ch4 PCA 主成分分析 - 特徵值向量

共線性問題、降維

1. mean-corrected or standardized data ? 標準化 or 原始資料 數據

[摘要統計結果 the means procedure – 變異數] 看到變異數相差比較多的話就要用標準化後的資料。

1. using eigen-value-greater-than-one要保留主成分

[ 相關矩陣特徵值 – 特徵值 ] 保留大於1的主成分

1. the percentage of total variance that can be accounted for?

[ 相關矩陣特徵值 – 累積 ] 看他需要看到累計多少就看第幾個數值

4.5. label 1 st principal component?? 命名

[ 特徵向量 ] 看prin1的哪幾個數值比較高 根據它們去命名

如果都差不多就命名為一般(聽力)

6. which ID has the largest 1st principal component score?? 主成分分數最大??

proc sort data=pca回傳 out=pca\_排序後輸出; by descending PRIN1; run;

proc print data= pca\_排序後輸出; run;

ch5 EFA 探測性因素分析 不知道共同因數是甚麼??

[ PCF 主成分因素分析 用主成分提取因素/PAF 主軸因素分析 識別淺再因素排除誤差變異]

1. pcf the data and identify the smallest number of common factors using eigen-value-greater-than-one rule 保留大於1的主成分

[ 相關矩陣的特徵值 – 特徵值 ] 找出有幾個大於1的主成分

1. using varimax rotations to identify the most plausible factor solution 最適當的因素

[ 旋轉法: Varimax - **旋轉的因子模型** ] 找出相對較大的數值並框起來

1. label the identified factors suitably and interpret the factors 命名+解釋
2. what is the percentage of total variance can be accounted for by the retained factors ? 保留因素解釋總變異百分比

[ 相關矩陣的特徵值 – 累積 ] 看保留了幾個就看累積的第幾個變數

1. examine the appropriateness of factor analysis for this data 這個資料適合因素分析嗎?

相關係數高、偏相關係數低

[ Kaiser 取樣適當性量數: 整體 MSA = ] KMO大於 0.6 就是適合的

1. examine the goodness of the factor solution 看RMSR 適合度

[ 均方根非對角殘差: 整體 ] RMSR < 0.1 都是好的

Ch6 CFA 驗證性因素分析 知道共同因數是啥

如果說標準差為2 表示所有都要一起乘以4 (2^2)

1. SMC ?

[ 平方複相關 – R平方 ] 觀察他有沒有大於0.5

1. CONSTRUCT RELIABILITY?

| **線性方程式的標準化結果** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **m** | **=** |  | **0.8102** | **(\*\*)** | **FX1** | **+** | **1.0000** |  | **EX1** |
| **p** | **=** |  | **0.7652** | **(\*\*)** | **FX1** | **+** | **1.0000** |  | **EX2** |
| **c** | **=** |  | **0.6665** | **(\*\*)** | **FX1** | **+** | **1.0000** |  | **EX3** |
| **e** | **=** |  | **0.8254** | **(\*\*)** | **FX2** | **+** | **1.0000** |  | **EX4** |
| **h** | **=** |  | **0.8311** | **(\*\*)** | **FX2** | **+** | **1.0000** |  | **EX5** |
| **f** | **=** |  | **0.8844** | **(\*\*)** | **FX2** | **+** | **1.0000** |  | **EX6** |

| **外因性變數之變異數的標準化結果** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **變數 類型** | **變數** | **參數** | **估計值** | **標準 誤差** | **t 值** | **Pr > |t|** |
| **誤差** | **EX1** | **THEEX1** | 0.34353 | 0.07547 | 4.5521 | <.0001 |
|  | **EX2** | **THEEX2** | 0.41444 | 0.07290 | 5.6854 | <.0001 |
|  | **EX3** | **THEEX3** | 0.55581 | 0.06821 | 8.1483 | <.0001 |
|  | **EX4** | **THEEX4** | 0.31867 | 0.04974 | 6.4065 | <.0001 |
|  | **EX5** | **THEEX5** | 0.30930 | 0.04941 | 6.2595 | <.0001 |
|  | **EX6** | **THEEX6** | 0.21786 | 0.04663 | 4.6725 | <.0001 |
| **潛在** | **FX1** | **PHI1** | 1.00000 |  |  |  |
|  | **FX2** | **PHI2** | 1.00000 |  |  |  |

一張含有 文字, 字型, 螢幕擷取畫面 的圖片

自動產生的描述

1. Which model is the best??

GFI 、 McDonald、BB NFI 以 0.9 為標準 大於0.9比較好

AGFI 大於 0.8 比較好

Ch7 Cluster analysis 集群分析

1. Use “RMS 標準差” and “R 平方” to identify the number of clusters

[ 群集歷史 – 新集群RMS標準差、半偏R平方、R平方、重心距離 ] 看從第幾個跳到第幾個比較有跳耀的感覺 題目可能會提示比較明顯的

若從集群數目3 跳到 集群數目2 比較明顯 則會分成三群

1. Name all clusters 命名

[ The MEANS Procedure ] 可以看到每個變數在分群中的平均值為多少

並運用他的高低去命名