



実践統計解析【第4回】

平均の差の検定の使い方

川瀬 雅也^{1*}・松田 史生²

皆さんが一番よく使う統計処理は「t-検定」ではないかと思う。今回は「t-検定」などの「2群の平均の差の検定」について考えたい。原理などは統計学の教科書にあるので、実際に使う時に必要なことにだけ触れる。

さて、例の2人はどうなっているだろうか。

B君：X教授のところへ行く時間だ。

Aさん：あっ、あの先輩！その前に聞いてもいいですか？

B君：どうしたの？

Aさん：今日の先輩って、めずらしくおしゃれな髪型でイケてますね。もしかしてデート……ですか？

B君：いやあ、そんないい話じゃないよ～。バイトの面接なんだよ。でもこんなしゅっとした格好をすることはめったにないから、いつもと違って何かあるってわかっちゃうよね。

Aさん：あっ、そうなんですか！服も髪も似合ってますよ（でも今日から私の髪型が変わったのには、先輩、気づいてくれないんだろうなあ、しょっちゅう変えるしなあ）。ところで先輩！検定には「両側検定」と「片側検定」がありますが、どうやって使い分けるんですか。

B君：えーと……。X教授に教えてもらおう。

検定とは

X教授：いらっしゃい。今日は検定の話だったね。

B君：はい。「t-検定」ですよ。僕もよく使っているので、今までとは違って答えられますよ。

Aさん：でも、実は……。と言う訳なんです。

X教授：なるほど。まずは、もっと基本的な話から始めようか。B君は「検定」というと「t-検定」と言ったが、実は大きな間違いだ。「検定」とは「データから推測したことが、統計的に正しいかどうか」を検証する手続きのことだ。「t-検定」は検定法の一つなんだね。

B君：そうなんですか。検定ってそういうことなんですか。

X教授：きちんと統計学を勉強しないことも多いから、知らないのも無理はないと思うな。あと、検定は統計学が発明した画期的手法なんだけど、とっつきにくいのも事実だ。けれど、一度検定の手順と使い方を理解してしまえば一気に世界が広がるから頑張っしてほしい。まず、「検定」を行うときに、「帰無仮説」をたて

るんだっただよね。これは「何を検証するか」をハッキリさせるためにたてるんだ。

B君：確かにそう習いました。でも、いつも疑問に思うことがあるんです。たとえば、データに差があるかどうかを調べるときにも、必ず、帰無仮説は「差がない」としますよね。どうしてでしょうか？

X教授：なるほど。もっともな疑問だね。「検定」の手順を追いながら説明しよう。B君の言う通り、帰無仮説（ H_0 ）では「差がない」とするのが一般的だ。帰無仮説に対する対立仮説（ H_1 ）は「差がある」となる。このように仮説を立てて検定するので「仮説検定」ともよばれる。「検定」では、帰無仮説を評価していることを忘れてはいけないんだ。もし、帰無仮説が成り立つと考えることが難しいときに、対立仮説が顔を出す。統計の授業では同時に二つの仮説を立てると教えられるから、学生が混乱しているのではないだろうか。

B君：そんな気がします。

X教授：手順と考え方を分けると理解しやすくなると思う。手順では、二つの仮説を立てるが、本当は“帰無仮説だけが存在して検定を受ける”とひとまずは、考えておくと、なぜ、帰無仮説を「差がない」とするかが分かりやすくなると思うよ。

「検定」の手順では、先程の二つの仮説を立てた後、有意水準を定め、有意確率と比較することになる。最終的に、帰無仮説を採用するのかどうかを判断するわけだ。一般的に「差がない」ことを否定するには差がある例が一つあればいいことは分かるね。

Aさん、B君：よく分からないです。

X教授：数学的に「差がない」とは、どんな時でも「差がない」ということを意味しているんだ。

B君：分かりました。差がある例を示せば、どんな時でもということにならないから「差がない」の否定になるんですね。

X教授：では「差がある」の否定はどうなるかな。

Aさん：「差がある」の否定は「すべての場合で差がない」ことを示さなければならないから、ある時、差がなくても、他の時に差が出たら「差がない」ことになりますよね。「差がある」の否定なんてできないですよ。

X教授：だから、帰無仮説としては「差がない」を使う訳なんだ。

B君：うーん。でも検定に使うデータは限られているか

*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）



ら、「すべての場合」って考えられないんじゃないですか。よく分からないんですけど。

X教授：統計学の基本を思い出して欲しい。標本と母集団ということを説明したね。標本は、君たちが実験で得るデータで、ある母集団からとられたと考えているわけだ。母集団は同じでも、標本（データ）は実験ごとに違ってくるね。

Aさん：私なんか、毎回違いすぎて困っています。

B君：慣れてくると、同じような値になってくるよ。

X教授：2人の話は誤差とも深い関係があるので、回を改めて考えようか。話をもとに戻すけど、検定で差があるかどうかを調べるのは「データの平均」でなくて、「(データの)母集団の平均」なんだ。だから統計的に考える必要があるわけだね。この点を意識せずにいると、とんでもない間違いを犯すことになるので注意が必要だ。

B君：なるほど、分かりました。母集団について考えるから、さっき、先生がおっしゃったように「差がある」の否定が難しいわけですね。

X教授：その通りだ。母集団を意識することで、両側検定にするのか片側検定にするのかの判断の基準も理解できる。次の話に、一気にいけないので、もう少し基礎の話の続けよう。

2群の平均の差の検定

X教授：「検定」にはどんな方法があるか知っているかい。

Aさん：t検定くらいです。

B君：僕もそうです。

X教授：ふつうは、そうだと思う。検定には(1)母平均の検定、(2)2群の平均の差の検定、(3)等分散検定、(4)適合度検定、(5)独立性検定などがある。この中で(2)が一番よく使う検定だと思うから、今回は(2)について考えよう。(1)はめったに行わないし、(3)は(2)の中で行うんだ。(4)、(5)については回を改めることにしよう。

Aさん：沢山あるんですね。

B君：でも、(2)しかやったことがない。

X教授：実際に、研究を始めるとそうだと思うね。(2)にはいくつか方法があるが、どの方法を選ぶかの基準は知っているかな。

Aさん：え、いくつも方法があるんですか？

X教授：それじゃ、方法の選択基準の話から始めよう。具体例があると分かりやすいので、次のデータを使うことにしよう。

このデータは、環境の違う2か所から採取したサンプル中の微生物数を調べた結果で、培養したときに現れた寒天プレート上のコロニー数で示してある。

A群：160, 168, 158, 165, 161

B群：153, 155, 149, 152

AとBで微生物数に差があると言えるかどうか調べよう。

まず、AとBの母集団が正規分布に従うのかどうかが重要になってくる。経験的に正規分布に従うとみなせる場合も多いと思うが、一度、試しに正規分布に従うかどうかの確認(正規性の検定)を行ってみよう。

```
> A <- c(160,168,158,165,161)
```

```
> B <- c(153,155,149,152)
```

```
> ks.test(A,"pnorm",mean=mean(A),sd=sd(A))
```

```
data: A
```

```
D = 0.23562, p-value = 0.8862
```

```
> ks.test(B,"pnorm",mean=mean(B),sd=sd(B))
```

```
data: B
```

```
D = 0.21017, p-value = 0.9798
```

これはKolmogorov-Smirnov検定とよばれる方法で、帰無仮説が「正規性がある」とおかれる。したがって、*p-value*をみるとA、Bとも0.05より大きいのでA、Bともに正規性があると見ていいことになるわけだ。ただ、生物工学分野のデータは正規分布に従うと考えていい場合が多いから、Kolmogorov-Smirnov検定はやらないのがふつうだと思っていいと思うよ。

ここで、*p-value*(有意確率)とは「得られたデータが帰無仮説によって説明できる確率のことで、言い換えると「データに偏りが偶然生じる」と考えていい確率となるんだ。この確率が有意水準以下なら、あまり小さな確率で帰無仮説を採用することになり矛盾が生じるので、帰無仮説を採用しない(棄却する)という判断になるわけだ。普通、有意水準は0.05を採用することが多いので、ここでもそうした。

もし、正規性がないと判断した場合はノンパラメトリック検定(Mann-WhitneyのU検定など)をとることになる。ノンパラメトリック検定については、別の回に説明することにしよう。

正規性が示されたら、次に、2群(AとB)の分散が同じかどうかを調べないといけない。つまり、等分散検定を行うわけだ。等分散検定ではF分布を使うので、F検定ともよばれる。

```
> var.test(A,B)
```

```
data: A and B
```

```
F = 2.608, num df = 4, denom df = 3, p-value = 0.457
```

等分散検定の帰無仮説は「等分散である。(分散に

差がない)」なので、 p -valueから分散が同じとみなせることが分かる。

2群とも正規分布に従い、かつ、分散が同じとみなせるときに、2人が言う t -検定、正確には“Studentの t -検定”を採用する。もし、分散が同じとはみなせない場合は、“Welchの t -検定”を採用することになるんだ。

最近の流れでいうと、生物分野のデータについては、経験的（過去の知見から）に等分散であるとして間違いがないとされる場合を除きWelchの t -検定を使うケースが主流になってきている。

B君： t -検定にも二つあるんですね。僕らの使っていたのは、今の話だとStudentの t -検定になるわけですね。

Aさん：これまでの話をまとめると、

○正規性のチェックで、正規性がなければノンパラメトリック検定、あれば、 t -検定

○等分散のチェックで、等分散ならStudentの t -検定、なければWelchの t -検定

となるわけですね。

X教授：その通り。先の例で「AとBの平均値に差はない」という帰無仮説について、Studentの t -検定を行ってみよう。

Aさん：検定を行う時に、さっき、お話しした両側検定と片側検定の使い分けがよく分からないんです。

X教授：この点もきちんと説明しながら、話を進めよう。

```
> t.test(A,B,var.equal=T)
```

```
Two Sample t-test
```

```
data: A and B
```

```
t = 4.3692, df = 7, p-value = 0.003277
```

となり、 p -valueが有意水準の0.05より小さいから、帰無仮説が棄却されてAとBの平均に差がある（正確には有意差がある）と考えてよいことになる。

統計の教科書では、検定統計量を計算して、自由度と有意水準から求めた t -分布のパーセント点の値と比較するとあったと思うが、最近では、誰も手計算で検定を行うことはないのだから、 p -valueを最初からみればいい。

もう1点、「有意差」という言葉に注意してほしい。「有意差がある」とすると差があると勘違いする人が多いのだが、これは間違いなんだ。「有意差がある」とは、「2群のデータの平均に生じた差が大きく、偶然生じたと考えるには無理があり、差がないとすることが難しい」ということを言っているわけで、「差がある」と断言しているわけではないんだ。

Aさん：難しいんですね。何か、分かったような、分からないような。

B君：でも、差があるんですね。

X教授：そこが問題だね。差があると積極的に言っているわけではないんだ。「差があるとしても問題ない」という気持ちなんだ。

前日も言ったけれど、検定結果も、他の統計処理と同じで「計算結果」なんだ。科学的に考えたとき、有意差があるとでも、素直に受け入れることができないこともあるだろうし、逆に、有意差がないと出た場合でも、この結論を受け入れない方がいい場合もあると思う。特に、生物関係のデータでは、このようなことが起こりやすいと思うんだが、計算結果がすべてではなく、科学的に考えたときにどうかということも重要になってくる。

このことを忘れずに、検定結果を見てほしいと何時も思うんだ。

B君：肝に銘じます。ところで、Welchの t -検定はどうやってやるんですか。

X教授：Rで、同じ例を使って計算すると、

```
> t.test(A,B,var.equal=F)
```

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

```
data: A and B
```

```
t = 4.622, df = 6.7008, p-value = 0.002712
```

とすればいい。

```
> t.test(A,B)
```

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

このようにRの $t.test$ のデフォルトはWelchの t -検定になっている。これは、先程、説明したようにWelchの t -検定が第1選択の方法になりつつあるためなんだ。

Aさん：ところで、今は両側検定ですか、それとも、片側検定ですか。

X教授：忘れるところだった。Rではデフォルトは両側検定なんだ。今まで見せた例は、すべて両側検定になる。

Aさん：と、言うことは、ふつうは両側検定が使われると考えていいわけですね。

X教授：その通りだね。というか片側検定は使い方が難しい。

Aさんの疑問「両側検定と片側検定の使い分け」を説明しよう。先程のデータを見てほしい。

A群：160, 168, 158, 165, 161 平均：162.40

B群：153, 155, 149, 152 平均：152.25



どちらが大きいと思う？

Aさん：もちろんA群です。ハッキリと大きさが分かるから、こんな場合は片側検定補法がいいわけですね。

B君：僕も、そう言いたかったんだ。

X教授：では、平均値の95%信頼区間を求めてみようか。

A群：158.55～166.25

B群：149.31～155.19

のように、信頼区間も重ならないね。

では、再現性を確認するために、B群の別プレートから数え直したコロニー数のデータを

C群：153, 156, 151, 159

として、A群と比べてみよう。どうなるかな。

Studentのt-検定での有意確率は0.02028で有意差があるとなって、95%信頼区間は150.63～158.87となる。少しだけA群の信頼区間と重なってしまうね。

B群とC群を比べると、Studentのt-検定のp-valueは0.2892と、「BとCの平均値に差はない」という帰無仮説を棄却できない、つまり、BもCも同じ母集団からとった標本であると示唆される。

つまり、基本的にStudentのt-検定にしてもWelchのt-検定にしてもまずは両側検定を行う。そして、経験的あるいは原理的に大小関係があらかじめ仮定できる場合だけ片側検定を行ってもいい。データの平均だけを見てから、片側検定にするのは間違いだ。

Aさん：やっと、ハッキリしました。Rですと、どうやって片側検定を行うんですか。

X教授：Rはさっきも言ったように、デフォルトは両側検定なので、片側検定を行うという命令を入力しないといけない。さっきのA群とB群のデータを使うことにして、 $A > B$ という場合を示してみよう。

```
> t.test(A,B,var.equal=T,alternative=c("greater"))
```

だ。alternative = c(" ")の形で指示をする。(" ")の部分にgreaterとすると $A > B$ 。lessとすると $A < B$ を対立仮説に持つことになるんだ。two.sideと入れると両側検定になる。

B君：学会で微妙なデータのポスターがあって、本当に差があるのか質問したんですが、「両側検定では有意差が少しの差で認められませんでした。片側検定では有意差が認められました。そこで、片側検定の結果

を採用して両者に有意差があると考えます」という答えでした。これは間違いですか。

X教授：聞く限り「大間違い」だね。片側検定をするとp-valueが両側検定の半分になる。そこで、有意差が出るほうの検定を採用するなどは本末転倒もはなはだしい。まったく統計が分かっていないものごとだ。まさに、計算結果を「水戸黄門の印籠」と間違ったやり方だね。君たちは、絶対にこのようなことをしないように。

Aさん、B君：はい、分かりました。

X教授：今までの例は“対応のない場合”の検定だったんだが“対応のある場合”というのものもある。

Aさん：まだ、あるんですか。

X教授：今日は時間も遅くなったから次回にしよう。

Aさん、B君：次回も、お願いします。

X教授メモー片側検定と両側検定

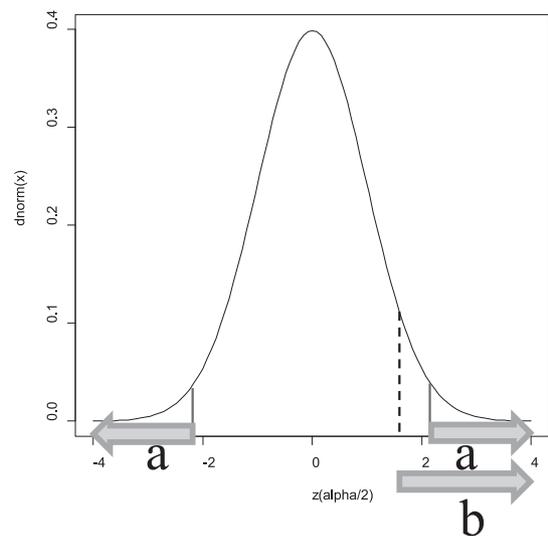


図1. 両側検定 (a) と片側検定 (b). aで示されている領域が両側検定の棄却閾、つまり、aの範囲がt-分布の両端にあり、その面積の合計が有意水準となる。bは $(A > B)$ の場合の片側検定の棄却閾。棄却閾に検定統計量が入ると帰無仮説が棄却されることになる。これを見ると片側検定の方が、検定統計量が棄却に行くに入る可能性が高くなる。つまり、有意差が出やすくなる。

標準的な参考書を紹介しておく

- 1) 向井文雄：生物統計学，化学同人(2011).
- 2) 山田剛史ら：Rによるやさしい統計学，オーム社(2008).
- 3) D.ザルツブルク：統計学を拓いた異彩たち，日経(2010).

(【第5回】は94巻12号に掲載予定です)