

試験所認定制度における技能試験

- 付与された値，Zスコア，ロバストな方法，そして四分位数法 -

顧問 藤井 賢三

[はじめに]

環境に配慮した企業活動のための ISO 14000 シリーズ，そして，自社の製品の品質保証に関する ISO 9000 シリーズの他に，ISO/IEC^{*1} ガイド 25 に基づく「試験所認定制度」が，各分野で検討されてきております。

この認定制度は，決して強制的なものではなく，あくまでも任意なものですが，WTO^{*2} の TBT^{*3} 協定により，今や国際的に技術レベルを保証する制度として定着しつつあります。

「試験所認定」を取得するためには，「権威ある認定機関」による品質システム審査と技術審査に合格しなければなりません。それには，

- ・ 品質システムについては，
一定の品質マニュアルがあること
測定分析体制が文書化されていること
- ・ 技術レベルについては
トレーサビリティが確保されていること
技能レベルが一定以上に維持されていること

などが必要で，書類審査と立入り審査の他に，多くの場合，当該認定機関が指定する「技能試験」への参加を義務づけられます。

現在我が国においては，J A B ((財)日本適合性認定協会)，J N L A (通産省工業技術院 標準部)，J C L A ((社)日本化学工業協会 日本化学試験所認定機構)，の三機関が，「認定機関」としての資格を取得していて，既に，金属材料，繊維製品，給水関連器具，電気製品，ゴム及びプラスチック，塗料及び顔料，化学工業製品，石油及び関連製品，等の分野で活動を開始しています。

環境分析分野に関しては，上記の三機関とも，試験所認定業務を志向しており，その中において日本環境測定分析協会は，「技能試験」のための専門機関として，いずれの「認定機関」ともタイアップして，「環境関連の技能試験用試料の提供と，測定結果の集計及び評価」を担当すべく検討を重ねてきております。

その準備段階として，平成 10 年度の水質技術委員会の共同実験を，「試験所認定制度における技能試験」に準じた仕様で行うこととし，筆者も同委員会の一員として参画しました。以下，その業務の中で得られた知見をご紹介します。

*1 IEC：国際電気標準会議

*2 WTO：世界貿易機関

*3 TBT：貿易の技術的障害

1. ISO/IEC ガイド 43-1 (JIS Q 0043-1)

同一の試料を複数の機関で測定して，その測定値を比較する，いわゆる「共同実験(クロスチェック)」は，各分野で，種々の目的に応じて実施されています。例えば，同業社間で行う技術研鑽のための共同実験，新しい測定方法の検討，また，商取引上の必要で，売手と買手の評価技術を比べておくような場合(しばしば二者だけで行われる)もあり，「試験所認定制度における技能試験」もその一つといえます。そして，この試験所間比較を「技能試験(試験所の実績を評定し，その継続的な実績を監視する目的で行う試験)」のために行う場合の進め方及び配慮事項等について，詳しく規定しているのが，ISO/IEC ガイド 43-1 (JIS Q 0043-1：ただし，このJISは英文の直訳に近く，用語の使い方，内容等に，今一つ難解な点が多い)です。

このJIS Q 0043-1の骨子を，私なりに解釈して整理したのがP3の図です。

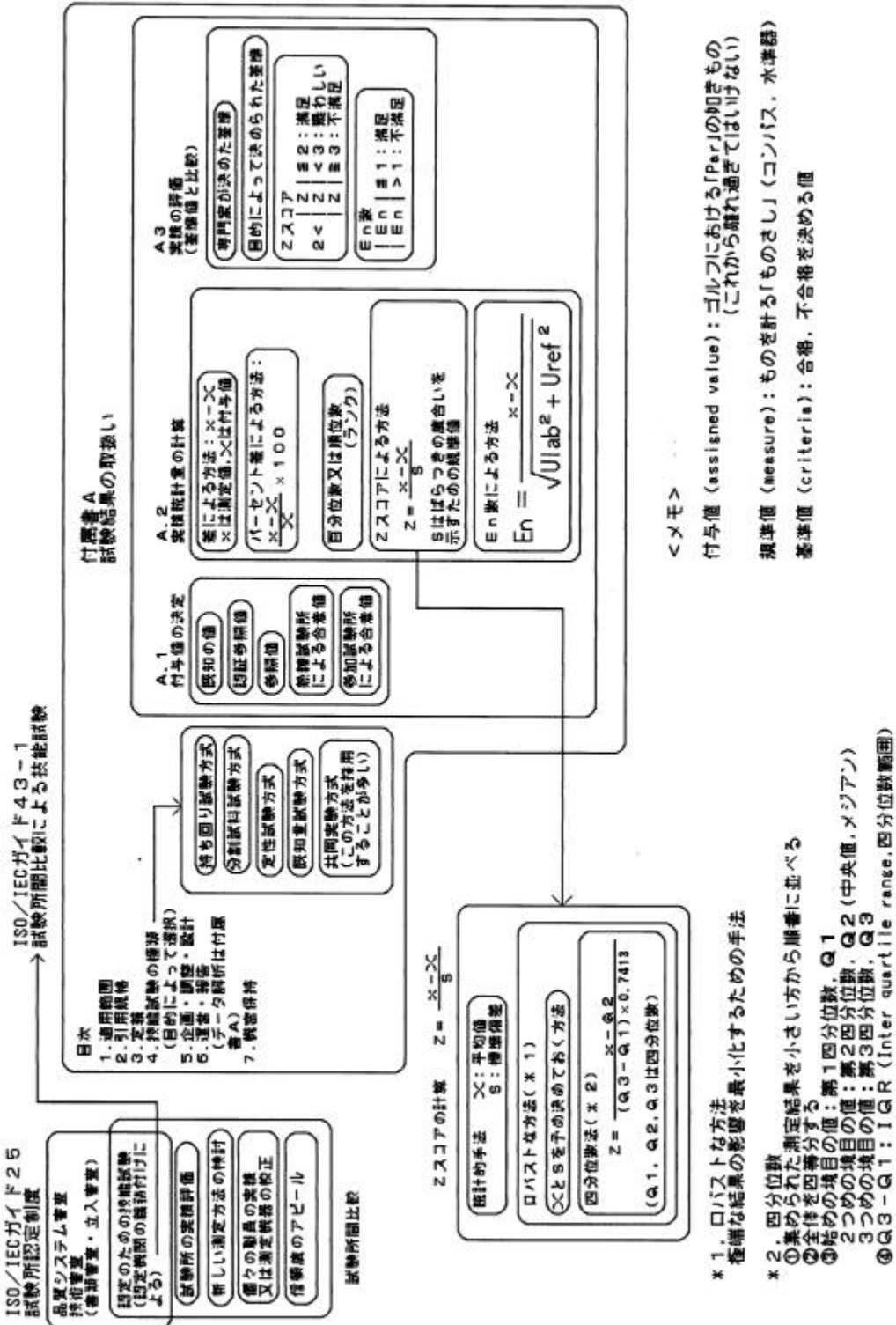
JIS Q 0043-1の本文中「4. 技能試験の種類」の項には，種々の技能試験の方式がまとめてあります。このうち4.3の共同実験スキームは，食品，体液，水質，土壌，及びその他の環境物質等の分野で採用されている方式で，これについて，

試料は，均質で，経時(日)変化のないこと。

同一の試料について反復測定(いわゆる $n=2$)の代わりに，濃度の少し違う試料を2個配布して測定させることにより，試験所内のばらつきの検定精度を高める($n=2$ の場合は，各自の選択意志が働いて，近似した値しか返ってこない)ことができる。

測定結果は，ロバストな方法(極端な値の影響を最小化する手法：後述)で処理する。

などの提示があり，このうち は，従来も気を配っていた事項ですが， については全く新しい知見で，私にとっても大変勉強になりました。



また、「5」には設計段階での配慮事項、「6」には報告値の取扱い・整理・評価等について、さらに、「7」には執行機関の守るべき機密保持、及び、参加機関の守るべき倫理的配慮、等についても触れられていて参考になりました。

2. 附属書A（技能試験データの取扱い方）

先に述べた如く、技能試験の種類はいろいろあり、「その目的によって選ぶこと」とされていますが、集められたデータのまとめ方について、その手順が、附属書Aに示されており、それによると、すべての技能試験に共通な手順として、

まず、「付与された値」を決める
を用いて実績統計量を計算する

の結果から個々の測定値の評価を行う

というステップを踏むこととなります。ここで出てくる「付与された値」及び「実績統計量」という用語はあまり聞いたことのない用語で当初は大変戸惑いました。

2.1 付与された値の決定

共同実験でのデータのヒストグラムを作ると、大凡「正規分布」に近い形（中央が多くて、両端が少ない：いわゆる山の形）をしています。この場合、測定値のままでは、それが山のどの辺にあるか判りません。そこで、山の中心近くに「標準になる値（X）」を決めて、Xとの差を求めると、山の中央付近にあるのか、あるいは大分離れたあたりにあるかが判りやすくなります。このXの値を「付与された値」と呼んでいて、

- 既知の値：たとえば、標準物質の希釈液を試料とした場合の計算値
- 認証参照値：（定量的試験について）限定的(definitive)な方法で決められた値
- 参照値：国家標準、又は、国家標準にトレーサブルな標準物質又は標準器を用いて測定された値
- 熟練試験所による合意値：当該分析方法について造詣の深い試験所が測定した値
- 参加試験所による合意値：その試験に参加した試験所からの測定結果の、平均値（場合によっては重みをつけたり、変換したりしてもよい）

中央値（メジアン）、モード等ロバストな方法で決められた値（後述）

などの中から目的に応じて適用するのがよいとされています。

この「付与された値」というのは、いわば、ゴルフでいう「par」のようなものと考えればよいと思います。ただし、ゴルフと違って技能試験では、+側、-側、何れもこの値から離れすぎてはいけません。

2.2 実績統計量の計算

次に、個々の参加機関の測定値(x)が、「付与された値(X)」からどの位離れているかを計算します。この場合、単に「生の測定値の差」のままでもいいのですが、何か「規準」になる数値を設定して、それに比較して「離れ具合」がどの位かを示す方が判り易いので、それに適した変換（計算）が行われます。この「変換した後の値」を「実績統計量」と呼んでいます。

43-1には次のような計算方式が例として示されています。

a) 差(x - X)：xは個々の測定値
Xは「付与された値」

b) パーセント差：

$$\frac{(x - X)}{X} \times 100$$

c) 百分位点又は順位(ランク)数：

全体の測定値を、例えば、小さい方から大きい方に並べた時に、端からどの位の所に位置しているか、又は、それを何等分かした時にどのランクに属しているか、といった表示の仕方。

d) Zスコア：

$$Z = \frac{(x - X)}{s}$$

ここで、sは、当該技能試験の目的に適した「ばらつきの推定値又は規準」の一つで一般的には、標準偏差又は変動係数などが使われる。

e) En数：

$$En = \frac{(x - X)}{\sqrt{U_{1ab}^2 + U_{ref}^2}}$$

U_{1ab} は参加者の測定値の不確かさ

U_{ref} は参照試験所の付与された値の不確かさ

従って分母は(x - X)の不確かさを示している。

2.3 実績の評価

2.2で「個々の参加機関の測定値」が、「付与された値」からどの位離れているかが数値で示されるようになりました。従って、その数値について、合格・不合格の基準を予め決めておけば（結果が出てから決めるのはよくない）、それに照合して評価を下すことができます。その基準としては、

- a) 専門家の合意による値を基準にする。
- b) 技能試験の実施仕様及び参加者の承認業務水準を考慮して決めた値を基準にする。
- c) Zスコアの場合は次の数値を基準にする。
 - |Z| 2 : 満足(satisfactory)
 - 2 < |Z| < 3 : 疑わしい(quetionable)
 - |Z| 3 : 不満足(unsatisfactory)
- d) Enの場合は次の数値を基準にする。
 - |En| 1 : 満足
 - |En| > 1 : 不満足

などの、基準値の決め方、及び、評価が示されています。

3. 四分位数法によるZスコアの計算

以上の如く、ISO/IEC ガイド 43-1(JIS Q 0043-1)には、種々のケースが記載されており、その技能試験の目的に応じて、最も適した方式を選ぶこととされています。そこで、日環協の平成 10 年度の共同実験は、A P L A C (Asia-Pacific Laboratory Accreditation Cooperation)及びJ N L Aで行った共同実験を参考にして進めることとし、その測定結果の集約も、それらと同じ方式で行いました。

その方式は、「四分位数」を用いて「Zスコア」を求める方式で、環境関連の試料の技能試験には最も適しているように思われます、

しかし、始めのうちは、そこで用いられている用語、即ち、「付与された値」、「Zスコア」、「ロバストな方法」、「四分位数」、等の意味が判らず苦労しましたが、とりあえず試料を配布し、集まった報告値を整理して行くうちに、少しずつそれぞれの結びつきと意味が判ってきました（このうち、「付与された値」については2.1で、また「Zスコア」については2.2でご説明しました）。残りの二つは、いずれも「Zスコア」の計算に関係している用語で、43-1には「ロバストな方法でZスコアを計算するのが望ましい」とあり、それを受けて、A P L A C及びJ N L Aでは「四分位数を使ってZスコアを計算」しておりますので、この四分位数を採用することが取りも直さず「ロバストな方法」とい

うこととなります。

そこで、「ロバストな統計量」という用語について考えてみましょう。

3.1 ロバスト(robust)な統計量

「ロバスト」という言葉は、なかなか判り難い用語です。JIS Q 0043-1では「頑健な」と訳されているので、ますます判らなくなってしまいます。

私達が共同実験を行う場合の参加事業所の数は、多い時で300社、少ないグループの場合は10数社位ですが、提供された報告値の中には、往々にして「きわめて飛び離れた値」が混じっています。このような異常値は、位取りのミスとか、希釈率を計算式に入れ忘れたような場合で、一見してその原因が判りますので、通常共同実験では担当者に連絡して野帳を見て貰い、その原因が判ればそれを修正してから統計量の計算を行います。

しかし、「試験所認定の技能試験」の場合は、そのようなミスも含めての試験ですから、簡単に修正を許すことはできません。そこで、このような異常値が混在していても、それによって影響されることがないように（仮りに影響があっても、できるだけ少なくてすむような）、計算手法が要求されるわけです。

要約すると、

「極端な値による影響を最小化するための手法」

がロバストな方法ということになります。このように意味が判ってくると、私達の身近にも「ロバストな方法」を使っている場合がいくつかあることに気が付きます。

騒音測定における、中央値 = L_{50} 、90パーセントレンジ = $L_{95} - L_5$ など

フィギュアスケートの技術点や芸術点を決める方法（最高点と最低点を除いて計算する）

嗅覚による悪臭測定の時に、6人のパネラーの値のうち最高と最低を除いた4人の値を平均する方法

などは、まさに「ロバストな方法」によるものといえます。

3.2 Zスコアの計算におけるロバストな方法の適用

2.2のd)で述べた如く、Zスコアは次式で求められます。

$$Z = \frac{(x - X)}{s} \dots\dots\dots$$

ここで、 x は測定値ですから「ロバストな方法」の入る余地はありません。 X と s は、通常の統計的処理の場合は、測定結果の平均値及び標準偏差を使いますが、これらの値は異常値の影響を受けますので、「ロバストな方法」でその影響がないようにしてやる必要があります。

このような場合の「ロバストな方法」としては、

X と s を参加者の測定結果から算出する場合

X と s を(全)参加者の測定結果から算出しない場合(例えば、付与された値とばらつきを予め決めておくような場合)

があり、APLAC及びJNLAの共同実験では、の計算手法として「四分位数法」を適用しています。

3.3 四分位数法によるZスコアの計算手順

次の(1)、(2)で四分位数を求めたあと、(3)で個々の参加機関のZスコアを計算します。

- (1) 各参加機関からの測定値を、最小値から最大値へと昇順に並べる。
- (2) 参加機関の総数(N)を4で割って、次の順番に該当する参加機関の測定値を求める。

小さい方から $N \times 1/4$ 番目の参加機関の測定値

(第1四分位数)： Q_1

小さい方から $N \times 2/4$ 番目の参加機関の測定値

(中央値又はメジアンと呼ぶ)： Q_2

小さい方から $N \times 3/4$ 番目の参加機関の測定値

(第3四分位数)： Q_3

この場合、 Q_1 、 Q_2 、 Q_3 は異常値に全く無関係に決められるから「ロバストな値」である。

- (3) 3.2の式に、

$$X = Q_2 \quad s = (Q_3 - Q_1) \times 0.7413$$

を代入して、 i 番目の参加機関のZスコア(Z_i)を、

$$Z_i = \frac{(x_i - Q_2)}{(Q_3 - Q_1) \times 0.7413} \dots\dots\dots$$

によって求める。(x_i は i 番目の参加機関の測定値)。

ここで、 $(Q_3 - Q_1)$ をIQR(四分位数範囲：inter quartile range)に置き換えると、

$$Z_i = \frac{x_i - (\text{中央値})}{IQR \times 0.7413} \dots\dots\dots$$

が得られる。

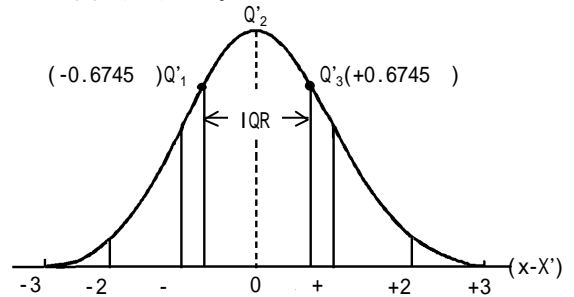
以上の手順を踏んで得られたZスコアは、 Q_1 、 Q_2 、 Q_3 が「ロバストな方法」で得られた値なので、Zスコアも同様に「ロバストな値」になります。従ってこの

方式によれば、異常値の影響を受けずに個々の測定値の評価を行うことができます。

3.4 Zスコアの計算式の $(Q_3 - Q_1) \times 0.7413$ の意味

四分位数を用いてZスコアを計算する際、分母の規準値(s)に $(Q_3 - Q_1) \times 0.7413$ を使います。参考文献4の付属書D、表D-1の(注)に説明が書いてありますが、少し専門的過ぎるようなので、私なりに解きほぐしてみました。

横軸を $(x' - X')$ (X' は平均値)として、標準偏差の正規分布曲線を描き、その中に四分位数等を記入すると図の如くなる。



正規分布における四分位数及びIQRの関係

$$\text{ここで、} IQR = Q_3 - Q_1 = 0.6745 - (-0.6745) = 1.349$$

技能試験の測定結果のIQRは $(Q_3 - Q_1)$

測定結果の標準偏差は異常値の影響を受ける可能性があるため、その代わりに、中央値が X で、IQRが $(Q_3 - Q_1)$ の正規分布の標準偏差()を準用する。

$$\text{この時、} 1.349 = Q_3 - Q_1$$

$$= (Q_3 - Q_1) / 1.349$$

$$= (Q_3 - Q_1) \times 0.7413 \text{ となる。}$$

ここで得られた を、ロバストな規準としてZスコアの計算式の分母(s)に使用しているのです。

(次号では、共同実験の実施例で説明します。)

[参考文献]

1. 試験所認定制度とその対応について 久代 勝 環境と測定技術 vol.26, No.4, 1999, P55
2. JIS Q 0043-1 (1998) 試験所間比較による技能試験 第1部 技能試験スキームの開発及び運営
3. 技能試験に参加して-APLAC-T001 松本保輔 環境と測定技術 vol.25, No.9, 1998, P55
4. 工業標準化に基づく試験事業者認定制度に係る技能試験 JNLA 報告書 JNPT 10-03 化学品・給水関連器具分野(水溶液の金属分析)
5. 現代統計実務講座 文部省認定社会通信教育用テキスト (財)実務教育研究所