

令和元年6月14日現在

機関番号：82626

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K18411

研究課題名(和文)測定値間の相関に着目した効率的な試験所間比較の設計

研究課題名(英文) Design of interlaboratory collaborative studies with using the correlation information between reported values

研究代表者

城野 克広 (Shirono, Katsuhiko)

国立研究開発法人産業技術総合研究所・計量標準総合センター・主任研究員

研究者番号：60509800

交付決定額(研究期間全体)：(直接経費) 1,200,000円

研究成果の概要(和文)：試験所(測定者)の測定能力の確認のために、同一のものを複数試験所で測定し、その結果を評価する試験所比較がよく行われる。本研究では、期間の短縮のために、全試験所をサブグループに分け、サブグループごとに異なるものを測定した場合の試験所間比較について、適切な実験計画と統計解析の方法を設計することを目的とする。本研究では、(1) サブグループを結びつけるリンク試験所の複数の測定値の間の相関の妥当性を確認する手法の開発し、(2) リンク試験所以外の試験所の測定能力についての評価を与える統計的検定をその検出力に着目して設計した。本研究の完成は品質の保証された計量基盤を提供するために貢献するものである。

研究成果の学術的意義や社会的意義

本研究の成果は、これまで統計学的に洗練されていなかったサブグループによって構成される試験所間比較の解析を行う手法である。この成果により、試験所間比較の設計において、サブグループを作ってアイテムを回付することが可能になり、短い期間で試験所間比較を終了させることができる。このことは、単に試験所間比較の実施コストを下げるだけでなく、試験所が測定能力を確認するサイクルを短くすることにつながるものである。この成果が試験所の第三者認証に活用されるよう、今後、関連するガイドラインなどに反映されることを目指すものである。

研究成果の概要(英文)：Interlaboratory studies, where we compare the measurement results for an identical item, have been implemented in various metrological discipline to check laboratories' measurement performances. However, in some interlaboratory studies, laboratories are divided into subgroups, and those in different subgroups measure different items for the purpose of efficiency. In this study, the appropriate design of experiment and the statistical analysis in such cases are discussed. We have developed (1) the method to confirm the validity of the reported correlation information from the laboratory linking multiple subgroups, and (2) statistical test to evaluate the non-linking laboratories with maximizing the power of test. This study will contribute to provide the quality infrastructure in terms of metrology.

研究分野：統計学・数理工学・品質管理・情報工学

キーワード：試験所間比較 測定の不確かさ 技能試験 適合性評価

様式 C-19、F-19-1、Z-19、CK-19（共通）

1. 研究開始当初の背景

複数試験所で同一あるいは同一とみなせるような対象を測定し、その結果を比較する試験所間比較は、試験所の測定能力を確かめる客観的な手段のひとつである。正しい測定は多くの産業の基盤であり、試験所の第三者認証の仕組みにも試験所間比較が取り入れられている。その実施には数年をかけることがある一方で、期間の短縮のため、全試験所をサブグループに分け、サブグループごとに異なるものを測定することも行われる。この場合、その違いが無視できるほど小さい複数の回付物を用意することが求められる。しかし、測定の種別や比較試験の規模によっては、その実現が難しいことがある。結果として、同じものを測定した試験所間のみで関連性を確かめたり（下図A参照）、統計学的背景の十分でない方法で全体を関連づけたりすることが行われている。

サブグループごとに異なるものを測定したときにも、全体を適切に関連づける解析を行うことが効率的である。しかし、これを困難にする原因のひとつは、参加する試験所の測定能力が確かなものであることをチェックする性質上、全ての試験所に同じレベルの測定精度を想定することができないことにある。

一つの事例としては、計量標準の国際的な同等性を確認するための比較試験がある。計量標準の国際相互承認協定[1]には、150を超える国家計量機関や国際機関が参加しており、その測定の同等性を確かめる必要がある。しかし、これらの機関の測定能力を一度に調査するのは現実的には不可能である。このため、国際的に高い水準にある測定能力を持つ機関のみが参加する試験所間比較のみを国際機関である国際度量衡局の諮問委員会が主催している。その他の計量機関はアジアやヨーロッパなどの地域ごと計量組織が実施する試験所間比較に参加することで同等性を確保することになっている。しかし、上に述べたように2種の比較試験を適切に関連づけることができていないのは多に議論の余地がある。

また、本研究には測定の不確かさの活用という側面がある。測定の不確かさは、測定値の信頼性を定量化したものである。国家計量機関はともかく、一般的な試験所については、試験所間比較の解析の目的で測定の不確かさの情報を用いることはまだ一般的ではない。しかし、同一品目を測定した場合に、その情報を活用した簡便な解析方法を応募者らが一連の論文で発表し、その活用に対する困難さは小さいものになっている。また、JIS Q 17025「試験所及び校正機関の能力に関する一般要求事項」[2]においては、試験所も測定の不確かさを評価する能力を持つことが求められており、その活用が期待されている。

2. 研究の目的

本研究は、サブグループごとに異なるものを測定したときに、全体を適切に関連づける方法を提案するものである。この目的のために、いくつかの試験所は複数のサブグループに所属するように試験所間比較を設計し、サブグループ間につながりを持たせる（図1B）。これらの試験所の測定結果を詳細に分析することにより、異なるサブグループに参加した試験所の間で、測定値を互いに比較できるようにする。本研究では、測定の不確かさと比べて、複数の回付物の間の測定値の違いが無視できないような場合でも、統計的に意味のある比較が可能な試験所間比較の設計を提案する。

本研究の最終的な目的は参加した試験所の測定能力についての評価指標を与えることである。単一の測定物を用いる場合、その品目に対する「真の値」の推定値が試験所間比較の「参照値」となり、測定値と参照値との差に着目する。複数の測定物が存在する場合も参照値を定め、その値との差に着目することが考えられる。従来研究において、参照値の定め方については複数の提案がある。本研究では、「統計指標の持つ意味が参加者に分かりやすいこと」と、下図の意味で十分な精度で「試験所の測定能力について評価可能な指標であること」の2点を考慮し、その中から最適なものを選定する。

具体的には、測定値が測定の不確かさで決定できる分布を持っているという仮説を検定できるように設計する。明確な統計モデルに基づくことで、分かりやすさと統計的な背景を持つ評価指標を与えることを両立しようとするものである。

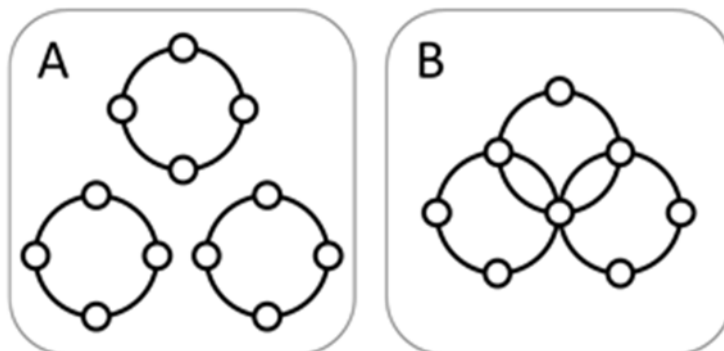


図1 試験所間比較の概念図

3. 研究の方法

本研究では、上に述べたように各試験所は各々に測定の不確かさを報告していることを前提とする。この測定の不確かさは、測定値の標準偏差に容易に変換できるものとする。また、2つの以上の報告値を報告する試験所からは、その2つの報告値の相関係数について報告されているものとする。

これらの前提の下、我々が考えたのは以下の手順である。

- (1) 1つのサブグループに他のサブグループよりも高い重要性を持たせ、サブグループには、測定能力が優れていると思われる試験所のみが参加できる。これをサブグループ C (Central) と呼ぶ。
- (2) サブグループ C に参加する試験所の試験所の測定能力について評価を与える。この場合、単一の測定物を用いる通常の試験所間比較の解析であり、従来の統計手法[3]により評価できる。これにより、「満足」な評価を得る試験所と、「不満足」な評価を得る試験所に分類する。これをサブグループ C に参加する試験所の最終結果とする。
- (3) それ以外のサブグループの一つをサブグループ R (Reginal) と呼ぶ。サブグループ C とサブグループ R の両方に参加している試験所をリンク試験所と呼ぶ。複数のサブグループ R に参加している非リンク試験所はないものとする。リンク試験所は、サブグループ C の評価においてかならず「満足」な評価を得ているものとする。（「不満足」であった場合には、サブグループ R の評価には報告データを用いない。）
- (4) リンク試験所の報告した相関係数の妥当性を確認するための、サブグループ C での報告値の全てと、サブグループ R でのリンク試験所の報告値を用いて、統計的検定を適用する。この統計的検定の結果、もし、妥当性が確認できなければ、その試験所をリンク試験所から外し、サブグループ R の解析には用いない。
- (5) リンク試験所が報告した値は報告した標準偏差と相関係数を持つことを前提として、非リンク試験所の測定能力について評価するための統計的検定を適用する。

この手順では、試験所の測定能力についての評価が「満足」なものであると確定した試験所があることにより、手順(2)で用いる従来の統計手法[3]の一つを拡張的に適用することができる。この従来の手法においては、測定能力についての評価が低い試験所の報告値に全体の解析が影響を与えることのないように、それらを外れ値として取り扱う。どの値を外れ値とするかを決定する際には、比較対象のデータの重みが揃っている方が取り扱いやすい。つまり、2つ値を報告している試験所と1つの値のしか報告していない試験所の報告値を比較して、測定能力についての評価を与えることは困難が大きい。このため、本研究では上の手順による解析を実施することとした。

本研究の具体的な課題は、手順(4)および手順(5)を実現する統計学的な手法を設計することである。上に示したように手順(4)の統計的検定は報告された相関係数の妥当性の確認のために実施される。手順(2)、(3)においてリンク試験所の測定能力についての評価は「満足」なものと確定している。これは、この試験所の報告する値がその報告した不確かさに対して妥当なものであることを評価する。しかしながら、2つ以上の値を報告した際のそれらの間の相関係数については妥当性の評価はなされていない。この相関の情報は、手順(5)の統計的検定のために重要な情報であるため、上の手順(4)においては、統計的検定を設計することを考える。

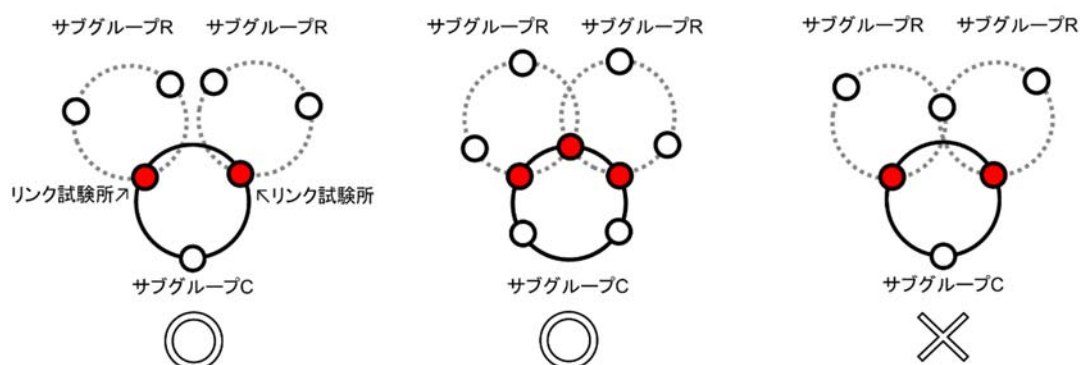


図2 リンク試験所は一つのサブグループ R の中に一つでもよい (左)。また、一つのサブグループ R にリンク試験所が複数あってもよいし、2つ以上のサブグループ R に一つのリンク試験所が参加してもよい (中)。ただし、サブグループ C に参加していない試験所が、複数のサブグループ R に参加することは、本研究では考慮しない。

4. 研究成果

(1) リンク試験所の相関係数についての統計的検定

リンク試験所 i ($i = 1, \dots, L$) がサブグループ C およびサブグループ R にて報告した値をそれぞれ x_i および y_i とする。それらの標準偏差で表した不確かさをそれぞれ $u(x_i)$ および $u(y_i)$ とする。また、 x_i と y_i の間の相関係数を ρ_i とする。さらに、サブグループ C で報告された非リンク試験所の報告値について、分散の逆数を重みとする重み付き平均を採ったものを \tilde{x} とする。ただし、外れ値は除いた後の重み付き平均とする。これらについて以下の統計モデルを立てる。また、 \tilde{x} の標準偏差で表した不確かさを $u(\tilde{x})$ とする。

$$z_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_i), \quad \tilde{x} \sim N(\mu_x, u(\tilde{x}))$$

ここで、

$$z_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}, \boldsymbol{\Sigma}_i = \begin{pmatrix} u^2(x_i) & \rho_i u(x_i) u(y_i) \\ \rho_i u(x_i) u(y_i) & u^2(y_i) \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

とする。 μ_x 、 μ_y は未知とする。この統計モデルをモデル L と呼ぶ。ベクトル $\mathbf{z}^* = (x^* \ y^*)^T$ の要素 x^* および y^* は、それぞれ μ_x 、 μ_y の推定量となる。

$$\mathbf{z}^* = \begin{pmatrix} x^* \\ y^* \end{pmatrix} = \left[\sum_{i=1}^L \boldsymbol{\Sigma}_i^{-1} + \begin{pmatrix} u^{-2}(\tilde{x}) & 0 \\ 0 & 0 \end{pmatrix} \right]^{-1} \left[\sum_{i=1}^L \boldsymbol{\Sigma}_i^{-1} \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} \tilde{x} u^{-2}(\tilde{x}) & 0 \\ 0 & 0 \end{pmatrix} \right]$$

以下の統計量を用いて、モデル L についての統計的検定を行うことを提案する。

$$\chi_{\text{pre}}^2 = \sum_{i=1}^L (z_i - \mathbf{z}^*)^T \boldsymbol{\Sigma}_i^{-1} (z_i - \mathbf{z}^*) + \frac{(x_i - x^*)^2}{u^2(\tilde{x})}$$

モデル L の下、上の統計量は自由度 $(2L - 1)$ の χ^2 分布に従う。このため、この統計的検定は有意水準を 5% とすると、以下が成り立つときモデル L であるという仮説は棄却できない。

$$\chi_{\text{pre}}^2 \leq \chi_{2L-1, 0.95}^2$$

ここで、 $\chi_{2L-1, 0.95}^2$ は自由度 $(2L - 1)$ の χ^2 分布の 95% 点である。

それぞれの試験所での相関係数の定量化について、技術的に十分な検証が行われていることは前提である。モデル L を統計学的な観点からも棄却できないことから、この手法により相関係数を含めたリンク試験所の報告値について、信頼できない理由を見出すことができない。このことから、サブグループ R における非リンク試験所の測定能力についての評価において、これらの報告値は信頼できるものとする。この成果を(査読付き)論文としてまとめ、書籍におさめられた (Shirono and Cox, *Advanced Mathematical and Computational Tools in Metrology and Testing XI*, World Scientific, Singapore (2018) 324-331)。

適用例 :

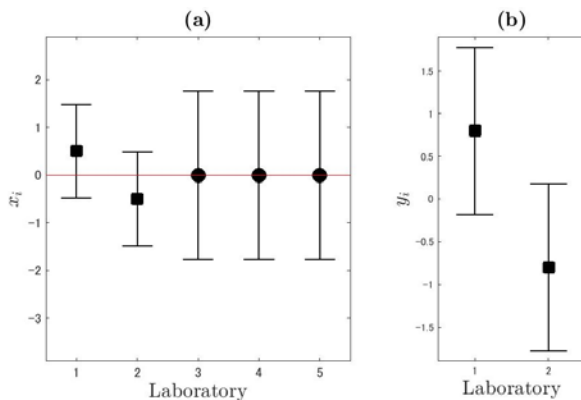


図3 試験所間比較のダミーデータ。(a) サブグループ C、(b) サブグループ R のリンク試験所の報告値。縦棒はそれぞれの標準偏差の 2 倍。

図3の(a)にはサブグループ C での報告値 x_i が、(b)にはサブグループ R でのリンク試験所の報告値 y_i が描かれている。それぞれ標準偏差の 2 倍を表す縦棒とともに描いた。ここで、リンク試験所 1 および 2 がそれぞれ相関係数 0.3 を報告したものとする。このデータに上の考えを適用すると、 $\chi_{\text{pre}}^2 = 5.7$ と求まる。自由度 $(2L - 1) = 3$ の χ^2 分布について、 $\chi_{3, 0.95}^2 = 7.8$ である。 $\chi_{\text{pre}}^2 \leq \chi_{3, 0.95}^2$ であるから、リンク試験所の報告した相関係数が妥当でないとする積極的なエビデンスは見つけれない。

(2) サブグループ R における非リンク試験所についての統計的検定
 リンク試験所 i ($i=1, \dots, L$) に対して、非リンク試験所 i ($i=L+1, \dots, n$) の測定技能についての評価を行う。このときには、以下のモデルを確かめる。

$$y_i \sim N(\mu_y, u(y_i)) \quad (i=L+1, \dots, n), \quad y^* \sim N(\mu_y, u(y^*))$$

ここで、 μ_y は未知であるものとし、 y^* は(1)の手順で定義したものである。これをモデル M と呼ぶ。モデル M の検定のためには、多くの基準が考えられる。このため、ある対象にしている試験所 j について、以下の対立仮説について、その検出力を調べることにした。

$$y_j \sim N(\mu_y + \theta_j, u(y_j)) \quad (\theta_j \neq 0)$$

y_i および y^* の線形結合で検定のための統計量を与えることを考える。

我々は、 θ_j の値に関わらず、上の対立仮説に対する検出力を最大にする統計量を決定することができることを見出した。具体的には以下の式で与えられる。

$$E_n = \frac{y_j - y_{\text{ref}}}{1.96 \sqrt{u^2(y_j) - u^2(y_{\text{ref}})}}$$

ここで、

$$u^2(y_{\text{ref}}) = \left(\frac{\tilde{y}}{u^2(\tilde{y})} + \frac{y^*}{u^2(y^*)} \right)^{-1} \left(\frac{\tilde{y}}{u^2(\tilde{y})} + \frac{y^*}{u^2(y^*)} \right)$$

$$\tilde{y} = u^2(\tilde{y}) \left(\sum_{i=1}^L y_i u^{-2}(y_i) \right), \quad u^2(\tilde{y}) = \left(\sum_{i=1}^L y_i u^{-2}(y_i) \right)^{-1}$$

である。実質的に y_{ref} が参照的な量となる。これは μ_y の推定量の一つである。この E_n はその絶対値が 1 以下のときには、有意水準 5 % でモデル M を棄却できず、試験所 j の測定能力についての評価は満足と与えられる。この成果は、国際会議にて発表し、プロシーディングが公開されている。(Shirono and Cox, IOP Conf. Series: Journal of Physics: Conf. Series 1065 (2018) 212006)。

さらに、我々はこの考えを実際に適用する際に問題となる点を解決するためのアルゴリズムを開発した。具体的には、上の考えでは試験所 j 以外の試験所が適切な報告をしていることを前提としているが、必ずしもこれは成り立たない。この点については、3 章で述べたように従来の統計手法[3]の一つを拡張する解決策を見出している。これらの成果をまとめ、近く査読付き論文として投稿する予定である。

適用例：

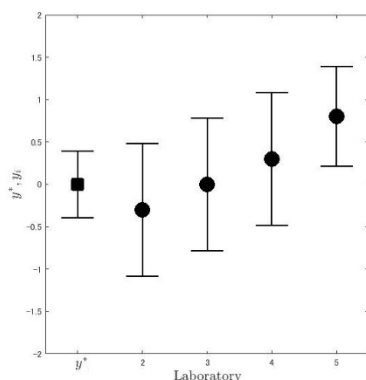


図 4 試験所間比較のダミーデータ。 y^* はサブグループ C の報告値とサブグループ R の報告値から得た重み付き平均。

図 4 には y^* および非リンク試験所の報告値 y_i が描かれている。それぞれ標準偏差の 2 倍を表す縦棒とともに描いた。ただし、 $L=1$ とする。このデータに上の考えを適用すると、試験所 2 から 4 は $E_n = -0.6, -0.2, 0.2$ となり、「満足」な評価を得る。しかし、試験所 5 については $E_n = 1.2$ となり、「不満足」という評価となる。現実に「不満足」と評価された場合には、当該試験所は是正処置を通常行う。特に、試験所の第三者認定においては、適切な是正処置が行われることが求められる。

本研究の手法では具体的な統計モデルを与え解析をし、また結果として参照値との差（を規格化したもの）が評価指標になっているために、分かりやすい指標になっていると考えている。また、近いアプローチの従来研究では測定値間の相関が弱いときに現実的でないかたよりが生じる問題があったが、本研究の手法ではこの問題が解消されていることが分かった。このように、分かりやすさと適切な評価を両立する手法が開発できたものと考えている。

本研究の学術的成果は今後関連のガイドラインや標準などの整備に活かし普及を目指す。この成果が普及することにより、試験所間比較による測定能力についての評価がより効率的かつ高い精確さで行えるようになる。この成果は、試験所の第三者認証に活用されることで、信頼性の高い計測の基盤を社会に提供することに貢献するものと考えている。

<引用文献>

1. BIPM. Mutual recognition of national measurement standards and of measurement certificates issued by national metrology institutes (MRA). Technical report, Bureau International des Poids et Mesures, Sèvres, France, 1999.
2. JIS Q 17025:2018 試験所及び校正機関の能力に関する一般要求事項、2018、日本規格協会、東京
3. M. G. Cox, The evaluation of key comparison data: determining the largest consistent subset. *Metrologia*, 2007, vol.44, 187-200.

5. 主な発表論文等

〔雑誌論文〕（計 0 件）

〔学会発表〕（計 2 件）

- ① Katsuhiro SHIRONO, and Maurice COX, “Statistical testing approaches in the analysis of a regional metrology organization key comparison”, *Advanced Mathematical and Computational Tools in Metrology and Testing XI*, 2017, Glasgow, United Kingdom.
- ② Katsuhiro SHIRONO, and Maurice COX, “Unilateral degree of equivalence maximizing the power of test in an analysis of a regional metrology organization key comparison”, *XXII World Congress of the International Measurement Confederation (IMEKO 2018)*, 2018, Belfast, United Kingdom.
DOI: :10.1088/1742-6596/1065/21/212006

〔図書〕（計 1 件）

1. Katsuhiro SHIRONO, and Maurice COX , "Analysis of a regional metrology organization key comparison: Preliminary consistency check of the linking-laboratory data with the CIPM key comparison reference value", in the book of *Advanced Mathematical and Computational Tools in Metrology and Testing XI*, World Scientific, Singapore, 査読有, 2018, 324-331
DOI: https://doi.org/10.1142/9789813274303_0032

6. 研究組織

(1)研究分担者 なし

(2)研究協力者

研究協力者氏名：Maurice Cox (National Physical Laboratory、英国)

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。