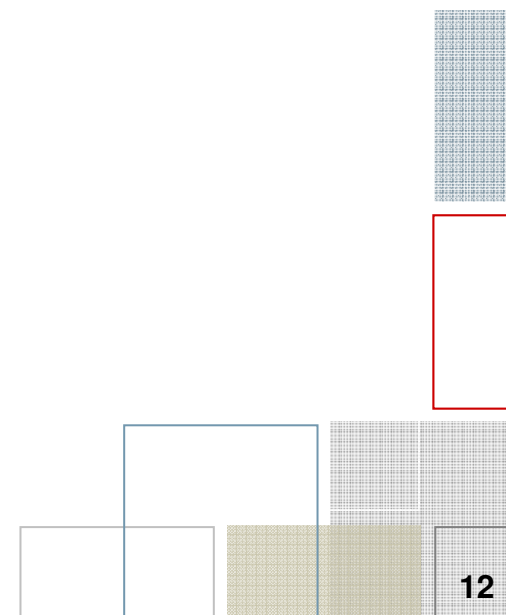
› 次數分配表

- 次數分配將列出該變數**所有可能值**（組別）與其**出現次數**的表列彙總，如**性別變數**中的可能值為**男與女**、**居住地變數**中的可能值為**北、中與南**。
- 此外，次數分配表對於**檢查資料輸入是否正確**非常重要，若出現不可能值，則代表原始資料輸入有誤，例如：性別變數出現不是男或女的數值。



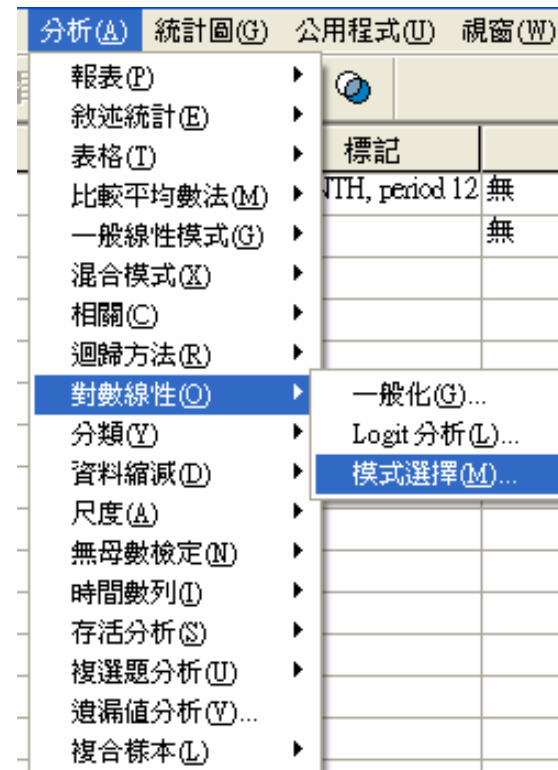
› 交叉分析(列聯表分析)

- 若是單一的分類變數，分成r類，可得到次數分配表。
- 若是兩個類別變數，則可進行交叉分析並進行卡方(Chi-Square, χ^2)檢定。
- 若變數一分為r類，變數二分為c類，可得r×c交叉表(Cross-tabulation)或稱交叉分析表，列聯表(Contingency table)。
- 例如
性別：男、女
政黨：國民黨、民進黨、親民黨、台聯黨



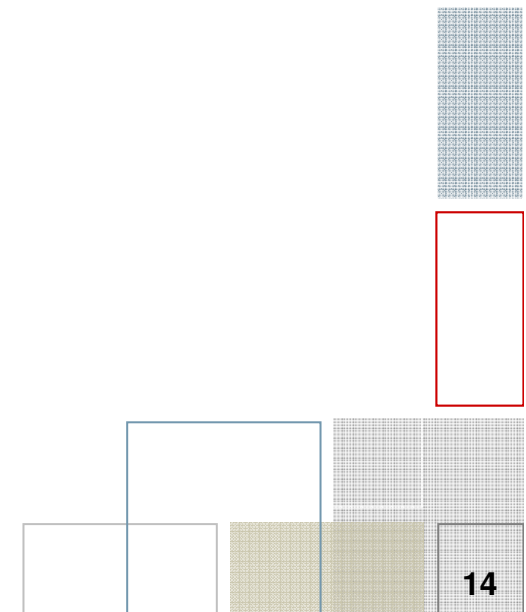
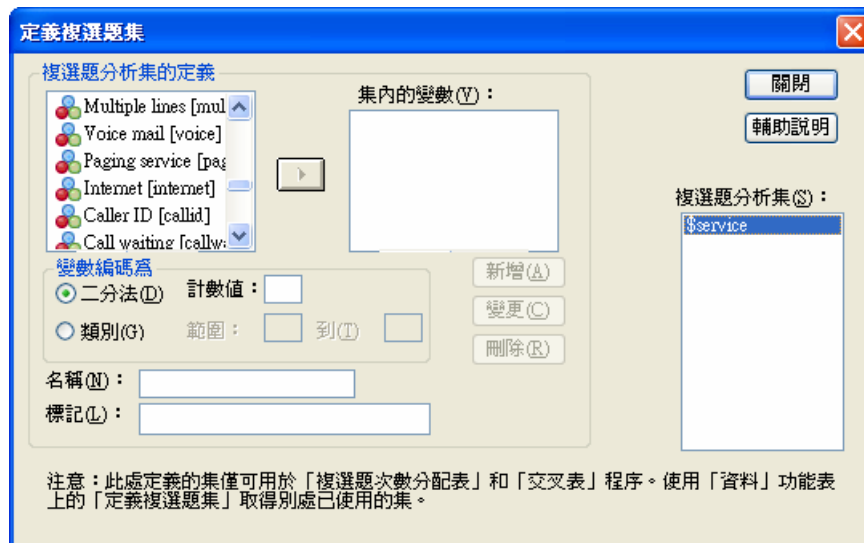
> 多重交叉分析(多重列聯表分析)

- 變數須為名目尺度(不連續變數)
 - 性別：男、女
 - 地區：如北、中、南
 - 政黨：國民黨、民進黨、親民黨、台聯黨
- 三個變數以上，但仍以三個為主



➤ 複選題分析

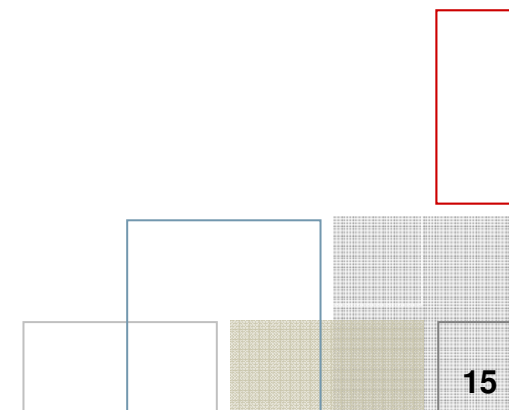
- 複選題只是請受訪者勾選**有**或**無**，亦即該選項只能在建立資料時，以名目尺度的1與0來呈現。
- 複選題經常被使用，但複選題所能做的統計分析只有**次數分配表**與**交叉分析表**等描述性統計。



› 卡方分析

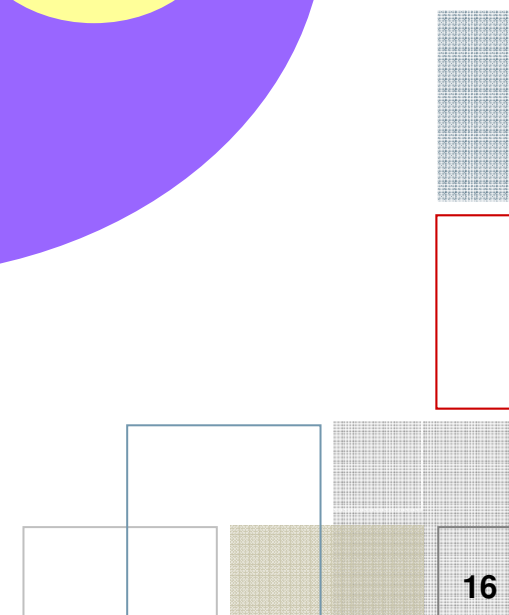
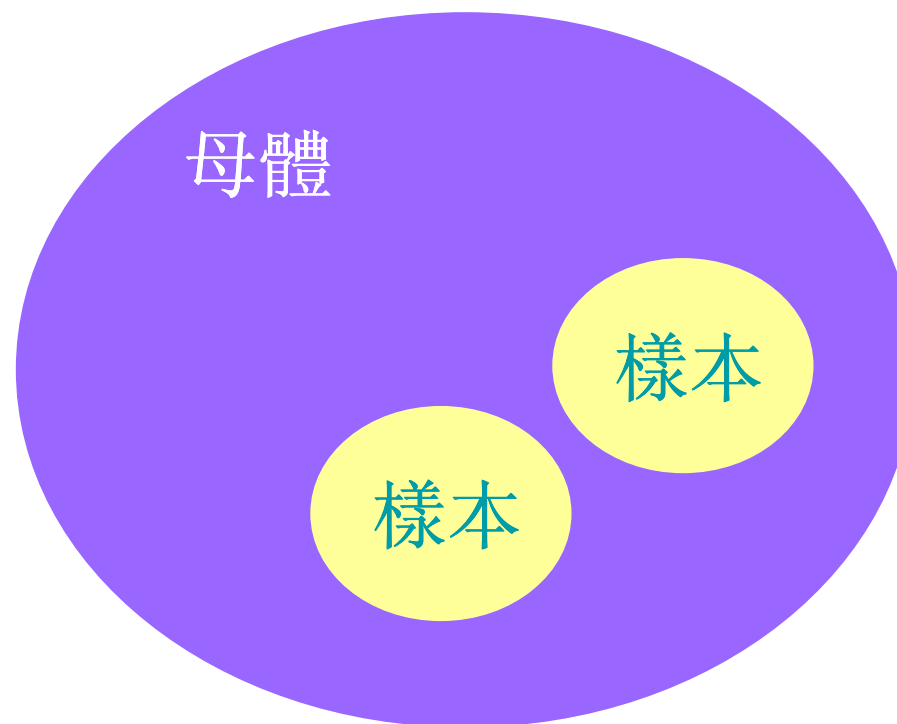
- 探討兩個類別變項關聯性的技術，稱為列聯表分析（即交叉分析）。
- 交叉分析是民調與市調最普遍被使用的統計方法。
- 如果搭配統計檢定技術，交叉分析則可以進行專業與學術的決策。
- 此種檢定稱為卡方檢定

	泛黑	泛白	未表態	總和
女性	0	30	20	50
	0%	100%	40%	50%
男性	20	0	30	50
	100%	0%	60%	50%
總和	20	30	50	100
	20%	30%	50%	



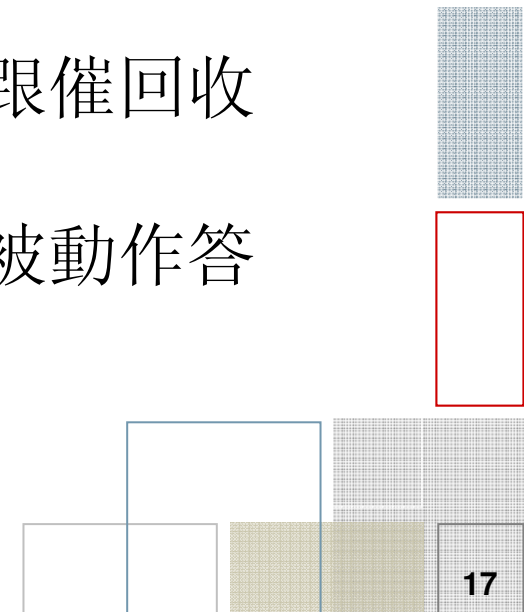
> 卡方分析

- 同質性檢定
- 適合度檢定
- 獨立性檢定



› 同質性檢定

- 目的
 - 檢定不同人口母群，在某一變項的反應是否具有顯著差異；亦即兩個樣本在同一變項中之分佈情形。
 - 檢測**雙樣本**在**單一變項**的分佈情形
- 適用時機
 - 郵寄問卷時，比較早期回收群及後期跟催回收群之人口統計變項。
 - 街頭訪問時，比較願意主動作答群及被動作答群之人口統計變項。



同質性檢定

- 調查男女大學生是否對政黨參與狀況有影響？

2 × 4 列聯表	國民黨	民進黨	親民黨	無黨籍	小計
男	75 15% 50% 75	75 15% 75% 50	50 10% 33% 62.5	300 60% 48% 312.5	500 50%
女	75 15% 50% 75	25 5% 25% 50	75 15% 67% 62.5	325 65% 52% 312.5	500 50%
小計	150 15%	100 10%	125 12.5%	625 62.5%	1000

$\chi^2 = (75-75)^2/75 + (75-50)^2/50 + \dots = 31 > 7.815$ (查表值); 拒絕 H_0 , 接受 H_1

意即大學生參加政黨的傾向具有性別差異之假設成立

期望值 f_e		X		Marginal total
		r1	r2	
Y	c1	AC/N	AD/N	A
	c2	BC/N	BD/N	B
Marginal total		C	D	N

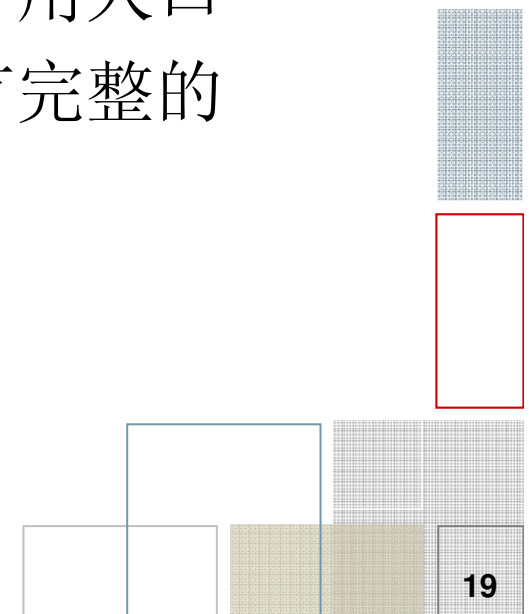
› 適合度檢定

- 目的

研究樣本是否抽樣母群分配相符合時，以卡方檢定進行之；每次檢定內容僅涉及一個變項。

- 適用時機

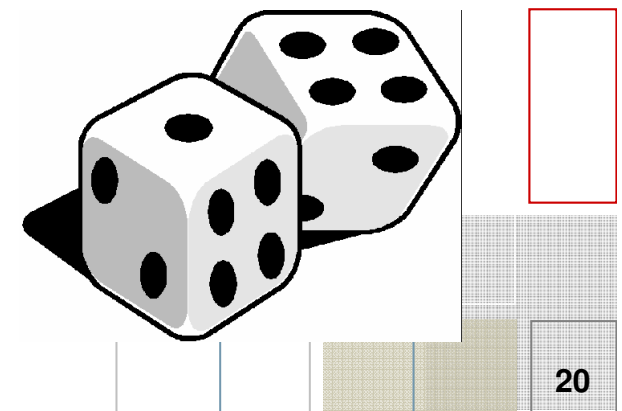
當研究者想知道樣本是否能代表母體時，用人口統計變項與母體資料比較。（如內政部有完整的人口統計資料）。



適合度檢定

- 某賭客懷疑莊家用的兩個骰子可能動過手腳，於是他暗地裡觀察了72局，紀錄了這72局中骰子組合點數各出現的次數…

骰子組合	2	3	4	5	6	7	8	9	10	11	12
次數	6	6	7	8	10	9	9	6	5	4	2



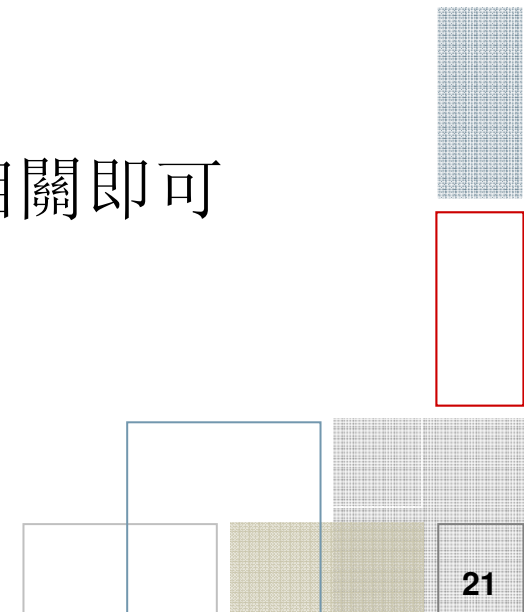
➤ 獨立性檢定

- 目的

- 同時檢測兩個類別變項（X與Y）之間的關係時，其目的在於檢測從樣本得到的兩個變項的觀察值，是否具有特殊的關聯。
- 檢測**同一個樣本**的兩個變項的關聯情形

- 適用時機

如研究者想知道學歷與性別之間是否有相關即可採用之。



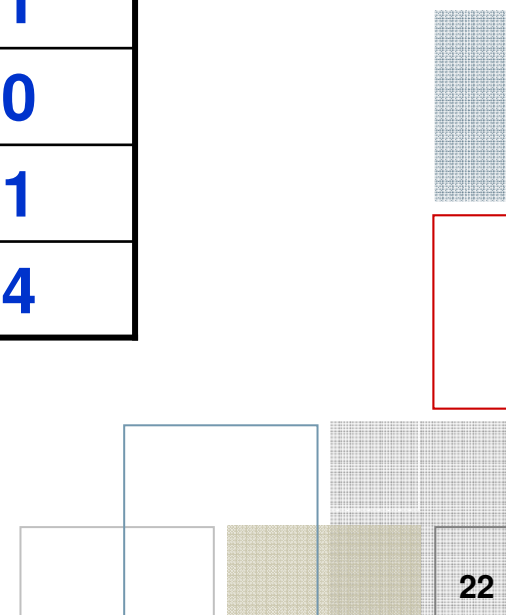
➤ 獨立性檢定

- 某學者想要了解教育程度的不同是否和工作薪資的多寡有關，於是蒐集了332位民眾的薪水資料…

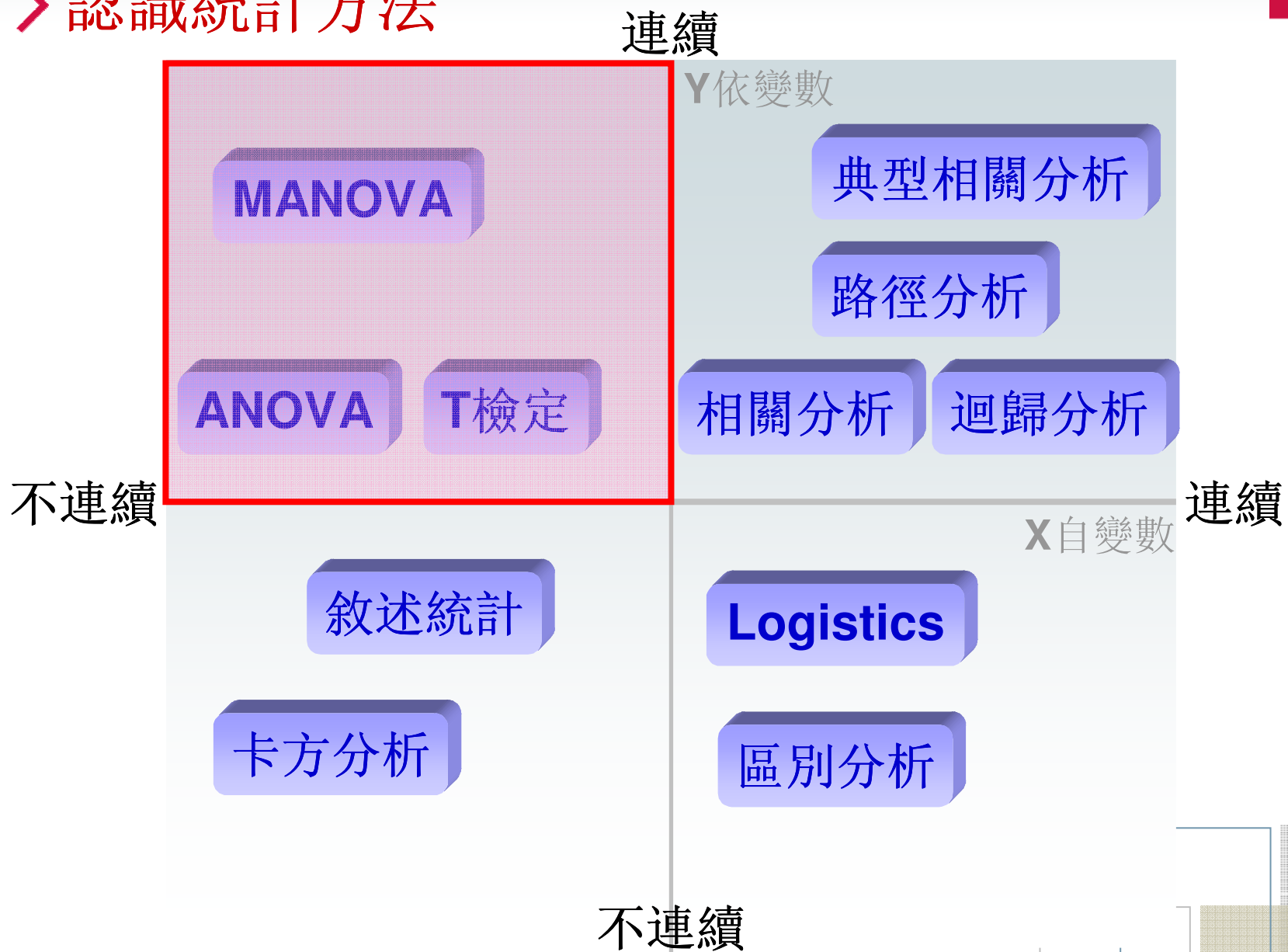
工作薪資

教育程度

	三萬以下	三萬到五萬	五萬以上
國小以下	36	28	11
國中	34	30	10
高中	23	22	21
大學以上	28	45	44

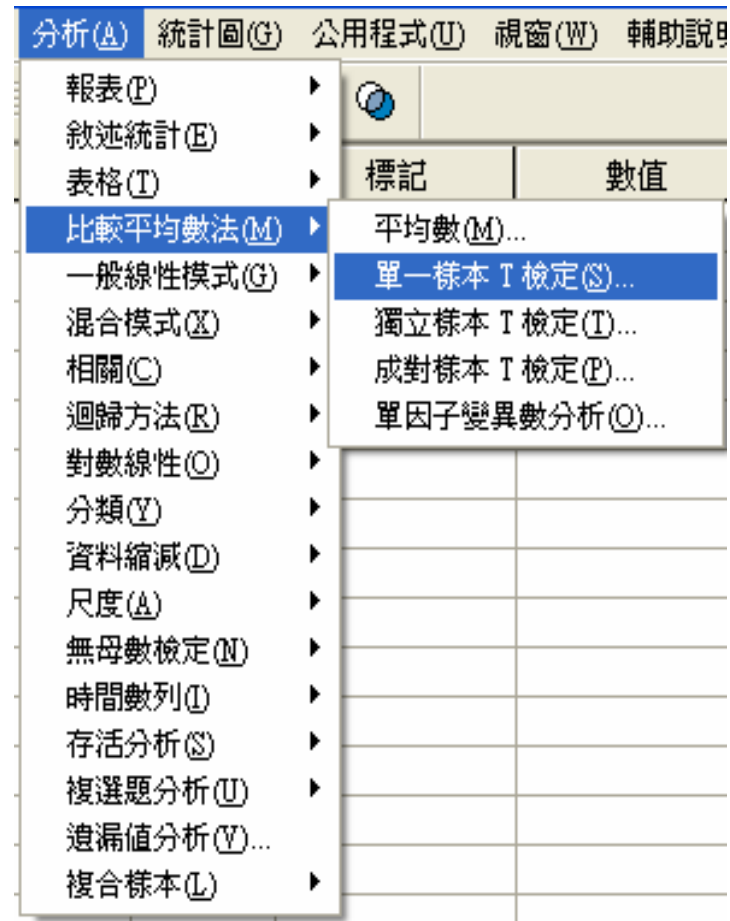


➤ 認識統計方法



➤ T檢定(t-test)

- 平均數的差異檢定
- 平均數比較除分為單母體與多母體檢定外，還可分為獨立樣本與相依樣本兩類。
- 相依樣本設計又可分為
 - 重複量數設計
repeated measure design
 - 配對樣本設計
matched sample design



➤ T檢定(t-test)

電信業者利用流失傾向給分來評估顧客忠誠度, 給分從0分到100分, 超過50分以上的門號就有可能會流失. 資料分析人員要評估不同忠誠度的顧客在服務使用上是否有差異.

➤ 單因子變異數分析(one-way ANOVA)

- 根據單一因子變數 (自變數)，來產生數量依變數的單因子變異數分析。變異數分析是一種假設，用於檢定**數個平均數是否相等**。這項技術是二個樣本 t 檢定的延伸。
- 除了確定平均數之間是否存在差異之外，您可能想進一步知道哪些平均數不一樣。比較平均數的檢定類型有兩種：**演繹式對比**和 **post hoc 檢定**。對比是在進行實驗之前設定的檢定，而 post hoc 檢定則是在實驗完畢之後進行。您也可以檢定不同類別之間的趨勢。

› 單因子變異數分析(one-way ANOVA)

甜甜圈在調理的過程中，是否會吸收不等量的油脂？我們在實驗中設定三種油脂類型：花生油、玉米油、和豬油。花生油和玉米油是不飽和脂肪，豬油是飽和性脂肪。除了判斷油脂的吸收量是否與所用的油脂種類有關之外，您還可以設定演繹式對比，來判斷飽和脂肪與不飽和脂肪的油脂吸收量是否不相同。

› 多重比較(事後比較)

- 一般在整體性之F檢定達顯著後才進行，比較方式為兩兩比較，但也可因研究目的的不同，只做平均數的多重比較，不必考慮整體性F檢定的結果。
 - Bonferroni
校正LSD多群比較產生的型一錯誤，校正方法為LSD的 $p\text{-value} \times N(N-1)/2$ 若在0.05以下結果顯著。
 - Scheffe method(雪費法)
適用於欲比較之各組之間人數不同或較複雜之比較時用之；對違反常態與同質假設不敏感，檢定結果具強韌性；對犯型一錯誤是最保守的方法。
 - Tukey HSD method(誠實顯著差異法)
較保守的檢定方法，適用於比較組數之各組人數相同。

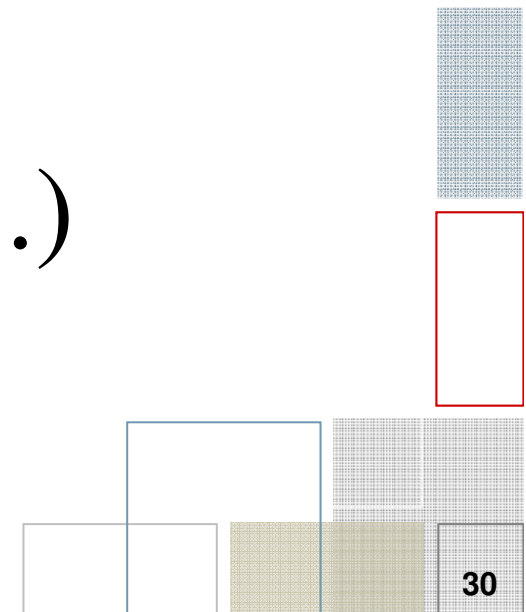
> 多重比較(事後比較)

- S-N-K (Student-Newman-Keuls)
比Tukey較不會犯型二錯誤
- Duncan
類似於S-N-K，但比較寬鬆，比較群組愈多愈容易顯著。
- Games-Howell method
使用於群組間N不相等，異質性及非常態，是蠻新的檢定法且結果具強韌性。
- Dunnett檢定
所有群組均與控制組平均數比較

➤ 多變量變異數分析(MANOVA)

- 在變異數分析中，如果要同時考驗數個依變項，**且具中度相關**，則採用MANOVA。
- MANOVA是ANOVA之擴展，當依變數有兩個以上時，採用多變量變異數分析可減少型一錯誤並同時了解群組間平均數在所有依變項的差異 (Bryman & Cramer, 1997)。

$$(Y_1, Y_2, \dots) = f(X_1, X_2, \dots)$$



› 多變量變異數分析(MANOVA)

在心臟病手術前，“血栓溶解劑”可以協助清除病人的動脈血栓。現有三種藥物alteplase，reteplase，和streptokinase。alteplase和reteplase是更新，更貴的藥物。一個區域性的醫院要確定新藥是否符合成本效益的原則，能取代原來的streptokinase。使用血栓溶解劑的好處是手術可以更順利，恢復期較短。如果較新的藥物更有效，那麼使用的病人就可縮短住院時間，將有助於彌補新藥的費用。

