

ISO 33405 (Reference materials – Approaches for characterization and assessment of homogeneity and stability) * の不確かさ評価の方法

産業技術総合研究所

工学計測標準研究部門 データサイエンス研究グループ

城野・高井

* ISO. ISO 33405:2024, Reference materials — Approaches for characterization and assessment of homogeneity and stability. Geneva: ISO, 2024.

ISO 33405 (Reference materials – Approaches for characterization and assessment of homogeneity and stability)

- 不確かさと小見出しに出てくるのは10章と付録Dのみ。
- 不確かさ評価のための規格というわけではない。

ISO 33405: 10 Evaluating measurement uncertainty

- $u_{\text{CRM}} = \sqrt{u_{\text{char}}^2 + u_{\text{hom}}^2 + u_{\text{sts}}^2 + u_{\text{lts}}^2}$
- u_{CRM} : 認証標準物質の値の標準不確かさ
- u_{char} : 値の特性づけに関する不確かさ
- u_{hom} : 不均質性の標準不確かさ
- u_{sts} : 定義なし。その代わりに u_{trn} が移送中の安定性に関する標準不確かさとされて定義されている。
- u_{lts} : 長期安定性に関する標準不確かさ

- u_{sts} と u_{trn} は書き間違いではないかと察する。
- 「値の特性づけに関する不確かさ (= 認証値の決定の不確かさ) 」と、「不均質性の標準不確かさ」、「移送中の安定性に関する」、「長期安定性の標準」と分けている。
- それぞれ、どのように算出するかについては、10章には詳しい記載がない！

青網掛けはISO 33405には直接記載のないコメント

ISO 33405: 10 Evaluating measurement uncertainty

- 不確かさの要因：
 - 測定の不確かさに関するすべての要因を考慮すること
 - 相関関係の情報について
- 包含係数：
 - 正規分布を仮定できる場合の包含係数
 - 有効自由度がいくらあればよいか。
 - 左右対称でない場合。
- 10章は、極めて、基本的なことが列挙されているに過ぎない。実際には、以下のように対応しており、7章以降は不確かさに関する記述もかなり多い。
 - 7章が不均質性の標準不確かさ
 - 8章が（移送中の不確かさを含む）不安定性の標準不確かさ
 - 9章が値の特性づけの標準不確かさ
- 不確かさの解説としては、10章9章、7章、8章の順が読みやすいので、その順番に解説する。

特性値の決定と不確かさ評価

- 以下のように細かく、トレーサビリティなどについての要求と、不確かさ評価のヒントが書いてある。
 - 一つの方法で一つの試験所が決める場合
 - (測定手順の有無の別に) 複数の方法で一つ以上のラボが決める場合
 - 純度の決定 / 順序尺度量 / 名義的性質 / 非認証の値
- 一つの方法で一つの試験所が決める場合はGUMに則って、 u_{char} を決めることになる。
- ただし、その際にはトレーサビリティ、ラボのバイアス、均質性や安定性との区分けについて、よく考慮する必要がある。
- 均質性との区分けについては後ほど説明。
- 複数の方法や複数ラボにより決める場合の詳細はAnnex Aにある。

特性値の決定と不確かさ評価

- 複数の値を y_i を用いて、平均値により、特性値を決める：

$$y_{\text{char}} = \frac{\sum_{i=1}^p y_i}{p} \quad (\text{A. 1})$$

- 標準偏差：

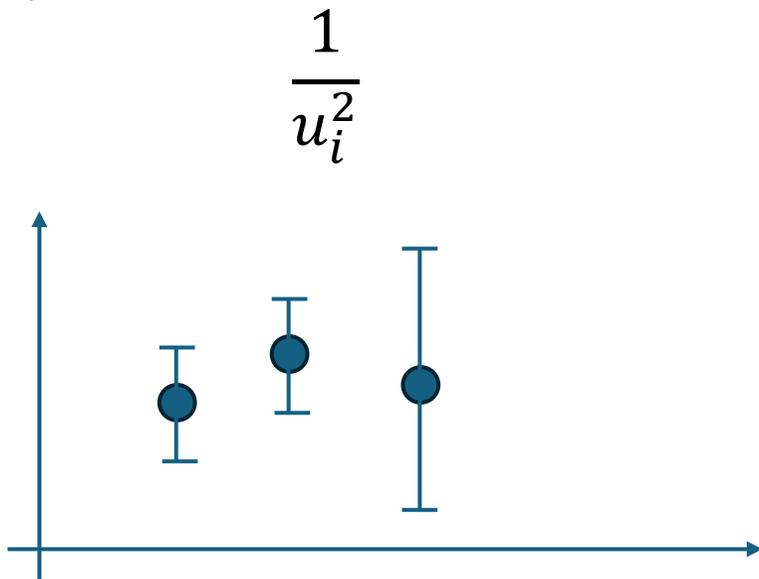
$$u_{\text{char}} = \frac{1}{\sqrt{p}} \sqrt{\frac{\sum_{i=1}^{p-1} (y_i - y_{\text{char}})^2}{p - 1}} \quad (\text{A. 4})$$

- 複数の方法を用いたり、複数のラボが参加した試験の結果を処理するための一つの方法。

- 不確かさの決定方法として(A.2)が使えるかどうかは要検討。
- この式には、直接にトレーサビリティソースを示唆する不確かさ要因がない。(複数のラボや複数の方法の違いの中心がゼロでない場合もある。)

特性値の決定と不確かさ評価

- 重み付き平均値平均についても記載あり。
- 重みは分散（=標準不確かさの2乗）の逆数とするのが、シンプルな場合



- この場合、 y_i が独立であれば、以下のようにするのが、統計学的にはよい方法：

$$u(y_{\text{char}}) = \sqrt{\frac{1}{\sum_{i=1}^p \frac{1}{u_i^2}}}$$

- 独立かどうかの検討が少々難しい。
- トレーサビリティソースを示唆する不確かさ要因はそれぞれ u_i の中に含むことができる。

均質性の評価：特性値の不確かさとの分離

- 特性値の決定の不確かさと、均質性の不確かさを分ける方法は、ISO 33405に明確には書いていない。
- 特性値の決定の実験と、均質性の不確かさの決定の実験を別々にやったとしても、特性値の決定の不確かさの中に均質性の不確かさが入っている。
- その点について、少し説明を加える。

ボトル	繰返し	測定値
1	1	122.44
	2	122.96
	3	123.45
2	1	117.60
	2	119.66
	3	118.96
3	1	126.84
	2	124.72
	3	123.14

※数値はISO 33405のAnnex C.1のTable C.1の一部（一部のみを選んでいるので、結果は異なる）

均質性の評価：特性値の不確かさとの分離

• 以下の手順で計算できますか？

- ① ボトルごとの平均値をとる。
- ② その平均値を繰り返しのデータのように扱って、標準偏差を計算する。
- ③ これを不均質性の標準不確かさとする。



NG

- 均質性の不確かさは、繰り返しの不確かさと別。
- 左のやり方では、単なる繰り返しのばらつきが大きくても、不均質性の標準不確かさも大きくなってしまう。
- 繰返しと不均質性の標準偏差は分けたい → **分散分析**という統計解析手法を使う。

均質性の評価：特性値の不確かさとの分離

- 分散分析した結果。この表では、ボトルの「分散」が不均質性の標準不確かさの2乗になっていれば分かりやすいのだが、そうではない。

ばらつき の要因	平方和	自由度	分散	F値
ボトル	59.4722	2	29.7361 = M_{between}	18.58855
繰返し	9.5982	6	1.5997 = M_{within}	
全体	69.0704			

F検定という統計手法のための指標、必ずしも適用しなくてはいけ
ないわけではない。

均質性の評価：特性値の不確かさとの分離

- ISO 33405 7.7.3
- Between-unit variance (n_0 : 繰り返し数) :
$$s_{bu}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) \quad (4)$$

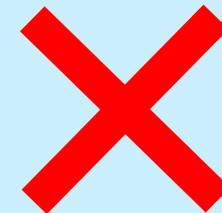
- 計算の結果

$$s_{bu} = 3.06$$

- u_{char} と u_{hom} との関係は、どうなるか？

- u_{char}

- 均質性の評価とは、別の実験で決める場合も多い。
- 例えば1つのボトルを選び、それについてのみ、精密に測定を実施する。
- そうであれば、不均質性の不確かさに対する考慮は不要か？



必要

均質性の評価：特性値の不確かさとの分離

- 1つのボトルを用いて測定していても、不均質性の不確かさに対する考慮は必要。
- 具体的には、標準不確かさとして、 s_{bu} の不確かさを考えなくてはならない。
- もちろん、それとは別に繰り返しや校正の不確かさなどの不確かさに関する考慮は必要。

- さらに一般化： m 個のボトルを用いる。それぞれ、 k 回測定を実施する。特性値決定の試験における1回の繰り返しのばらつきの標準不確かさを u_{rep} とする。繰り返し・均質性以外の要因の標準不確かさを u_{other} として、以下のように見積もる：

$$u_{char} = \sqrt{u_{other}^2 + \frac{u_{rep}^2}{k} + \frac{s_{bu}^2}{m}}$$

均質性の評価：特性値の不確かさとの分離

- u_{hom}
- もし、ボトル間よりも複雑なばらつき
の構造がなければ、以下のように定め
られる。

$$u_{\text{hom}} > s_{\text{bu}}$$

例えば、測定日、測定者、ロット、製
造工場が複数あると複雑になる。

- 特に、ボトル（などのユニット内）で
の不均一な構造が無視できない場合、
その標準不確かさを s_{wu} として：

$$u_{\text{hom}} = \sqrt{s_{\text{bu}}^2 + s_{\text{wu}}^2} \quad (9)$$

- 実際には、ユニット内の不均質性は採
取の量や、測定原理に依存するもの
と思われるため、画一的な表現は難し
い。
- 採取方法が決まっている場合は、繰り
返しのばらつきの中に含まれるはず。
繰り返しのばらつきの標準偏差の評価
は以下の通り：

$$s_{\text{rep}} = \sqrt{M_{\text{within}}}$$

- s_{wu} はこれより小さいのだが、その厳
密な算出は、困難。

均質性の評価：複雑な評価方法

- ISO 33405には、上位のばらつきの構造がある場合などについて、書いている。
 - 測定日や測定者、その他の測定パラメータによる繰り返し以外のランダムな効果が入る場合
 - 製造ロット、製造工場、2段の小分けなど、標準物質の性状に影響するパラメータがある場合
 - もっと複雑な場合、また、異なる解析方法（制限つき最尤法REML及びベイズ統計）について
- 解析方法の記載は正直に言って不十分。必要な場合は統計学に基づいて、解析を進めていく。

均質性の評価：負の分散の対応

- 分散分析の結果、不均質性の標準不確かさを求めるための分散（＝標準不確かさの2乗）が負の値になり、計算が不能になることがある。

ボトル	繰返し	測定値
1	1	121.30
	2	128.74
	3	119.91
2	1	120.87
	2	121.32
	3	119.24
3	1	122.44
	2	122.96
	3	123.45

※数値はISO 33405のAnnex C.1のTable C.1の一部（一部のみを選んでいるので、結果は異なる）

均質性の評価：負の分散の対応

- 負の値になり、平方根が取れない。

$$\frac{M_{\text{between}} - M_{\text{within}}}{n_0} = -0.27989$$

ばらつき の要因	平方和	自由度	分散	F値
ボトル	14.31742	2	7.158711 = M_{between}	0.895019
繰返し	47.99033	6	7.998389 = M_{within}	
全体	62.30776	8		

均質性の評価：負の分散の対応

- その場合、先に話した通り、ユニット間の分散をゼロとするのが、基本的な計算方法。

$$s_{bu}^2 = \max\left(\frac{M_{\text{between}} - M_{\text{within}}}{n_0}, 0\right) = 0$$

- Annex C.2の例では特別な方法を用いてるが、個人的には統計学的な根拠が薄弱で使用はお勧めできない。一方で、化学物質の標準物質の不確かさ評価ではしばしば使われている。
- 7.7.7に示された解析方法である、REMLやベイズ統計なら、統計学としては自然な枠組みで、負の分散の処理ができる。

均質性の評価：その他

- 最小サンプル数
 - 標準物質を頒布するときの均質性評価に用いる最小サンプル数やサンプリング方法などについても記載あり。
- 外れ値の対応
 - まず原因の検討。
 - 検討の結果、測定値やあるユニットについての結果を除くこともあり。
 - 検討の結果、再度測定することもあり。

安定性の不確かさ評価：導入

認証標準物質の不確かさモデル：

$$u_{\text{CRM}} = \sqrt{u_{\text{char}}^2 + u_{\text{hom}}^2 + u_{\text{sts}}^2 + u_{\text{lts}}^2}$$

この中で安定性に関する不確かさ成分は2つある。

u_{sts} ：輸送中の安定性に関する標準不確かさ

u_{lts} ：長期安定性に関する標準不確かさ

先述の通り u_{sts} と u_{trn} は書き間違いではないかと思われる。

輸送安定性と長期安定性 (8.2.2, 8.3.4)

安定性には2つの種類がある (8.2.2)。

長期安定性：規定された保管条件下で、有効期間を通じて物質の特性値が維持されるかに関する安定性。評価は配付前に行う。通常12か月以上の試験が計画される。

輸送安定性：使用者の構内への輸送に起因する追加的な影響に関する安定性。「短期安定性」とも呼ばれる。試験期間は妥当な輸送期間を反映する（例：国際輸送が3週間なら、3～4週間の試験）。

不確かさモデルの中で、 u_{trn} は輸送安定性に、 u_{lts} は長期安定性に対応する。

輸送安定性に問題がある場合、 u_{trn} として認証値の不確かさに追加するか、あるいは輸送条件を厳密に管理することで影響を無視できるようにするかの選択になる。

通常は輸送条件が十分に管理されている前提としており、輸送安定性試験は著しい変化がないことの確認に主眼が置かれている。

輸送安定性試験と不確かさ (8.3.4.1)

輸送安定性試験は、予想される輸送条件下での安定性を評価する試験である。

試験条件の設定にあたっては、物質の輸送に許容される最大期間において、国際輸送時に十分起こりうる極端な温度を含めることが望ましい。

類似物質の安定性について知見があり、管理された輸送条件が必要と分かっている場合（例：ドライアイスでの輸送）、試験は意図された輸送条件に限定してもよい。

輸送安定性試験は、定量的な変化率の精密な推定よりも、著しい変化がないことの確認に重点がある。

したがって、 u_{trn} の不確かさ評価は、変化が検出されなければゼロとされることがある。検出された場合にはその変化量を u_{trn} として認証値の不確かさに含める。

長期安定性試験と不確かさ (8.3.4.2)

長期安定性試験は、製品の寿命にわたって規定された保管条件下における安定性を評価するために実施される。

実時間の長期試験は通常12か月以上続くが、加速試験を併用することで期間を短縮することも可能である。認証書の有効期間は、安定性試験の結果とモニタリング計画によって保証される。

市場投入までの期間が重要な場合、長期安定性試験を12か月未満に限定し、頻繁なモニタリングによって認証前の限られたデータを補完することが認められている。

長期安定性からの不確かさ u_{ITS} は、安定性試験の結果から推定される。試験で有意な傾向が検出されなかった場合でも、検出限界を考慮した不確かさ成分を計上する場合がある（後述の式(10)）。

加速試験を用いる場合、モデル外挿の不確かさが追加で生じるため、不確かさの見積もりはより複雑になる。

安定性の不確かさをゼロにできる場合 (8.7.1)

安定性の不確かさ成分をゼロにできる条件：

- ① 妥当な技術的理由があること
- ② 有効期間中の変化が $u_{\text{CRM}}/3$ 未満であると示されること
- ③ 上記が実験と観測によって裏付けられていること

安定であることが明白に見えるだけでは不十分で、定量的な裏付けが必要。

その場合でも、悪影響が及ぶ前に物質を使用すること、および予期しない変化を迅速に検出可能なモニタリングを採用することが必要である。

すなわち、不確かさが0であってもモニタリングは省略できない。

安定性の統計的評価：線形回帰（8.5.2）

不確かさ評価の前提となる統計的評価の手順：

- ① 線形回帰モデル $y = b_0 + b_1 \cdot t$ を適用する
- ② 残差を検査する（非線形性、外れ値、等分散性）
- ③ 傾き b_1 の標準誤差 $s(b_1)$ を算出する
- ④ 傾きのゼロからの有意性を t 検定で判定する

各測定 of ランダム誤差が独立かつ同一分布の場合に適用される。中間精度が繰返し性より大きい場合は独立性の仮定が崩れるため、8.5.5 に従った処理が必要になる。

線形回帰は安定性評価の最も基本的なツールであり、この結果（特に β_1 と $s(\beta_1)$ ）が不確かさ評価に直接使われる。

均質性評価が分散分析をベースとするのに対し、安定性評価は回帰分析がベースである。ただし、ユニット間不均質性などの追加的なばらつき要因がある場合は、混合効果モデル等のより複雑な処理が必要になる（8.5.5）。

傾向の有意性と不確かさへの対処 (8.5.2.6, 8.6)

統計的に有意：t検定で傾き $\neq 0$ と判定される場合。

技術的に顕著：有効期間にわたる予測劣化量が、対象特性値の標準不確かさと比較して無視できない場合。

顕著な傾向がある場合の不確かさへの対処法：

- a) その特性値を認証しない
- b) 認証値の有効期間を短縮する
- c) 劣化の不確かさを認証値の不確かさに追加する
- d) 認証値と不確かさを時間の関数として提供する
- e) b)～d) の組合せ

c)は一般的な方法で、補正されていないかたよりを不確かさに含めることと本質的に同等。

d)は放射性崩壊のようにモデルが確立されている場合に適用される（例： ^{242}Pu 濃度の認証で、崩壊率と半減期の不確かさを提供する方法）。

有意な傾向がない場合の不確かさ推定 (8.7.3)

ケース1：安定性の技術的根拠がある場合

→ 安定性の不確かさは0としてよい。ただしモニタリングは必要。

ケース2：技術的根拠が不十分な場合

→ 予測不確かさを推定し、認証値の不確かさに含める。

$$\text{式(10)} : u_{\text{ITS}} = s(b_1) \cdot (t_{\text{m1}} + t_{\text{cert}})$$

$s(b_1)$: 傾きの標準誤差

t_{m1} : 値付けから初回モニタリングまでの期間

t_{cert} : 認証書の有効期間

式(10)の考え方：有意な傾向は検出されなかったが、 $s(b_1)$ 程度の大きさの傾きが存在する可能性は排除できない。その傾きが $(t_{\text{m1}} + t_{\text{cert}})$ の期間継続したときの変化量を不確かさとする。

モニタリング間隔 t_{m1} が長いほど、この不確かさは大きくなる。頻繁なモニタリングは不確かさの低減にも寄与する。

有意な傾向がある場合の不確かさ評価 (8.7.4)

アプローチ1：時間依存の認証値を提供する

認証値は傾向の最良推定を反映した関数とし、不確かさもばらつき要因 (8.7.2) を考慮した関数とする。

アプローチ2：時間依存でない認証値を提供する

有効期間($t_{m1} + t_{cert}$)にわたる劣化量を推定し、一様分布等を仮定して標準不確かさに変換する。これを予測の不確かさと合成する。劣化が一方向のみの場合、不確かさは非対称になることがある。

不確かさのばらつき要因 (8.7.2) としては、a) 測定 of 繰返し性、b) 測定シリーズごとの変動、c) ユニット間の不均質性がある。これらが適切に回帰モデルに当てはめられていれば、モデル係数の標準誤差に反映される。

アプローチ1は、放射性崩壊のように劣化モデルが十分に確立され、使用者が時点ごとの値を計算できる場合に適している。

アプローチ2の場合、劣化量そのものとその劣化量の予測不確かさの両方を合成する必要がある。前者は一様分布で標準不確かさに変換する。

安定性の不確かさ評価：まとめ

安定性不確かさ u_{ITS} , u_{STS} の評価フロー：

① 安定性を評価する（試験 or 過去知見）

輸送安定性 → u_{STS} / 長期安定性 → u_{ITS}

② 有意な傾向があるか判定する（t 検定）

傾向なし + 技術的根拠あり → 不確かさは0

傾向なし + 根拠不十分 → 式(10)： $u_{\text{ITS}} = s(b_1) \cdot (t_{m1} + t_{\text{cert}})$

で推定

傾向あり → 8.6の対処法を選択し8.7.4で評価

③ モニタリングによる継続的な確認

不安定性判定：

$$|x_{\text{CRM}} - x_{\text{mon}}| \leq k \cdot \sqrt{u_{\text{CRM}}^2 + u_{\text{mon}}^2}$$

④ モニタリング結果に基づき不確かさを更新する