



実践統計解析【第7回】

多重比較って何？

川瀬 雅也^{1*}・松田 史生²

これまでの回で、平均の差の検定の話は一応終わったとお考えではないだろうか。実は、もう一つ大事な話が残っている。たとえば、三つ以上の条件から得られた結果の平均値に有意差がないか比較する場合がある。それから、マイクロアレイ解析のように2条件間で数千以上の異なる遺伝子の発現量を比較することもよく行われている。このようなとき、どうやって平均を比べるべきなのか。今回は、この問題を考えてみたい。

t-検定は繰り返すべからず？

Aさん：B先輩、昨日の映画帰りの件、調べてきましたよ。片思いの相手と高校3年間のクラス替えで1回は同じクラスになれる確率は1学年10クラスの場合 $1 - (1 - 0.1)^3 = 0.271$ であっています。

B君：こっちも調べたよ。偶然同じクラスになる確率が0.1なんだから $0.1 \times 3 = 0.3$ のはずっていうCさんの計算法は、確率じゃなくて期待値だね。

Aさん：いつもクールなC先輩が、映画のクラス替えのシーンにあんなに熱くなっちゃうなんて、博士課程ってそんなに大変なんですか？

B君：そうみたいだね、そっとしてあげよう。けどどさ、この計算は今やっている研究に出てくるんだよね。大腸菌の野生株と変異株の間で発現量が異なる遺伝子を、3連の実験でマイクロアレイを使って探索しているんだけど、有意水準を5%にしてStudentのt-検定をすると、発現量に有意差がない遺伝子も5%の確率で有意差があるという結果に偶然になってしまうというのはいいよね。

Aさん：以前、X教授に教えてもらいましたね。

B君：でも、大腸菌のマイクロアレイでは約4000遺伝子でt-検定を行うから、さっきのCさんの方法で、 $0.05 \times 4000 = 200$ 遺伝子を本当は発現量に有意差がないのに有意差あり、と誤って判定する期待値が計算できるんだな(擬陽性)。つまり有意水準5%でt-検定をして300遺伝子に有意差があったとしても擬陽性の期待値が200遺伝子なんだから、その内 $200/300 = 66\%$ が擬陽性ってこと(False Discovery Rate, FDRという)になってしまって解析結果が使い物にならない。

Aさん：なるほど、検定の繰り返しには要注意ですね。じゃあどうしたらいいんでしょう？

B君：……。ところでこないだの3種類の菌の抗生物質

の生産量(mg/L)を調べた結果はどうなったの？

Aさん：(都合が悪くなるとううなんだから……。) さっき出しました(表1)。

表1. 菌体による抗生物質生産量(mg/L)

| 菌体 | 1回目 | 2回目 | 3回目 |
|----|------|------|------|
| A | 10.1 | 9.8 | 9.5 |
| B | 8.8 | 8.9 | 9.3 |
| C | 11.1 | 10.9 | 11.5 |

Cが一番生産性が高いと思うんですが、統計的に確かめるにはどうすればいいんでしょうか？

B君：うーん、やっぱりStudentのt-検定かな？

Aさん：でも、今まで2組のデータの間のt-検定しか出てきませんでしたよ。

B君：だから、2組ずつ組を作ってt-検定して、後で、まとめればいいんだよ。この間読んだ論文でもそうしていたし。

Aさん：かなり不安……。二つまとめてX教授に聞いてみましょう。

検定を繰り返すと……

Aさん：X教授！……。ということなんですが、どうしたらいいんでしょうか？

X教授：まずは3群の比較からいこうか。そのB君の見た論文は大間違いだ。だが、B君を責められんな。最近皆、注意するようになってきたが、少し前までは、平気でB君の言ったような方法で正しいと思っていたんだからな。今でも間違った検定で通っている論文を見かける。

ここで「t-検定は2群の平均の差の検定にしか使えない」ことを肝に銘じてほしい。3群以上の場合には、StudentにしてもWelchにしてもt-検定は使えないんだ。

Aさん、B君：そ、そうなんですか。

Aさん：でも、どうしてt-検定を繰り返してはいけないんですか。

X教授：t-検定を行う時に有意水準を定めたことを思い出して欲しいんだ。今、仮に有意水準を5%にしたとしよう。2群間のt-検定を繰り返すとすると、Aさん

*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)



はA, B, Cの3種の菌で比べたいからA-B, A-C, B-Cの3回t-検定を行うことになるね。1回の検定で有意差が出ない確率は $(1-0.05)$ だから、3回で一度も有意差が出ない確率は $(1-0.05)^3 = 0.857$ となる。つまり、有意水準が5%のつもりだったが、3回繰り返すことで、有意水準は $1-(1-0.05)^3 = 0.143$ と、14.2%になってしまったわけだ。本当は有意差なしと判定すべき時に、有意差ありとしてしまう確率が大きくなっているんだ。

B君：驚きです。繰り返すだけで、有意水準が変わるなんて。肝に銘じます。

Aさん：思い出したんですけど、3群以上の場合は分散分析を使えばいいって聞いたことがあります。

X教授：なるほど。でも、今回の統計処理の目的は、どの菌が一番生産性が高いかを統計的にハッキリさせたいわけだろう。分散分析で可能かな。

B君：えーと……。

Aさん：意地悪しないで、教えてくださいよ。

X教授：よし、よし。では、分散分析の話から始めようか。

分散分析

X教授：分散分析には、主なものとして一元配置分散分析、繰り返しのない二元配置分散分析と繰り返しのあつ二元配置分散分析があるんだ。一元配置分散分析とは、今回のデータのように、A, B, Cの比較だけを行おうとする場合で、比較変数が菌体の種類という1種類だけの場合をいうんだ。もし、菌体の種類に加えて、実験の各回の比較も行うなら、変数に実験回が加わり、2種類になるから二元配置分散分析になる。おそらく、生物工学の分野で使う可能性があるのは一元配置分散分析だと思うので、今回はこの方法だけを説明しよう。

分散分析では、Aさんの言うように3群以上のデータを取り扱うことができるんだ。そして、検定の1種だということも忘れないでほしい。検定というと、大抵の場合、t-検定を思い浮かべるが、他の検定もあるんだ。分散分析が検定であることを思い出すと、分散分析の帰無仮説は何だろうと考えると思うが、

Aさん、B君：そう言われると……

X教授：Aさんのデータを使うことにすると、帰無仮説は「A, B, Cの生産量の平均はすべて等しい」となるんだ。では、対立仮説は？

Aさん：分かりません。

B君：降参です。

X教授：「A, B, Cの生産量の平均はすべてが等しいわけではない」だ。つまり、分散分析で有意差が出たとき「どれかの平均に有意差があると見ていいが、どれかは分からない」というわけだ。

Aさん：有意差があると分かっても、その先、進めない

んですか。

X教授：その通りだね。場合によっては、分散分析の結果で十分なきもある。品質管理なんかの時、普通は有意差が出ないことが前提で、全ロットに有意差がないことを確認するときは分散分析で十分だね。

Aさん：私の場合は、分散分析では不十分なんですね。

X教授：そうだね。いい機会だし、不十分だけど、きちんと分散分析ができることも大事かもしれないので、練習として分散分析を行ってみよう。

```
> A <- c(10.1,9.8,9.5)
```

```
> B <- c(8.8,8.9,9.3)
```

```
> C <- c(11.1,10.9,11.5)
```

三つのデータをまとめるんだ。

```
> data <- c(A,B,C)
```

データにラベル(どの菌のデータか)を貼り付けるんだ。

```
> MB <- c(rep("A",3),rep("B",3),rep("C",3))
```

```
> MB2 <- factor(MB)
```

この操作でデータの割り付けができたんだ。確認すると、

```
> data
```

```
[1] 10.1 9.8 9.5 8.8 8.9 9.3 11.1 10.9 11.5
```

```
> MB2
```

```
[1] A A A B B B C C C
```

```
Levels: A B C
```

では、分散分析を実行しよう。

```
> anova(lm(data~MB2))
```

```
Analysis of Variance Table
```

```
Response: data
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
```

```
MB2     2 7.2022  3.6011 42.645 0.0002839 ***
```

```
Residuals 6 0.5067  0.0844
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

で有意差があると出た。

分散分析にはいろいろな関数が用意されていてanovaの他に、aovと言う関数でも実行できる。

```
> res <- aov(data~MB2)
```

```
> summary(res)
```

```
      Df Sum Sq Mean Sq F value    Pr(>F)
```

```
MB2      2 7.202 3.601 42.65 0.000284 ***
Residuals 6 0.507 0.084
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

桁の切り上げが関数により違うが、この点を考慮すると同じ結果だとすぐにわかるね。

Rのようなソフトをきちんと使うには慣れが必要なので、参考書^{1,2)}を見ながら、自分で練習してみることが大事だ。

Aさん：はい、結果を見るとCの生産力が高いと言えませんか？

B君：でも、Bだけが低いのかもしれないし。

X教授：よく気が付いたね。二つの可能性があるね。分散分析だけでは、結論を出せないときがよくあるんだ。

では、どの菌が一番生産力が高いのかを調べることにしよう。これには、多重比較という方法を使うんだ。

Aさん：多重比較って、初めて聞きます。

B君：僕も。

多重比較

X教授：多重比較は3群以上ある場合、有意水準を上げずに個々の群と群の平均値を比較する方法と言われてる。分りやすいいえば、A、B、Cのどの群とどの群の間で、平均値に有意差があるのかを、有意水準を5%なら5%に保ったまま比較する方法と考えてもらうといいと思う。

Aさん：便利な方法があるんですね。修士課程のB先輩はなぜ、知らなかったんですか。

B君：学部の統計学で習わなかったし……、先生もうるさく言わなかったし……

X教授：まあ、多重比較まで統計学の講義で教えられることは少ないと思う。最近では、論文の審査で指摘されることが多くなってきて、多重比較も浸透してきたけどね。

B君：多重比較はどう計算すればいいんですか？

X教授：実は、多重比較は先に説明した比較を行う手法の総称で、いくつもの方法があるんだ。いちばん基本的なものがBonferroni法だ。最初に見たAさんのデータの検定でいうと3回t-検定を行う必要があるね。

Aさん：はい。

X教授：有意水準を5%のままにすると間違っていることは説明済みだ。そこで、有意水準を検定を繰り返す回数で割ってやる。つまり、 $0.05 \div 3 \approx 0.0167$, 1.67%とするんだ。こうすると、さっき説明した問題が解決される。これがBonferroni法だ。ただし、この方法は判定基準が厳しくなり検出力が落ちるので、今はあまり使われない。よく使われるのは、すべての群間を総当たりで比較するTukey-Kramer法と、コントロール群とその他の実験群を比較するDunnnett法だと思

う。AさんのデータはTukey-Kramer法で比較するのがいいと思う。

Aさん：多重比較といっても、いろいろあって、方法の選択を間違えといけないんですね。

B君：統計を甘く見てはいけないということだね。

Aさん：自分も知らなかったくせに。

X教授：2人は仲がいいね。この調子で切磋琢磨だね。

Aさん：いや、えっと、そのB先輩が……

X教授：(あれ?) まず、最初に、注意してほしいのは、Tukey-Kramer法はデータに正規性が仮定していることだ。これは、Dunnnett法でも同じなんだ。今回は、正規性のないノンパラメトリックな場合の方法は説明しないことにするから、必要になってきたら自分で勉強してほしい。

では、実際にTukey-Kramer法で検定してみよう。さっき説明した一元配置分散分析に続いて行う形が一般的だ。必ず、分散分析を行わなければならないことはないが、入力コマンドを見てもらえばわかるように、分散分析の結果を利用しているから、分散分析の時と同じ手順は必要になってくる。Tukey-Kramer法ではTukeyHSDという関数を使う。

> TukeyHSD(aov(data~MB2))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = data ~ MB2)

\$MB2

| | diff | lwr | upr | p adj |
|-----|-----------|------------|-------------|-----------|
| B-A | -0.800000 | -1.5280046 | -0.07199542 | 0.0345917 |
| C-A | 1.366667 | 0.6386621 | 2.09467125 | 0.0028858 |
| C-B | 2.166667 | 1.4386621 | 2.89467125 | 0.0002385 |

結果を見ると、有意水準5%ならA-B、A-C、B-Cのすべてに有意差があるね。つまり、表1から判断してCが一番生産力が高いと言ってもいいかな。(統計的な結論の出し方とは少し異なるが、生物工学的にはこう言っている)

Aさん：分かりました。実験結果もよくて、研究を進められることが分かりました。有難うございます。

B君：コントロールと比較することも、よくあるのでDunnnett法も教えてください。

X教授：まず、デモ用のデータを用意しよう(表2)。

表2. Dunnnett法のためのデモデータ

| 群 | データ | | |
|---------|------|------|------|
| Control | 7.0 | 8.5 | 8.9 |
| A | 9.1 | 8.7 | 9.8 |
| B | 10.5 | 11.1 | 10.2 |



Dunnett法では multcomp というパッケージを使うので、まずこのパッケージをインストールしよう。

```
> install.packages("multcomp")
『パッケージを ‘各自のPCでの保存場所’ にインストールします。このセッションで使うために、CRANのミラーサイトを選んでください。』
```

と出て、ミラーサイト一覧が出てくるので、一番近くのサイトを選ぶとダウンロードとインストールが始まる。

```
『ダウンロードされたパッケージは、以下にあります。
C:\各自のPCでの保存場所』
```

と出れば成功だ。続いてデータの入力だが、これもいろいろな方法を知っておくのもいいと思うので、データフレームを使う方法を紹介しよう。

```
> dundata <- data.frame(cat = factor(c(rep(1,3),rep(2,3),
rep(3,3))),
labels = c("control","A","B"),re = c(7.0,8.5,8.9,9.1,8.7,
9.8,10.5,11.1,10.2))
> plot(re~cat,data=dundata)
```

とするとデータの分布がプロットされる(図1)。これを見ると何となくBとControlに有意差がありそうに思えるね。

```
> library(multcomp)
> dres1 <- aov(re~cat,data = dundata)
```

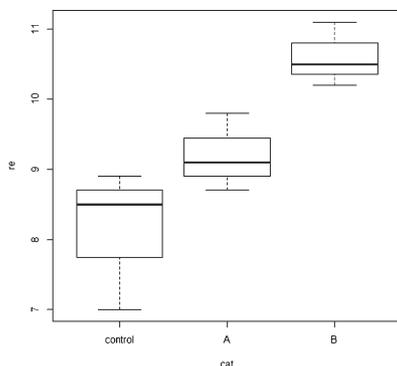


図1. 表2のデータの分布

まず、分散分析の操作を行うんだ。

```
> dres2 <- glht(dres1, linfct = mcp(cat = "Dunnett"))
> confint(dres2, level = 0.95)
```

Dunnett法を指定して、さらに、信頼区間を指定する。

```
> summary(dres2)
Simultaneous Tests for General Linear Hypotheses
Multiple Comparisons of Means: Dunnett Contrasts
Fit: aov(formula = re ~ cat, data = dundata)
Linear Hypotheses: Estimate Std. Error t value Pr(>|t|)
A - control 0 1.0667 0.5818 1.833 0.19382
B - control 0 2.4667 0.5818 4.240 0.00971 **
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

となって、図からの予想通りControlとBの間に有意水準5%で有意差が認められる結果となった。

Aさん：入力法は慣れるまで大変ですけど、使えるようになったら便利ですね。

X教授：それからもう一つ、マイクロアレイデータの解析では、B君の言うように擬陽性が出てしまうことが避けられないから、そのかわりにFDRが許容範囲以下になるように有意水準を調節するという考え方をする。ただし、多重検定補正法にはいくつかの考え方があり、どれが正しいとも言えないので、実験結果にFDRとその計算法を明記することが大事だ³⁾。最近大規模データ解析がはやっているようだが、統計解析をきちんとしないと誤った結論を得てしまう。統計学の勉強は、手法の特徴や得られる結果の解釈に重点を置く時代になってきていると思う。手法の特徴が分かれば、正しい手法の選択ができるようになるからね。

B君：(帰り道で) Cさんに教えてあげなきゃね。

Aさん：B先輩、でも、配属された研究室に好きな人ができたら好きな人と同じ研究室になる確率は1、ですよ？事後、ですけど。

B君：Cさんそれで納得するかなあ……

参考文献

- 1) 金 明哲: Rによるデータサイエンス, 森北出版 (2007).
- 2) 山田剛史ら: Rによるやさしい統計学, オーム社 (2008). この他にも、Rに関する書籍は多数あるので、一度見て、これなら分かりそうと思うものを選んでほしい。
- 3) B君の説明していた方法はBH法という。Benjamini, Y. et al.: *J. R. Stat. Soc. B*, 57, 289 (1995).

(【第8回】は95巻6号に掲載予定です)