淺談高職統計量數的應用與教學

數學科　謝宗儒老師

壹、前言

在現今資訊爆炸的時代，我們每天都要面對並解讀大量的數字或文字訊息。然而，若是沒有一套有系統的處理方式來幫助我們有效率並快速的呈現出有用的資訊，對於龐大而未經整理的資料集，我們並沒有理解的能力。

統計學是近百年剛興起的一門學科，它可以藉由一些統計技巧與方法，將龐大的資料彙整並提供相關訊息給決策者作為決策參考。其實，在日常生活當中，我們可以很輕易的看到許多統計方法的使用。例如：新聞報導、報章雜誌、成績分析…等。因此，具備基本的統計知識與統計分析技巧已成為學生學習的重點之一。所以，本篇文章將從現今職校統計教材抽取部分主題做一更深層的探討，以提供教師在教學或學生學習上的補充方向。

貳、正文

一、統計學的意義

統計學主要應用在面對雜亂無章的數據資料中，經由資料的蒐集、整理與分析，提煉出資訊和知識，作為合理決策的科學方法。由此可知，統計學的應用其實相當廣泛。例如：保險科學、生物醫學、商業應用、經濟學、心理學…等，都能發現統計應用的影子。

二、統計學的分類

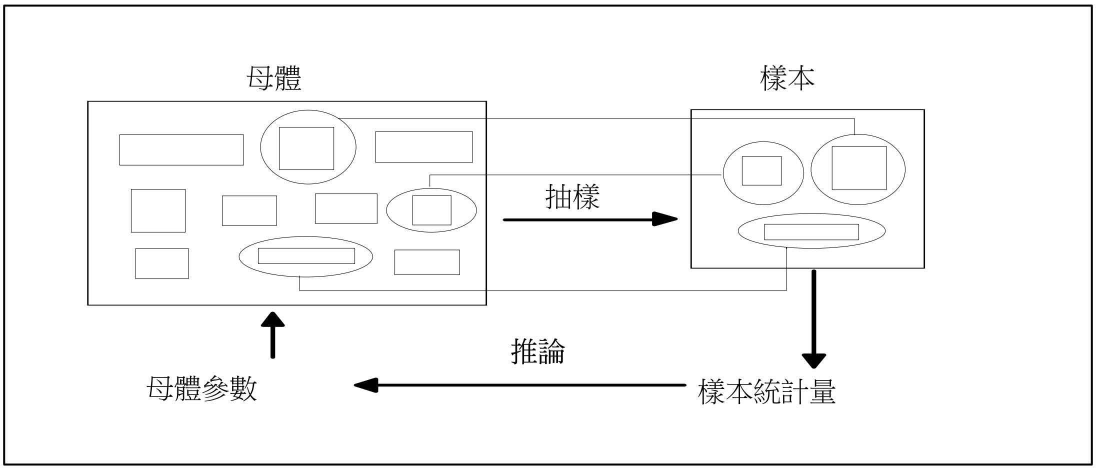
統計學就處理問題的不同大致上可區分為兩大區塊，分別是敘述統計以及推論統計。以下就這二部分做一個簡單的介紹。

1. 敘述統計

敘述統計主要是經由資料的蒐集、整理分析，並使用適當的統計方法，表現出資料的特性。在高中（職）所探討的內容包括集中趨勢量數（平均數、中位數、四分位數、眾數…等）、差異分析與圖表的呈現方法。

2. 推論統計

推論統計探討的是利用機率理論，以及尋求最適當的抽樣方式蒐集樣本，並且根據蒐集到的樣本特性，再經由假設與檢定推論未知母體的方法。由於這已經牽涉到更高深的數學理論及數學計算，所以這部分在高中（職）領域並不會探討到，通常留到大學時才會討論。但我們仍可以利用下面的圖一表來更清楚了解統計推論的基本概念。



圖一：母體與樣本間的關係（應用統計學 林惠玲 陳正倉著 雙葉書廊 2002 ）

三、常見的統計量計算

在這部分我們將探討在高中（職）校領域中幾個常見的統計量數，並深入探討其中的意涵。

1. 平均數

平均數應該是大家最早接觸到的第一個統計計算方法。然而，在一般學生的認知當中，

$$\text{平均數} = \frac{\text{觀察值數字總合}}{\text{觀察值個數}}$$

這樣的想法其實不全然是正確的，因為平均數又可往下細分為 ： 算術平均數、加權平均數、 幾何平均數、 調和平均數 、截尾平均數…等。依照不同的條件與情況，所使用的平均數就不一樣。而這樣的情況往往也使得大家誤用統計方法，造成解讀上的錯誤。在這，我們先將焦點集中在較常見的算術平均數及加權平均數上。

首先，我們希望藉由下面的例題來說明算術平均數及加權平均數在計算上的統計涵義和這兩個計算方法之間的差異性。

例題一：小明此次段考成績如下：國文 85分、數學 90分、英文 70分、社會 75分，試求小明段考成績的平均數。

這個基本的計算問題對大部分的學生都能輕易拿分，大家的算法不外乎是

$$\text{平均數} = \frac{\text{觀察值數字總合}}{\text{觀察值個數}} = \frac{85+90+70+75}{4} = 80$$

但若我們將題目在加上下列的條件：國文每週6節課、數學每週4節課、英文每週4節課、社會每週2節課，此時的平均數就變為

$$\text{平均數} = \frac{\text{觀察值數字總合}}{\text{觀察值個數}} = \frac{85\times 6+90\times 4+70\times 4+75\times 2}{6+4+4+2} = 81.25$$

從上例的兩個算法來觀察，相同的成績卻造成不同的平均數，原因是為何呢？其實，在統計的概念中，可將平均數與期望值的概念畫上等號，我們將期望值簡單定義如下：

期望值= 預期得到的報酬 × 可能發生的機率

因此，第一個算法我們可將它改寫為

$$\text{平均數} = \frac{\text{觀察值數字總合}}{\text{觀察值個數}} = \frac{85+90+70+75}{4} = \frac{1}{4}\times 85+\frac{1}{4}\times 90+\frac{1}{4}\times 70+\frac{1}{4}\times 75$$

看出兩者之間的差異性了嗎？其實，一般平均數的算法我們往往已經先假設了它們發生的機率是相等的，而加權平均數則是針對各觀察值的發生機率下去計算，因而造成兩種平均數結果的不同。然而，或許有學生會有疑問，那麼哪種成績的計算方式對學生比較有利呢？我認為這就得看學生各科之間成績的表現了，若是每科分數相差不多或是所占機率較低的學科分數較高，那麼第一種加設各科發生機率相同的算法可能比較占優勢；換句話說，若該生成績是所占機率較高的學科分數較高，那麼，第二種加權的算法就會比較占優勢了。

最後說明一個平均數的重要概念，經由統計方法呈現的數據都只是一個參考數值，它不一定會發生。最常見的一個例子：

例二：投擲一個公正的骰子，試求點數出現的期望值

因為是公正骰子，故六面出現的機率皆為二分之一。

$$\therefore \text{期望值} = \frac{1}{6}(1+2+3+4+5+6)=3.5$$

我們求得期望值為 3.5，理論上可以解釋為，當我們長期重複實驗下來，骰子平均出現的點數為 3.5，但我們都曉得這是不可能會發生的。

2. 變異數與標準差

另一個在統計上常用的統計量數為變異數，其主要用途在衡量一組資料的離散程度。簡單來說，當變異數小的時候，代表各數值之間的差異低，平均數代表性高；變異數大，代表各數值之間的差異高，平均數代表性低。但這部分因應課綱的規劃，也做了些微的變動，卻也造成許多教師與學生的教學與學習上的困擾，新舊計算方式列出如下。

$$\text{標準差} = \sqrt{\frac{\Sigma\left(\text{觀測值}-\text{平均數}\right)^2}{\text{觀測值個數}}}$$

（舊版本計算公式）

$$\text{標準差} = \sqrt{\frac{\Sigma\left(\text{觀測值}-\text{平均數}\right)^2}{\text{觀測值個數}-1}}$$

（新舊版本計算公式）

造成困擾的首要因素，我們發現第一次課綱的變動並非所有教科書都統一公式，因而造成學生學習上的困擾，不曉得哪個公式才是正確的。第二，兩個公式之間的差異大部份的教科書都沒詳細交代，特別是為何新版本的公式分母要減一。第三，教師缺乏專業進修，一開始的變動相信很多若沒特別翻書查閱的數學教師也許連這兩者的差異在哪也分辨不出來，更別談論他們是如何教導給學生們了。

其實這兩個公式在統計學上的定義還是有差異的，舊版本的計算方法是用於當蒐集的資料為母體資料時，我們直接除以母體觀察值個數即可；當蒐集的資料為樣本資料時，我們必須除以觀察值個數再減一。事實上，在更深入的推論統計學探討中，我們常用樣本變異數來估計母體變異數，為了不使估計產生低估現象，因而將觀察值個數再減一。更進一步的，根據不偏性的原理，我們可以證明這兩者其實是相等的。

【證明】

$$
\begin{aligned}
& E(S^2) \\
&= E\left[\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right] \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)\right\} \\
&= \frac{1}{n-1}(n-1)\sigma^2=\sigma^2
\end{aligned}
$$

至於是否一定要減一，這並沒有特別規定，我們也可以減去其他數字，最後只要做係數上的調整仍可得到相同的結果。

四、幾個有趣的統計問題

1. 這樣算考的不錯嗎?

事件：老謝在某次的數學考試拿了 55 分，由於害怕回家後遭受一陣毒打，於是在爸媽看成績前老謝特別強調：『這次考試全班平均才 40 分，我可是高於班平均 15 分喔』。

一般在比較成績的好壞我們常拿該科分數與班平均做比較，這事件的情況看起來該生考得應該算是不錯的成績，但只憑單一統計量數就輕易的下結論是很危險的。特別是處理平均數這類容易受到極端值影響的統計量數，為了更進一步的確定，我們最好再參考該班的成績變異數或是標準差來判斷此平均數的代表性高低。如此才能進一步的評斷該生此次的考試表現是否真的優於班上平均表現。

2. 你願意接受這份工作嗎?

事件：老謝應徵工作，公司規模：老闆、經理 2 位、組長 3 位、員工 5 位。老闆說：公司平均月薪 6 萬元，受訓期間 1 萬五千元，很快會加薪。

這是在一般報紙求才很常見的甄才手段，打著高平均月薪來吸引人才應甄。但大部份的情況都是進入公司後才發現，一切與當初的預期是相差甚多的。雖然誠如所說，平均月薪六萬，但分佈可能如下表所示。

	人數	月薪	總計
老闆	1 人	24 萬	24 萬
經理	2 人	10 萬	20 萬
組長	3 人	4 萬	12 萬
職員	5 人	2 萬	10 萬

這是很明顯平均數受到極端值影響的例子，所以一般對於統計觀念不甚了解者，往往都會被統計數據所蒙騙。

3. 賭或不賭

事件：假設教室裡有大約 50 位同學，今天老謝和同學們打賭。在教室裡，至少有兩位同學的生日是同一天，若輸的人要給贏的人一百元，你會和老謝打賭嗎?

首先，假設生日只有 365 種可能。考慮一般的情形，若教室裡有 n 個人且至少有人同一天生日的機率為 P，則

1-P = 教室裡 n 個人的生日都不相同的機率

$$= \frac{365\times364\times\cdots\times(365-n+1)}{365^n}$$

所以，當 n=50 時，我們可以求出 P 的近似值大約為 97% 。這個結果其實出乎大家的預料之中，在一般的直覺當中，這樣發生的機率應該是相當低的，但透過一些簡單的數學計算，卻發現可能發生的機率卻高達 97%。而這就是給決策者做決定的重要參考依據。

參、結論

具備基本的統計學知識已成為未來學習的趨勢，由近期課綱的改革可以發現，統計學的內容逐漸朝著內容的廣度去發展。值得注意的是，在課程內容裡加入了信賴區間的主題，然而，就一個學統計的過來人來看，我認為這部分可以著重在數據的解釋

即可，切勿將學習重點放在公式的引導和計算題上。因為，這部分的學習歷程必須經過較扎實的統計理論和機率論的訓練才能理解其主要意涵，並非短時間就能速成。所以，教師在教學時可搭配報章雜誌的統計數據，以實際的生活例子為主，課本理論為輔，使學生了解日常生活中處處皆可見到統計應用的實例。

肆、參考文獻

一、原著/墨爾 譯者/鄭惟厚。統計學的世界。(台北市：天下遠見，2002.12)
二、鄭惟厚。統計學的世界。(台北市：天下遠見，2007.09)