

試験所間比較による技能 試験の概要と統計的方法

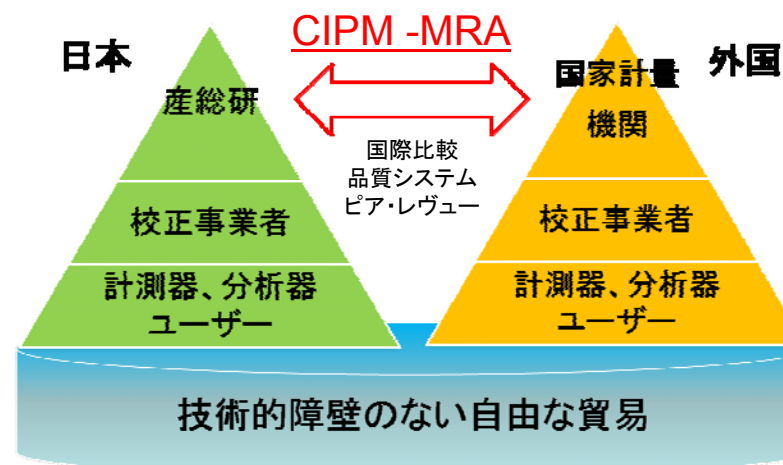
産業技術総合研究所 計測標準研究部門

物性統計科 応用統計研究室

城野 克広 <k.shirono@aist.go.jp>

1. MRAと試験所間比較

CIPM-MRA(国際相互承認協定)



冊子 “計量標準の国際相互承認” (NMIJ) 参照

MRAの広がり

“計量に関する様々な国際取り決めの貿易、法制及び標準化との関連性に関するBIPM, OIML 及びILAC の共同声明並びに宣言” (2006年)にみるようにMRAは他分野へも広がっている。

CIPM、OIML、ILACは、

☞ 各国政府に対し、法律、法規制などで計量が求められるとき、CIPM MRA、OIML MAA、ILAC取り決めに署名した組織を支持する。

☞ 標準化組織、取締機関及び貿易機関に対し、この取り決めに留意し、促進し、活用する方法と手段を開発するために三つ組織と協力することを求める。という内容が明記されている。



BUREAU
INTERNATIONAL DES
POIDS ET MESURES



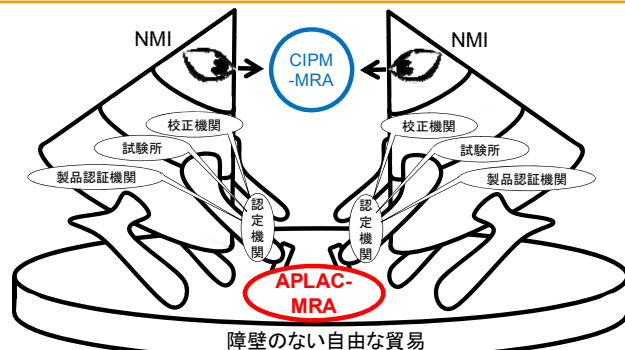
ORGANISATION
INTERNATIONALE DE
METROLOGIE LEGALE



INTERNATIONAL
LABORATORY
ACCREDITATION
COOPERATION

ILAC-MRA

ILAC-MRA(やAPLAC-MRA)に認定機関が署名(日本ではIAJapanとJABが署名)することが、認定された試験所/校正機関が発行する試験報告書/校正証明書の国際的な受入れができるための必要条件と言える。



独立行政法人 産業技術総合研究所

5

ILAC-MRAにおける試験所間比較

試験所認定に係る様々な基準において技能試験の必要性に触れている。特に試験所・校正機関のための標準であるISO/IEC 17025と、認定機関のための標準であるISO/IEC 17011に以下のような記載がある。

IEC/ISO 17025:2005 5.9項

- 試験所・校正機関は、請け負った試験・校正の有効性の監視のため品質管理手順をもつ。(中略)次の事項を含むのがよいが、これらに限定されない。
- b) 試験所間比較または技能試験プログラムへの参加

IEC/ISO 17011:2004 7.15.3項

- 認定機関は、利用できるものがあり適切である場合、その認定校正機関・試験所が技能試験又はその他の試験所間比較に参加することを奨励しなければならない。

独立行政法人 産業技術総合研究所

6

ワン・ストップ・テスト

MRAは、“ワン・ストップ・テスト”のための協定とも言える。



Before ワン・ストップ・テスト



ワン・ストップ・テスト

独立行政法人 産業技術総合研究所

7

2. 試験所間比較を用いた技能試験

独立行政法人 産業技術総合研究所

8

ISO/IEC GUIDE 43 試験所間比較による技能試験

試験所間比較の用途、設計および実行について、企画・調整する際に考慮すべき基本的な原則を規定する。

計画は、スキームの開始前に合意され、文書化される。統計的設計、試験品目の管理などについても文書化する。(方法・手順については、通常参加者が選択する。)

調整者が(あるいは外部契約によって)試験品目の準備・調整をし、適切に包装、参加者に輸送する。順守すべき詳細な指示書を提供する。

試験所から受け取った結果を解析し、付与された値及びその不確かさの決定、実績統計量の計算、実績の評価などについての報告を参加者に返す。

報告書は規定の日程内で速やかに利用可能とする。

参加者は、実績評価が間違っていると考えるならば、調整者に照会する。スキームの進展に活発に寄与するために、試験所からのフィードバックを奨励する。

GUIDE43-2には、試験所認定機関による技能試験スキームの選定および利用について述べている。

ISO/IEC GUIDE 43 (JIS Q 0043)参照

Guide 43から17043へ

ISO/IEC Guide 43は2011年2月をもって廃止となり、ISO/IEC 17043が新規格として有効になった。新しい文書はISO/IEC 17025、ILAC G13、IUPAC Harmonized Protocolという関連の文書とよく一貫性を保つものである。

大きな変更点

- ・審査機関に対する考慮
 - ・校正スキームに対して付与された値のトレーサビリティの要求
 - ・下請けの限界
 - ・付与された値の不確かさの報告に対する要求
 - ・報告書における計測方法の情報や技術的コメントを含むことへの要求
- D. W. Thohen"ISO/IEC 17043: the new International Standard for proficiency testing" Accred. Qual. Assur. 13(2008) 727より

本資料ではGuide 43の記述に基づいた解説を行う。

企画・調整および設計

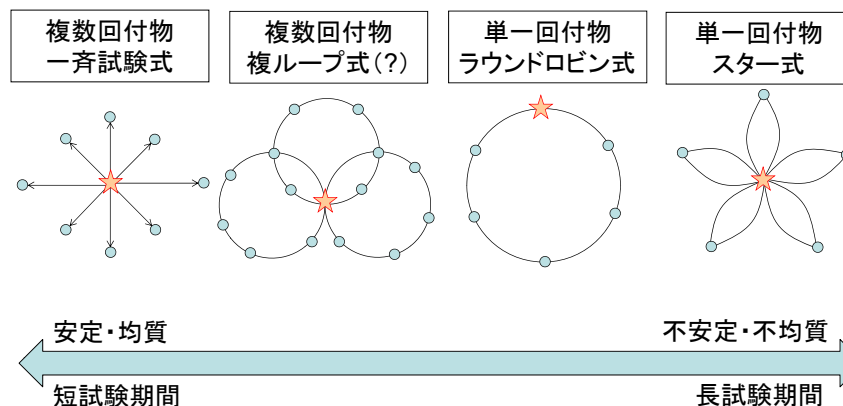
Guide 43では、コーディネーター(技能試験のプロバイダー)の事前の計画を立てる際のガイドを用意している。

5.1.2 調整者は、これらの他の人員と協議して、特定の技能試験に必要なスキームを開発することが望ましい。計画は、スキーム開始前に合意され、文書化され、典型的には次の情報を含む。

- 技能試験スキームを辞しする機関の名称および所在地...
- 選択された品目および試験の性質、並びにそれらの選択を裏付ける考察に関する簡単な記述...
- 付与された値の決定および外れ値の検出法を含め、用いられる統計解析の概要...
- 試験実績評価の手法の根拠

企画・調整および設計

試験の形態には以下のようなものがある。時間的制約と被測定物の安定性・均質性・コストによって形態を定めることになる。



付与された値・実績の評価

まず、“付与された値の決定および外れ値の検出法を含め、用いられる統計解析の概要”と“試験実績評価の手法の根拠”に着目する。

付与された値とは何か？ 試験実績の評価はどのように行うのか？

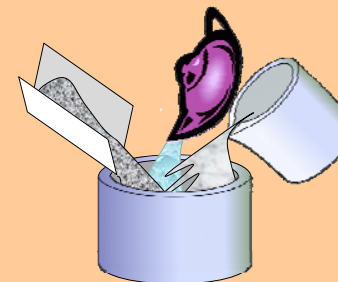
☞ ISO/IEC Guide 43 (JIS Q 0043)を補足規格であるISO13528 (JIS Z 8405) “試験所間比較による技能試験のための統計的方法”に記載の方法を紹介する。

付与された値

付与された値とは技能試験の“正解”のこと。

例えば、

硬化コンクリートのセメント量の定量 (JIS Z 8405 5.2.3項)



“この場合、コンクリート試料片は、各成分(セメント、骨材、水)の量を測定し、各コンクリート試料を固めるために互いに混ぜ合わせて作成する。この手法の利点は、試料片を作成する精度がセメントが含有量の定量より高いことによる。”

→ この場合試料作成時のセメント含有量が付与された値

付与された値

この例のような分かりやすい定式化によって付与される値が与えられることは少ない。ISO13528には以下の5つの方法が提案されている。

技能試験前に決定する場合

- 定式化(先スライドのような物理的関係から決める方法。)
- 認証参照値(不確かさの付いた標準物質を購入する。)
- 参照値(認証参照物質を用いて値付けしておく方法。)

技能試験後に決定する場合

- 熟練試験所による合意値
- 参加試験所による合意値

→ どの方法でも不確かさはある！！

付与された値

技能試験の後に付与された値を決定するときには、ISO13528付属書Cに記載のロバストな平均値が用いられる。

- ① 外れ値を検出し、
- ② その代わりに外れ値の検出値を代入、
- ③ それらの平均と標準偏差を計算する。
(詳細は参考資料)

標準偏差はロバストな標準偏差と呼ぶ。

(注) 統計的基礎を備えた他の方法でもよい。(例えばメジアン。)

実績統計量

実績の評価は**実績統計量**に基づいて行われる。

実績統計量とは、**zスコア**とか **E_n 数**とかのこと。

zスコア

$$Z = \frac{x - X}{\hat{\sigma}}$$

E_n 数

$$E_n = \frac{x - X}{\sqrt{U_{\text{lab}}^2 + U_{\text{ref}}^2}}$$

他の指標に関してはISO13528 (JIS Z 8405)を参照 されたい。

zスコア

$$Z = \frac{x - X}{\hat{\sigma}}$$

ISO/IEC Guide
43 付属書Aより

$|z| \leq 2$: 満足、 $|z| > 2$: 疑わしい、 $|z| \geq 3$: 不満足

x は報告値。 X は付与された値、 $\hat{\sigma}$ は「**技能評価のための標準偏差**」。

例えば、

ナッツ製品中のアフラトキシン (JIS Z 8405 6.2.2項)



“法定限度が10 µg/kgである場合、アフラトキシンの試験に使用する方法の再現精度の変動係数は50 %を越えないことを規定した法令がある。”

→ **この場合、技能評価の標準偏差としても5 µg/kgを設定するのがよい。**

zスコア

法令など規定値がある場合は珍しい。ISO13528には以下の5つの方法が提案されている。

技能試験前に決定する場合

- 規定値(先スライドのような法令など。)
- 達成期待レベル
- 一般的なモデル(その実験の再現精度に関するよく知られたモデルがある場合にそれを用いる。)
- 精度評価実験結果(技能試験スキームに使用する測定方法を標準化している場合にその標準的不確かさを使用する。)

技能試験後に決定する場合

- 単一技能試験スキームによって得たデータを用いる場合

zスコア

単一技能試験スキームによって得たデータを用いる場合はロバストな標準偏差が用いられる。

(注) 統計的基礎を備えた他の方法でもよい。とくに(四分位範囲×0.7143))は、付与された値がメジアンるときよく使われる。(参考資料)

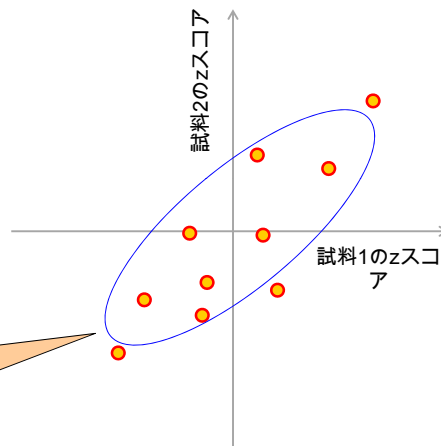
zスコアへの批判

☞ 「ロバストな標準偏差」などによって技能評価のための標準偏差を定めるとき、試験所の成績が達成期待レベル以上に揃うと、達成期待レベルに足る結果を出した試験所のzスコアが2以上になることがありうる。

他の手段、例えば規格化された試験法で「精度評価実験結果」を用いれば、上記の問題は起きにくい。

zスコア

単一ラウンドの比較試験において、複数の回付物や試料を測定することはよく行われる。その際には両方についてのzスコアを平面上にプロットしたユーデン・プロット(散布図)による評価が行われる。



P%の試験所がそのだ円の中に入っていることが期待される信頼域だ円を任意のPに対して描くことができる。詳細はJIS Z 8405参照。

En数

$$E_n = \frac{x - X}{\sqrt{U_{\text{lab}}^2 + U_{\text{ref}}^2}}$$

$|E_n| \leq 1$: 満足、 $|E_n| > 1$: 不満足

ISO/IEC Guide
43 付属書Aより

x、Xは報告値と付与された値。 U_{lab} と U_{ref} は各々の拡張不確かさ。

E_n 数への批判

- ➡ 拡張不確かさは二乗和してはいけない。つまり、甘い指標。
- ➡ 認定だけなら、 U_{lab} が大きいほど通りやすい逃げ得の指標。
(ちなみに、CIPM基幹比較のDOEは、 E_n 数と違い(多少は)攻め得。)

試験方法が一通りの場合、不確かさ評価方法も統一化されていれば、後者の問題は起きにくい。

実績の評価

技能試験の実績は、実績統計量による採点のほか、「専門家の合意」、「目的への適合性」、「参加者の合意」などに基づいて確定する。長期にわたって実績が監視されることもある。

ISO 13528の4.1.1

“優秀なスタッフがあり、運営が良好である試験所においても、異常な結果が発生することがある。...(略)...ここに規定する基準(ここでは、zスコアや E_n 数のこと)を用いて、検討対象の測定方法を実施するのに不適格であるとして、試験所を不良としてはならない。技能試験を試験所判定に使用する場合は、その目的に合った適切な基準を設定しなければならない。”

均質性・安定性

話を戻して。“選択された品目および試験の性質、並びにそれらの選択を裏付ける考察”は様々あるが、まずは回付物の均質性・安定性が心配なところ。

ISO13528では均質である確認を以下のようにする。

$$s_s \leq 0.3\hat{\sigma}$$

s_s : 試料間標準偏差
 $\hat{\sigma}$: 技能評価のための標準偏差

試料間標準偏差は分散分析によって算出する。

均質性・安定性

ISO13528では安定である確認を以下のようにする。

$$|\bar{x}_{..} - \bar{y}_{..}| \leq 0.3\hat{\sigma}$$

$\bar{x}_{..}$: 均質性試験時の全試料、全繰返しの平均値

$\bar{y}_{..}$: 均質性試験時から技能試験における遅延と同程度の時間をおいて測定した全試料、全繰返しの平均値

$\hat{\sigma}$: 技能評価のための標準偏差

均質であることが確認できないとか、安定であることが確認できない場合の対処については言及がない。

3. 試験所間比較における留意点

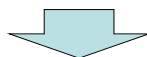
技能試験に大切なこと

「実績統計量」は最後の手段。

他の方法で、不具合のある試験を運営している試験所を見出すことの方がずっとよい。

ISO 13528の4.1.1

“優秀なスタッフがおり、運営が良好である試験所においても、異常な結果が発生することがある。”



ほとんどの技能試験では試験条件に関する
アンケートが実施されている。

技能試験に大切なこと

- ☞ なるべく間違いが起きやすい要素を的確に。しかし、暗に試験方法に対する指示を与えないように。普段の測定を行えるようなアンケートを作成に努めるべき。(そのためにクレームがつかない程度にあえて冗長にするという手段もあるかも。)
- ☞ 試験中の写真を撮るなどが効果的な場合も。
- ☞ 特にJISなどに規格化された測定の技能試験では、測定に対する指示をほとんど与えないことも。



試験結果のフィードバックという観点からも強く実施が望まれる。

均質性・安定性の問題

均質性・安定性と統計的方法に対する以下の誤解が多い。

- (1) 「不安定であっても統計的手法を用いれば、適切なzスコアが計算できる」という誤解。
- (2) 「不均質な場合は、試験所の報告値のばらつきに均質性のばらつきを足して技能評価の標準偏差を計算する」という誤解。

(1) 不安定であっても統計的手法を用いれば、適切なzスコアが計算できる？

例えば、コーディネータ側で不安定性に関するばらつきがよく調査できているものとし、その標準不確かさ $\hat{\sigma}_{stab}$ とする。試験所の報告値からロバストな標準偏差から算出した技能評価のための標準偏差を $\hat{\sigma}$ とする。

で、どうする？

均質性・安定性の問題

$\hat{\sigma}$ には不安定性の不確かさが含まれているから、

$$Z = \frac{x - \bar{X}}{\sqrt{\hat{\sigma}^2 - \hat{\sigma}_{stab}^2}} \quad ?$$

× × × **不正解** × × ×

x(報告値)には不安定性によるずれも含まれている。
分母だけ小さくしてはいけない。

結局、どうしようもない。不安定性のずれを含んだままzスコアを計算することになる。これは曖昧な指標となる。
一斉試験スキームでの不均質性に関しても同様。

均質性・安定性の問題

(2) 不均質な場合は、試験所の報告値のばらつきに均質性のばらつきを足して技能評価の標準偏差を計算する？

例えば、不均質性に関するばらつきの標準不確かさ σ_{sample} とする。試験所の報告した値の標準偏差を σ とする。

➡ もし、試験所間のばらつきの標準偏差を σ_{lab} とすると、

$$\sigma^2 = \sigma_{sample}^2 + \sigma_{lab}^2 + \dots$$

すでに不均質性の不確かさは入っている。
だから、もう足さない。

➤ 「各試料のずれを報告値から補正し、一方 σ^2 からは σ_{sample}^2 を引く」が正解。ただし、ずれをどう算出するかも厄介さがある。何もしないとこれも曖昧な指標に。

4. おわりに

おわりに

同等性評価においては、

ピア・レビュー & 品質システム
> 試験所間比較による技能試験

という意識が重要！

で、あればこそ、逆に比較試験の意味と発展性がある。

＋ 継続的な比較と弛みない改善

➡ 最大のフィードバックが得られるよう標準化すべきところ(試験方法、不確かさ評価方法)を見極める。

関連文書紹介、参考資料

関連文書

CIPM基幹比較に関する文書。

☞ MRA本文

CIPM “Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes” (http://www.bipm.org/utis/en/pdf/mra_2003.pdf)

☞ 基幹比較のCIPMガイドライン

CIPM “Guidelines for CIPM key comparisons” (<http://www.bipm.org/utis/en/pdf/guidelines.pdf>)

☞ 基幹比較のCIPM推奨の方法

M. G. Cox, “The evaluation of key comparison data”, *Metrologia* 39(2002)589-.

関連文書

技能試験に関し、以下の2点を紹介した。技能試験を設計される場合には参考に手元においておきたい。

☞ ISO/IEC 17043

Conformity assessment – General requirements for proficiency testing
 2010年2月に発行された。JIS Q 0043はこの規格ではなく、ISO/IEC Guide 43の翻訳であり、本資料はJIS Q 0043を基に作られている。

☞ ISO13528 (JIS Z 8405)

Statistical methods for use in proficiency testing by interlaboratory comparisons

ロバストな平均、ロバストな標準偏差

ロバストな平均値とロバストな標準偏差の計算方法がISO13528 附属書CにアルゴリズムAとして記載されている。

試験所数 p 、報告値 x_i ($i = 1, \dots, p$)のとき、

- (1) x^* を x_i の($i = 1, \dots, n$)メジアン。
 $|x_i - x^*|$ を計算し、そのメジアンを s として $s^* = 1.483s$ 。
- (2) $\delta = 1.5s^*$ 。もし、 $x_i < x^* - \delta$ であれば、 $x_i^* = x^* - \delta$ とする。
 もし、 $x_i > x^* + \delta$ であれば、 $x_i^* = x^* + \delta$ とする。それ以外は、 $x_i^* = x_i$ 。

$$(3) \quad x^* = \sum_{i=1}^p x_i^* / p, \quad s^* = 1.134 \sqrt{\sum_{i=1}^p (x_i^* - x^*)^2 / (p-1)}$$

を新たに定義し、前の x^* と s^* に比べ、有効数字の3ケタ目が変わるときには、(2)に戻る。変動しなければ、 x^* がロバストな平均、 s^* をロバストな標準偏差とする。また付与された値の標準不確かさは $u_x = 1.25 \times s^* / \sqrt{p}$ 。

ちなみに通常ロバストというのはアルゴリズムや手順につけられる形容詞であることは承知した上で、便宜のためにロバストな平均値などと呼んでいる。

メジアン、四分位範囲による標準偏差

ロバストな平均値とロバストな標準偏差と並んでよく用いられる付与された値と技能評価のための標準偏差の算出方法。

試験所数 p 、報告値 x_i ($i = 1, \dots, p$)のとき、

- (1) Grubbs検定(JIS Z 8402-2)などで外れ値を検出し除いたのち、 x_i の($i = 1, \dots, n$)メジアンを付与された値 X とする。
- (2) x_i の($i = 1, \dots, n$)のうち $n/4$ 番目の大きさのものと、 $3n/4$ 番目の大きさのものの差IQRを計算する。
- (3) 正規分布では上側25 %点と75 %点の差の0.7413倍が標準偏差に当たることから正規化四分位範囲NIQR = $0.7413 \times \text{IQR}$ を技能評価のための標準偏差とする。

ロバストな平均値とロバストな標準偏差と比べると、「とっつきやすい」という長所があるが、「試験所の報告値が正規分布している」という点が、受け入れがたい場合もあるだろう。

均質性と安定性についての 補足情報

1. ISO 13528(JIS Z 8405) 附属書Bの詳細

ISO 13528 附属書B

ISO 13528(JIS Z 8405)は「試験所間比較による技能試験のための統計的方法」が記された国際規格である。この規格の附属書Bは「試料の均質性試験及び安定性試験」と題して、“技能試験に使用する試料が均質及び安定であることを適切に確認するための方法について規定”している。



均質??不均質??

均質性試験の手順

“B.1 均質性試験の手順”に均質性確認の方法が規定されている。

- a) 均質性試験を実施する試験所を選択する。均質性試験に使用する測定方法や測定する対象試験物質の特性について、試料間の不均質性について感度の高いものを選択すれば、すべての測定対象に対して均質性試験を実施なくてもよい。

...

➤ 繰返しの不確かさが小さい試験方法ほど均質性の影響は出やすいので、なるべく精度の高い測定方法で熟練した試験機関がチェックすべきである。
 ➤ 複数の試験項目がある場合には、もっとも均質性の影響が出やすいものだけで、この規格に適合する。

均質性試験の手順

その他にも、かなり具体的な均質性確認の方法が規定されているが、これより精度の高い試験を行うことも実際上は可能なので、必ずしもこれに縛られる必要はないと感じるが...

B.1 の方法

- ① 試験所に実際に送る一つのサンプルは2回以上の試験が可能なものであるとする。
- ② 10個以上の g 個のサンプルをランダムに選ぶ。
- ③ なるべく試験の繰返しの差が出にくいように、1サンプルにつき2つの試験部位を選ぶ。
- ④ $2g$ 回の試験をランダムに行い、分散分析する。

“試験の繰返しの差が出にくいように”は、「試験部分間差分を最小にする」という記述に対応する。ランダムに選ぶべきでは？（後述）

均質性試験の評価基準

いずれにせよ、「1つのサンプル内での繰返しの標準偏差と、サンプル間の標準偏差を求めよ。」ということである。これらをそれぞれ、 s_w 及び s_s とする。下の規定は先に示した通り。

“B.2 均質性試験の評価基準”

試料間標準偏差 s_s を技能評価のための標準偏差 $\hat{\sigma}$ と比較する。次の場合にはこの試料が十分均質であるとみなす。

$$s_s \leq 0.3\hat{\sigma}$$

この式の係数0.3の根拠は、この基準が満たされる場合、試料間標準偏差によって生じる技能試験の標準偏差が約10 %を超えないということである。

均質性試験の評価基準

もしも、均質性が技能評価の標準偏差の0.3倍よりも大きかったらどうするかについては、よく理解できない記述が多い。

- a) 試料作成手順を検査し、改善の可能性を調べる。
- b) 多数の試料を技能スキーム中の各参加者に配布し、各試料について測定結果を取得するように要求する。
 （これについては後述。）

試料内標準偏差が次のように増加する??という記述がある。

$$\sigma_{r1} \leq \sqrt{\sigma_r^2 + s_s^2}$$

これを“繰返し回数選択のための σ_r に置き換える”としているが、何のための操作なのか私にはよく分からない。

均質性試験の評価基準

- c) 次の式によって技能評価のための標準偏差 $\hat{\sigma}$ を計算して、試料間標準偏差を技能試験のための標準偏差に使用する。

$$\hat{\sigma} = \sqrt{\hat{\sigma}_1^2 + S_S^2}$$

ここに $\hat{\sigma}_1$ は試料の不均質性に許容度を含まない技能試験の標準偏差である。

計算できないし、解決策になっていない??

試料のばらつきが技能評価のための標準偏差に入ってはならない。

記述の内容からして、a) ~ c)のいずれかをういよと理解される。
b)の前半の記述は有意義であり、これについては後述する。

均質性試験に使用する式

分散分析の方法について、B.3に規定されている。ここで使用する分散分析は一元配置である。

分散分析

例えば、10個のサンプルをそれぞれ2回試験した平均値、 $\bar{x}_1$ 、 $\bar{x}_2$ 、...、 $\bar{x}_{10}$ が得られたとして、下のようにその分散を計算する。

$$u^2 = \frac{1}{10-1} \left\{ (\bar{x}_1 - \bar{\bar{x}})^2 + \dots + (\bar{x}_{10} - \bar{\bar{x}})^2 \right\}$$

$\bar{x}_1$ 、...、 $\bar{x}_{10}$ は繰返しの変動の影響も受けるので、この分散にはサンプル間のばらつきの他に、繰返しの変動も入ってしまう。結局、複雑な計算をしないとサンプル間のばらつきだけを取り出すことができない。そのための計算手法を分散分析と呼ぶ。

均質性試験に使用する式

先に述べた「なるべく試験の繰返しの差が出にくいように」選ぶのが疑問。分散分析を行うときにはランダムにサンプリングすべき。 m 個のサンプルをランダムに選び、それぞれ n 回試験したときに使える分散分析表を下に貼付する。

要因	S (変動)	f (自由度)	V (分散)	分散の期待値
試料間 (S)	$S_S = \sum_{i=1}^m \sum_{j=1}^n (\bar{x}_i - \bar{\bar{x}})^2$	$f_A = m-1$	$V_A = S_A/f_A$	$\sigma_r^2 + n \cdot \sigma_S^2$
繰返し(e)	$S_e = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$	$f_e = m(n-1)$	$V_e = S_e/f_e$	σ_r^2
総和	$S = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{\bar{x}})^2$	$f = mn-1$	σ_r : 繰返しの標準偏差 σ_S : 試料間のばらつきの標準偏差	

安定性試験の手順

“B.4 安定性試験の手順”に安定性を事前に確認する手順が記載されている。

- ① 均質性試験を行った試験機関が、均質性試験と同じ測定方法を行えるようにする。
- ② 実際の技能試験で想定される程度の期間を空ける。
- ③ 3個以上の g 個のサンプルをランダムに選ぶ。
- ④ なるべく試験の繰返しの差が出にくいように、1サンプルにつき2つの試験部位を選ぶ。
- ⑤ $2g$ 回の試験をランダムに行い、全平均値 $\bar{y}$ を計算する。

➤基本的には間を空けて、均質性試験と同様の試験を行うということである。もちろん時間変化のもっとも顕著な試験項目を選ぶべきだろう。

安定性試験の手順

安定である確認はまたも、0.3を基準にする。

$$|\bar{x}_{..} - \bar{y}_{..}| \leq 0.3\hat{\sigma}$$

$\bar{x}_{..}$: 均質性試験時の全試料、全繰返しの平均値

$\bar{y}_{..}$: 均質性試験時から技能試験における遅延と同程度の時間をおいて測定した全試料、全繰返しの平均値

$\hat{\sigma}$: 技能評価のための標準偏差

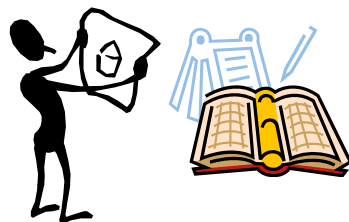
安定性が確認できない場合は「試料作成及び保管手順を検査し、改善の可能性を確認する。」とあるのみ。規定にはないが、実際上は、測定を速やかに行うように指示するのがよいだろう。

均質性と安定性についての 補足情報

2. 実験計画について

実験計画について

均質でない場合は試料の個数を増やして、その平均値を報告することで、均質性が与える問題をクリアすることができる。これは繰返し性についても同様である。平均値に関する統計について説明し、予備実験や過去の結果から、実験計画を立てる方法について説明する。



この章の内容
はISO 13528
の記載ではあ
りません。

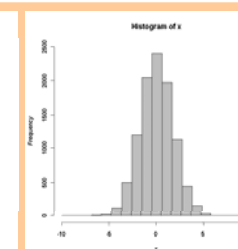
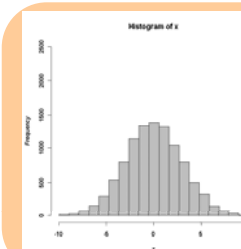
平均値の標準偏差

重要なのは以下の性質のみである。

もし $x_1, x_2, \dots, x_n$ が同じ標準偏差 σ の分布に従うなら、それらの平均値 $\bar{x}$

$$\bar{x} = (x_1 + \dots + x_n)/n$$

の標準偏差は $\sigma/\sqrt{n}$ であることが知られている。



左図が x_1, x_2, x_3 が従う分布とすると、

$$\bar{x} = (x_1 + x_2 + x_3)/3$$

が従う分布は右図のようなもの。同じ分布からの平均値は、分布の平均の周りに小さく鋭く分布する。

技能試験で知りたいものとは？

技能試験の報告値 x_i は、
全サンプルの母平均 μ に、
試験所に特有のかたより L_i 、
サンプルのばらつきに起因するずれ S_i 、
繰り返すごとにずれることによるかたより e_i
が加わったものと考えられる。

$$x_i = \mu + L_i + S_i + e_i$$

「 L_i が平均0、標準偏差 σ_L の正規分布に従っているかどうか？」というのがZスコアによる技能評価の本質。言うなら、

S_i や e_i には興味がない！！

報告値の標準偏差

報告値 x_i は1か所の試験所で、 m 個のサンプルについて、各 n 回ずつの試験が行われた平均値として報告されるものとしよう。報告値 x_i はどのような標準偏差を持つだろうか？

x_i の3つの成分について順に考える。 L_i については、「 L_i は平均0、標準偏差 σ_L の正規分布に従っている」と仮定しよう。

他のふたつの成分(S_i 、 e_i)はどうだろうか？

報告値の標準偏差

☞ S_i は m 個のサンプルから求めた平均である。

「一つのサンプルが持つずれ S_{ij} が平均0、標準偏差 σ_S の分布に従っている」と仮定しよう。先に示したように m 個の S_{ij} の平均である S_i は標準偏差 $\sigma_S/\sqrt{m}$ の分布に従う。

☞ e_i は総計 nm 回の繰返し試験から求めた平均である。

「一回の繰返しのたびに e_{ij} だけずれるとし、 e_{ij} は平均0、標準偏差 σ_e の分布に従っている」と仮定しよう。 e_i は m 個のサンプルにつき n 回、つまり nm 回 e_{ij} の平均であり、標準偏差 $\sigma_e/\sqrt{(nm)}$ の分布に従う。

報告値の標準偏差

報告値 x_i の標準偏差 σ は、これら3つを合成して、

$$\sigma = \sqrt{\sigma_L^2 + \frac{\sigma_S^2}{m} + \frac{\sigma_e^2}{nm}}$$

σ を推定できれば、「 x_i が平均0、標準偏差 σ の正規分布に従っているか？」は検定できる。

$$\sigma_L^2 \gg \sigma_S^2/m + \sigma_e^2/nm \text{ なら、 } L_i \sim x_i, \sigma_L \sim \sigma$$

上の検定と「 L_i が平均0、標準偏差 σ_L の正規分布に従っているか？」が同じとみなせる！

最低サンプル数と繰返し数

予備試験では、1か所の試験所で、 m' 個のサンプルについて、各 n' 回ずつの試験が行われた平均値として報告されたものとしよう。



予備試験の報告値 $\{x'_1, \dots, x'_{p'}\}$ から、ロバストな標準偏差や四分位法を用いて決定した技能評価のための標準偏差 $\hat{\sigma}_{\text{予備}}$ は以下の分散を推定したものである。

$$\hat{\sigma}_{\text{予備}} \sim \sqrt{\sigma_L^2 + \frac{\sigma_S^2}{m'} + \frac{\sigma_e^2}{n'm'}}$$

最低サンプル数と繰返し数

均質性試験の分散分析から求まるサンプル間の標準偏差 s_S と繰返しの標準偏差 s_W はそれぞれ、 $s_S \sim \sigma_S$ 、 $s_W \sim \sigma_e$ と考えられるから、予備試験の結果と合わせて、

$$\hat{\sigma}_L \sim \sqrt{\hat{\sigma}_{\text{予備}}^2 - \left(\frac{s_S^2}{m'} + \frac{s_W^2}{n'm'} \right)}$$

と、試験所間のかたより L_i の従う分布の標準偏差 σ_L を推定することができる。



σ_L に比べて $\sigma_S/\sqrt{m}$ と $\sigma_e/\sqrt{nm}$ を小さくしたい！！

最低サンプル数と繰返し数

例えば、
 $\sigma_S^2/m < 0.1 \times \sigma_L^2$ かつ $\sigma_e^2/nm < 0.1 \times \sigma_L^2$
 ならば、十分に条件を満たすだろう。

つまり、

$$\frac{s_S^2}{m} < 0.1 \times \hat{\sigma}_L^2 \quad \text{かつ} \quad \frac{s_W^2}{nm} < 0.1 \times \hat{\sigma}_L^2$$



$$m > 10 \cdot \frac{s_S^2}{\hat{\sigma}_L^2} \quad \text{かつ} \quad nm > 10 \cdot \frac{s_W^2}{\hat{\sigma}_L^2}$$

コメント

この条件は、繰返し回数のガイドラインとほぼ同じ条件となる。ただし、附属書Bに書かれたサンプル間のばらつきが無視できない時の繰返し回数のガイドラインとは違うので注意。(附属書Bの背景が把握できない。)

実際にはJISの規定などで、繰返し回数が定まっている場合もあると考えられる。2つの式には両方とも m が入っているので、繰返し回数を増やさなくてもサンプル数を増やせば、JISの規定通りの試験で繰返しのばらつきの問題を解決することができる。