



利用Excel對學生分數進行簡單統計分析方法

文·圖 | 陳國勳



前言

本文旨在介紹常用的基本統計分析方法及其相關概念，並通過學校常用辦公軟件Excel，實現對學生成績進行統計分析，並能通過函數、工具箱及圖表等，幫助老師快速地了解學生成績情況，本文會介紹的功能有統計敘述、次數表分配表、正態分佈圖、t-檢定等常用方法。

統計分析方法相關概念

設有n項成績資料 $X = x_1, x_2 \dots x_{n-1}, x_n$ 有以下相關概念。**平均數**：所有觀察值的總和除以個數，是資料的平均點，公式 $\bar{X} = \frac{1}{n} \sum x_i$ 。**全距**：所有觀察值的最大值減去最小值，代表分散程度，公式是 $\max(X) - \min(X)$ 。**變異數、標準差**：用來衡量觀察值與平均數的距離，變異數公式 $S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{X})^2$ ；而標準差公式 $S = \sqrt{S^2}$ 。**中位數**：依數值大小排列居中的數，若數值個數為奇數，則中位數為中央之數。若數值個數為偶數，則中位數為中央二數的平均。**眾數**：觀察值中出現最多的的數值或類別。**頻數**：也稱次數，是對資料按某種標準進行分組，統計出各個組內含個體的個數。**常態分佈**：公式是 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ ，其中 σ 是標準差即等於 S ，而 μ 是平均數即等於 \bar{X} 。**t-檢定**：用來判定兩組資料平均數的差異，分為單一樣本、獨立樣本及成對樣本；獨立樣本公式是 $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$ ，其中 \bar{X}_1 及 \bar{X}_2 、 S_1^2 及 S_2^2 、 N_1 及 N_2 分別是兩組資料的平均數，變異數及個數；成對樣本公式是 $t = \frac{\bar{d}}{s_d\sqrt{n}}$ ，其中 \bar{d} ， s_d 分別是配對樣本差值之平均值及標準差。

Excel進行示例

Excel是最常用的辦公軟件，具處理資料、統計分析、製作表格、圖表製作等強大功能，其他辦公軟件檔案亦很容易轉換成Excel檔案格式，故此，將介紹幾個常用的函數與統計工具。

敘述統計的公式方法

描述統計（descriptive statistics），又稱敘述統計，是統計學中描繪或總結觀察量的基本情況的統計總稱。首先將成績資料 $X = x_1, x_2 \dots x_{-1}, x_n$ 進行排序，然後通過公式函數計算出相應的結果，如圖1，其中MAX函數找出**最大值**，MIN函數找出**最小值**，**全距**是最大值減最小值；而**分段數**是將成績資料分成k個段即 $2^k \geq n$ ，首先用COUNT函數計出筆數n，然後再用SQRT函數開方，最後用ROUNDUP函數作四捨五入處理計出每一分段的間距；AVERAGE函數計出**平均值**及MEDIAN函數計出**中位數**以及MODE函數計出**眾數**，STDEV.S函數計出**標準差**及利用VAR.S函數計算**變異數**，具體公式如下圖所示。

	A	B	C	D	E	F	G	H
1	序號	第一次大測成績				成績分析	結果	公式
2	1	40				最大值	98	=MAX(B2:B33)
3	2	45				最小值	40	=MIN(B2:B33)
4	3	45				全距	58	G2-G3
5	4	50				分段數	6	=ROUNDUP(SQRT(COUNT(B2:B33)),0)
6	5	51				分段間距	9.7	=G4/G5
7	6	55				均值	74	=AVERAGE(B2:B33)
8	7	60				中位數	78	=MEDIAN(B2:B33)
9	8	65				眾數	78	=MODE(B2:B33)
10	9	68				標準差	15.1	=STDEV.S(B2:B33)
11	10	68				變異數	227	=VAR.S(B2:B33)
12	11	72						
13	12	75						
14	以下省略...							

圖 1：成績敘述統計的公式方法

利用分析工具箱進行敘述統計

在Excel中內建有分析工具箱，可以完成進階的資料統計分析，包括日常會用到的敘述統計、t-檢定、變異數分析、相關系數及迴歸分析等…而加入方法：1)檔案->選項->增益集->分析工具箱如圖2；2)資料->資料分析工具箱如圖3；3)在視窗口設定輸入/輸出範圍及平均數信賴度，如圖4所以，分析結果如圖5。

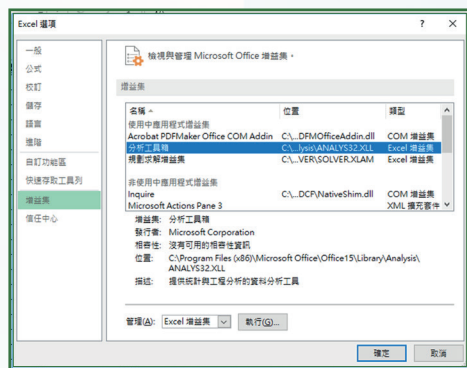


圖 2：分析工具箱增益集

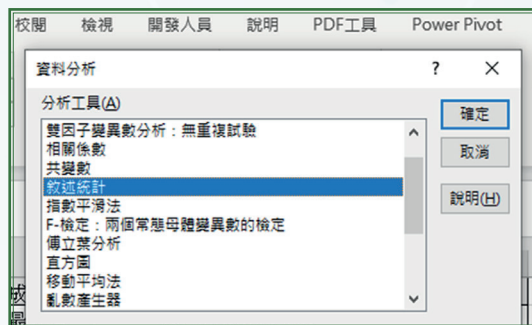


圖 3：分析工具箱中敘述統計功能

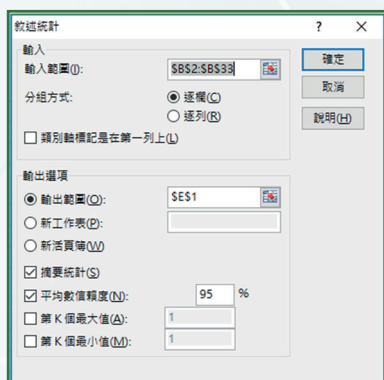


圖 4：敘述統計設定

欄1	
平均數	73.53125
標準誤	2.660942
中間值	78
眾數	78
標準差	15.05256
變異數	226.5796
峰度	-0.195396
偏態	-0.771406
範圍	58
最小值	40
最大值	98
總和	2353
個數	32
信賴度(95.0%)	5.427027

圖 5：敘述統計結果

通過以分析工具箱與上述函數的結果一致，如對函數不太熟悉的老師可用此方法，此外，還可得到峰度與偏態（度），它們代表資料的分佈型態，如下圖可大概了解學生成績的分佈情況。

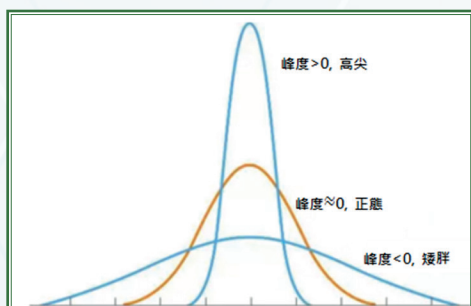


圖 6：峰度形態

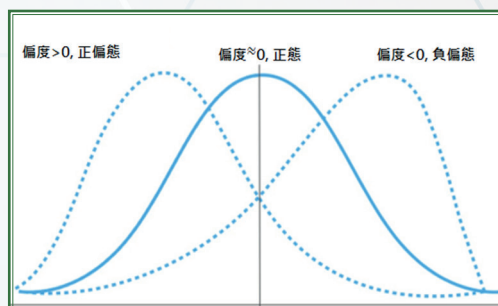


圖 7：偏度形態

建立次數表分配表與正態分佈圖

建立次數表分配表需要確定各組界，從而計算每組次數，而各組界的組寬 C ，是利用全距 R 除以分段數 K 即 $C=R/K$ ，組界的第1組為 a_1 是最小值，第2組為 $a_2 = a_1 + C$ ，第3組 $a_3 = a_2 + C$ ，第 N 組 $a_n = a_{n-1} + C$ ，形成差數列。每組界內人數，是利用FREQUENCY函數，將全部成績及組界作為參數，計算出每組界內的人數（組次數），最後，利用NORM.DIST函數計算出各組別作圖的正態分佈值。

J	K	L	M	N
組別	組界	組界代表範圍	組次數	正態分佈值
1	=G3	0-40	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M2,G7,G\$10,FALSE)
2	=K2+G\$6	41-49	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M3,G7,G\$10,FALSE)
3	=K3+G\$6	50-59	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M4,G7,G\$10,FALSE)
4	=K4+G\$6	61-69	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M5,G7,G\$10,FALSE)
5	=K5+G\$6	70-78	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M6,G7,G\$10,FALSE)
6	=K6+G\$6	79-88	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M7,G7,G\$10,FALSE)
7	=K7+G\$6	89-98	=FREQUENCY(B2:B33,K2:K8)	=NORM.DIST(M8,G7,G\$10,FALSE)

圖 8：次數表分配表與正態分佈值公式方法

建立出正態分佈圖的方法：1)選定組界代表範圍、組次數、正態分佈值；2)插入->組合式圖表->建立自訂組合式圖表，步驟如圖9左圖;從而可以更清楚看出同學成績的分佈情況，是集中在高分、合格、不合格的區域，是集中，還是分散等狀況。同樣可以通過正態分佈圖驗證出，本例子的峰度是接近正態、偏度是負偏態，與函數結論一致如下右圖所示。

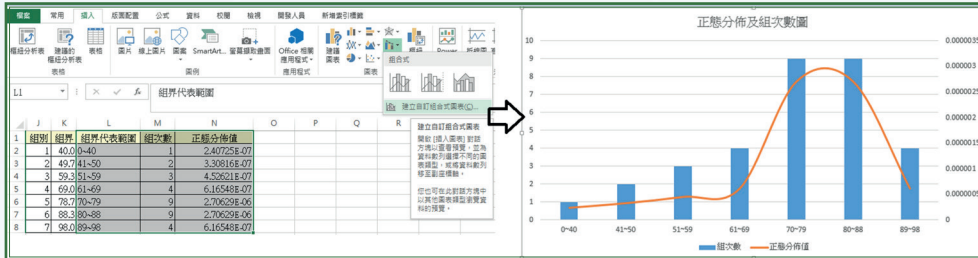


圖 9：建立正態分佈圖方法

t-檢定

t-檢定是用來檢定兩組樣本平均值有沒有明顯的差異，本文是利用兩個班同學對同一大測分進行獨立樣本t檢定，然後，同一班同學在第一次大測成績（前測）及第二次加強複習後的大測（後測）進行成對樣本t檢定，所用函數是t.test，另外，大家也可以利用分析工具箱進行詳細的t-檢定。

	A	B	C	D	E	F	G
1	序號	A班大測成績	B班大測成績			F-value	所用公式
2	1	40	61		獨立雙樣本t檢定(變異數不相同)	0.180268271	=T.TEST(B33,C2:C33,2,2)
3	2	45	62				
4	3	45	54				
5	4	50	41				
6	5	51	64				
7	6	55	83				
8	7	60	79				
9	8	65	65				
10	9	68	95				
11	10	68	92				
12	11	72	63				
13	12	75	81				
14	13	75	45				
15	14	75	39				
16	15	77	69				
17	以下省略...						

圖 10：獨立樣本 t 檢定函數方法

	A	B	C	D	E	F
1	序號	第一次大測成績	第二次大測成績(加強複習后)		成對雙樣本 t 檢定	所用公式
2	1	40	46			0.000501516 =T.TEST(B2:B33,C2:C33,2,1)
3	2	45	49			
4	3	45	53			
5	4	50	50			
6	5	51	61			
7	6	55	63			
8	7	60	69			
9	8	65	62			
10	9	68	78			
11	10	68	66			
12	11	72	72			
13	12	75	78			
14	13	75	73			
15	14	75	79			
16	15	77	86			
省略					

圖 11：成對樣本 t 檢定函數方法

上述例子是利用公式從行t檢定其中首兩參數是兩組樣本資料，第三個參數是雙尾（一般假定是否相等都選擇雙尾），第四參數“1”是代表分別是成對樣本，“2”是代表分別是獨立樣本，從t檢定得出的P值，若 $P < 0.05$ 在存在顯著性差異，若 $P < 0.01$ 則存在極顯著性差異，可看出圖10計算出的P值 $0.18 > 0.05$ ，故兩班同學成績不存在明顯的顯著性差異，而圖11計算出的P值 $0.00050156 < 0.01$ 則存在兩次測驗成績極顯著性差異。

實例分析

本實例目的是比較新冠肺炎疫情停課後復課高二級學生（19/20年度）下稱樣本1及正常上課的高二級學生（20/21年度）下稱樣本2，在第三段的電腦科測驗成績及考試成績進行比較，分析停課對學生的成績有否顯著性的差異。本實例樣本1有158人，樣本2有172人，而測驗在5月尾，即復課的兩星期後進行，而考試則在7月進行的。本例利用EXCEL自帶分析工具箱中的t檢定功能進行分析，步驟如下圖：

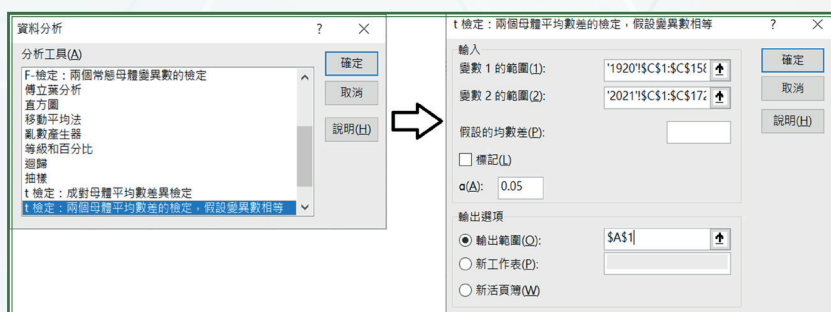


圖 12：t 檢定功能步驟

	樣本 1 考試成績	樣本 2 考試成績
平均數	84.66455696	86.9127907
變異數	80.95045553	84.91047872
觀察值個數	158	172
Pooled 變異數	83.01497982	
假設的均數差	0	
自由度	328	
t 統計	-2.23923165	
P(T<=t) 單尾	0.012905405	
臨界值：單尾	1.649512493	
P(T<=t) 雙尾	0.02581081	
臨界值：雙尾	1.967222827	

圖 13：測驗成績 t 檢定

	樣本 1 測驗成績	樣本 2 測驗成績
平均數	90.42405063	85.80581395
變異數	66.1948319	73.32324085
觀察值個數	158	172
Pooled 變異數	69.91116705	
假設的均數差	0	
自由度	328	
t 統計	5.012319048	
P(T<=t) 單尾	4.40749E-07	
臨界值：單尾	1.649512493	
P(T<=t) 雙尾	8.81497E-07	
臨界值：雙尾	1.967222827	

圖 14：考試成績 t 檢定

結果分析

從測驗成績t檢定（圖13）可以看出，樣本1及樣本2的P值（雙尾）為8.81497E-07，小於0.05存在顯著性差異，而考試成績t檢定（圖14）可以看出，樣本1及樣本2的P值（雙尾）為0.025，小於0.05同樣存在顯著性差異，但因為測驗成績的T值（圖13）是5.0123，而考試成績的T值（圖14）是-2.23，因為考試成績的T值比測驗成績的T值更接近0，所以考試成績的差異比測驗成績小。分析原因是測驗成績是復課後不久進行的，採取了宜鬆不宜緊的策略，故此差異較大，而經過復課後兩個月的授課，學生基本上已恢復狀態，故此，在考試成績上已和正常上課時的差異相對比較少。

小結

筆者介紹了基本的統計分析方法及其概念，並說明了如何利用EXCEL進行了敘述計、建立次數表分配表、正態分佈圖，最後，利用實例進行t-檢定差異性分析，希望本文能幫助同工對學生成績進行基本統計分析。

參考文獻

- 王智立、黃國忠（2018）。**量化研究與統計分析:SPSS操作與範例說明**。台灣地區：華泰文化。
- 邱皓政（2019）。**量化研究與統計分析：SPSS與R資料分析範例解析**。台灣地區：五南。
- 張屹、周平紅（2015）。**教育研究中定量資料的統計與分析——基於SPSS的應用案例解析**。北京：北京大學出版社。
- 教育及青年發展局——資訊科技教育網。取自<https://portal.dsej.gov.mo/webdsejspace/site/itedu/index.jsp>
- 維基百科——敘述統計。<https://zh.wikipedia.org/wiki/%E6%8F%8F%E8%BF%B0%E7%BB%9F%E8%AE%A1%E5%AD%A6>

陳國勳

聖保祿學校電腦科教師