



間違いから学ぶ

## 実践統計解析【第5回】

### 正しい統計記述とは？

川瀬 雅也<sup>1\*</sup>・松田 史生<sup>2</sup>

皆さんが論文を書く際、どのように統計処理の結果を書かれているだろうか。案外、気にされず、研究室の流儀で書いているのではないだろうか。統計処理を行った方がいいが、必要な情報が書かれておらず、本当に統計的なデータの吟味を行ったのか疑ってしまうような論文によく出会う。今回は、論文で何を書かなければならないかについて触れることにする。

Aさん：先輩おはようございますっ。あっあの、相談したいことがあるんですが……。あれ？先輩、無精ひげが……

B君：いやあ、徹夜だよ。やっと、学会報告できるようになったんで、図を作っていたんだ。統計処理結果をどう書くか、過去の論文を参考にしたらみんな書き方がバラバラで、えらい時間がかかったんだ。

Aさん：それで、書けたんですか？

B君：一応ね。でも、先生に見せたら……。

Aさん：X教授に教えてもらいましょう。

B君：えー。先生にどう言われたか聞かないの？

Aさん：どうせボロカスに言われたんでしょ？

B君：……。

#### 統計法の記述

X教授：いらっしゃい。今日是对応のあるt検定の話だったね。

Aさん：実は……なんです。

X教授：なるほど。大事なことから、まず、論文などの統計に関する記述法を説明しようか。まず、B君はどんな図を書いて先生からどうボロカスに言われたの？データと図を見せてくれないか。

B君：ボロカスを強調しなくても……。データはaとbという2種類の物質を添加したときの酵素活性です。

表1. 酵素活性にあたる添加物の効果

a	1.2	1.5	1.9
b	0.7	0.9	1.0

学会発表で使うので図にして、説明のところに統計処理の情報も書いたんです。実際に、作った図がこれ

です。

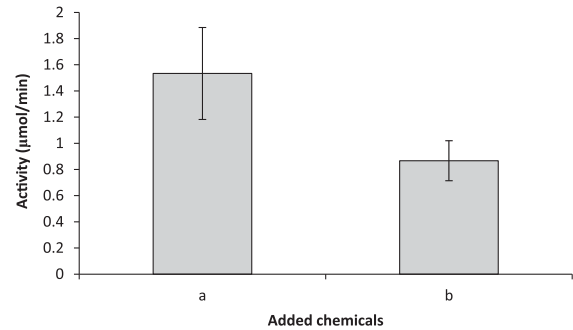


Fig. 1. The effect of added chemicals on enzyme activity. Bars show the SD of data. \*P < 0.05.

先生に見せたら、説明文は means ± SD, \*P < 0.05 で、使った統計処理法を書くように言われました。

X教授：なるほど。

Aさん：直してもらったのに、どうして悩んでいるんですか。思ったより、ボロカスじゃないし。

B君：最後に、他の人がどう書いているのか、ちゃんと勉強して書きなさいと言われたんだ。でも、他の人の書き方がまちまちで、完全に混乱しているんだ。

Aさん：ボロカスじゃなくて、頭がボロボロですね。

B君：……。

X教授：二人はいいコンビだね。話を戻そう。確かに、B君の言うように、人によって統計に関しては書き方がまちまちだ。誰かのものを参考にしようと思えばするほど混乱するのも事実だと思う。決して、B君の出来が悪いわけではない、と思うが。

統計に関しては、どんな情報を書く必要があるかをまとめてみようか。

1. 採用した方法。ここでは、Student's t-testかな。
2. 有意水準
3. データの表し方とデータ数。ここでは、平均±SDだけでは不十分で、データ数も書かなければいけない。
4. 判定結果。ここでは\*印の意味かな。

あと、図の書き方だが、研究室である程度作法が決まっていると思うので、それに従うのがいいとは思いますが、気になる点も二つほどあるので指摘しておこう。

まず、有意確率の記述だがPと大文字にしているが、

\*著者紹介 <sup>1</sup>長浜バイオ大学(教授) E-mail: m\_kawase@nagahama-i-bio.ac.jp

<sup>2</sup>大阪大学大学院情報科学研究科(准教授)



本来はpと小文字にする方がいいと思う。大文字は、確率一般に使われる記号で、混乱を招く場合があるから注意が必要だ。pを $p$ のようにイタリックにすることも多いが、これも必ずと言う訳ではない。もう一つは、グラフのバーの付け方だが、棒の中は見えないので、ふつうは上側だけ見えるようにするんじゃないかな。これは、発表する学会で決まっているときもあるから、もう一度、確認する方がいいと思うよ。論文では投稿規定をよく読むことだね。

以上の点を、書き直すと、

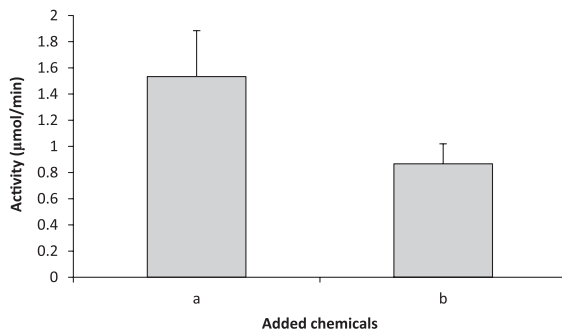


Fig.2. The effect of added chemicals on enzyme activity. Student's t-test was used to compare two groups. Difference was assessed with two-side test with an alpha level of 0.05. Values are means  $\pm$  SD of three independent experiments. Asterisk indicates a significant difference ( $p < 0.05$ ).

少し、細かく書きすぎているかもしれないが、この程度の情報は欲しいと思う。日本語で書く場合も、必要なものは同じだ。

Aさん：論文や報告書の場合はどうなんですか。やっぱり、図に、これだけ書かないといけないんですか。

X教授：論文の場合は、Materials and methodsの項にStatisticsという項を作って方法の詳細を書けば、図の下にそれほど多くのことを書く必要はない。

Statisticsには、The data were analyzed for statistical significances using Student's t-test. Difference was assessed with two-side test with an  $\alpha$  level of 0.05.と書いて、図の下には、Values are means  $\pm$  SD of three independent experiments. Asterisk indicates a significant difference ( $p < 0.05$ ).と書けばいいんだ。

B君：統計処理で共通している部分は、まとめてMaterials and methodsの項に書けばいいんですね。何となく分かってきました。

Aさん：さらっと流しているけど、私たちの指導教員も間違っていたんですね。

B君：そこを、わざわざ、つつかなくても。

X教授：まあ、そういうことだね。

この部分の書き方はメモ（X教授メモ1を参照）を

見ればわかるが、本当にきちんと書いている人の方が少ないと思う。これは、統計の教育が単に、計算ができればいいというスタンスで行われてきたからだ。今でも、書き方まで指導しているというのは少ないんじゃないかな。

たった100個の図表についてみただけで、思ったより記述の内容がバラバラだろう。そして、情報の不足も結構見られる。統計に関する記載は、研究そのものとは関係ないと思われるかもしれないが、おろそかにすると論文の信用に関わることもあるから注意が必要だ。

Aさん、B君：驚きの結果ですね。僕たちは、間違いなく書けそうです。

X教授：よかった。後、もう1点。データの表し方だけど、なぜ、SD（標準偏差）を使っているんだね。

B君：どの論文も使っているの。

X教授：よく考えてみよう。君が比較したいのは平均値の違いじゃないかな。

B君：そうです。

X教授：では、平均値の広がりを表すのは標準偏差かな。

Aさん：標準誤差ではなかったでしたっけ。

X教授：その通り。一般にはSDが使われているんだが、本当はSE（標準誤差）を使う方が理にかなっているんじゃないかと思うんだ。だが、必ずこうしなければならないということではないので、何を、自分が表したいかよく考えて、SDを使うかSEを使うかを判断すればいいと思う。

Aさん、B君：分かりました。

### —X教授メモ1—

生物工学関係の有名雑誌に掲載された論文の図表から無作為に100個選んで調べてみたところ、エラーバーが記載されていないもの：7つ（大規模発酵など実験回数が1回場合はやむを得ない）、エラーバーの意味を記載していないもの：33個、実験の反復数の記載のないもの：33個だった。おおよそ全体の半分弱の記載に問題があり、特に発酵結果の表で記載忘れが目立つようだ。日本を代表する国際誌として対応が必要かもしれない。

また、12個のStudent's t-testの事例のうち、両側検定と明記されたもの：17% (2/12)、unpairedと明記されたもの：25% (3/12)、等分散を議論したもの：なし (0/12) だった。検定法としてはANOVA (5/100)、Tukey's test (4/100) などもよくつかわれているようだ。有意確率 $p$ が小文字のものは71% (15/21)と広く認識されているようだが、このあとで話題になる標準偏差と標準誤差の選択では標準偏差が86% (49/57)と多数派だった。

## 検定法の選択

X教授：ところで、統計処理の内容について見てみたいんだけど。

B君：何か間違っていますか。

X教授：このデータの処理法として、本当に、Student's t-testでいいのかな。

B君：はい。等分散検定をしたら、

```
> a <- c(1.2,1.5,1.9)
```

```
> b <- c(0.9,0.7,1.0)
```

```
> var.test(a,b)
```

```
data: a and b
```

```
F = 5.2857, num df = 2, denom df = 2, p-value = 0.3182
```

で等分散だと見ていいと出ましたから。

Aさん：aもbも三つしかデータがないから、本当に、正規分布に従う母集団からとってきたとしても、母集団の分散をデータから推定できるんですか。

B君：へっ。何も気にしていなかった。

X教授：いいところに気が付いたね。多くの場合、何も気にせずに、データは正規分布に従う母集団から無作為にとってきたと思いこんでいるんだ。

よく考えると、少ないデータから計算した不偏分散がの集団の分散の推定値だと見ていい保証はないんだ。もちろん、疑えばきりがないが。試しに、Student's t-testの場合とWelch's t-testの場合を比べてみよう。まず、Student's t-testでは

```
> t.test(a,b, var.equal=T)
```

```
Two Sample t-test
```

```
data: a and b
```

```
t = 3.0151, df = 4, p-value = 0.03935
```

確かに、 $p < 0.05$ で有意差が認められるね。では、Welch's t-testでは、

```
> t.test(a,b)
```

```
Welch Two Sample t-test
```

```
data: a and b
```

```
t = 3.0151, df = 2.7306, p-value = 0.06424
```

$p > 0.05$ で有意差は認められない。

B君：どういうことなんですか。

X教授：データ数が少ないときにはよくおこる現象だね。

データ数が少ないから、いくら注意して実験を行っても、偶然、値が偏ってしまうこともありうる。このデー

タはまさに、その状況なのかもしれないね。等分散検定で等分散だとなっても、データ数が少ないときは、そのまま進めてもいいのか注意が必要だ。

Aさん：どうすればいいんですか。

X教授：今は、パソコンで簡単に計算できるから、Student's t-testとWelch's t-testの両方を行ってみて同じ結果になれば、Student's t-testの結果を信用しているのではないかな。もし、違っていたなら、データ数を増やす必要があるかなと思う。

B君：やってみます。

数日後、

B君：データを増やした結果を持ってきました。

```
> a <- c(1.2,1.5,1.9,1.8,2.1,2.0)
```

```
> b <- c(0.9,0.7,1.0,0.7,0.8,0.6)
```

というデータで、等分散検定では $p\text{-value} = 0.09084$ で、等分散と見ていいとなりました。

Student's t-test. では、

```
> t.test(a,b,var.equal=T)
```

```
Two Sample t-test
```

```
data: a and b
```

```
t = 6.405, df = 10, p-value = 7.782e-05
```

で有意差があり、また、Welch's t-testでも、

```
> t.test(a,b)
```

```
Welch Two Sample t-test
```

```
data: a and b
```

```
t = 6.405, df = 6.8195, p-value = 0.0004072
```

のように有意差が出ました。

データを増やすとスッキリしました。

Aさん：本当ですね。

X教授：よく、実験を繰り返したね。統計は、単なる計算結果だが、その結果をどう見るかが大事なんだ。今回のように、三つのデータだけで結論を出そうとすると、もしかすると間違っているかもしれない。そこで、慎重に、二つの検定法の結果を比べて実験を繰り返すことで、正しい結論にたどり着けるんだ。今回は、三つのデータでStudent's t-testの結果と同じ結果に落ち着いたが、何時もこうなるとは限らないからね。

B君：分かりました。

Aさん：もし、実験を繰り返すことが難しい場合は、どうすればいいんですか。



X教授：ケースバイケースだが、間違いを起こさないという観点からすればWelch's t-testを選択する方がいいだろうね。Welch's t-testの方が厳しい目の判定になるから。ただ、最初のケースの場合だとWelch's t-testで有意差がないと出ても、これで有意差がないと結論を出すのは早いと思う方がいい。Student's t-testでは、ギリギリだが有意差が出ているから。つまり、「得られたデータからはハッキリと有意差があるかどうかの判定はできなかった」という結論に落ち着くのではないかな。

B君：こんな時は、ハッキリさせるために実験を繰り返す方がいいんですね。

X教授：その通りだね。

Aさん：だから、再現性がうるさく言われるんですね。

B君：それだけじゃないけど、統計処理の面からは、データを増やすことになるのかな。

X教授：そう考えてもいいだろう。

### 対応のある2群の平均の差の検定

X教授：ずっと、先送りになっている対応のある場合について話しておこうか。

Aさん：そうでした。

B君：お願いします。

X教授：対応のある場合の例として、5種類のサンプル中の有効成分量をD法とE法で分析したとき、同じ結果を出すと言えるかどうかという場合を考えてみよう。ここでは、同じサンプルの分析値同士を比べないといけないね。こういった比較の相手が決まっている場合が、「対応のある場合」に相当するんだ。

表2. D法とE法による分析結果

	S1	S2	S3	S4	S5
D法	51.0	55.2	52.6	61.2	55.4
E法	51.8	55.6	50.9	59.6	54.5

まず、帰無仮説は「D法とE法で、分析結果に違いはない」となり、対立仮説は「D法とE法の分析結果は異なる」となる。有意水準0.05で検定してみよう。

対応のある場合は、まず、両データの差を求めるんだ。

```
> D <- c(51.0,55.2,52.6,61.2,55.4)
```

```
> E <- c(51.8,55.6,50.9,59.6,54.5)
```

```
> diff <- D-E
```

```
> diff : データの差を新しい変数diffとする
```

```
[1] -0.8 -0.4 1.7 1.6 0.9
```

この差の母平均が0とみなせれば、有意差はないという結論になると考えるんだ。そこで、Rでは、

```
> t.test(diff)
```

```
data: diff
```

```
t = 1.17, df = 4, p-value = 0.307
```

```
alternative hypothesis: true mean is not equal to 0
```

となり、p-value = 0.307だから帰無仮説を採用して、有意差はないとの結論になる。alternative hypothesisは対立仮説のことで、この部分で平均が0と比較していることを確認ができる。

B君：よく分かりました。

Aさん：私たちが対応のある場合の検定をすることはありますか。

X教授：あまりないと思うな。ただ、可能性は0ではないので、知識として知っておく方がいいと思う。

B君：あと、実験を繰り返すと、とんでもなく外れたデータが出る時があるんです。どう扱えばいいんですか。

Aさん：私なんか、しょっちゅう出ますよ。

X教授：実験のウデの問題もあるだろうけど、今日は時間がないので、次回に話をしよう。

Aさん、B君：お願いします。

Aさん：先輩、次回も楽しみですね！

### —X教授メモ2—

Microsoft Excelで有意差検定を行う場合に使うTTEST関数の書式は下記の通り。

TTEST (配列1, 配列2, 尾部, 検定の種類)

尾部が1の時は片側、2の時は両側検定となる。また、検定の種類は1が対応のある2群、2が対応のない等分散の2群 (Student's t-test)、3は対応のない非等分散の2群 (Welch's t-test) の検定となる。したがって、反復数n=3のデータで有意差検定をする場合は

TTEST (配列1, 配列2, 2, 3)

としておくのがもっとも間違いがない。

### 参考書

- 1) 奥田千恵子：医薬研究者のための統計記述の英文表現、金芳堂 (2004).  
\*現在は第3版。英語で統計的な記述を行う時参考になると思う。

(【第6回】は95巻2号に掲載予定です)