

量測精確度的統計推論

賀全慶

中山科學研究院系統維護中心 技正

摘 要

本文針對量測精確度的評估方法做簡單介紹，內容包括評估資料蒐集的方式、分析的模型與方法、結果的確認與信賴區間的推導。目的在於協助不論是從事檢測或校正的人員，當開發一個新的量測方法或是擬對現行量測方法進行分析評估時，一個簡單有用的分析技術與觀念。

關鍵詞：量測精確度、準確度、變異數分析、信賴區間

1. 前言

不論是檢測還是校正，基本上都是由操作人員利用一定的程序（方法）與設備，在某個時間點、某種環境下對待測（校）件施以量測的動作。所以，一般談到「量測系統」都會論及這個系統的組成要素，包括：(1) 人員、(2) 設備、(3) 設備的校正結果、(4) 環境、(5) 時間與 (6) 校正的程序（方法）。這些因素都會影響校正結果的品質。

量測的目的在於求得待測件的真值。由於量測系統中各項組成要素都存在某些不可避免的隨機變異，這些隨機變異造成量測結果的不確定性，因此在大多數情況下，待測件的真值其實是個未知數。雖然真值不可知，但吾人希望量測的結果非常接近真值，這是吾人實施量測的目的，因此，量測的品質就成為關注的焦點。

在統計學與量測學的領域中，用來評估量測品質的指標稱為「精準度 (Accuracy)」。ISO 5725-1【1】將此精準度指標又區分為「精確度 (Precision)」與「準確度 (Trueness)」兩部分。所謂「精確度」是指各次量測結果彼此接近的程度。以統計學的術語來說，就是衡量多次重複量測結果的變異數。眾所皆知，吾人希望量測結果的變異數越小越好，因為變異數越小，代表量測的結果資料越集中，代表量測的精確度越高。但是，光有精確度不足以完整表達量測結果的品質，吾人還需要知道量測結果與（約定）真值之間的差距，此即「準確度」。一般以大量重複量測結果的平均值與（約定）真值之間的差距做為評估準確度的指標，吾人期望此差距為 0。在統計學的術語上，準確度又稱為「偏差 (Bias)」。

量測的結果需要考慮精確度，主要是因為使用相同的待測件、在相同的條件下進

行重複量測，每次得到的結果不一定會相同。這是因為在任一個量測系統、任何一次量測過程中都有無可避免的隨機誤差存在。這些會影響量測結果的因子在執行量測的過程中不可能完全受到控制。在實務上，因為這些不可控因子會影響量測的結果，因此在解讀量測結果時，必須將此影響列入考慮。當量測結果與指定值之間的差異小於隨機誤差時，吾人認為量測結果與指定值之間沒有差異。

假設量測時使用相同的待測件，量測的品質特性相同且只有一項，量測的方法(程序)也經過標準化、文件化，則會影響量測品質的因素只剩下前述的 (1)~(5) 項。量測的「精確度」是以重複量測結果的變異性來衡量，而此變異性又分為兩類，一為「重複性變異 (Repeatability Variation)」，指得是當因素 (1)~(5) 均維持不變的條件下執行重複量測所得到的量測變異數，一般以 σ_r^2 表示；另一為「重現性變異 (reproducibility Variation)」，指得是當因素 (1)~(5) 至少有一項變動的條件下，執行重複量測所得到的量測變異數，一般以 σ_R^2 表示。一般而言， $\sigma_R^2 \geq \sigma_r^2$ 。

本文僅就「精確度」的評估部分做介紹，「準確度」的部份將另文探討。

2. 量測精確度的評估

量測精確度的評估須透過多家實驗室共同參與量測實驗而得。假設有 p 個實驗室，使用相同的待測件與標準化的量測方法，對相同的品質特性進行進行 n 次重複量測。假設沒有離群值，所得量測結果的資料格式如表一。其中，

y_{ij} ：第 i 個實驗室第 j 次量測的觀察值；

$y_i = \sum_{j=1}^n y_{ij}$ ：第 i 個實驗室所有量測值的總和；

$\bar{y}_i = \frac{y_i}{n}$ ：第 i 個實驗室所有量測值的平均；

表一、量測結果的資料格式

	實驗室			
	1	2	...	p
重複觀察數	y_{11}	y_{21}		y_{p1}
	y_{12}	y_{22}		y_{p2}
	.	.		.
	.	.	...	.
	y_{1n}	y_{2n}		y_{pn}
總和	$y_{1.}$	$y_{2.}$	...	$y_{p.}$
平均數	$\bar{y}_{1.}$	$\bar{y}_{2.}$	...	$\bar{y}_{n.}$
變異數	S_1^2	S_2^2	...	S_p^2

$$S_i^2 = \frac{\sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2}{n-1} : \text{第 } i \text{ 個實驗室所有量測值的變異數；}$$

此外，定義

$$y_{..} = \sum_{i=1}^p \sum_{j=1}^n y_{ij} : \text{所有觀察值的總和；}$$

$$\bar{y} = \frac{y_{..}}{np} : \text{所有觀察值的總平均。}$$

對任何一個實驗室而言，該 n 個觀察值必須是在重複性條件（亦即前述的條件(1)~(5) 均維持不變）下、於短時間內量測而得，而且每一次的量測活動必須互為獨立。此外，參與實驗的 p 個實驗室係由實驗室母群體當中隨機挑選而得。

基本上，這樣的實驗安排符合典型的一因子變異數分析--隨機效應模型。其變異數分析表如表二。

表二、一因子變異數分析表：隨機效應模型

變異來源	平方和	自由度	均方	均方期望值	統計量
實驗室	SSL	$p-1$	$MSL = \frac{SSL}{p-1}$	$E(MSL) = \sigma_r^2 + n\sigma_L^2$	$F = \frac{MSL}{MSR}$
隨機誤差	SSE	$p(n-1)$	$MSE = \frac{SSE}{p(n-1)}$	$E(MSE) = \sigma_r^2$	
總和	SST	$pn-1$			

其中，

$$SST = \sum_{i=1}^p \sum_{j=1}^n (y_{ij} - \bar{y})^2$$

$$SSL = \sum_{i=1}^p \sum_{j=1}^n (\bar{y}_{i.} - \bar{y})^2 = n \sum_{i=1}^p (\bar{y}_{i.} - \bar{y})^2$$

$$SSE = \sum_{i=1}^p \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2 = (n-1) \sum_{i=1}^p S_i^2$$

此時，重複性變異的估計值為

$$S_r^2 = MSE = \frac{\sum_{i=1}^p S_i^2}{p}$$

實驗室間變異的估計值為

$$S_L^2 = \frac{MSL - MSE}{n}$$

重現性變異的估計值則為

$$S_R^2 = S_L^2 + S_r^2$$

3. 變異數分析結果的確認

一因子變異數分析有三大基本假設：(1) 常態母體，(2) 所有母體變異數都相等（變異數齊一性），(3) 觀察值互相獨立。其中有關獨立的部分，在進行實驗時已有要求，因此在完成變異數分析後，尚須對前兩項假設條件進行確認。

有許多技術可以用來檢定觀察值是否符合常態分配，最常用的莫過於常態機率圖【2】。首先，計算所有觀察值的殘差， e_{ij} ，如下。

$$e_{ij} = y_{ij} - \bar{y}_{i.}, \quad \begin{cases} i = 1, 2, \dots, p \\ j = 1, 2, \dots, n \end{cases}$$

將殘差由小到大排序，然後計算殘差的期望值， $E[e_{(k)}]$

$$E[e_{(k)}] = \sqrt{MSE} \left[Z \left(\frac{k - 0.375}{n + 0.25} \right) \right]$$

其中， $E[e_{(k)}]$ 表第 k 個排序殘差的期望值， $Z(x)$ 表標準常態分配的 x 百分位數。繪製殘差與其相對應期望值的散佈圖，如果圖形顯示點的分布呈現一條直線，則可以合理的判定觀察值符合常態分配的假設。如果覺得目視結果判斷不易，可以計算殘差與其期望值的線性相關係數並與臨界值做比較【3】，若計算所得結果大於臨界值，則判定觀察值符合常態分配的假設。

其他如卡方檢定中的適合度檢定或無母數統計推論中的 Kolmogorov- Smirnov 適合度檢定都可以用來檢定常態母體的假設。

至於變異數齊一性的假設，最常用的是 Bartlett 檢定【4】，程序如下。

$$\text{假設檢定：} \begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2 \\ H_1 : \text{Not } H_0 \end{cases}$$

計算卡方檢定統計量： $\chi_o^2 = 2.3026 \times \frac{q}{c}$

其中，

$$q = (n-1) \left[p \log_{10}(MSE) - \sum_{i=1}^p \log_{10} S_i^2 \right]$$

$$c = 1 + \frac{p+1}{3p(n-1)}$$

若 $\chi_o^2 > \chi_{p-1, \alpha}^2$ 則拒絕接受 H_0 。

4. 量測精確度估計值的信賴區間

表一所得到的觀測結果只是眾多可能結果之一，因此所得到的重複性變異或重現性變異都只是一個點估計值。吾人可以利用其統計特性求得其 $(1-\alpha) \times 100\%$ 的信賴區間。

對重複性變異而言，由於

$$\frac{p(n-1)MSE}{\sigma_r^2} \sim \chi_{p(n-1)}^2$$

可得

$$\frac{p(n-1)MSE}{\chi_{p(n-1), \alpha/2}^2} \leq \sigma_r^2 \leq \frac{p(n-1)MSE}{\chi_{p(n-1), 1-\alpha/2}^2}$$

對重現性變異而言，

$$S_R^2 = S_L^2 + S_r^2 = \frac{(n-1)MSE + MSL}{n}$$

$$\frac{p(n-1)MSE}{\sigma_r^2} \sim \chi_{p(n-1)}^2 \Rightarrow MSE = \frac{\sigma_r^2 \chi_{p(n-1)}^2}{p(n-1)}$$

$$\frac{(p-1)MSL}{\sigma_r^2 + n\sigma_L^2} \sim \chi_{p-1}^2 \Rightarrow MSL = \frac{(\sigma_r^2 + n\sigma_L^2) \chi_{p-1}^2}{p-1}$$

$$S_R^2 = \frac{\frac{\sigma_r^2 \chi_{p(n-1)}^2}{p} + \frac{(\sigma_r^2 + n\sigma_L^2) \chi_{p-1}^2}{p-1}}{n} = \left(\frac{\sigma_r^2}{np} \right) \chi_{p(n-1)}^2 + \left(\frac{\sigma_r^2 + n\sigma_L^2}{n(p-1)} \right) \chi_{p-1}^2$$

$$= u_1 \chi_{p(n-1)}^2 + u_2 \chi_{p-1}^2$$

$$u_1 = \left(\frac{\sigma_r^2}{np} \right), \quad u_2 = \left(\frac{\sigma_r^2 + n\sigma_L^2}{n(p-1)} \right)$$

此為兩個獨立卡方變數的線性組合，不具封閉形式的機率密度函數，因此無法求得 S_R^2 確切的信賴區間【4】。但是，吾人可以求得 σ_L^2 / σ_R^2 的信賴區間，而這是一項有意義

的指標，它代表實驗室間的變異佔重現性變異的百分比。如果此百分比值大，代表重現性的變異主要來自實驗室之間的差異，實驗室內量測值的變異相對較小，代表實驗室內重複量測結果的一致性高。如果此百分比值小，代表實驗室之間的一致性高，但實驗室內重複量測的重複性較差，這是否代表標準量測方法需要改善（亦即，開發更好的量測方法）值得探討。

$$\frac{\frac{MSL}{n\sigma_L^2 + \sigma_r^2}}{\frac{MSE}{\sigma_r^2}} \sim F_{(p-1), p(n-1)}$$

$$P\left(F_{(p-1), p(n-1), 1-\alpha/2} \leq \frac{MSL}{MSE} \frac{\sigma_r^2}{n\sigma_L^2 + \sigma_r^2} \leq F_{(p-1), p(n-1), \alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(L \leq \frac{\sigma_L^2}{\sigma_r^2} \leq U\right) = 1 - \alpha$$

$$L = \frac{1}{n} \left(\frac{MSL}{MSE} \frac{1}{F_{(p-1), p(n-1), \alpha/2}} - 1 \right), \quad U = \frac{1}{n} \left(\frac{MSL}{MSE} \frac{1}{F_{(p-1), p(n-1), 1-\alpha/2}} - 1 \right)$$

$$P(L\sigma_r^2 \leq \sigma_L^2 \leq U\sigma_r^2) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{L\sigma_r^2}{L\sigma_r^2 + \sigma_r^2} \leq \frac{\sigma_L^2}{\sigma_L^2 + \sigma_r^2} \leq \frac{U\sigma_r^2}{U\sigma_r^2 + \sigma_r^2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{L}{1+L} \leq \frac{\sigma_L^2}{\sigma_R^2} \leq \frac{U}{1+U}\right) = 1 - \alpha$$

因此， σ_L^2/σ_R^2 的 $(1 - \alpha) \times 100\%$ 信賴區間為

$$\left(\frac{L}{1+L}, \frac{U}{1+U} \right)$$

5. 範例

假設有八個實驗室參與煤礦中含硫量測定的實驗，表三為八個實驗室得到的測定結果，表四則為其變異數分析表。從變異數分析表得知，各實驗室量測的平均值是不一樣

的。

圖一顯示各實驗室量測值的 95%信賴區間，從圖一中可以看出實驗室 D 與 F 兩者有著顯著的差異。圖二為常態機率圖，由於圖二中各點的散佈呈現一條直線，吾人可以合理的判定量測資料符合常態分配的假設。圖三則為變異數齊一性檢定的結果，從 Bartlett 檢定的 p 值可知，各實驗室量測值母體的變異數可以視為相等。這些結論支持使用變異數分析法的合宜性。

根據表四的結果，可得：

$$S_r^2 = MSE = 0.000188$$

$$S_L^2 = \frac{MSL - MSE}{n} = \frac{0.002434 - 0.000188}{5} = 0.00045$$

$$S_R^2 = 0.00045 + 0.000188 = 0.00064$$

σ_r^2 的 95%信賴區間為：(0.000122, 0.000329)

σ_L^2 / σ_R^2 的點估計值為 0.70，而其 95%信賴區間為：(0.43, 0.92)

表三、煤礦中含硫量的測定值

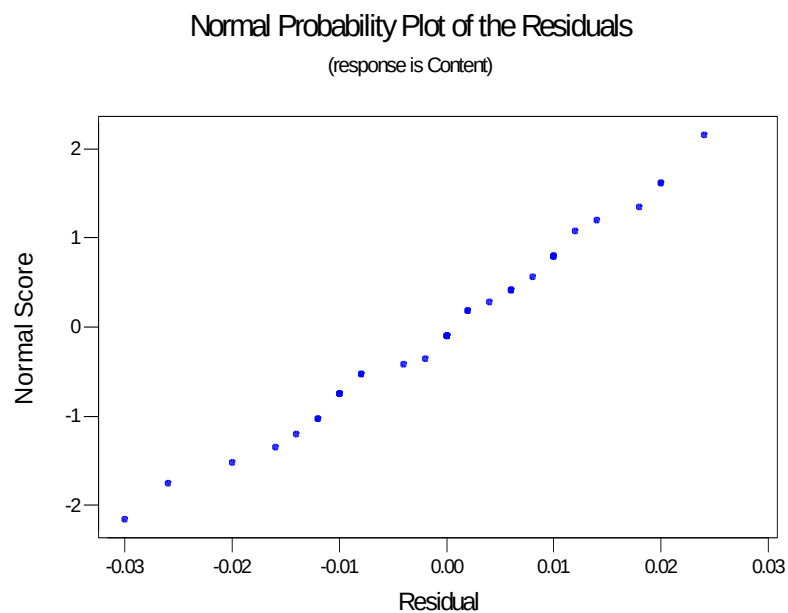
實驗室	A	B	C	D	E	F	G	H
測定值	0.71	0.69	0.66	0.67	0.7	0.73	0.71	0.7
	0.71	0.67	0.65	0.66	0.69	0.74	0.71	0.65
	0.7	0.68	0.69	0.65	0.66	0.73	0.69	0.68
	0.71	0.68	0.67	0.68	0.71	0.72	0.7	0.66
	0.69	0.67	0.68	0.65	0.69	0.73	0.69	0.69

表四、變異數分析表

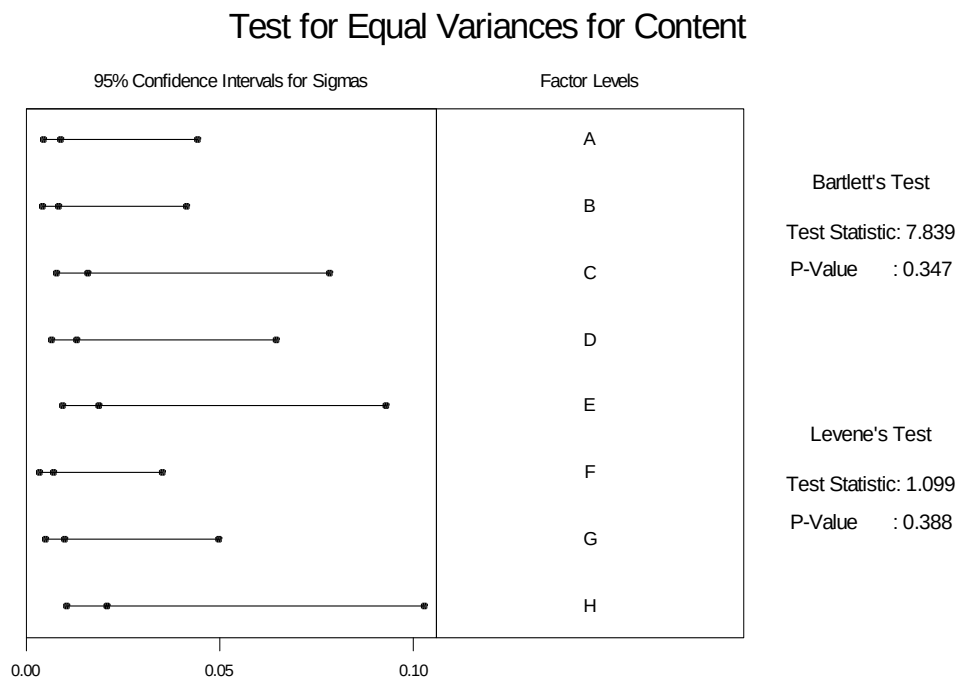
變源	SS	自由度	MS	F	P-值	臨界值
實驗室	0.017038	7	0.002434	12.98095	8.85E-08	2.312738
誤差	0.006	32	0.000188			
總和	0.023038	39				

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
A	5	0.70400	0.00894	(---*---)
B	5	0.67800	0.00837	(---*---)
C	5	0.67000	0.01581	(---*---)
D	5	0.66200	0.01304	(---*---)
E	5	0.69000	0.01871	(---*---)
F	5	0.73000	0.00707	(---*---)
G	5	0.70000	0.01000	(---*---)
H	5	0.67600	0.02074	(---*---)

圖一、各實驗室量測值的信賴區間



圖二、常態機率圖



圖三、變異數齊一性的檢定

6. 結語

本文針對量測精確度的評估方法，從資料的蒐集、分析、評估到區間估計做一簡單的說明。實務上，在執行量測精確度評估的時候，有許多事項需要注意與要求，包括待測件、測試方法、測試時間、參與實驗室的選擇、重複量測數、觀察值的處理（如：離群值的檢測）...等，詳細的要求可參考 ISO 5725-2 【5】。該規範採用多項待測件、多個實驗室參與與各實驗室所獲重複觀察值數目可能不等的方式評估量測精確度。本文則採最簡單的形式，僅取一項待測件且每個實驗室都獲得相同觀測數目的方式做說明，目的在讓讀者了解精確度分析的一般做法。

同時，本文也採用與該規範不同的評估方式，選用一般統計學上常用的實驗設計與變異數分析技巧，因為可以執行這些運算與分析的軟體垂手可得（例如：EXCEL），增加使用者的實用性與方便性；此外，本文還增加有關精確度指標信賴區間的推導，讓整個分析工作更趨完整。

7. 參考文獻

- 【1】ISO 5725-1, *Accuracy (Trueness and Precision) of Measurement Methods and Results -- Part I: General Principles and Definitions*, 1994.
- 【2】Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W., *Applied Linear Regression Models*, 3rd ed., Irwin, 1996.

- 【3】 Looney, S. W. and Gullledge, T. R., "Use of the Correlation Coefficient with Normal Probability Plots", *The American Statistician*, pp. 75-79, 1985.
- 【4】 Montgomery, D. C., *Design and Analysis of Experiments*, 2nd ed., John Wiley & Sons, 1984.
- 【5】 ISO 5725-2, *Accuracy (Trueness and Precision) of Measurement Methods and Results -- Part II: Basic Methods for the Determination of Repeatability and Reproducibility of a Standard Measurement Method*, 1994.

Evaluation of Precision of a Measurement Method

Chuanching Ho

Integrated Logistic Support Center
Chung Shan Institute of Science and Technology

Abstract

An approach is proposed in this paper to evaluate the precision of a measurement method. It includes the requirements for data collection, model for analysis, validation of the assumptions of the analytical method and derivation of the confidence intervals for repeatability variation and the ratio of between laboratories variation to reproducibility variation. The purpose is to provide the practitioners a simple tool to evaluate the precision of a measurement method, be it newly developed or applying an existing procedure.

Keywords: Precision, Trueness, ANOVA, Confidence interval