実験化学者のための

誤差入門　第２版

もうボスに怒られない！

想定読者

- 化学をはじめとする実験系の大学生、大学院生の方
- 誤差や統計の言葉は聞いたことがあるが、雰囲気を知らない方

概要

- 真の値はわからない。実験値は不確かだ。
- 複数回まったく同じ測定をして、ばらつきの程度を把握する。
- ばらつきのあるデータを図示できるようになる。
- 不確かな実験値を使って、差異を主張したり、計算したりできる。
- 自分で計測したデータで検出下限を見積もれるようになる。

誤差が気になる———真の値はわらかない。

大学に入り実験をするようになって戸惑ったことがある。それまでは、ペーパーテストの問題を解いている間は、計算で使う確実な値が与えられていた。1 g の物体は間違いなく 1 g だった。ところが、ためしに 1 円玉の重量を電子天秤で計測してみると、以下のような結果になった。

1 回目	2 回目	3 回目	4 回目	5 回目
0.980 g	0.944 g	0.977 g	1.001 g	0.948 g

平均値は 0.970 g。答えが 1 g と知っていれば、四捨五入して 1 g だといえるが、答えを知らないものを計測した結果だったら？その 1 円玉は、本当に 1 g ちょうどではなく、0.970 g なのではないか？そもそも、何桁まで信頼できる値なのだろう。1 g、1.0 g、0.97 g、0.970 g。どこで四捨五入すればいいのか、四捨五入するのが正しいことなのか、どうして分かる？

実のところ、真の値はわからない。測定には不確かさがまとわりつく。しかし、不確かさから逃れられないといっても、不確かさを手懐ける方法がある。それを紹介するのが本稿の役割である。

誤差の類型———なんで誤差なんか生じるんだ！

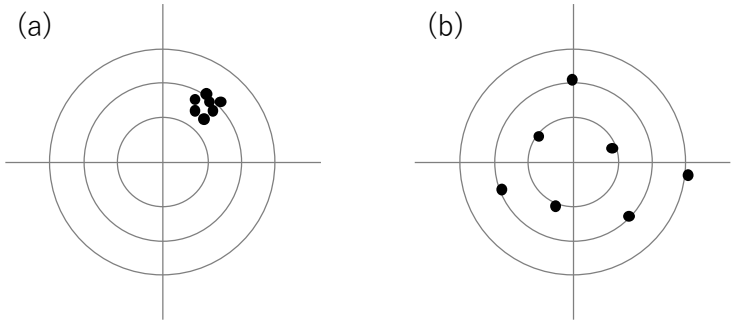
まず、真の値なる答えと、計測して得られた測定値があると考えよう。真の値はただ一つの答えであるが、測定値は前述の例のようにばらつく。真の値と測定値との差を、誤差と呼ぶ。誤差の由来は大きく 3 群に分類できる。

分類	原因の例
過失	装置の故障、設定ミス、作業者の操作ミス、勘違い
系統誤差	較正ずれ、サンプルへのごみの付着
偶然誤差	ノイズ、避けられないばらつき

過失を誤差の原因に挙げるのは馬鹿らしい気もするが、どうしたって人は間違える。**系統誤差**は何度測定しても同じように影響してくる誤差である。0.1 g のごみが付着した 1 円玉の重量を測定すれば、1.1 g になっても不思議でない。過失や系統誤差を極力排除したとしても、**偶然誤差**のせいで測定値がばらつくのはやむを得ない。偶然誤差から逃れるには、十分にばらつきの小さい測定手法を用いるか、測定回数を増やして統計的に処理するしかない。

精度と確度———何となく「精度が良い」とか言ってない？

系統誤差と偶然誤差は、誤差の出方が違う。下図は、同心円の中心が真の値、マーカーが測定値を表したものである。下図(a)は測定値がほぼ同じ位置に集まっているが、真の値からの差が大きい。一方で、下図(b)は測定値がばらついているものの、測定値の平均をとると真の値に近そうだ。



簡単に誤差といっても、(a)と(b)の場合では様相が異なる。そこで、測定値のばらつきが小さいことを**精度 precision** が高い、真の値からの距離が近いことを**確度 accuracy** が高いと表現することが多い。

確度を高める———最後に較正したのはいつかな。

化学実験で確度が悪いときは、装置の**較正 calibration** ができていない場合が多い。電子天秤を例にとると、サンプルステージに変な力がかかっていない状態でゼロ点を補正するのはもちろん、定期的にメーカー等の点検

を受けることを勧める。電子天秤の示す値が何かおかしいと思ったら、先生や上司に提言してみよう。ところで、費用もかかれば手間もかかる作業をしてもらうには、説得できるデータが必要になりそうだ。どうやって較正がズレているデータを取得すればよいだろうか。

較正のためには、答えがわかっている標準物質を用意して、これを測定する。米国では NIST、日本では産総研が標準物質を管理する親玉である。ここではそんな大仰なものに関わりたくないので、身近なもので正確な重量が分かっている物質を使おう。それは、水だ。水の密度は温度で変わるが、幅広い温度に対して水の密度が調べられている。その値は図書館にある理科年表や化学便覧に掲載されているため、信頼性の高い情報を容易に入手できる。ホールピペットのような定容性の良い器具で、室温と平衡になった水を量り取り、重量を計測する。これを繰り返した平均値が文献値と大きく異なるならば、較正がずれている可能性が高い。

逆に、電子天秤がよく較正されている条件では、この手法はマイクロピペットの較正に使うことができる。たとえば、マイクロピペットを容量 1000 μL に設定して、慎重に分取した水の重量を計測することで、水の密度の文献値から実際には何 μL を分取しているのかを求めることができる。

ほぼすべての分析装置には適切な較正方法が定められている。較正方法については、装置のマニュアルを読むか、製造・販売元に問い合わせる。

精度を高める———n 増ししといて！

測定手法には精度の限界がある。市販の精密電子天秤では 1/10〜1/100 mg まで表示があるが、一番下の桁の数値はたいていあてにならない。前項で述べたマイクロピペットで 1000 μL の水の重量を計測するのを数十回繰り返しても、まったく同じ値が得られることは稀だろう。しかし、恐れることはない。ほとんどの場合、偶然誤差は正規分布に従う。1 回 1 回の測定値はばらついたとしても、統計的に実験値の不確かさを見積もることができる。そして、単に測定回数、実験回数を増やすだけで、不確かさを

小さくできる。

実験値は測定値の単純平均を用いる。ばらつきの程度は**標準偏差**であらわされる。標準偏差の説明は統計の教科書に必ずあるので参照してほしい。実験値の不確かさは、測定回数（標本サイズ）n と標準偏差 SD を使って計算される**標準誤差** SE = SD / sqrt(n) で求められる。n が分母にあることから、単に測定回数を増やすだけで不確かさを小さくできる。測定回数を増やす慣用表現として「n 増し」が使われることがある。冒頭で示した 1 円玉の重量を測定したデータで計算した結果を以下に示す。

	値	Excel で求めるときの関数
平均	0.970	=Average(データ)
標準偏差	0.024	=STDEV(データ)
標準誤差	0.011	=STDEV(データ)/sqrt(count(データ))

表に示した計算結果から、1 円玉の重量の実験値 0.970、不確かさ 0.011 で 0.970±0.011 g と書けるが、**普通はそう書かない**。不確かさが 0.011 g であるときに、0.001 g の位の値に意味がない。そこで、不確かさを有効数字 1 桁に丸めて、実験値 0.97±0.01 g と表記する。0.97±0.01 g とまったく同じ意味で、0.97 (1) g という表記の仕方もある。

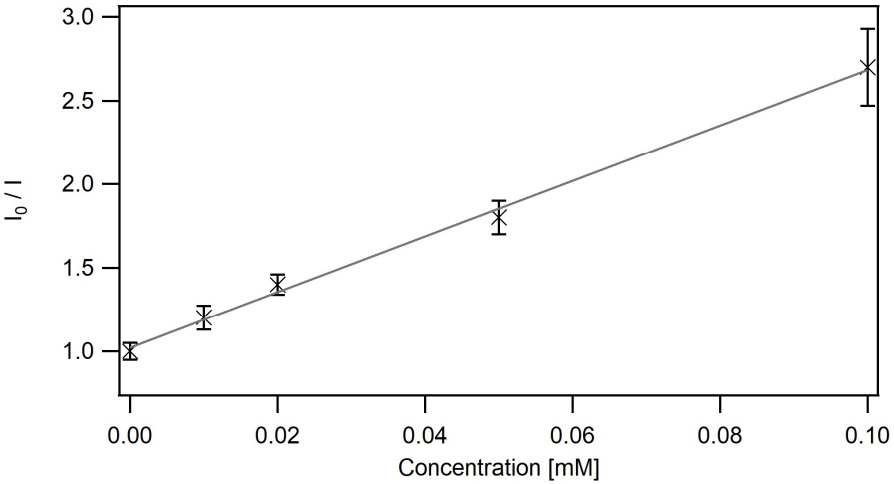
なお、不確かさの有効数字は必ずしも 1 桁にしなくてよい。たとえば、プロトンの大きさを調べた文献［*Science*, **2013**, *339*, 417］では、不確かさを有効数字 2~3 桁で示している。普通の化学実験ではそこまで精密な測定はしないので、有効数字 1 桁で十分だろう。

ところで、実験値が a±b という形で数値を報告されたとしても、a と b が何を意味しているのかは自明でない。たいてい a は単純平均だが、b は標準偏差なのか、標準誤差（不確かさ）なのか、あるいはその定数倍なのか、報告者が明記するべきである。

エラーバー———俺たちは雰囲気でグラフを描いている

ばらつきの大きさがわかっている測定値を使ってグラフを描く際には**エラーバー**を用いることが多い。散布図や棒グラフの形で平均値をプロットして、平均値から縦軸方向または横軸方向にヒゲを伸ばす。ばらつきの大きさは標準偏差や標準誤差を用いることが多く、95%信頼区間（統計学の教科書を参照）を用いることもある。

ところで、下の図のエラーバーは標準偏差だろうか、標準誤差だろうか。あるいはそれ以外の何かを示しているのだろうか。著者による説明がなければ、何を指しているのかわからない。グラフの余白でもレポートの実験項でもいいから、「エラーバーは標準偏差」と書くようにしたい。



エラーバーが描かれていると嬉しいこととしては、プロット間の差異が有意差（後述）なのかを大まかに見て取ることができる。

ばらつきのある値を図示する方法としては、ほかに**箱ひげ図（box plot）**がある。箱ひげ図では 4 分位点ごとに線を入れるため分布がわかりやすい。箱ひげ図もひげを最大値最小値で描くのか、95%位点で描くのか、あるいは別の方法で外れ値を除くのかといった書き方の自由度があるため、グラフを描く際には条件を記載してほしい。

確かに差があることを示す———それって有意差？

2 つの試料 A と B を同じ測定にかけて得られた結果から、A よりも B の方が良いだとか、A も B も変わらないだとか主張することがよくある。実験値が出てくるところまでは、同じ手順を踏めば誰でも同じ結果を得ることができるはずである。しかし、A と B が違うだとか同じだとかいう判断をするのは、解釈する人のさじ加減ではないか。

最終的な価値判断は置いておいて、2 つの実験値が異なるか否かは統計検定を行うことで誰でも同じ結果を得られる。たとえば、2 標本（2 試料について同じ条件で測定したデータセット）の平均値が異なることを示すことを考えよう。たいていの場合、測定値のばらつきが正規分布に従うと仮定できるので、**t 検定**を行えばよい。統計検定には特殊なツールは必要ない。Excel でも実施できる。具体的な手順については「Excel t 検定」などで検索すればよい。あるいは、「統計学入門」のような名前の書籍を読もう。

統計検定に関連して**有意差**という言葉が使われる。なんとも雰囲気の良い言葉だが、どのような意味を持つのだろうか。t 検定から得られる値に p 値がある。前述の例でいうと、標本 A と標本 B の平均値が偶然一致する確率が p 値である。よく使われるのが $p<0.05$ であり、このとき平均値が偶然一致する確率が 5%（$p=0.05$）なら十分に低い確率だから、標本 A と標本 B の平均値は異なると主張するような使われ方をする。このような内容が<u>有意水準 5%で標本 A と B の平均値に</u>**有意差あり**に込められている。ちなみに、下線部は省略されることが多い。有意水準は 5%でなく 1%などほかの値でもよいが、伝統的に 5%が用いられることが多い。また、t 検定は平均値の検定であるから下線部後半も省略される。省略されるからといって、有意差が統計検定の結果を述べる言葉であることを忘れてはいけない。1 回の測定だけでばらつきを評価せずに有意差を主張するならば、それは単にそう思っているだけである。

統計的に有意差があったとしても、それが何を意味するのかを考えるのは研究者の役割である。以下に有意差に意味がない 2 つの例を挙げよう。

1. 標本サイズが大きすぎないか

   標準誤差の紹介で述べたように、標本サイズ（測定回数）を大きくすれば、不確かさは小さくなる。t 検定でも、標本サイズを大きくすれば平均値の差が小さくとも有意差と判定される。いま気にしているのは、3 回の測定で有意差にならなかったからといって、1000 回測定して有意差にすることに意味のある実験なのか？※意味のある実験もある。

2. 研究の目的から逸脱していないか

   学術研究では機構解明のため差が言えれば十分であったとしても、新しい計測技術を構築する立場からは不十分であることがある。あらかじめ差があるとわかっている良品 A と不良品 B を新技術で測定して得られた指標をについて、標本サイズ 100 で有意差が得られたとする。新技術は完成したといえるか？もし新技術の目標が 100 秒以内に差を判定することで、1 測定に 60 秒かかっていたら、判定には 6000 秒かかることになり、新技術の性能が不足しているとわかる。また、なんとなく有意差が得られたと述べたが、有意水準 5%で有意差があったとしても、偶然 5%の確率で見逃してしまうことになる。見逃しがどの程度許されるのかは、新技術で何をしたいかによって決まる。

不確かさのある数値同士の計算———誤差は伝播する。

実験値が不確かさを持つのだから、実験値を用いた計算結果も不確かさを持つ。計算結果にどのような形で不確かさが伝わるのかを**誤差の伝播測**と呼ぶ。伝播測は実験値を用いて計算する関数が与えられれば自分で計算できるが、面倒なので四則演算くらいは伝播測を先に計算しておくとラクだ。ありがたいことに「誤差の伝播測」で検索すると、先人が計算結果を示してくれている。代表的なものとしては、大阪大学 石島研究室（http://www.fbs.osaka-u.ac.jp/labs/ishijima/）の理論・計算ページがわかりやすい。

誤差の伝播測を使えば、10±3 が 1.0±0.2 の何倍かといった計算ができ

るようになる。計算結果は 10±4 倍である。

定量下限と検出限界———ノイズじゃないの？

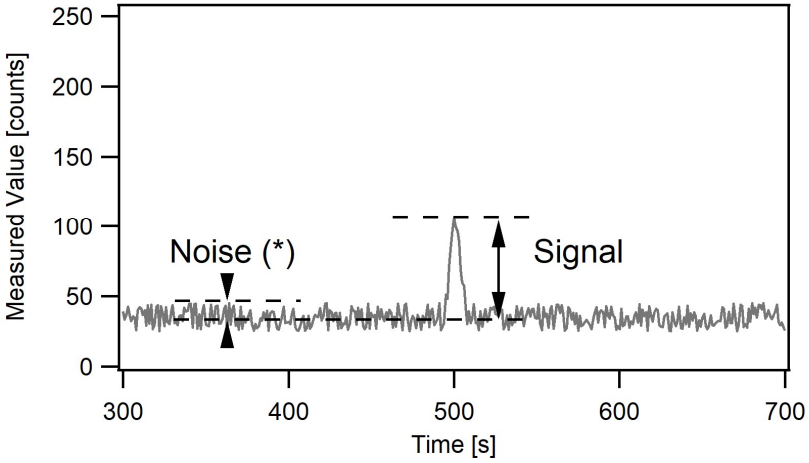
ここまで読み進めてきた読者は測定とばらつきが不可分なものであることを承知しているはずだ。いくら細かく測定したとしても測定値が必ずばらつきを伴うとしたら、私たちはどうやって「検出できた」と自信を持てるのだろうか。

ある測定系において、ある値以上の信号が測定されたときに「検出できた」といえる臨界点を**検出限界**と呼び、「測定値 A は測定値 B の X 倍」といった形で測定値を定量的に扱える下限を**定量下限**と呼ぶ。実際のところ、「検出限界」や「定量下限」をどうやって算出するかは流儀がいくつかある。「精度」や「確度」といった言葉が何となく使われていることを思い出せば、もっともな話である。

慣習的には、ブランク測定の標準偏差 σ を使って、**検出限界 3 σ、定量下限 10 σ** とすることが多い。思考停止的ではあるが、とりあえずこの定義で計算しておけばボスには怒られないだろう。重要なのは、その値がどうやって算出されたものか把握しておくこと、その算出法が研究の目的に適したものであるか判断することである。たとえば、測定手法や装置間の性能を比較するとき、異なる定義のもとで算出された検出限界を比較しても意味がないことは明らかである。いくつもの定義をここに列挙することはしないので、より正確な理解を求める場合は分析化学会誌の記事［*ぶんせき*, **2010**, *5*, 216］を参照してほしい。本稿で用いた検出限界 3 σ、定量下限 10 σ は用いられることが多いが好ましくないと指摘されている。

検出限界の求め方———信号、ノイズ、具体的にどの値？

実際にありそうな測定データから、言葉の定義を覚えるとわかった気になるが、それだけでは自分の研究に使えない。次の図のように、横軸が Time、縦軸が何らかの Measured Value である場合を考えよう。



縦軸は検出器が出力する電圧だったり電流だったりする。測定した全範囲で一定のオフセットがかかっている。このような現象は検出器の暗電流などによって生じるので、差し引いてしまえばよい。Time が 500 s あたりに信号が検出されていることがわかる。信号の影響がない領域、すなわち検出されたピークから十分に離れた Time の領域で差し引きするオフセット（平均値, 35）やノイズの大きさ（標準偏差 σ, 5.7）を計算する。検出限界を 3 σ とすれば、ノイズの平均値に 3 σ を足した値を超える Measured Value が得られたとき「検出できた」といえる。縦軸は検量線によって重量［g］や濃度［mol/L］といった単位系に変換できるので、測定したい単位系（重量［g］や濃度［mol/L］）でこの測定の検出下限を表示できる。単位系の変換には誤差の伝播に注意しよう。

次に、SN 比を求めてみよう。検出されたピークの縦軸の値（106）をそのまま使ってはいけない。バックグラウンドの信号（先に求めたオフセット, 35）を差し引いた値（75）が Signal の値である。したがって、SN 比は 75/5.7 = 13 となる。

注意したいのは、**SN 比の定義も複数存在する**ことである。本稿で用いたのはクロマトグラムなど化学分野で使われる計算方法である（Agilent Technologies のカタログを参照）。SN 比を Wikipedia（SN 比）や JIS

（K0211）で調べると、標準偏差ではなく標準偏差の 2 乗を使って SN 比を定義している（$\sigma_{Signal}^2/\sigma_{Noise}^2$）。また、測定値が電子データではなくロール紙に記録されていた時代はノイズの大きさを標準偏差ではなくノイズ領域の最大値と最小値の差分として定義していた。SN 比だとか検出限界のような一見どこでも通じそうな言葉で表されるものが、分野でも時代でも変化している。論文や報告書のように後に残る文書には、せめて自分がどのようにそれらの値を定義しているか記載しておくべきであろう。

むすび

とりあえず、自分の実験で得られた測定値をどのように書けばいいのかがわかった。有意差という言葉と統計検定の関係を知り、不確かさを持つ数値の計算もできるようになった。あなたが化学系の学生なら、この時点で同級生 10 人に 1 人のレベルで誤差を適切に扱えると思ってよい。

実験値の不確かさを正しく扱えるということは、どれだけ信頼できる実験値を自分が出しているのかを理解しているということだ。検出限界だとか SN 比だとか、それっぽい言葉を知っていると何となくわかった気になる。しかし、科学の言葉であれば定義が存在するはずだし、定量的に扱えてしかるべき言葉である。それっぽい言葉を知っているだけでなく、現実に即して概念を運用できるか否か、それが素人と専門家を隔てる高い壁のひとつである。本稿を執筆する過程で、著者は素人から専門家への脱皮を目指している。

専門家の著作には遠く及びはしないものの、本稿では読者と同じく未知の視点から誤差の取り扱いを試みた。近接分野で成長の途上にあるであろう読者諸氏にとって、経験を外れた極端な系や、新規な技術での実験結果を考察する入り口として、本稿が役に立てば幸いである。

本稿の内容に関するご意見、疑問、誤りの指摘などは twitter @bunkaiken にご連絡ください。

2019/8/11　　分光解析研究会

ALISS.JP