factor analysis

這是我們第一個section 是在介紹factor analysis分析的動機與目的

Factor analysis 的目的是在做啥, 我們看先看個例子, 在這個數據當中有兩個變數,爾數據是在量測國小學生的閱讀能力跟鞋子的大小, 看這張圖能發現有很強的correltaion如果再回歸當中把 鞋子的大小當predictor 把閱讀能力當作response 然後用regression解釋的話當鞋子越大閱讀能力就越高,換句話說當你買大一點的鞋子,就能提升你的閱讀能力,直覺上是不可能的事情. 在這例子當中把回歸用因果論去解釋是蠻荒謬的事情, 就算回歸fit 出來 很高只能表示 Response 跟predictor 的correlation很強, 這個狀況下就有回歸當中所謂的latent variable 的存在

這個例子可以想到latent variable 是年紀 當年紀越大(接受的教育越多),所以閱讀能力跟著提升. 而factor analysis 就在處理如何從觀察到的變數萃取出 latent variable. 然後我們稱這latent variable 為factor

那我們現在來看 factor analysis 能夠做甚麼

第一個 dimension reduction (跟pca 一樣) 但這並不是主要的目的

第二個 是希望把很多的變數中的relationship用比較少的factor 去描述他.

不管在做任何分析之前我們會有做模型的假設, 然後在這裡我們是用 Orthogonal Factor Model 去描繪factor 跟 Observer variable 的關係, 然後假設是這樣的形式, 他假設變數之間有線性關聯, 然後這些F就是所謂的COMMON FACTOR 剛才的例子就是 age. L 為factor loading, 代表這個factor影響X的大小, 假設X1為問卷題目1, F為問項的概念,當l越大問項跟該概念有高密切的關係.