

教育行政類科適用教材

教育統計

 韓哲 老師編著



科舉大師數位科技股份有限公司

目錄

第一章 緒論	1
第二章 描述統計	15
第三章 推論統計	40
第四章 實驗設計	99
第五章 相關迴歸	147
第六章 無母數與卡方	180

第一章 緒論

在日常生活中測量的概念，大家其實都耳熟能詳，像是身高、體重、氣溫、飲料容量、智力、成績等等，皆與測量的概念相關。另外在不同工作領域上也會有其測量的特別方式。例如：在醫護人員會測心跳、體溫等等生理相關的反應；體育教練則會測量選手的跑步時間、心跳率或各種運動能力相關表現。測量與我們生活息息相關，不過要說明測量的概念卻不太容易，更別提一般人對於測量工具僅止於如何應用，而非去瞭解測量背後的原理，因此接下來的內容將逐一說明測量所需擁有的基礎觀念。

第一節 測量與測驗

壹、測量與測驗

一、測量的定義～賦予符號標記的過程，來說明個別差異

所謂的測量有兩種含義，第一種是用數字去呈現人事物中某個屬性的數量；第二種是利用分類的方式去界定人事物中某個屬性是隸屬哪種類型。前者是所謂的度量化(scaling)，後者則是分類(classification)的工作。

二、測驗的定義

在心理或教育相關測驗中，即是利用測量去對人的某個屬性，進行分類或是用數字來反映該屬性，比如智商、學習動機或是性格等。因此想要測量這些屬性的專家，就必需要設計出能夠真實反應出這些屬性的測量工具或測量方式。而測驗正是被發明出來測量人類各種屬性的科學化的工具。

所謂科學化的測量工具必需包含下列幾個條件：測量前必先有理論能夠解釋被測對象的性質，同時在測量過程也需要符合標準化，而測量的結果必需是可靠及正確的，沒有前述的條件，測量結果解釋就沒有太大的意義，像是利用尺量桌子長度，假若尺的測量結果不正確及可靠，就算有量出結果，在解釋上也沒有多大意義。這一些關於測驗的觀念將會在

往後的章節當中逐一說明。

貳、測量的基本形式

一、測量方式

(一) 結構化與非結構化

測量結構化與否的測量反映了測量的標準化程度。結構化測量如結構化問卷，就是具有一定作答格式與內容的測量工具，施測者必需完全依照測量工具所提供的標準刺激去收集被測者的反應。而非結構化測量則是依照受測者不同的情況，施測者可以適時介入測量過程並主導問題方向。這類測量的標準化程度較低，但是資料的豐富性較高。

(二) 封閉式與開放式測量

由研究者預設受測者回答的內容及範圍。預先設定題目的選項，稱為封閉式問卷(close-ended questionnaire)，受測者完全根據研究者提供的選項來作答，如薪資收入等。反之，當某些問題的答案是無法藉由選項呈現，如對特定事件的看法等，這類問題需採開放式的作答方式。稱為開放式問卷(open-ended questionnaire)。

二、測量結果種類

(一) 類別性測量(categorical measurement)

類別測量的主要功能在於區分差異。確認受測者的屬性、類別。類別性測量是最常被使用的測量方式，例如性別、宗教、職業等等。類別性測量的條件，第一是題目選項間必需完全互斥。第二是選項要能夠周延的包含所有可能選擇。

(二) 連續性測量

連續性測量主要在進行特定屬性程度上差異的測量，以反應某些概念或現象的程度強弱之分，像是對特定事件的認同程度差異。以下是態度量表常見的幾種連續性測量方式：

1. 李克特式量表

Likert (1932) 提出，假設受測者作答的量表值為等距量尺（不同選項間有相同差異量），可以針對不同人在同一題或是同一個人在不

同題目間之分數進行相互比較，另一方面，受測者在答題過程沒有反應偏差的現象出現，像是選擇題目選項時，一直勾選無意見。為目前使用最為廣泛的作答量尺，又因為假設受測者回答結果是等距量尺，故可以進一步分析題目的特性，例如：可以計算內部一致性信度，來反應題目的可靠性。

2. Thurstone 量表

Thurstone 和 Chave (1929) 首先提出，假定每個題目選項的量表值有程度上的不同，透過一群熟悉相關測驗主題的專家（50 位以上），對每個題目進行評分，最後依據專家的評分計算中位數，最後從專家評分結果中挑選適當題目編成量表，並給予受測者作答，再依作答結果加總平均，所得結果即反應態度分數。

3. Guttman 量表

Guttman (1945) 提出，認為量表題目具有同質累積特性，故亦有人稱為「累積量尺」，一個量表僅能測量一個變項，每一個題目都是單向性和次序性，且同一量表題目皆為同質性，所以能排出每個題目的態度程度差異，並且在建立量表過程可以透過數值挑選題目。

4. 語意差別法(亦稱形容詞揀選和語意分析或形容詞檢核表法)

Osgood, Suci, & Tannenbaum (1957) 使用形容詞檢查表要求受測者基於施測者提供的形容詞，檢選出符合描述自己的描述，再進行語意分析（因素分析）。該測驗目的原本是用來測量態度與概念的意義，但是後來成為測量性格的實用工具。例如：Pervin (1967a,b) 使用上述測量方式，發現自認與學院環境越不相似者，他們的大學生活越不滿意，且很有可能輟學。

5. 強迫選擇法

是一種利用兩個立場相反的描述句，一句代表正方立場，一句代表反方立場。要求受測者自兩句當中挑出較接近自己想法的選項。並將正方選項勾選題數加總得到該量表的總分。優點是明確指出個人立場，避免反應心向的問題及反應。缺點是量表題目數量較一般李克特式量表多。

參、反應心向

一、反應心向(response set)的定義

反應心向是指受測者在作答上具有某種固定的作答傾向。比如受測者傾向回答特殊的答案，稱之為差異反應或離異反應(deviation)。或是傾向回答同意或傾向回答不同意的答案，稱之為順從心向(acquiescence)或唯唯諾諾。如果受測者有習慣性的漏答題目，也可歸為反應心向的一種，稱為粗心(careless responding)，或是題項遺漏(omitting items)。

如果受測者在填答時希望討好施測者或想要透過好的分數以使他人對自己有好的印象，稱之為偽善(faking good)。相反的，如果受測者想要透過分數造成負面印象或藉此獲得他人注意、同情或幫助。則稱之為偽惡(faking bad)。而如果受試者以中間值或不置可否的答案來作答的話則稱之為中庸傾向(mediocre)。

在反應心向當中最常見的就是社會期許或稱為社會偏愛心向(social desirability)，也就是受測者傾向以社會大眾所期許的選項來作答，並避免使用社會不贊同的作答方式。社會期許性跟偽善及偽惡不同之處在於，社會期許是一種無意識的自動化行為。而後兩者則是有意的欺瞞。

二、反應心向的處理

(一) 廢卷處理法

研究者藉由受試者填答的趨勢來判斷是否存在特殊的作答反應，像是極端反應就是一種受測者傾向選擇極端的答案作，例如在五點量表中大量選則一或五，或明顯偏向回答正向答案或負向答案。研究者可將該答案卷以廢卷處理，不列入資料分析中。

(二) 事前估計法

將測驗題目在施測之前，交由其他專業人員評估題目的社會期許性程度。並在進行結果分析時，將代表此社會期許程度的數值作為控制變項。或是在預試分析時透過因素分析找出具有高度社會期許性的題目，並考慮將這些題目刪去，以去除社會期許的影響，然後再正式施測。簡單來說，事前估計法就是利用統計上的方式去控制反應心向的影響。

（三）事後估計法

處理反應心向的另一種做法，是認為反應心向需要被測量與估計以進行統計處理。因此有些學者會主張將反應心向的得分作為抑制變項 (suppressor)，並放入迴歸方程式當中，利用統計分析進行控制。這樣能夠使其他預測變項能夠在排除反應心向的情況下進行估計與分析。

（四）使用其它測驗形式

為避免反應心向的問題，可以選用較不受影響的測驗題型，例如非文字測驗。以投射測驗、繪圖測驗或語句完成測驗、自由反應測驗等。避免使用自陳量表。或者可以減少量尺的選項，像是將七點量表改為五點或三點量表。即可以減少反應心向的偏離情況。

第二節 測量與統計

壹、心理測量的基本概念

在心理測量當中，我們會根據心理測量的結果，利用描述統計將測量結果的資料整理為有意義的訊息。並且利用這些訊息依據統計觀念，作進一步推論預測。

貳、測量當中的統計概念

在心理測量當中有三個基本概念：變異、相關跟預測。其中，變異的概念是最為重要。測驗的目的就是利用量化的方式來表達個別差異。相關則是利用統計的方式來回答某兩件事情是否有關聯的方式。而預測可說是相關的延伸和應用，預測能夠回答更具體的問題，像是我們想要知道智商跟學業表現是否有相關存在？那我們是否能夠利用智商去預測個體在學業上的表現？可以透過統計的迴歸分析來回答這類的問題。

參、前言

教育統計是一門很詭異的學科，很多想進教育行政的同學，剛開始遇到統計時，心裡面總是充滿焦慮與慌張，總認為念教統是需要超強數學能力，才有能力去學好它。其實教統屬於應用統計，並非理論統計，只要你具備基礎的數學能力，教統的學習旅程，就會比較寬廣，更重要

的是你念的是教育行政，並非統計或數學。

另一方面公職錄取率並不高，競爭對手的能力也逐年提升，主要是因為越來越多人想要進入公職，其中在準備時，讓許多人頭疼不已，因為很少人知道該如何準備，許多人裹足不前，其實裡面有許多內幕，你需要有經驗的人引領。以老師的經驗，可以跟各位同學說，關鍵在於理解教統的核心觀念，絕對可以幫助各位同學獲得適當的分數，就算是對手能力逐年提高也不用擔心。

另一方面，考公職的準備過程，需要相當的毅力，因為沒有良好態度與毅力堅持，再好的協助，也是枉然。更重要的是，老師過去的學生考上後，再遇到老師時，都會跟老師說～老師我考上了，同學表情散發出一股獨有的自信與雨過天晴的開朗，也往往讓老師心有所感，並獻上深深祝福，因為又有一位同學憑藉自己的毅力，述說自己的故事，這才是我教書最大的成就感，也是我願意持續自己熱情的原因。

這一堂教統旅程即將開始，希望聽課的同學，能 Enjoy 課程，跟你的學長姐一樣，對你周遭的人述說一個新的故事。

一、數學基礎

大多數同學在學習數學時，往往深感恐懼，主要是因為數學太過抽象，學習方式有採取脈絡獨立的學習，讓大多數同學深感挫折，並興起數學不好，統計就會不好。其實學教統的數學基礎，僅需要少數觀念具備，我們就可以繼續學習旅程下去。

(一) 國中的數學大致上可以區分成三大塊：

1. 數的解析

- (1) 質數、因倍數的啟發～加速計算技巧（11 到 40 的平方）
- (2) 數列級數～如何尋找數字背後的規律（1、1、2、3、5、8、、）
- (3) 機率～開門有獎的故事（樹狀圖、貝氏定理）

2. 代數

- (1) 常數與變項～ abc v.s. xyz
- (2) 乘法公式～帕斯卡三角形及其延伸
- (3) 聯立方程求解～代入消去法（行列式～克拉馬定理）

(4)函數～一個 X 不可以對應到多個 Y （估計誤差來源）

(5)配方法～找極大、極小值（偏微分）

(6)直角坐標系～斜率（現在高中數學才教）

3.幾何

(1)函數與圖形在直角坐標系的關係～函數的幾何關係

(2)圖形意義～看圖要注意橫軸、縱軸的意義，並由圖形中數字的分布情況，找出符號表達意涵

看完上述國中數學的相關統計基礎，可能有同學已經昏了過去。各位同學放心，接下來要說明的是數學學習因應方式，剛剛你可以睡著，但現在請稍稍清醒一下：

a.符號文字化，文字符號化

b.複雜數字簡單化

c.能說出來最棒

d.老師說的，你要注意

e.練習至上

二、統計定義

統計學(statistics)：

是指蒐集、整理、分析、呈現及解釋樣本資料的特性，並且在多少信心或可以忍受的風險下，將樣本資料所獲得的結果，用來推論母體性質與事實，進而做出決策的一門科學。換言之，是研究量化資料蒐集、處理、呈現，並分析其規律性的科學。

心理與教育統計學(Statistics for Psychology and Education)乃是統計學的應用延伸，藉由數理邏輯所發展出來的統計原理和公式，作為“研究心理學與教育學問題”的工具。

三、根據母群性質分類

(一)(有)母數統計(parametric statistics)：

需要得知母群分配特徵(受母群分配特性限制)的統計方法，常用於等距、比率量尺等連續變項資料，例如： Z 、 t 、 F 等考驗法

(二)無母數統計(Nonparametric statistics)：

不需瞭解母群分配特徵(不受母群分配特性限制)的統計方法，常用於**名義**、**次序量尺**等**間斷**變項資料，例如：K-W 二氏、W-U 二氏考驗。

四、統計內容分類：

(一) 描述統計 (descriptive statistics)：

利用測量、計算、描述、劃記和排列次序等方法，將一群資料加以整理、摘要、濃縮，使其呈現結果「**具意義化**」，並讓人容易了解的統計方法，例如：教育部想要統計全國學校的師生比率、老師想要瞭解全班學生的平均成績。

(二) 推論統計 (inferential statistics)：

透過描述統計將原始資料進行處理後，例如：我們想知道全國大專學生體重過重的人數，由於受限於**時間與經濟**因素，所以我們可以從全國大專學生(**母群**)中隨機抽取具有代表性的學生(**樣本**)，再根據樣本的資料來推論母群的性質，並說明進行決策的信心程度或所冒之**風險**程度。

(三) 實驗設計 (design of experiments)：亦稱為 ANOVA

透過研究方法配合，來確認因果關係，故為了驗證某種研究假設，必須操弄**自變項**，控制**無關變項**，以觀察自變項對**依變項**所產生影響，例如：我們要研究不同的教學方法對學生考試表現的影響，則「教學方法」視為自變項，學生「考試表現」為依變項，同樣會影響學生表現的「作答時間長短」就屬於無關變項，因此必須被控制住，像是所有考生必須以相同時間限制內作答完畢。

(四) 相關、迴歸 (教育測驗的統計基礎之一)：

同時關心 X 與 Y 變項間的關係，希望透過樣本的性質，建構模式並應用推論統計技巧，推論母群是否具有該模式存在，例如：由高中在學成績預測大學聯考成績，並且依據該樣本資料所建立的迴歸線，來推論母群是否有迴歸線存在。像是利用過去大學指考成績與學生入學狀況，建立模式，供作新一年考生的落點分析。

(五) 無母數統計 (卡方檢定 ~ Pearson, 1900)

心理與教育領域所研究領域的資料，並不一定可以使用有母數統計，來進行推論，所以需要另外的統計程序進行檢驗，例如：學期中與學期末對普心的喜歡態度改變狀況。

五、研究變項的意義及探討

（一）變項的意義

人、事、物的特徵中，往往有可以透過變動的量，來加以表達，但是因為量一直不斷在改變，所以我們會利用概念代表這個不斷變動的量，而這個概念我們稱為變項，例如：班上學生的成績就是一個變項，它可以代表全班學生的成績變動。

（二）概念與構念的意涵

- 1.概念=事物共同特徵的想法或名稱，以約定俗成的文字表達出存在心中的想法，具有多樣性、普遍性、工具性以及複合性，例如：貓。
- 2.構念=又稱為假設性構念的學術用語，既不可直接觀察也非具體存在，乃是學者基於學術目的，為了說明假設實體所發明或借用的名詞，例如：成就動機。

（三）變項的種類

1.一般教育研究中常見變項

（1）自變項

研究者於實驗情境所操弄的條件，相對於依變項而言，藉由自變項的改變，觀察對於依變項的影響，例如：教學方法對成績的影響。

（2）依變項

隨著自變項改變而改變的變項。例如：成績隨著焦慮程度變化而改變。

（3）中介變項

和實驗處理可能無關，但是實驗處理會透過中介變項對依變項產生重大影響，例如：學習動機是一種中介變項，教學方法對成績的影響，是透過動機間接影響而成的。

（4）調節變項

和研究問題可能沒有關係，但是卻會影響自變項與依變項間的關係，可以透過特定方式加以控制，例如：IQ 就是一種調節變項，它會影響教學方法與成績，所以我們可以利用統計方法控制住 IQ，避免 IQ 影響研究結果的推論。

(5)混淆變項

即使研究者對調節變項或是中介變項做好安排及控制，仍有其他的複雜因素影響實驗結果，例如：想探討數學學習的性別差異，卻忽略情緒干擾研究結果，此時情緒干擾就是一個混淆變項。

(6)無關變項

研究情境中，能為研究者操控而完全控制，不再對實驗結果產生影響的變項，例如：教學方法對成績影響，其中測量依變項結果受作答時間影響，故限制作答時間相同。

2.主動變項與屬性變項的概念

主動變項在於其可主動為研究者所**操弄**；屬性變項則為不可操弄，而僅止於可**測量**的變項，例如：睡眠時間可以透過剝奪，加以操弄；性別則無法操弄僅能測量。

3.心理學變項與社會學變項的概念

心理學變項意指描述個人心理特質；社會學變項則以個人在社會團體所獲得的各種特徵而言，例如：學習動機就是屬於心理學的變項；社會階層就是社會學變項。

4.連續變項與類別變項的概念

連續變項可以用連續數值表示其特徵的變項，例如：時間、成績；類別變項則是以相同本質來對變項加以歸類，例如：IQ 區分成高、中、低組。

5.預測變項和效標變項:

預測變項(predictor)：用來預測的變項；效標變項(criterion)：被預測的變項。例如：用高中成績來預測大學聯考成績，高中成績(X)就是預測變項，而大學聯考成績(Y)就是效標變項。

6.研究變項的界定方式

(1) 界定研究變項的意義與方式

a. 概念型定義

依假設或概念的界定標準，也就是概念界定概念，而非依觀察或測量所得的定義。例如：睡眠品質對成績的影響。

b. 操作型定義

依可觀察或測量特徵以界定研究變項，包含了研究程序及活動，用以了解操弄變項的實際程序活動，難以涵蓋較多的範圍，例如：睡眠時間的多與少對心理學成績的影響。

(2) 操作型定義的種類

a. 測量的操作型定義

根據如何測量研究變項而下定義，也就是告訴研究者如何測量和觀察研究變項的操作型定義，例如：IQ 分數是透過 IQ 測驗所獲得，最終依據分數的差異來進行分組。

b. 實驗的操作型定義

根據如何操作研究變項而下定義，也就是在告訴研究者操作研究變項的詳細過程或活動內容，例如：透過剝奪睡眠時間來進行分組。

六、測量的概念

(一) 測量的定義

賦予個體特徵的數量化過程，也就是以數量描述個體特徵的個別差異程度之過程。例如：智力（IQ）經過智力測驗測量結果為 120 分。

(二) 量尺的概念

1. 量尺定義

使用數字來表示研究者所關心的變項特徵，用以表現不同程度差異，而不同數字代表不同意義，例如：IQ 測驗結果為 120。

2. 量尺的種類(Stevens, 1946, 1951)

(1) 名義量尺(nominal variable)：

又稱類別變項，使用數字或文字，來辨別任何事物或類別之

不

同。例如：球員背號、學號、顏色類別。

(2)順序量尺(ordinal variable)：

又稱為等級或次序變項，除了分出不同之外，尚能表明各類別的次序、大小或優劣。例如：名次、等第、軍階。

(3)等距量尺(interval variable)：

除了分類、比較次序外，因為它具有「相等單位」(equality of units)的特性，因此還能算出大小差異量。例如：溫度、亮度、音量、等，我們可以認為 $12^{\circ}\sim 24^{\circ}$ 的差異量等於 $24^{\circ}\sim 36^{\circ}$ 的差異量。

(4)比例量尺(ratio variable)：

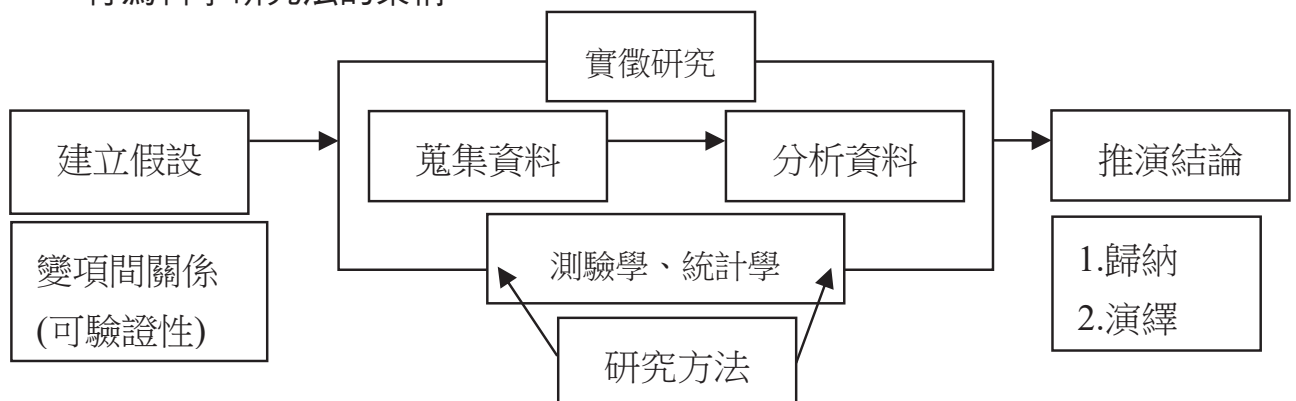
除了分類、比較次序、算出差異量外，因為它具有「絕對零點」(absolute zero)，因此可以求出比例，例如：2 公尺是 1 公尺的 2 倍。何謂絕對零點？是指測量所得之結果是具有絕對性。例如：身高、體重、年齡、等，皆是測量出多少，就代表多少。

表 1-1 四種量尺特性比較表

特性 量尺	不同	大小	差異量 (相等單位)	比例關係 (絕對零點 ~具有相同 起點)	合適的例子	可能使用的統計方法
名義量尺	√				背號，學號	無母數
次序量尺	√	√			等第，名次	無母數
等距量尺	√	√	√		溫度，亮度	母數
比例量尺	√	√	√	√	身高，體重	母數

肆、統計在心理學用途

一、行為科學研究法的架構



二、假設(hypothesis)

對研究現象所提出的暫時性解釋。

- (一) 現象：一種概念；指出變項間的關係(差異性、關聯性、因果性)。
- (二) 暫時性解釋：有待實徵驗證的。

三、形成知識的方式-研究方法

(一) 科學方法的概念

1. 結合演繹法與歸納法，稱為歸納-演繹法。俱備公正客觀的性質。先由觀察個別現象，根據所得結果來判斷研究假設的真偽。
2. 科學方法的五大簡化步驟：
 - (1) 選擇及界定問題。
 - (2) 陳述研究的問題。
 - (3) 推演和執行研究程序。
 - (4) 蒐集且分析所得資料。
 - (5) 提出研究結果與報告。
3. 科學研究必須具備描述、控制、預測與解釋的研究目的；同時也必須具備系統性、實徵性、可證驗性與可重複性的特點。
4. 科學研究在社科研究，仍有可能的限制：
 - (1) 研究題材複雜，難以精確預測及控制人類行為的複雜性。
 - (2) 直接觀察常流於觀察者角度之解釋，易有主觀而失客觀的弊端發生。
 - (3) 精確複製常因特殊社會情境脈絡而有所限制，難有準確的複製。
 - (4) 研究結論不若自然科學推論普及程度，行為常存在變異而打破結論一致性。
 - (5) 實驗控制不易且研究變項眾多，另有測量工具利於假設構念而設計，難有具體界定。

四、心理與教育統計學的功能

(一) 進行科學研究的工具