実践統計解析【第11回】

誤差の伝播

川瀬　雅也[1]*・松田　史生[2]

Aさん：先輩，1年生の学生実験でTAをしたら，溶液の作り方で難しい質問をされて答えられなかったんです．

B君：それは，次回にばっちり応えて後輩にかっこいいとこ見せないとね．

Aさん：0.1 mMの塩化ナトリウム水溶液を100 mL作る実験だったんです．手順書には10 mMの原液の溶液を100 mL作ってから，100倍に希釈すると書いてあります．

B君：普通の手順だね．

Aさん：ところが，デキル感じの学生が，天秤で5.8 mgの塩化ナトリウムを量り取り1 Lの溶液を作っちゃなぜいけないのか？と言い出して．

B君：え，5.8 mg量れるの？

Aさん：なぜか微量天秤があるんです．希釈の操作がない分，誤差が小さくなるんじゃないかって．これからX教授に教わりに行くので，つきあってくれますよね？

B君：いま，実験中なんだけど……．データがばらついて困ってるんだよね．

Aさん：先輩！かわいい後輩にはかっこいいところ見せなきゃいけないんでしょ？

皆さんも，低濃度の溶液を作るとき，まず，濃度の高い溶液を作ってから希釈されていると思う．このとき誤差について考えたことはあるだろうか．

誤差のおさらい

X教授：今日は，どうしたのかな？

Aさん：実は，こういう次第なんです．

X教授：なるほど，質問した学生さんは，誤差がどんなものか，1年生だからまだ，分かっていないとは思うが，よく考えているのは確かだね．君たちは，どう思う？

Aさん：このやり方を教わって，ずっと実行してきたので，何の疑いも持っていませんでした．

X教授：まず，誤差について復習をしてみようか．誤差とは，たとえば，ある測定を行ったとき，真の値と測定値との差であることは分かっているね．

Aさん：はい．

X教授：では，操作を繰り返したときの誤差はどうなるかな．

Aさん，B君：分かりません．

X教授：これが，今日の本題と関係する「誤差の伝播」という問題なんだ．

誤差の伝播[1)]

X教授：まず，少し準備が必要になるんだ．君たち，偏微分は知っているかな．

Aさん：微分ですか？

B君：まさか，偏微分って，熱力学で出てくる奴じゃ．あれ，見るだけで……

X教授：何を言っているんだ．そんなに難しいことではない．単に，長い説明文の代わりに，簡単な数式で説明しているだけじゃないか．

Aさん：数学は苦手なんで．

X教授：そんなことを言うと数学の先生が怒ると思うな．数学と数式は違うからね．大事なことは，数式は自然科学の共通言語で，避けてはいけないということと，数式を理解するための最小限の数学は勉強してほしいということだね．

Aさん：生命科学でも，当然，必要ですよね．数学と無縁になりたかったけど．

X教授：まあ，気楽に．では，偏微分を説明しよう．B君が言ったように，熱力学で必ず習うと思うが，復習の意味も込めて，まず，次の，2変数関数を考えよう．

$$f(x, y) = ax + by$$

この関数の偏微分を行ってみよう．xについては

$$\frac{\partial f}{\partial x} = a$$

同様に，yについては

$$\frac{\partial f}{\partial y} = b$$

＊著者紹介　[1]長浜バイオ大学（教授）　E-mail: m_kawase@nagahama-i-bio.ac.jp
[2]大阪大学大学院情報科学研究科（准教授）

となるんだ．つまり，微分する変数を含まない項は定数と見なせばいいわけだ．では，

$$f(x,y)=ax^2+by^2+cxy$$

ではどうなるかな．

Aさん：$\dfrac{\partial f}{\partial x}=2ax+cy$

$$\frac{\partial f}{\partial y}=2by+cx$$

です．

X教授：その通り．簡単だろう．

Aさん，B君：はい．学部の時は熱力学とごっちゃで難しかったんですが…

X教授：先に進むよ．測定値がn個あり，その値が$(x_1,x_2,\cdots,x_n)$だったとする．これらの測定値から計算される結果をyとして，計算式を$y=f(x_1,x_2,\cdots,x_n)$とする．そして，i番目の測定値の分散をσ_{xi}^2とすると，計算結果yの分散σ_y^2は

$$\sigma_y^2=\sum_i\left(\frac{\partial f}{\partial x_i}\right)^2\sigma_{xi}^2$$

となるんだ．分散が標準偏差σ_yの2乗だというのは大丈夫だよね．この式が成り立つのは，毎回の測定が独立であることが条件だということも忘れてはいけない．

Aさん：この式を示すために偏微分を説明したんですか．

X教授：そうだよ．数式で書くと短いが，文章にすると「計算結果の分散(σ_y^2)は，各測定値の分散(σ_{xi}^2)に，各変数の測定値の計算式の偏導関数の二乗($\left(\dfrac{\partial f}{\partial x_i}\right)^2$)を定数として掛けたものの総和($\sum\limits_i$)となる」となるんだ．

B君：何か，哲学的な文章ですね．

Aさん：先輩，格好つけてもだまされませんよ．

X教授：それで本題のAさんを悩ましている希釈の問題だね．この問題を考えるときに必要な数値があるんだ．

塩化ナトリウムの分子量：58.5
微量天秤の標準偏差σ_w：0.5 mg
1 mLホールピペットの標準偏差：0.02 mL
100 mLメスフラスコの標準偏差：0.1 mL
1 Lメスシリンダーの標準偏差σ_v：1 mL

と仮にしよう．

まず，質問した学生さんの言うように1段階で溶液を作ると，量り取る量w (= 5.8 mg)に0.5 mgの誤差が入るし，溶解の際にも溶液全量v (= 1 L)に0.001 Lの誤差が入ると考えていい．

溶解操作の関数はw (= 5.8 mg)の塩化ナトリウムをv (= 1 L)に溶かす場合

$$f=\frac{w}{v}$$

と書くことができる．溶解操作全体の分散は$\sigma_{一段階}^2$はさっきの式から，

$$\sigma_{一段階}^2=\left(\frac{\partial f}{\partial w}\right)^2\sigma_w^2+\left(\frac{\partial f}{\partial v}\right)^2\sigma_v^2$$

となり，溶解操作の関数の偏微分はそれぞれ

$$\frac{\partial f}{\partial w}=\frac{1}{v}\qquad\frac{\partial f}{\partial v}=-\frac{w}{v^2}$$

だから，溶解操作全体の分散は

$$\sigma_{一段階}^2=\frac{1}{v^2}\sigma_w^2+\frac{w^2}{v^4}\sigma_v^2=$$

$$\frac{1}{1.000^2}\times0.5^2+\frac{5.8^2}{1.000^4}\times0.001^2=$$
$$0.25+0.0000336=0.5^2$$

という計算から（2項目の値はものすごく小さい），

$$\sigma_{一段階}=0.5(mg/L)=0.5/58.5=8.55(\mu M)$$

標準偏差のことを誤差というので，学生の言う一段階法では，8.55 μMの誤差が入ってくるわけだ．

B君：100 μMを作るのだから誤差は8.55%ですね．

X教授：この計算から，誤差の原因となっている操作を特定できる．分散の計算式の1項目は，微量天秤で5.8 mgを測りとる操作の分散，2項目は1 Lの水を測りとる操作の分散だ．計算結果を見ると，2項目の寄与はほぼ無視できる．

Aさん：ってことは，全体で8.55%の誤差のほとんどが，微量天秤で秤量する操作のせい．ということですね．

X教授：そういうことだね．次に希釈を行う二段階法を調べてみよう．最初に量り取る量はw = 58.5 mg，溶液量はv = 0.1 Lで，ここからu = 0.001 Lを測りとるので，ここまでの操作の関数は

$$f = \frac{w}{v}u$$

となる．これを，さらに，v = 0.1 Lの水で希釈するから，$f = \frac{w}{v^2}u$ となる．0.1 Lの水で希釈する同じ操作を2回していると考えるわけだね．この時の全体の分散$\sigma^2_{二段階}$は

$$\sigma^2_{二段階} = \left(\frac{\partial f}{\partial w}\right)^2 \sigma_w^2 + \left(\frac{\partial f}{\partial v}\right)^2 \sigma_v^2 + \left(\frac{\partial f}{\partial u}\right)^2 \sigma_u^2$$

また，

$$\frac{\partial f}{\partial w} = \frac{u}{v^2} \qquad \frac{\partial f}{\partial v} = -\frac{2wu}{v^3} \qquad \frac{\partial f}{\partial u} = \frac{w}{v^2}$$

だから，溶解操作全体の誤差は

$$\sigma^2_{二段階} = \frac{u^2}{v^4}\sigma_w^2 + \frac{4w^2u^2}{v^6}\sigma_v^2 + \frac{w^2}{v^4}\sigma_u^2 =$$

$$\frac{0.001^2}{0.1^4}\times 0.5^2 + 4\frac{58.5^2 0.001^2}{0.1^6}\times(1\times10^{-4})^2 + \frac{58.5^2}{0.1^4}\times(2\times10^{-5})^2$$

$$0.0025 + 0.000137 + 0.013759 = 0.13^2$$

$$= 0.13\ (\mathrm{mg/L})$$

つまり，誤差2.18 μMで2%強になる．

どうかな（この部分の計算は参考文献1，p. 68–70の例題を改変して作った）．

Aさん：誤差が1/4くらいになっていますね．希釈した2段階の方の操作が多いのに，誤差が小さくなるわけですか．

X教授：どうしてだと思う？

B君：計算式を見ると，さっきの1段階法で誤差のもとになっていた，1項目（微量天秤で58.5 mgを測りとる操作）の値が非常に小さくなっていて，むしろ，3項目のホールピペットで1 mL測りとる操作の寄与が一番大きいですね．それでも，2項目（0.1 Lの水を測りとる操作）とあわせても，全体の分散は2段階法のほうが，1段階法よりも小さいので，

Aさん：2段階法の方の誤差が小さくなる，ってわけですね．あのカシコそうな1回生，次回，ギャフンと言わしてやろう．

X教授：Aさん，最近，とても積極的だね？誰のせいだろうね．それはともかく1段階法は，微量天秤で5.8 mgを測りとる操作の分散が大きくて，全体の分散も大きくなってしまう．誤差が伝播してしまっているわけだね．

B君：でも2段階法と同じ微量天秤を使っていますよね？

X教授：2段階法では10倍の58.5 mg 測りとっているだろ．このせいで「微量天秤で58.5 mgを測りとる操作の分散」は「微量天秤で5.8 mgを測りとる操作の分散」の1/10になったんだね．これが誤差の改善に効いているみたいだね．

B君：なるほど，一段階法の微量を測りとる操作が誤差のもとだったわけですね．でも2段階法でも全体の誤差は1/10になっていませんよね．

Aさん：それは，ほかの操作が誤差の要因になっているからですよ．2段階法の場合3項目のホールピペットで1 mL測りとる操作が一番大きな誤差のもとになっていますよね．

X教授：その通りだね．じゃあ，2段階法で，たとえば1 mLホールピペットがなかったので，10 mLメスピペットで代用したとしよう．その標準偏差が，5倍の0.1 mLだったとすると，全体の誤差はどうなるかな．

Aさん：（しばらく計算して）誤差は0.58 (mg/L)で9%くらいになってしまいました．これだと1段階法と変わらないですね．

B君：そっかー，もともと寄与が大きい操作の誤差を大きくしてしまうと，実験全体に伝播してしまうんだね．これは気を付けないと．

X教授：教訓は二つあるね，まず，データのばらつきが問題になるときは，実験操作全体でどこが誤差に大きく寄与しているのかを突き止める必要があるね．一番怪しいのは，微量を測りとる操作だね．

Aさん：たしかに，微量天秤で5.8 mgを測りとるよりも，多めに測りとってそれから希釈したほうが誤差が小さくなりそうですよね．

X教授：それから，測りとる器具を適正に選ぶ必要があるということだね．

B君：10 μLを測りとるときにも，ついつい面倒になって1 mL用とか200 μL用のピペットマンを使っちゃうことがあるけど，20 μL用を使うのがベストなわけか……反省です．

X教授：誤差の原因がわかったら，その操作の精度を上げる努力をすると，必ず全体の精度が向上する．たとえば2段階法で，1 mLホールピペットの操作を訓練して，標準偏差を0.01 mLと半分にできたとしよう．

Aさん：（しばらく計算して）誤差は0.07 (mg/L)で1%

弱と半分くらいになっていますね．ホールピペットと微量天秤の操作の寄与が同じくらいになっています．

B君：実験がうまくなるコツですね．

Aさん：ところで，誤差の計算で公式的なものはないんですか．

X教授：一般的な公式ではないが，それぞれの分散がσ_A^2, σ_B^2であるAとBの四則演算については，次の公式がある．

和と差の分散は同じで$\sigma_A^2+\sigma_B^2$となるんだ．積の分散は，$A^2B^2\left(\frac{\sigma_A^2}{A^2}+\frac{\sigma_B^2}{B^2}\right)$，商$\frac{A}{B}$の分散は$\frac{A^2}{B^2}\left(\frac{\sigma_A^2}{A^2}+\frac{\sigma_B^2}{B^2}\right)$となるんだ．

Aさん：これを使うと測定結果から導き出した結果に，どの程度の誤差があるか見積もることができますね．X教授ありがとうございました．じゃ先輩研究室に帰って復習しましょ．

B君：だから，実験があるんだってば…でも今日教わったことをうまく使って誤差の少ない実験ができるよう頑張ります．

X教授：それよりも，女の子の気持ちを推し量る誤差が大きすぎるみたいだね．ま，頑張りたまえ．

B君：？？？？

Aさん：有難うございました．

少し進んだ話題

X教授：今回の話題は，実験誤差の取り扱いということで，いきなり偏微分から入って面食らった読者もいるかと思うが，偏微分の計算法は意外と簡単だと思ってもらえたのではないかな．いい機会なので，偏微分の応用についても紹介しておこうと思う．

最適化を行う際に用いられるラグランジュの未定乗数法と呼ばれる手法で，統計学（たとえば，主成分分析）でもよく使われる．

変数の数は幾つでもいいが，簡単な2変数の場合を解説しておく．詳しくは解析学のテキストを参照してほしい．この方法は，$g(x,y)=0$という条件の下で$f(x,y)$を最大化（あるいは最小化）する方法で，まず，次のような関数

$$L(x,y,\lambda)=f(x,y)-\lambda g(x,y)$$

を考える．そして，

$$\frac{\partial L}{\partial x}=\frac{\partial L}{\partial y}=\frac{\partial L}{\partial \lambda}=0$$

を解けば，$f(x,y)$を最大化（あるいは最小化）する(x,y)の組を得ることができるわけだ．

例）$g(x,y)=x^2+y^2-5=0$

$$f(x,y)=2x+y$$

とする．$f(x,y)$を最大にする(x,y)の組を求めよう．

$$\begin{aligned}L(x,y,\lambda)&=f(x,y)-\lambda g(x,y)\\&=2x+y-\lambda(x^2+y^2-5)\end{aligned}$$

$$\frac{\partial L}{\partial x}=2-2\lambda x=0$$

$$\frac{\partial L}{\partial y}=1-2\lambda y=0$$

$$\frac{\partial L}{\partial \lambda}=x^2+y^2-5=0$$

これを解くと，$(x,y)=(\pm 2, \pm 1)$のとき，$f(x,y)=2x+y$の極値を与え，$(x,y)=(2,1)$のとき$f=5$で最大となる．また，$(x,y)=(-2,-1)$のとき$f=-5$で最小となる．各自で確認していただけるだろうか．

次回は，重回帰分析を取り上げてみたい．

参考文献

1) 化学同人編集部：実験データを正しく扱うために，科学同人 (2007).