

常用25種統計方法

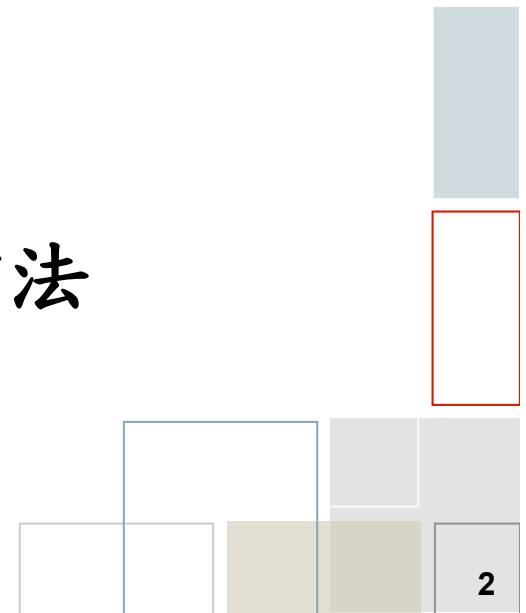


分析方法的基本分類

統計方法的
組合運用



認識統計方法



統計方法的分類

類別變數

連續變數

關係
無因果

敘述性統計
複選題分析
(多重)交叉分析
卡方 (同質、獨立、適合度)檢定

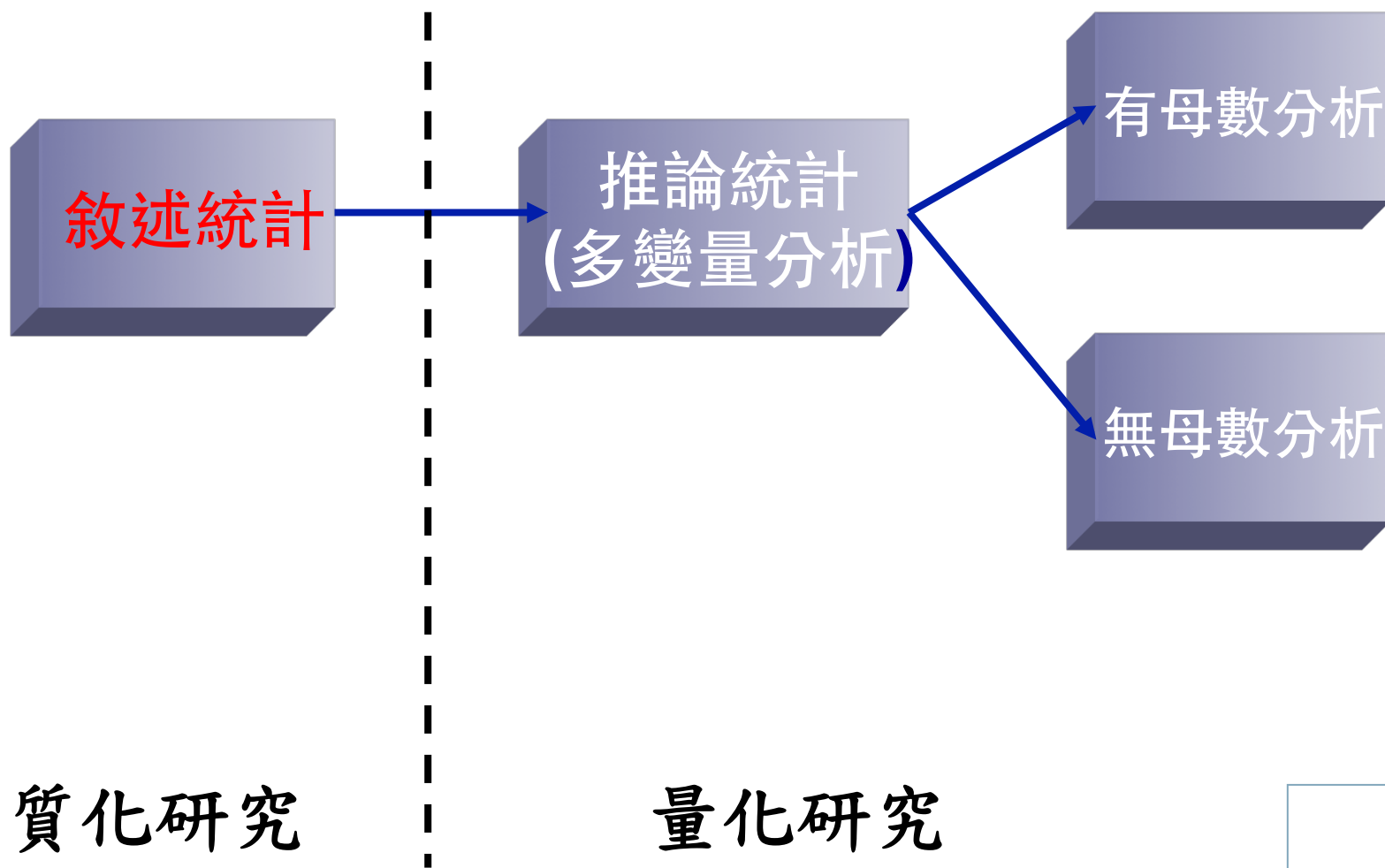
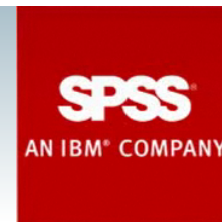
主成份分析
因素分析
集群分析
(偏)相關分析

關係
有因果

二元羅吉斯迴歸
區別分析

偏相關分析
迴歸 (路徑)分析
(多變量)變異數分析
偏最小平方迴歸

SPSS的統計應用



敘述統計

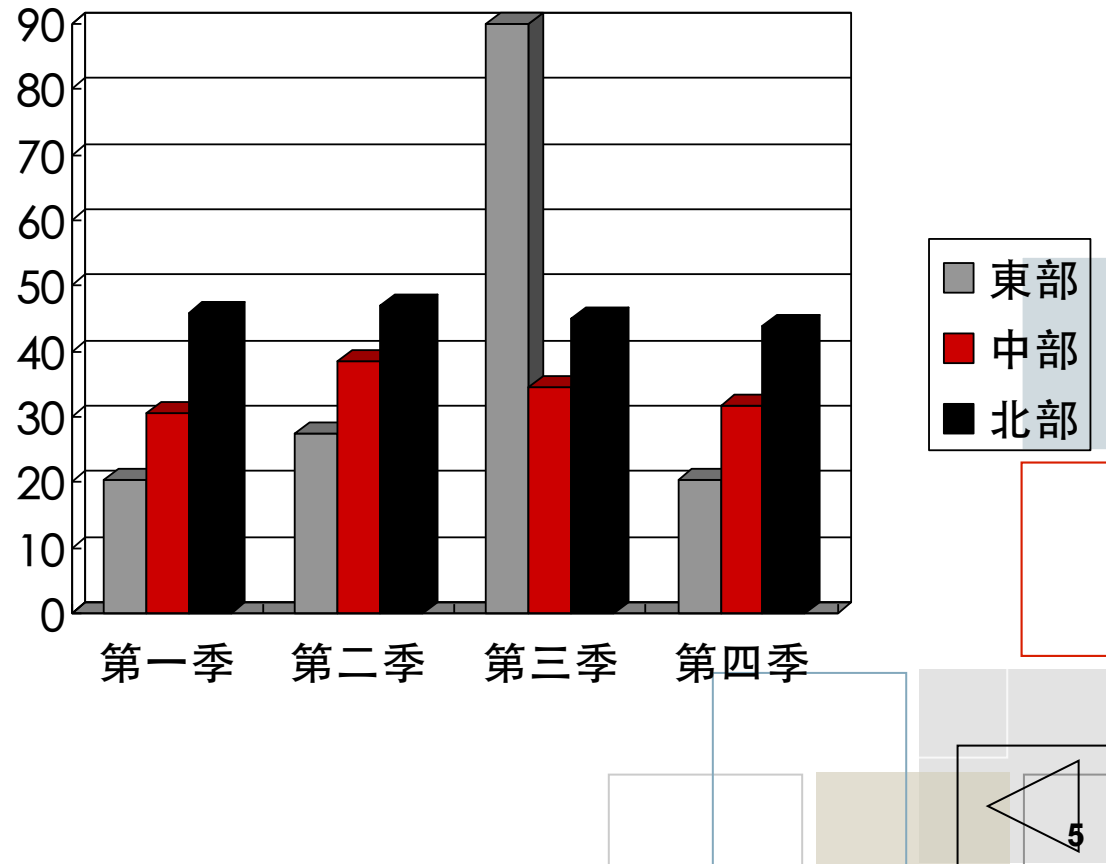
推論統計

有母數分析

無母數分析



- 次數分配表
- 交叉分析
- 多重交叉分析
- 複選題分析



敘述統計

推論統計

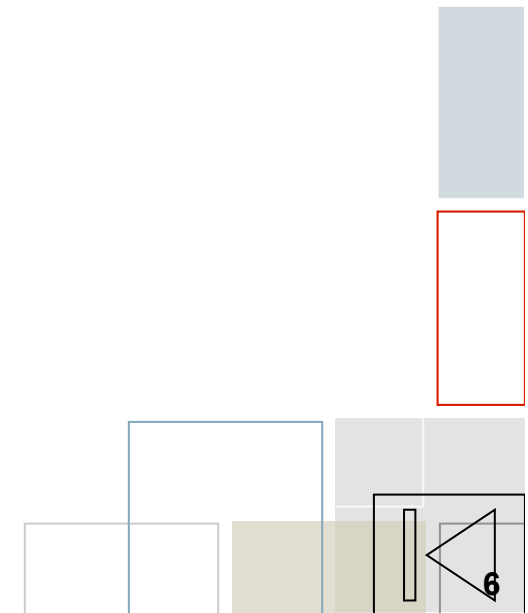
有母數分析

無母數分析



- 項目分析 (t 檢定)
- 信度檢定
- 變異數分析(ANOVA)
- 多重比較(差異分析)
- 多變量變異數分析(MANOVA)
- 相關分析
- 偏相關

- 迴歸分析
- 路徑分析 (多個迴歸分析)
- 偏最小平方迴歸
- 羅吉斯分析 (logistic)
- 集群分析
- 區別分析
- 主成份分析
- 因素分析



敘述統計

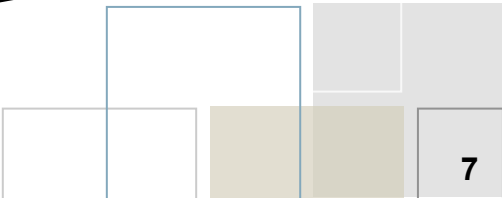
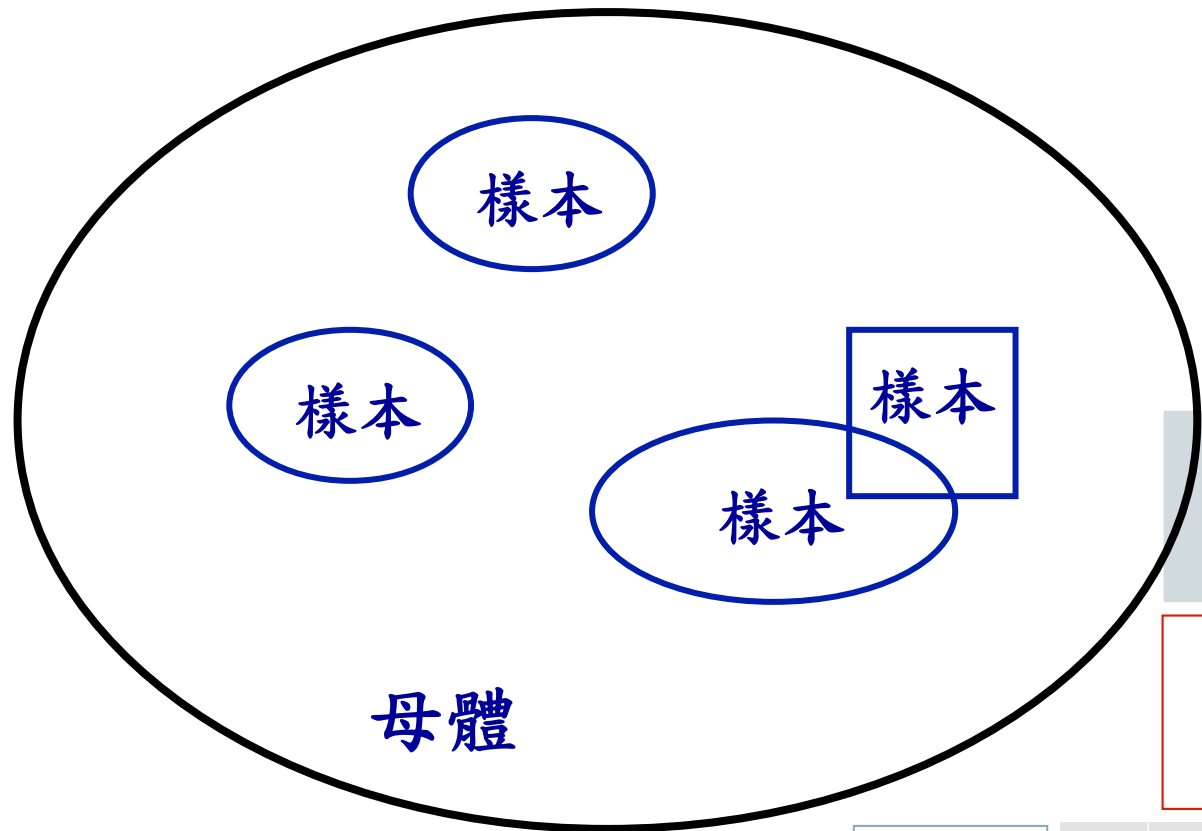
推論統計

有母數分析

無母數分析



- 卡方分析
 - 同質性檢定
 - 適合度檢定
 - 獨立性檢定



連續
Y(依變數)

MANOVA

偏最小平方迴歸

路徑分析

ANOVA t檢定

(偏)相關分析

迴歸分析

不連續

連續

X

敘述統計

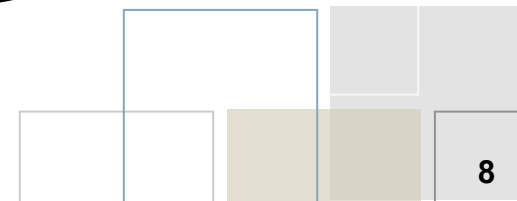
羅吉斯分析

(自變數)

卡方分析

不連續

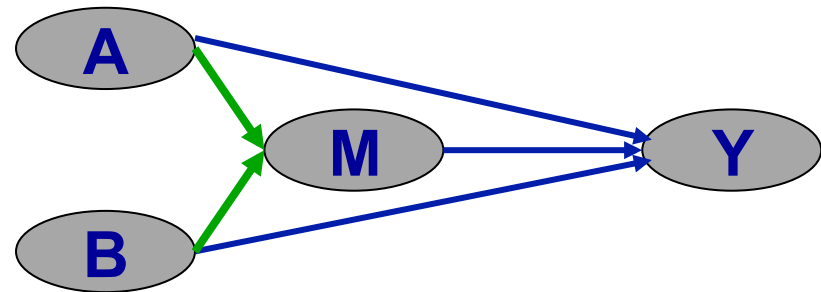
區別分析



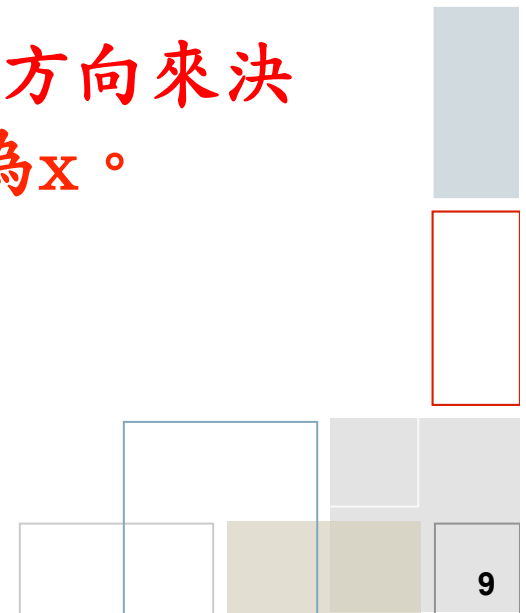


到底何者是 x、
何者是 y 呢？

人口統計變項肯定為 x，其餘構面視其位置而定；
形成某構面的題目可視為該構面之 x_n 。

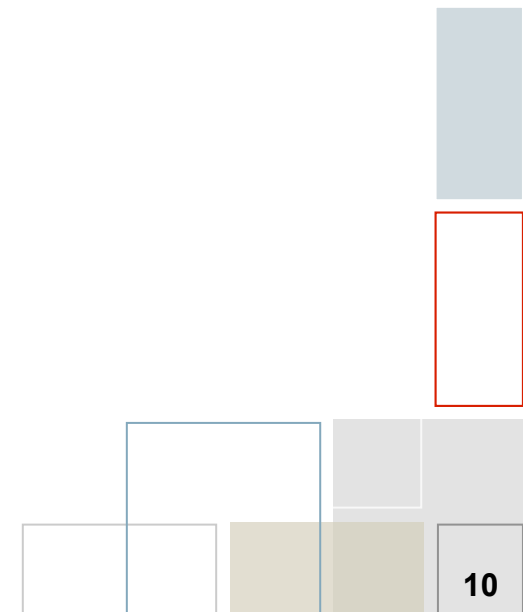


依照箭頭的方向來決定 y，箭尾為 x。



主成份與因素分析相同之處

1. 二者皆可用來將原有之 P 個變項減為 m 個成份分數或因素分數($m < p$)，以符合精簡(parsimony)的原則。
2. 當觀察變項數目多且共同性高之情況下，資料採此二種方法分析，所得結果近似。



主成份與因素分析相異之處

- 因素分析模式屬於統計的模式，變項之測量含有誤差(error)部分，而主成份分析模式屬於數學模式，故主成份分析模式為共同因素分析模式的一個特例。
- 主成份分析的重點在解釋資料的變異量，而因素分析之重點在解釋變項間的相關。
- 主成份分析中，成份是觀察變項的線性組合；而在因素分析中，觀察變項為各因素之線性組合加上誤差。



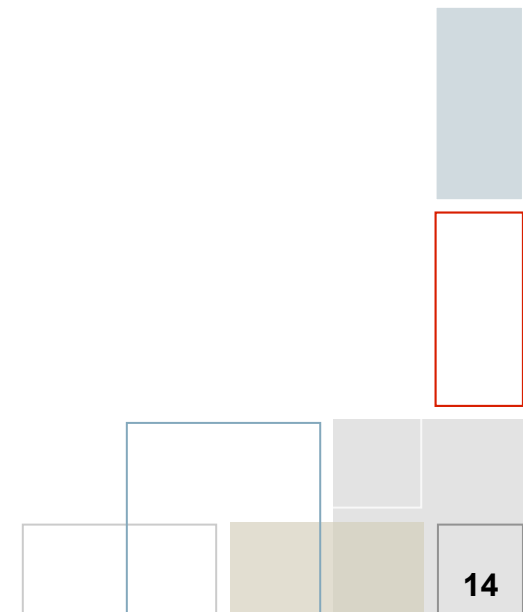
主成份與因素分析相異之處

- 主成分分析中不需要有假設(assumptions)，因數分析則需要一些假設。因數分析的假設包括：各個共同因數之間不相關，獨特因數 (specific factor) 之間也不相關，共同因數和獨特因數之間也不相關。
- 主成分分析中，當給定的共變異數矩陣或者相關矩陣的特徵值是唯一的時，主成分一般是獨特的；而因數分析中因數不是獨特的，可以旋轉得到不同的因數。



主成份與因素分析相異之處

- 在因數分析中，因數個數需要分析者指定（spss 根據一定的條件自動設定，只要是特徵值大於1的因數進入分析），而指定的因數數量不同而結果不同。在主成分分析中，成分的數量是一定的，一般有幾個變數就有幾個主成分。



主成份與因素分析相異之處

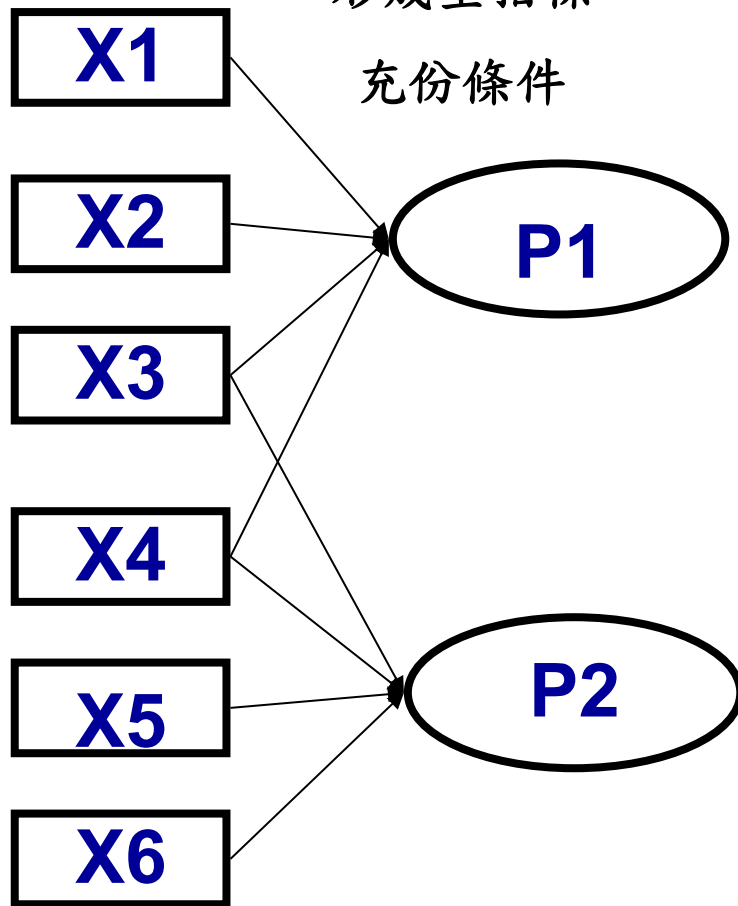
- 和主成分分析相比，由於因數分析可以使用旋轉技術幫助解釋因數，在解釋方面更加有優勢。
- 大致說來，當需要尋找潛在的因數，並對這些因數進行解釋的時候，更加傾向於使用因數分析，並且借助旋轉技術幫助更好解釋。而如果想把現有的變數變成少數幾個新的變數（新的變數幾乎帶有原來所有變數的資訊）來進入後續的分析，則可以使用主成分分析。當然，這中間情況也可以使用因數得分做到。所以這中間區分不是絕對的。



主成份 vs. 因素分析

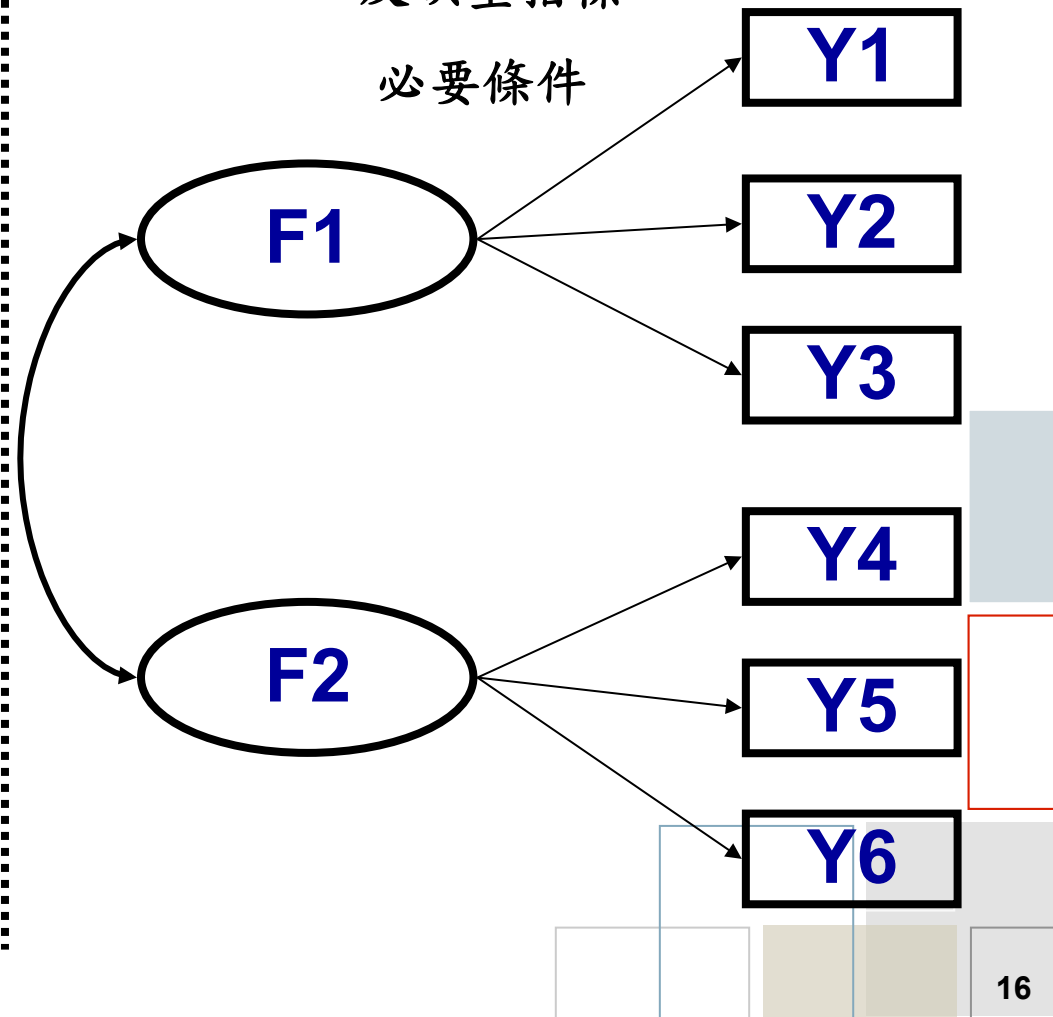
Formative indicator

形成型指標
充份條件



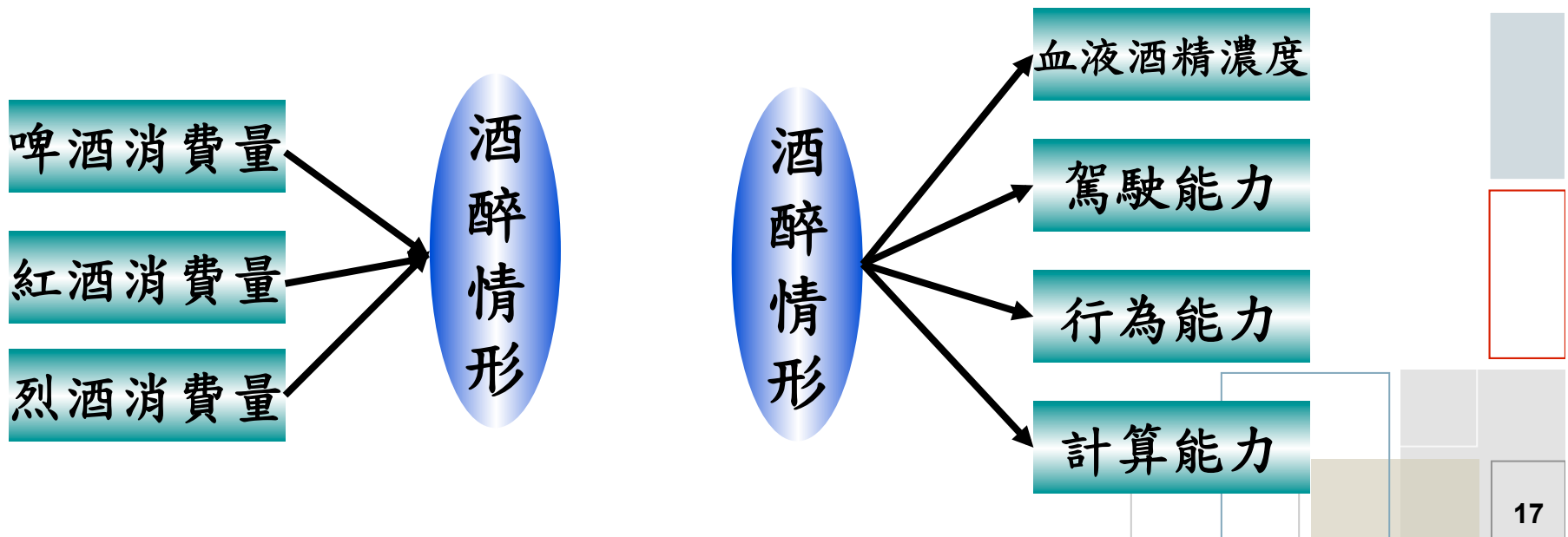
Reflective indicator

反映型指標
必要條件



充份條件 vs. 必要條件

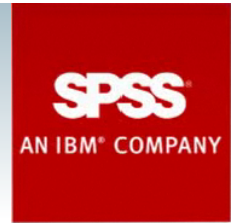
- 充份條件 (*sufficient condition*) :
 - 該做什麼事，才足以 (*suffice*) 達成目標？
- 必要條件 (*necessary condition*) :
 - 一旦達到目標了，必然 (*necessarily*) 會有那些現象？



使用因素分析的理由

- 當變項數目多且共同性高時，主成份分析與共同因素分析之結果十分近似。
- 因素分析可以轉軸易於解釋命名。
- 主成份分析之負荷量與真實的負荷量比較之下，有高估的現象。
- CFA 在複雜情境下（如偏態或峰度），與PCA 結果之比較，CFA 較能提供穩定的數值。

因素分析(針對變數分類)



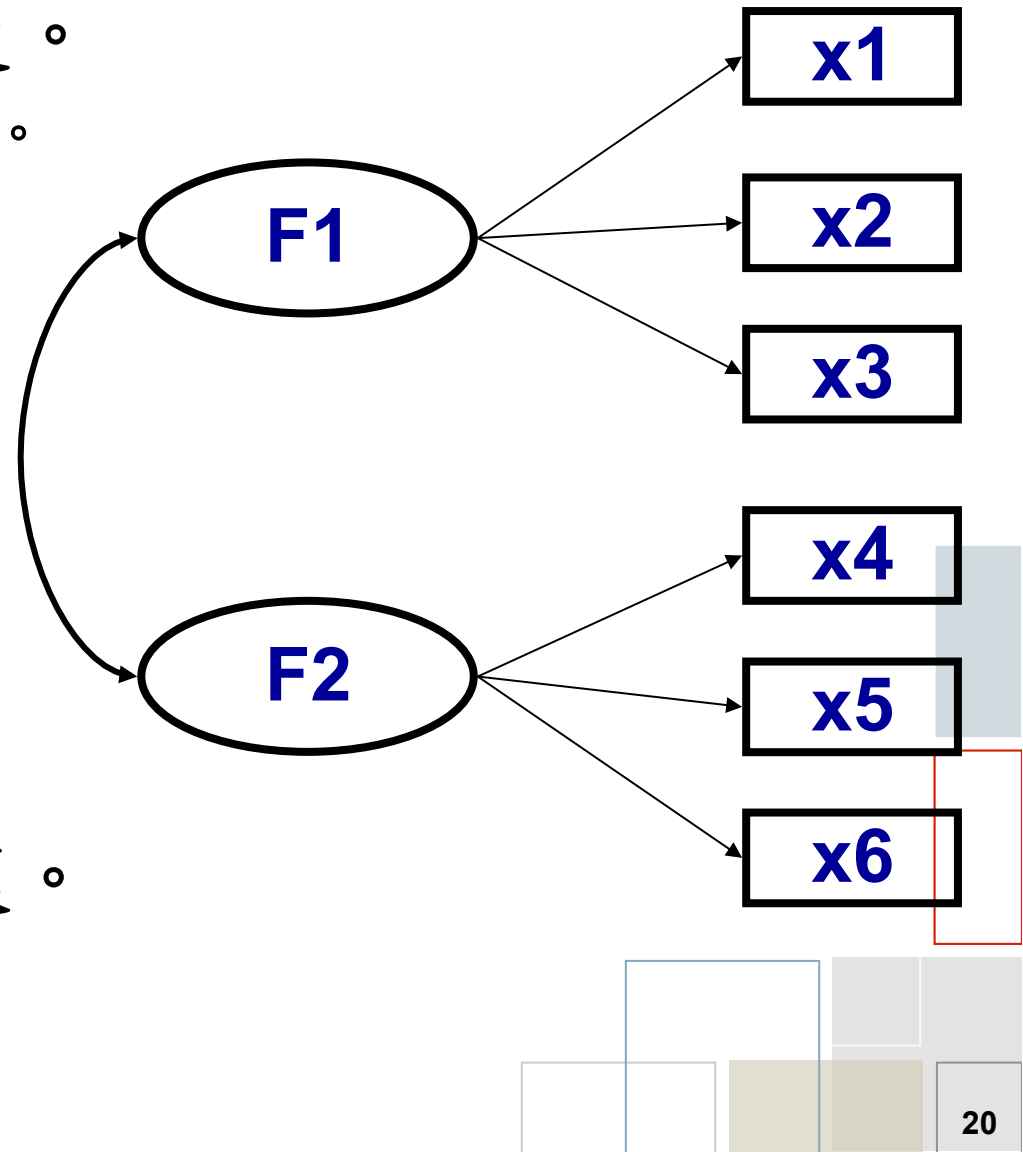
- 為了要證實研究者所設計的測驗的確在測某一潛在特質，並釐清潛在特質的內在結構，能夠將一群具有共同特性的測量分數，抽離出背後**潛在構念**的統計分析技術，即為因素分析。
- 主要目的是把數個不易解釋，但卻**彼此相關**的變項，轉化成少數幾個有概念化意義，而彼此獨立性大的因素；並可大幅縮減欲分析的變項，使之成為少數幾個因素，以利於資料分析的進行。



主要的功能



- 建立量表的建構效度。
 - 收斂效度及區別效度。
- 簡化測量的內容。
- 用來協助測驗編製，進行項目分析，檢驗試題的優劣好壞。



集群分析(針對人群分類)



- 目的
 - 將資料分成幾個相異性最大的群組，而群組內的相似程度最高。
- 分析工具
 - 二階段集群分析
 - 200個樣本以內採階層式集群分析
(實務上100個判斷就有問題了)
 - 200個樣本以上採 **K-means**
- 集群分析(cluster analysis)分組完後，可用區別分析驗證之

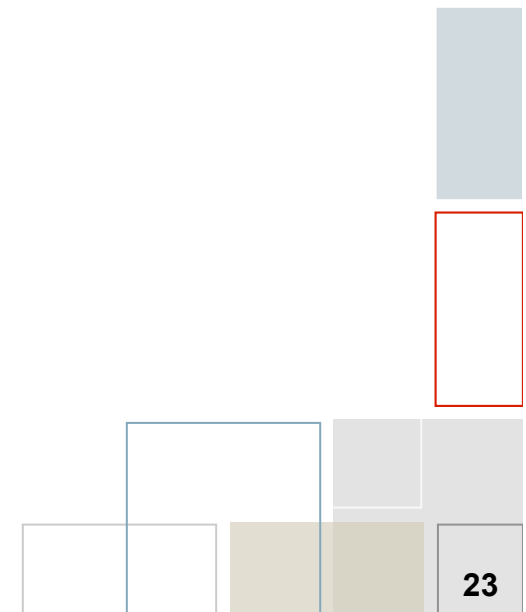


樣本大小決定

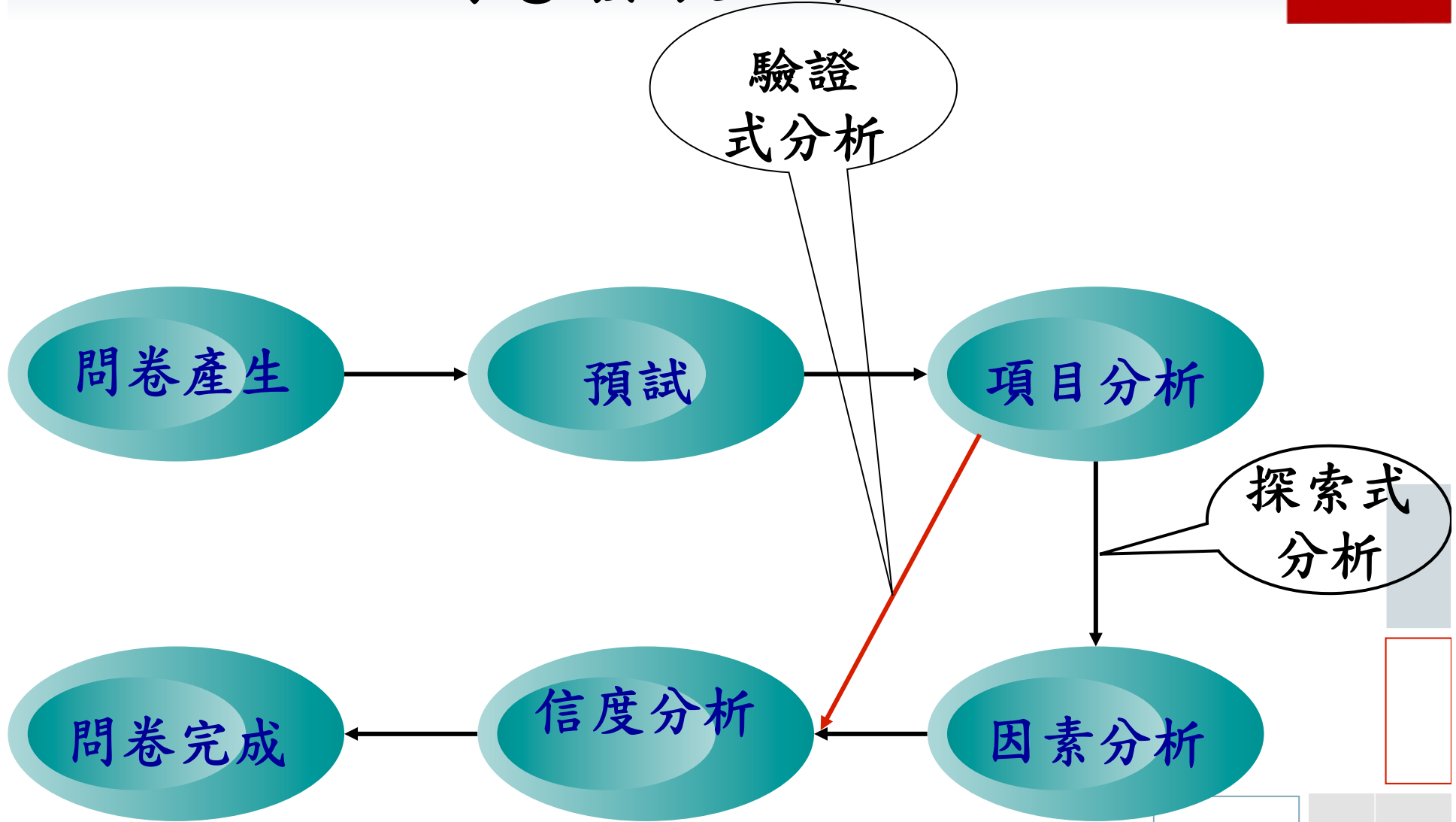
- 經驗：以問卷中最大的構面中的題項數目為主，以5~20倍為抽樣數目，依母體大小決定之(multivariate data analysis, 5th ed., Hair, Jr. et al, p.98~99)。
- 計算：

$$n = \frac{1.96^2 s^2}{e^2}$$

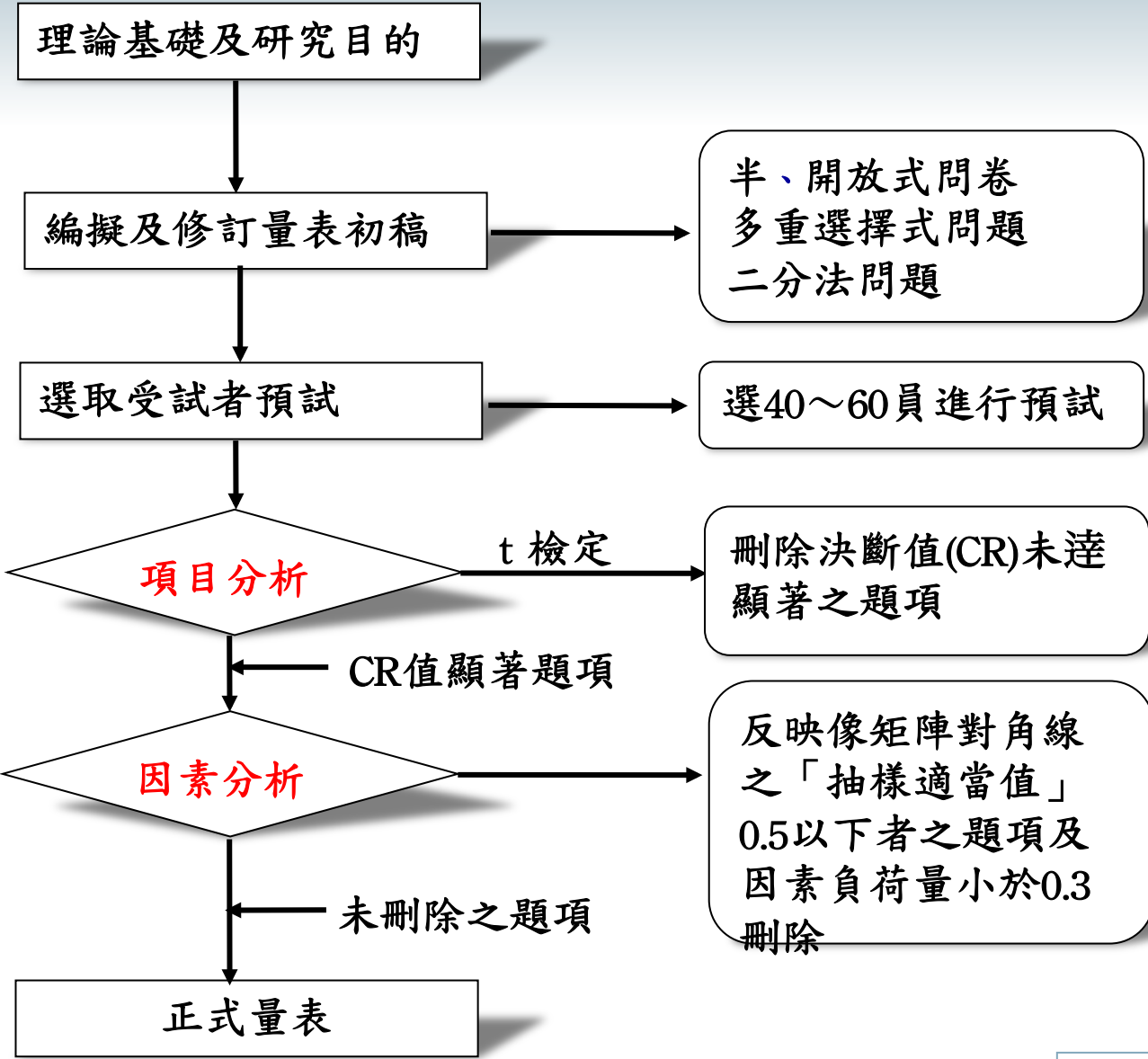
$$n = \frac{1.96^2 p(1-p)}{e^2}$$



問卷發展流程



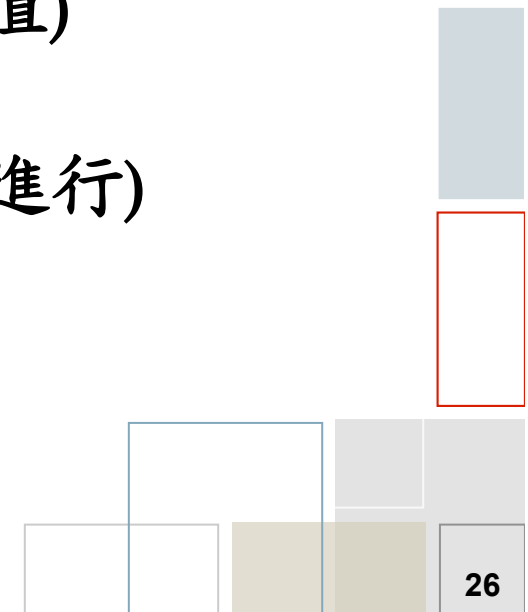
問卷發展流程



預試(pretest)



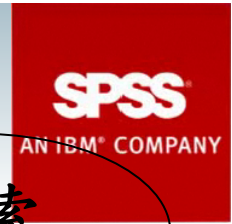
- 目的:
 - 了解問項題意是否清楚及修正錯字等
 - 刪除不具鑑別力的題項，來提升問卷品質
- 可供應用的統計方法:
 - 次數分配(檢查有無輸入錯誤或遺漏值)
 - 項目分析(刪除不具鑑別力的變數)
 - 因素分析(縮減變數的數量以利分析進行)
 - 信度分析(求得因素之內部一致性)



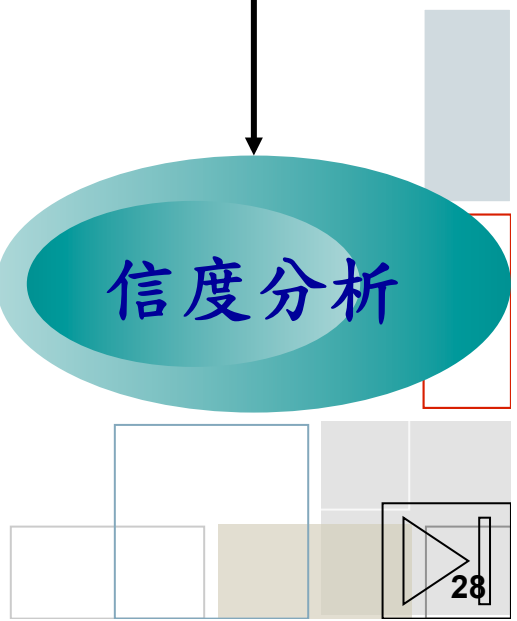
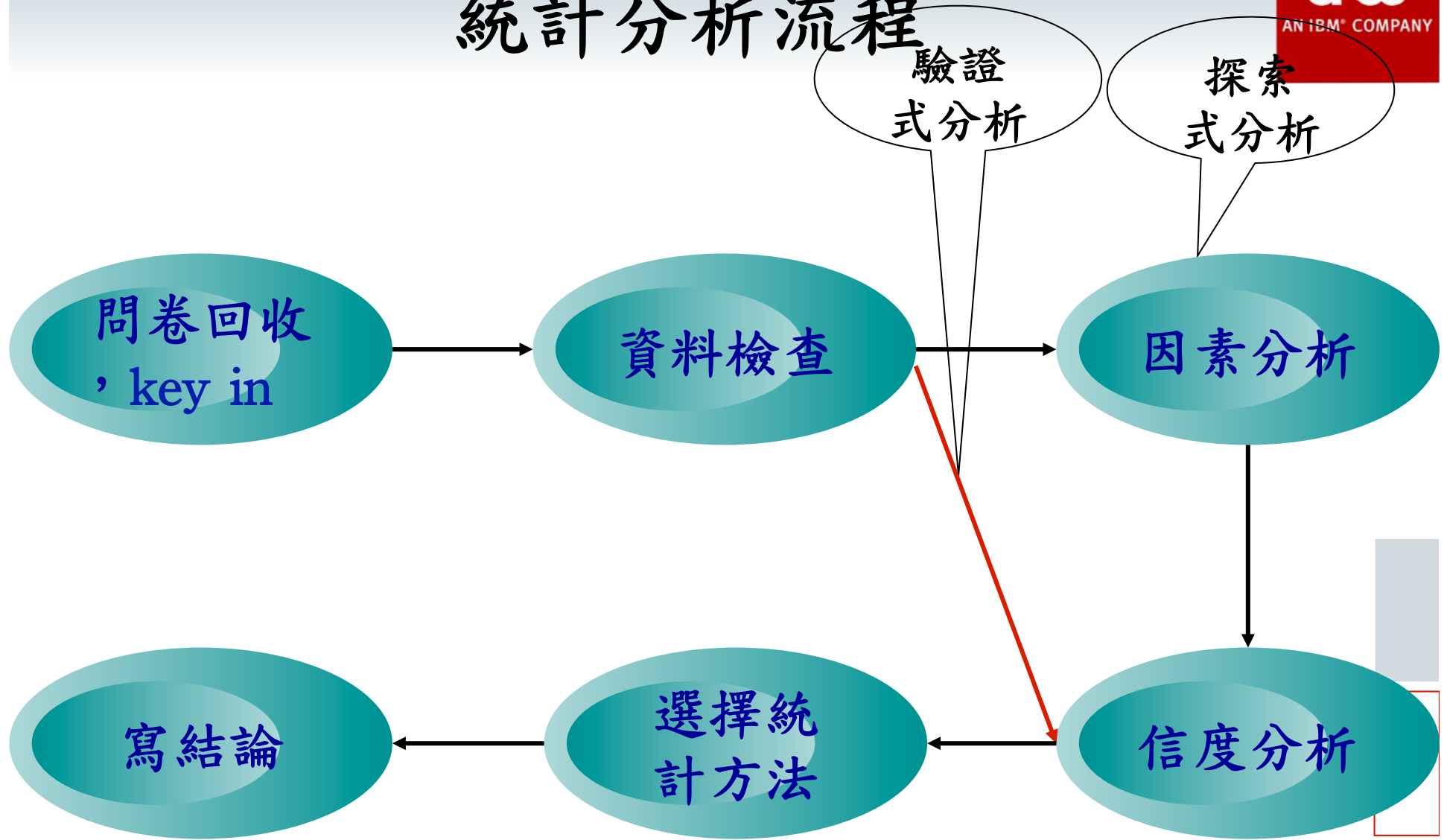
一般統計方法運用的方式

- 資料鍵檔（用 excel 建檔）
- 檢查資料（檢查資料輸入是否有誤）
- 遺漏值處理
- 刪除資料（tip: 由後往前刪）
- 項目分析（問卷搜集完後）
 - 目的：減少題項，找出有鑑別力的題目
- 因素分析（探索式）
 - 目的：再度縮減題項，並根據預試資料作初步的信度分析。

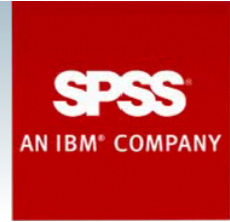




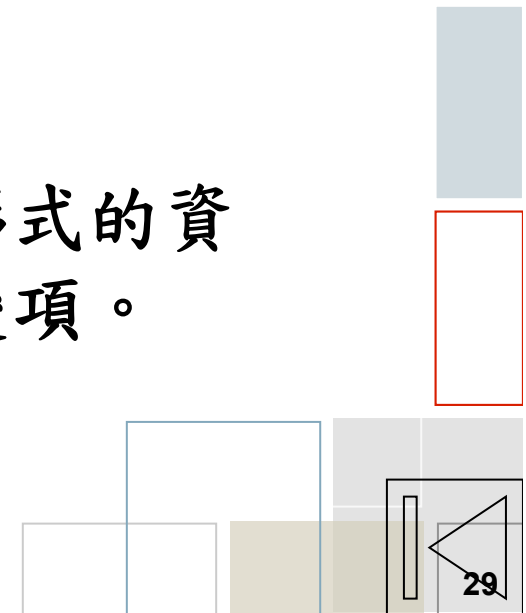
統計分析流程



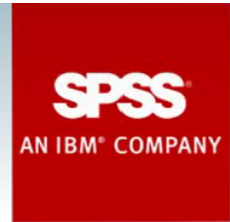
卡方分析(chi-square test)



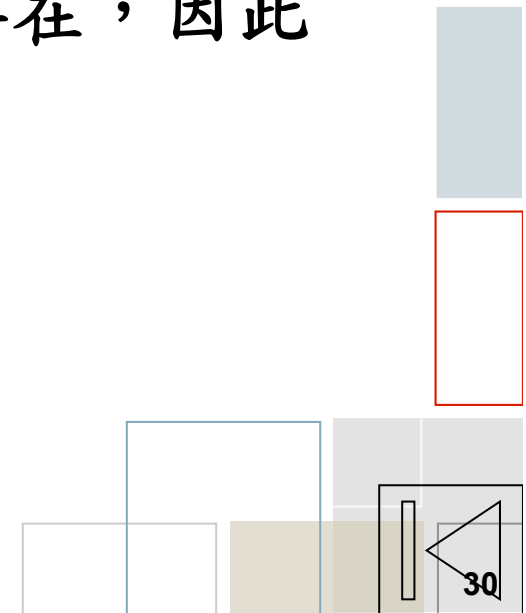
- 類別資料的產生
 - 原發性類別資料(自然分類)
當被測定的變項的本質是名義性的屬性，例如性別資料。
 - 操弄性類別資料：
以人為操作的手段所獲致的分類性資料，例如實驗操弄的分類結果。
 - 虛擬化類別資料：
由其他類型的資料型態轉換成類別形式的資料，例如由連續變項轉換來的類別變項。



t 檢定 (t-test)



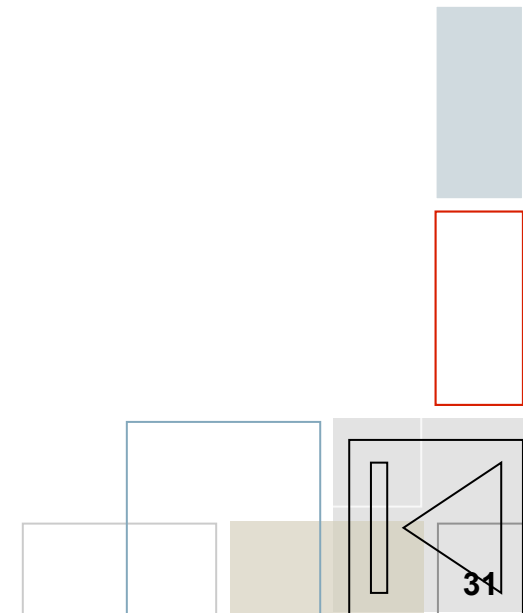
- 平均數的差異考驗
- 雙母群考驗
 - 如果研究者想同時考慮不同情況之下的平均數是否有所差異，例如男生與女生的平均數的比較，此時即牽涉到多個平均數的考驗；不同的平均數，代表背後具有多個母數的存在，因此被稱為多母數的平均數考驗。



變異數分析—ANOVA

- 平均數差異檢定：基本原理是計算兩個數值以上（平均數）之間的差異，如果差異夠大，大於統計上的隨機差異，便可能獲得顯著的結果，拒絕虛無假設、接受對立假設。平均數差異與隨機差異的比值，決定了統計的顯著與否。

$$Y_1 = X_1 + X_2$$

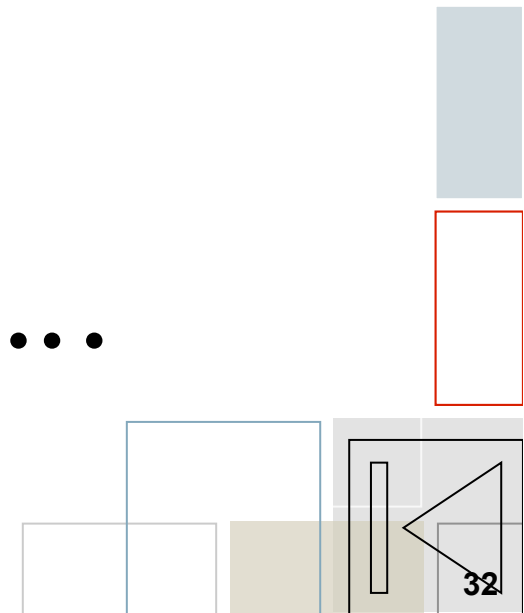


多變量變異數分析(MANOVA)

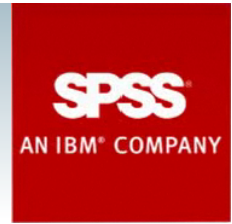


- 在變異數分析中，如果要同時考驗數個依變項，**且具中度相關**，則採用MANOVA。
- MANOVA是ANOVA之擴展，當依變數有兩個以上時，採用多變量變異數分析可減少型一錯誤並同時了解群組間平均數在所有依變項的差異(Bryman & Cramer, 1997)。

$$Y_1 + Y_2 + \dots = X_1 + X_2 + \dots$$



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。

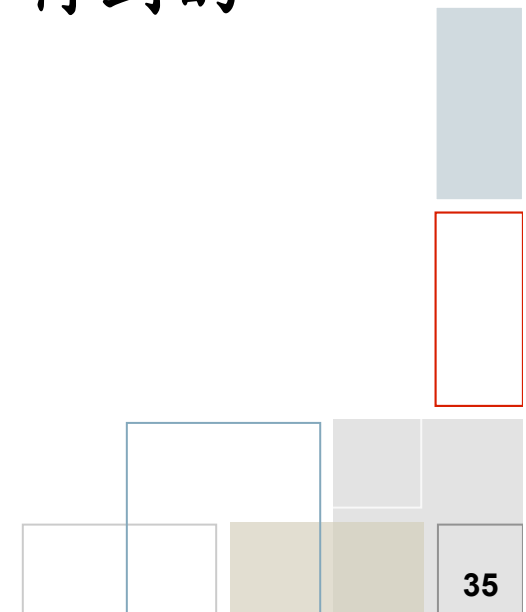
多重比較(事後比較)



- **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。
- **Dunnnett檢定**
 - 所有群組均與控制組平均數比較

相關分析(皮爾森相關)

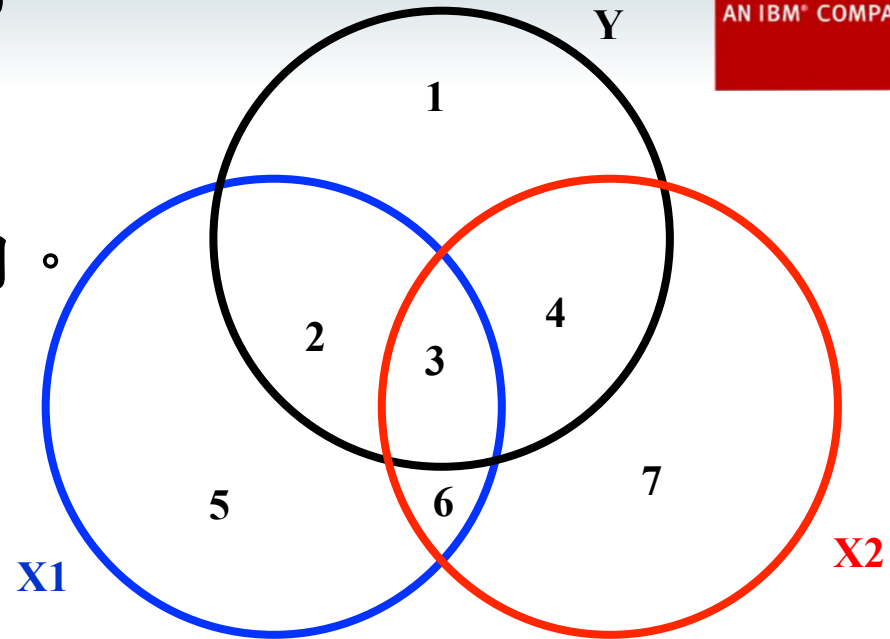
- 相關 (**correlation**) 是用以檢驗兩個變項線性關係的統計技術，以相關係數 (**coefficient of correlation**) 來表示其相關程度。
- 皮爾森相關係數是一個標準化的關聯係數。其原理是先計算出兩個變項的共變量，再除去兩個變項的標準差，加以標準化，得到的一個去除單位的標準化分數。
- 相關係數介於**-1**至**1**之間。



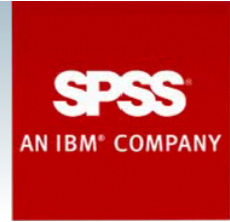
偏相關



- 使用時機：
 - 證明中介效果是否存在時使用。
- 零階相關 (Zero-Order)
 - Y和X1的皮爾森相關
 - $(2+3)/(1+2+3+4)$ 。
- 偏相關 (Partial Correlations)
 - Y和X1的偏相關 $(2/1+2)$ ；Y和X2為 $(4/1+4)$ 。
 - 如標準化迴歸係數
- 半偏相關 (Semi-Partial Correlations)
 - 移去X1,X2對因變數的共同影響後, X1與因變數的相關 $(2/1+2+3+4)$ ；X2與因變數的相關 $(4/1+2+3+4)$ 。
 - 係數愈大代表該變數對Y的影響愈大。



迴歸分析(Regression)



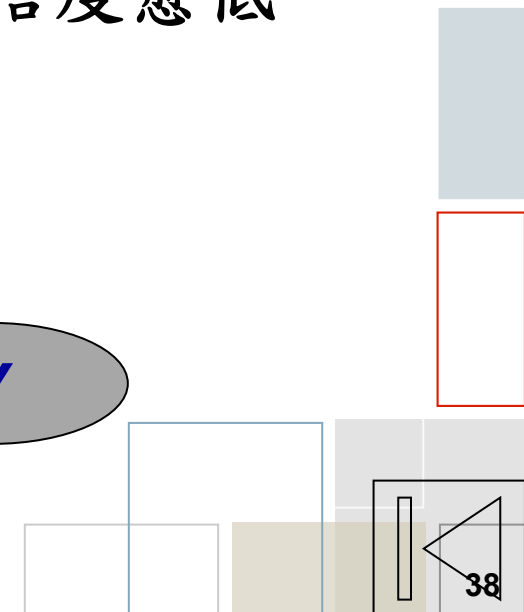
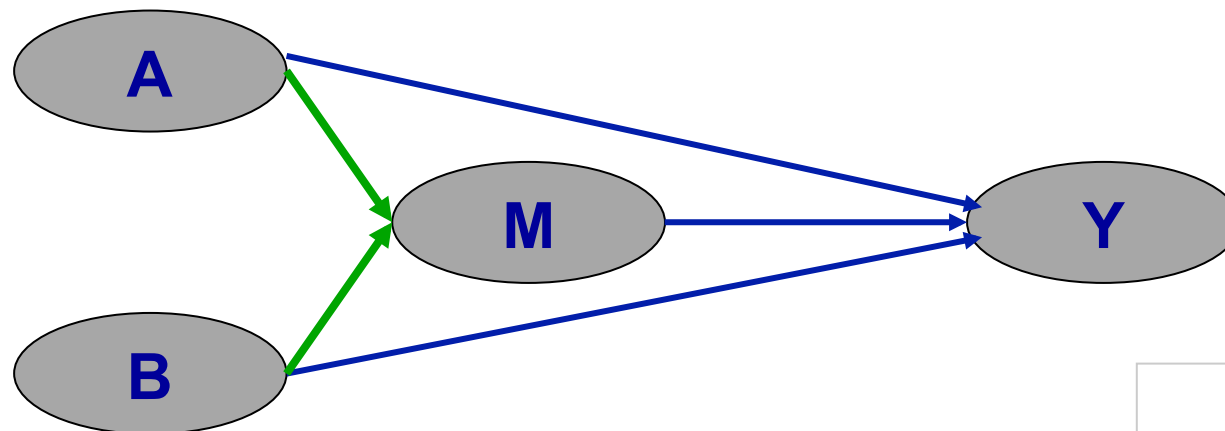
- 基本條件: 連續變項之間的關係
- 線性關係 (linear relationship) ，指兩個變項的關係可以被一條最具代表性的直線來表達之時，所存在的關連情形。
- 迴歸分析的結果無法證明 x 和 y 之間有因果關係存在。
- 因果關係須滿足三個條件：**1.** 除了評估變數之外，其餘變數保持不變；**2.** x 、 y 有顯著相關；**3.** 因(x)必發生在果(y)之前。

- 迴歸方程式
$$Y_1 = aX_1 + bX_2 + \dots$$

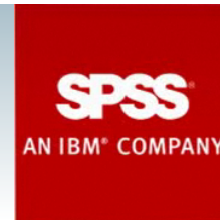
路徑分析 (path analysis)



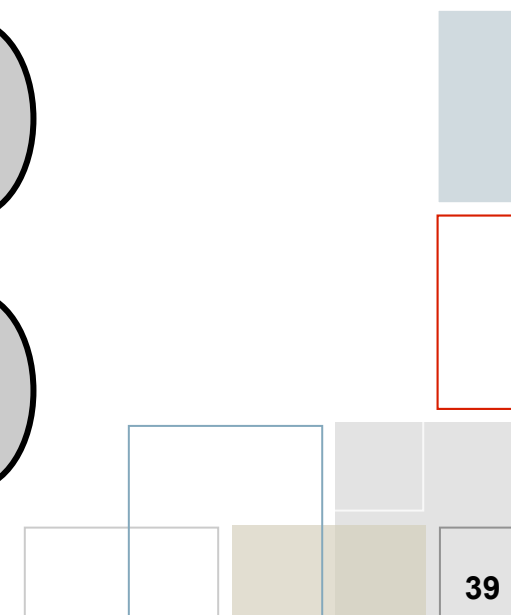
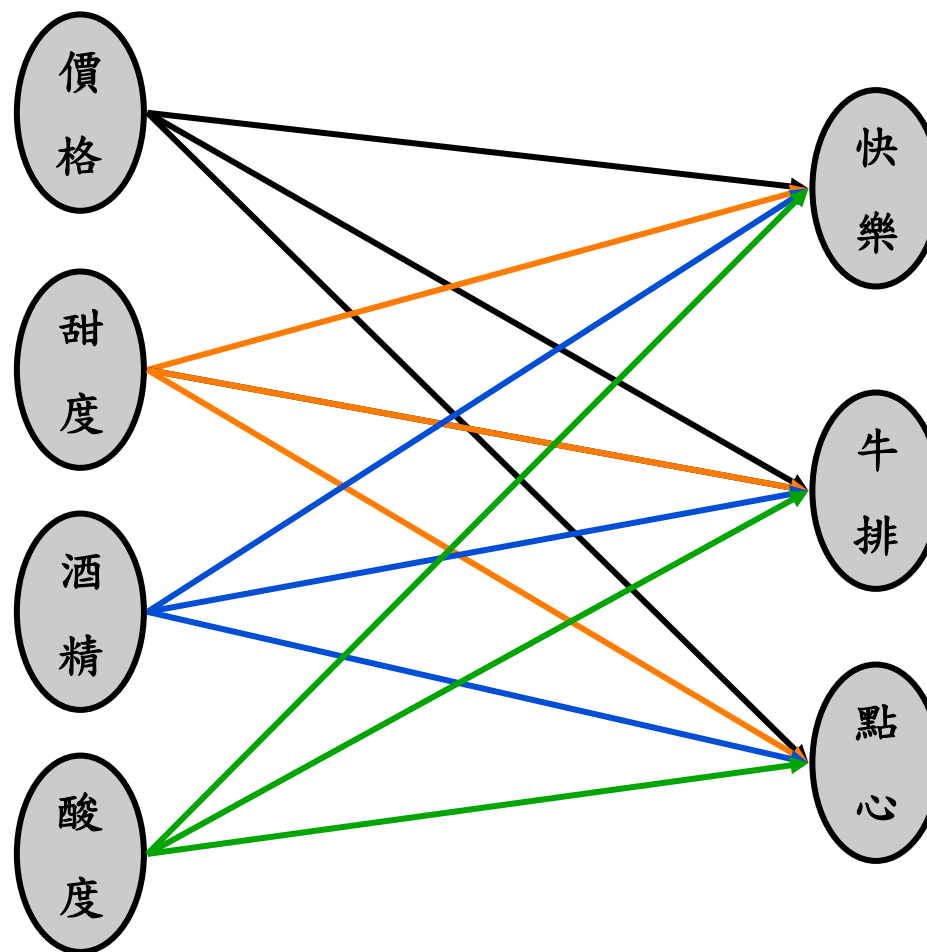
- 路徑分析由一系列的迴歸分析所組成，除了借用迴歸方程式的原理，並透過假設性的架構，將不同的方程式加以組合，形成結構化的模式。
- 優點:可用來證明中介變數的存在與否
- 缺點：執行n次迴歸，其信賴數係數 α 成為**0.95ⁿ**。因此，執行愈多次迴歸，其模型之可信度愈低。



多因多果及小樣本的分析利器－ PLSR



- 偏最小平方迴歸=多元線性回歸分析+典型相關分析+主成分分析

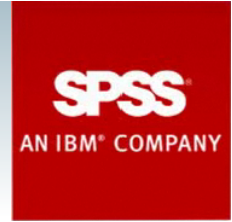


PLS的特點

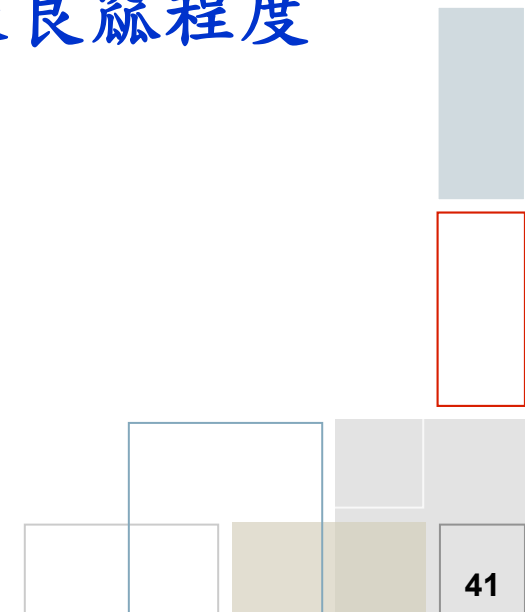
- 能夠在自變量存在**嚴重多重共線性**的條件下進行迴歸建模
- 允許在**樣本點個數少於變量個數**的條件下進行迴歸建模
- PLS 在最終模型中將包含原有的所有自變數
- PLS 模型更易於辨識系統信息與誤差（甚至一些非隨機性的誤差）
- PLS 模型中，每一個自變量的迴歸係數更容易解釋



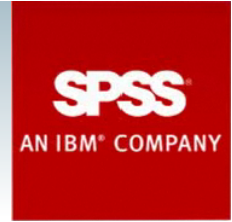
項目分析與信度估計



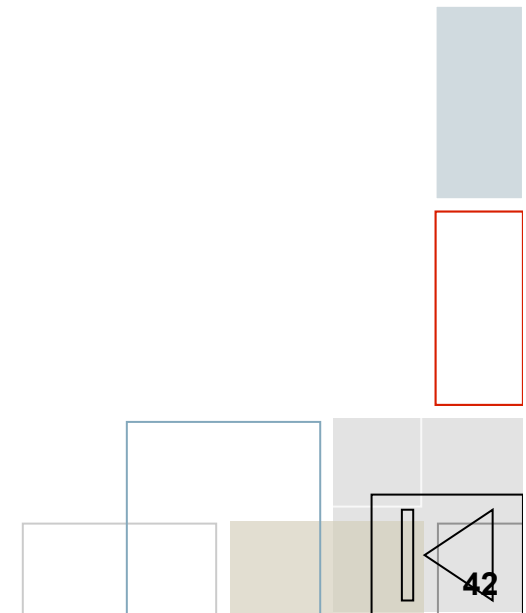
- 預試分析
 - 目的在確認量表題目的堪用程度
 - 最重要的工作為項目分析，試探性的信度分析，以作為題目改善的依據
- 信效度檢驗
 - 提供各項客觀指標，作為測驗與量表良窳程度的具體證據。



二元 Logistic 分析



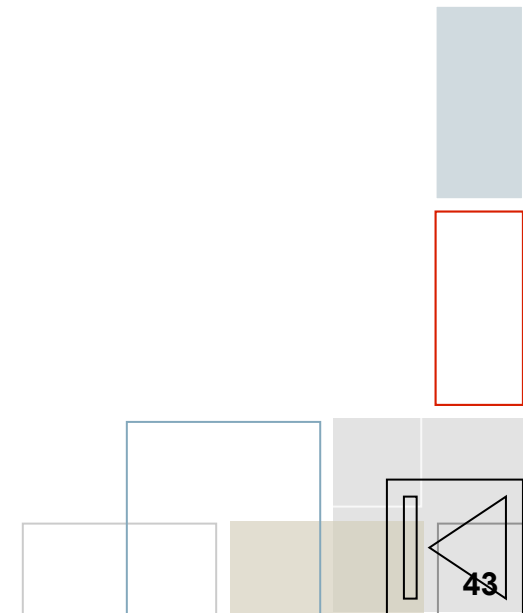
- 適用於兩群體的分類
- 不需符合一般多變量的嚴格假設及較具強韌性
- 與區別分析同樣具有正確的統計檢定能力及整合非線性影響的能力
- 能應用於各種範圍的特徵
- 適合用於建構決策模型



區別分析

- 區別分析(**discriminate analysis**)的主要目的在計算一組「自變項」的線性組合對「間斷變項」加以分類，並檢查再分組的正確率。

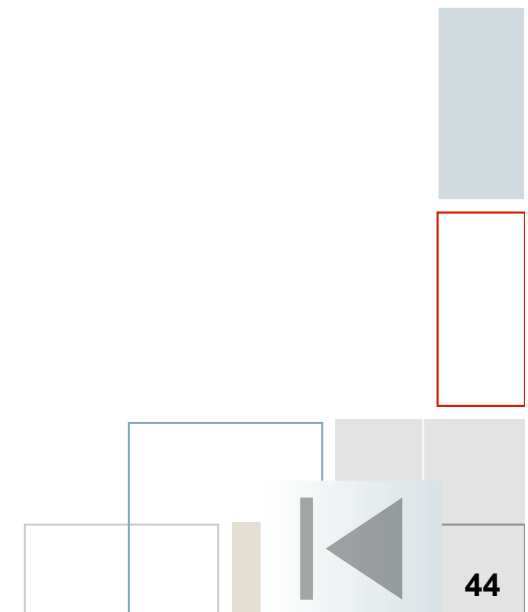
$$Y_1 = X_1 + X_2 + \dots$$



交叉分析(列聯表分析)



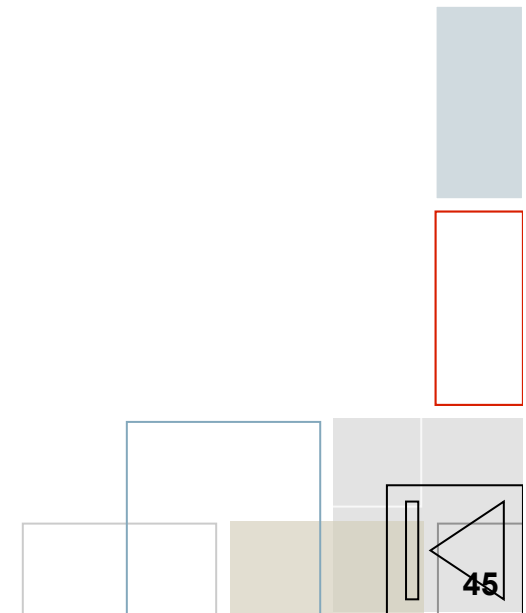
- 變數須為名目尺度(不連續變數)
 - 性別：男、女
 - 傳播媒體：電子媒體、平面媒體、網路媒體
- 只能有兩個變數



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
 - 傳播媒體：電子媒體、平面媒體、網路媒體
- 三個變數以上，但仍以三個為主



複選題分析 (Multiple Response)

- 複選題分析提供多元化的資訊以供決策參考

請問您會想換行動電話的原因(複選)

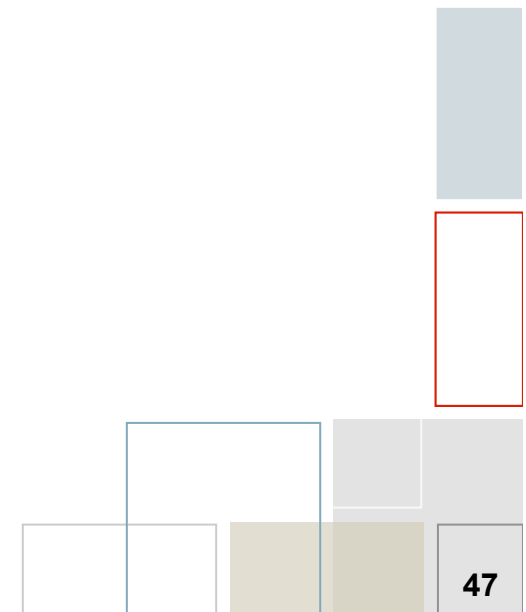
- 故障 機型老舊 換手機門號 新產品推出
 電信公司服務不好 有促銷活動 其他_____

以下哪些因素能吸引您更換手機(可複選)

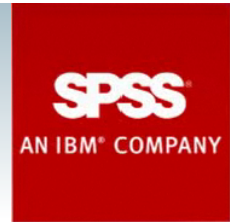
- ₁ 價格便宜 ₂ 功能齊全 ₃ 造形酷炫 ₄ 廠牌 ₅ 通話費率
₆ 國際漫遊 ₇ 服務態度 ₈ 促銷活動 ₉ 申請手續 ₁₀ 通話品質
₁₁ 體積小 ₁₂ 售後服務 ₁₃ 帳單可靠 ₁₄ 其他_____

複選題處理程序

1. 資料輸入(以個別題目型態輸入)
2. 定義集群
 - 點選統計分析→複選題分析→定義集合
 - 定義集內的變數
3. 完成虛擬複選題變項
4. 次數分配表分析
5. 交叉表分析



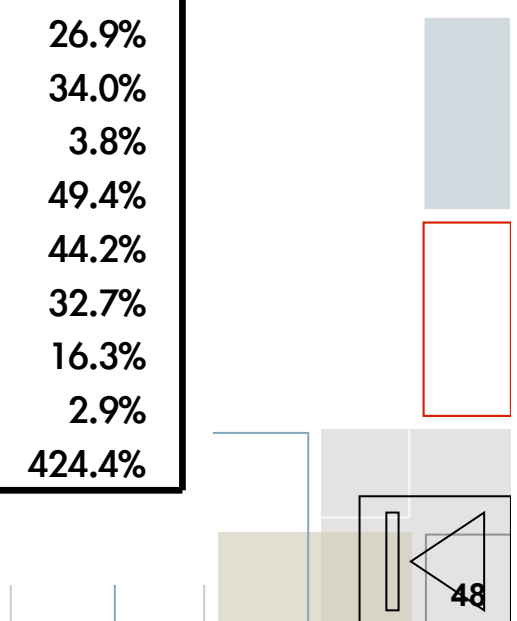
複選題實務上的意義



\$c7 次數

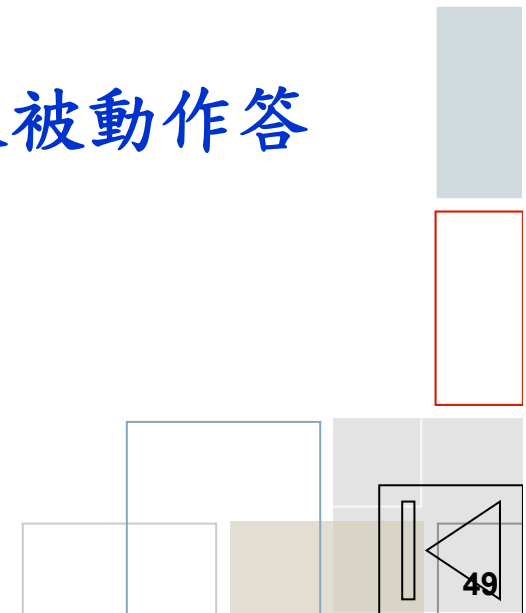
		反應值		觀察值百分比	
		個數	百分比		
\$c7 ^a	c7m1	價格便宜	180	13.6%	57.7%
	c7m2	功能齊全	163	12.3%	52.2%
	c7m3	造形酷炫	72	5.4%	23.1%
	c7m4	廠牌因素	91	6.9%	29.2%
	c7m5	通話費率	140	10.6%	44.9%
	c7m6	國際漫遊	22	1.7%	7.1%
	c7m7	服務態度	84	6.3%	26.9%
	c7m8	促銷活動	106	8.0%	34.0%
	c7m9	申請手續方便	12	.9%	3.8%
	c7m10	通話品質	154	11.6%	49.4%
	c7m11	體積小	138	10.4%	44.2%
	c7m12	售後服務	102	7.7%	32.7%
	c7m13	帳單可靠	51	3.9%	16.3%
	c7m14	其他	9	.7%	2.9%
總數			1324	100.0%	424.4%

a. 二分法群組表列於值 1。



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。
- 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。
 - 受訪者中包含不同團體。
 - 網路問卷與紙本問卷的比較。



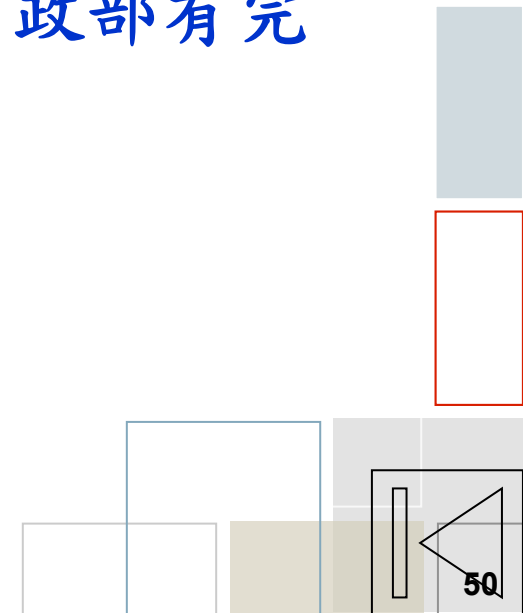
適合度檢定

- 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。

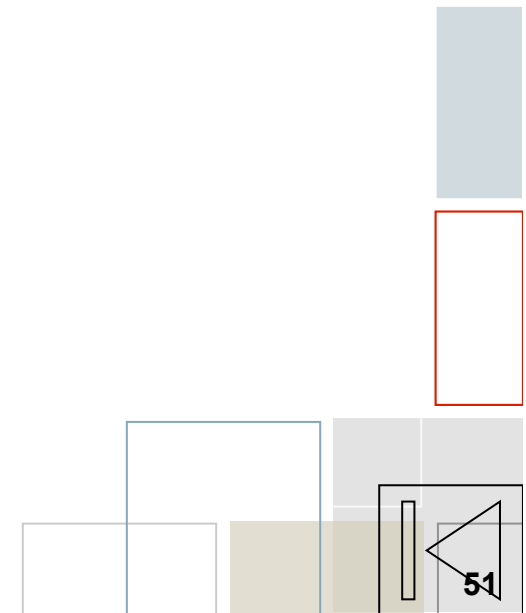
- 適用時機

- 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）。



獨立性檢定

- 目的
 - 想要同時檢定兩個**類別變項**之間的關係是否相關時，採用卡方檢定。
- 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。



離散與連續變數的互換

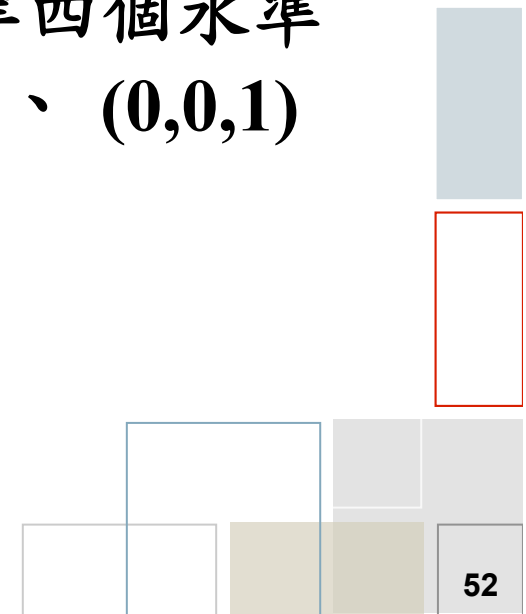


■ 連續轉不連續

- 採人為操弄，將連續變數分類。
- 例：將全班成績改為高分組、中分組、低分組

■ 不連續轉連續

- 虛擬變數(Dummy Variables)=水準數(**n**) -1
- 例：地區:東部、北部、中部、南部等四個水準
- DV可設為(0,0,0) 、 (1,0,0) 、 (0,1,0) 、 (0,0,1)
- 數量最多的設為(0,0,0) 。



介紹完畢

SPSS 祝您寫作愉快