



χ^2 検定の使い方？

川瀬 雅也^{1*}・松田 史生²

6月に入り、春にM1になったAさんも大学院生活に慣れてきて、持ち前の好奇心がふつふつとわき上がって来たようだ。B君も留年はせずM2になった。この2人に少しは成長のあとを見てとることができるだろうか。

Aさん：B先輩，早速ですが，質問していいですか。
 B君：大学院生活に大分慣れてきたみたいだね，質問はいつでも，ドンとこいだ。
 Aさん：新4回生にも，そう言っていますよね，大丈夫ですか？
 B君：だんだん，言いたいことを言うようになってきたね？
 Aさん：前からです。日本人の血液型の構成はA：B：AB：Oが，大体，38：22：9：31って聞きましたけど，この研究室はどうなのでしょうか。
 B君：急に血液型に興味が出たの？
 Aさん：この間，資料を見ていたら，血液型を見分ける乳酸菌の話を見つけて，面白そうだなと思ったんです¹⁾。それで，この研究室の血液型の構成はどのかなって，急に気になりだして。
 B君：今のメンバーだけじゃサンプル数不足だから，過去のメンバーのデータも集めてみようか。
 Aさん：そんなことできるんですか？
 B君：秘書のYさんが血液型占いに凝っていてデータを持っているはずだよ。Aさんも血液型聞かれたでしょ？

Yさんはやはり50名のデータを持っており，それによると，A型が15名，B型が9名，AB型が10名，そしてO型が16名であった。

Aさん：日本人の典型的な比率と同じようにも思えるし，少し違うようにも思えますが，どうでしょうか。
 B君：対応のあるt検定を使えばいいんじゃないかな。
 Aさん：うーん，なんか違うような気がするんですよね。
 B君：信用してよ，それじゃ，X教授に僕が正しいことを証明してもらいに行こうよ。

適合度検定

X教授：やあ，久しぶりだね，今日は何かな？
 Aさん：ご無沙汰しています，それで，実は……です。
 X教授：なるほど，B君はなぜ，対応のあるt検定を選んだのかな。
 B君：それはですね，A型の数値同士を比較しないといけないので，対応があると考えたんです。
 X教授：数値に対応があるという点に気が付いたのは，進歩したと言っていいと思うが，後一步，詰めが足りないな。
 Aさん：いつものことですね，こないだも……。
 X教授：まず，対応のあるt検定だが，同じものを異なる二つの方法で測定したときに，方法により得られる数値に違いがあるかどうかを見るような場合に使用する方法で，同じものや事項から得られるデータというのが肝心なところなんだ。
 血液型という見方で言えば，対応のあるt検定でも良さそうに思えるが，君たちの研究室にいた人のデータと，日本人の血液型構成を調べた検査対象者は同じではないだろう。
 B君：確かにそうです。
 Aさん：データを得た対象が違うから対応のあるt検定は使えないんですね，先輩，分かりました？
 B君：はい，大学院に入ると，急に強くなったね。
 Aさん：どんな方法を使えば分かるんですか？
 B君：スルー！
 X教授：統計学の教科書には必ず解説のある，適合度検定だな，この表（表1）を見てくれるかな。
 度数 (n_i) はメンバーの総計 (N) 50名の内訳で，期待確率 (p_i) は日本人の構成比というのは分かるね。

表1. 研究室メンバーの血液型の構成

	A型	B型	AB型	O型	合計
度数	15	9	10	16	50
期待確率	0.38	0.22	0.09	0.31	1
期待度数	19	11	4.5	15.5	50

*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)



期待度数は、もし、標準的な構成比通りなら、各血液型の度数はいくらくらいかを計算したもの、つまり、 Np_i の値だ。これを見るとどうかな。

Aさん：少し、AB型の人が少ないように思えますが。

B君：確かに。

Aさん：他に言うことないんですか。

X教授：まあまあ、あまり先輩をいじめないようにね。次に、教科書的に言えば、 χ^2 値を計算することになる。

$$\chi^2 = \sum_i \frac{(n_i - Np_i)^2}{Np_i}$$

で、度数と期待度数の類似度を示す指標だ。いろいろな場面で χ^2 値が顔を出すので、計算法くらいは覚えておいてもいいかもしれないね。

B君： χ^2 値ですか。

X教授： χ （エックス）ではなくて、ギリシャ文字の χ （カイ）だよ。統計学で出てこなかったかな。

B君：統計学では、半分くらい確率と確率分布の話で、あとはt-検定の計算練習をしたところで終わりましたから、適合度検定は勉強していないんです。

Aさん：私もです。

X教授：どうしても数学の先生は確率分布に力を入れるからな。ここでは教科書的な計算は基本的にはパソコンにさせるので、Rに次のように入力してみよう。

```
>chisq.test(c(15,9,10,16),p=c(0.38,0.22,0.09,0.31))
```

Chi-squared test for given probabilities

data: c(15, 9, 10, 16)

X-squared = 7.9441, df = 3, p-value = 0.04718

【注】 `>chisq.test(c(15,9,10,16),c(19,11,4.5,15.5))`

とするとまったく異なる結果となる。これは、最初に入力したような確率の場合は、与えられたデータが指定の確率になるかどうかを見ているので、データはあくまでも4個というようにRは判断するが、注のように期待度数を入れると、期待度数もデータとRは認識しデータ数は全部で8個となる。データ数の違いは自由度の違いとなるので、 χ^2 検定の判定に用いる χ^2 分布の形が異なってくる。このため、結果が異なってくるわけである。

適合度検定をRで行う場合は、比率は普遍であると考えて確率を指定する方法をとるようにしてほしい。

とになって、標準的な日本人の血液型構成比と一致しているというのは難しいという結論になるね。ただし、有意確率が0.05に非常に近いので「カイ二乗近似は不正確かもしれません」という警告が出るので、この結論としては「君たちの研究室のメンバーの血液型の構成比は、標準的な日本人の血液型構成比とよく一致しているとは言えない」という程度でいいのではないかな。有意確率がもっと下がるか、上がるかすると、この警告は出ないから、有意水準と有意確率の値が近いときは、結論の付け方に注意が必要だ。

Aさん：分かりました。

X教授：適合度検定では χ^2 値を使っているが、もう一つ、この χ^2 値を使う検定があるんだ。

独立性検定

B君：何ですか。教えてください。

X教授：独立性検定という検定法だ。適合度検定と独立性検定では、どちらも χ^2 値を使うので、合わせて χ^2 検定とよばれることが多いね。

B君：初耳です。

Aさん：独立性検定と言われても、さっぱり、どんな検定かイメージができません。

X教授：まず、分割表というのを説明しようかな。社会科学の諸分野ではクロス集計表とよばれることが多いものだね。

たとえば、新薬Aと偽薬Bを多数の患者さんに飲んでもらったときに効果があったかを調べたとしよう（表2）。お薬を飲んだという安心感で症状がよくなるプラセボ効果というのがあるので、何の効果もない偽薬との効果を比較するんだ。

B君：新薬Aは効果があるように見えますけど、偽薬もそこそこ効いてますね。

X教授：独立性の検定というのは、今B君が言ったみたいに、行項目（ここでは新薬Aか偽薬Bか）と列項目（効果有、無）の間に関連性があるかどうかを検定するもので、この例では、新薬Aは偽薬Bより効果があっ

表2. 新薬Aと偽薬Bの効果

	症状の改善	
	あり	なし
新薬A	200	125
偽薬B	56	250

この表は2行2列であるので2×2分割表とよばれる。一般に、m行n列の場合はm×n分割表となる。

X教授：有意確率（p-value）を見ると有意差があるこ

たかということだね。

独立性検定の帰無仮説は、行項目と列項目は独立している、つまり、関連性はないということなので、この点はしっかりと押さえておく必要がある。

実際に計算してみよう。

```
> data <- matrix(c(200,125,56,250),ncol=2,byrow=T)
> data
  [,1] [,2]
[1,] 200 125
[2,]  56 250
```

ncolは列数で、byrow=Tはデータが行方向のデータであることを示している。この命令を入れてやると分割表と同じ形になる訳だ。

```
> chisq.test(data)
Pearson's Chi-squared test with Yates' continuity
correction
data: data
X-squared = 120.42, df = 1, p-value < 2.2e-16
```

となって、有意差があると出るので新薬Aは偽薬Bにくらべて効果があると言っていいことが分かるね。

X教授：適合度検定は理論から導かれた予測値と、実験で観測された実測値が当てはまっているかどうかを議論できるので、生物工学分野でも出芽酵母の四分子解析などで大活躍したんだ。独立性検定は、例のように新薬の効果を調べる時などによく使われている。2群間の有意差検定に比べると、最近はある程度活用例を見かけないけど、使い次第ではとても有力な方法になる。たとえば、新たな形質転換法の有効性を示すとか、マイクロアレイ解析で遺伝子発現に変化があったのか調べる時に使えるんじゃないかな。ぜひ自分の研究にも活用してほしい。

ノンパラメトリックな手法

Aさん：ところで以前、2群間の有意差検定の時に、t検定は母集団が正規分布に従うことが前提だ、と教わりました。けど、データが正規分布にならないケースもあると思うんです。その場合どうしたらいいか教えて貰えませんか。

X教授：確かに、そこまで踏み込んだ統計学の講義は少ないよね。ノンパラメトリックな手法を用いれば、データに正規性がない場合でも検定を行うことができる。

2群間の有意差検定の場合はWilcoxon検定がよく使われている。Wilcoxon検定は比較する2群の分布がどんなものでもいいけど、同じ形であることが前提になるんだ。おそらく生物の研究対象であるなら、まったく分布の形が異なる2群を比較することはないと思うので、この方法を知っておくだけで十分だと思う。

B君：確かにそうだと思います。

X教授：Wilcoxon検定の原理は、データを群に関係なく小さなものから順に並べて、それぞれの順位が重なっているかどうかを見るということだ。生物工学分野ではほとんどの場合2群間で対応がない場合のWilcoxonの順位和検定というものを使う(Mann-WhitneyのU検定という時もある)。

例として、次の2群のデータ間に有意差があるかどうかを考えよう。

V群：20,18,15,13,10,6

W群：17,16,12,9,8,4

データ数が少ないので、正規性の検定をしてもハッキリしたことが分からないから、一度、箱ひげ図(図1)を書いて分布を見てみよう。

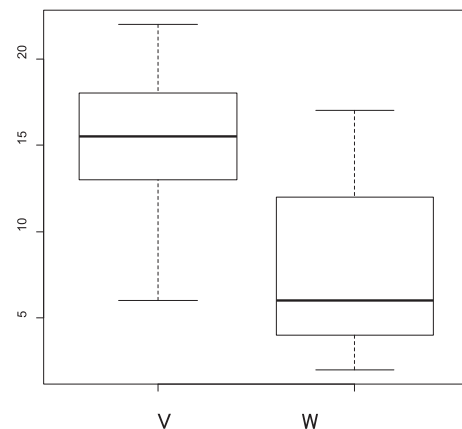


図1. VおよびWの箱ひげ図

```
> V <- c(22,18,15,13,16,6)
> W <- c(17,5,2,7,12,4)
> boxplot(V,W)
```

Wは少し正規分布からずれているように思うね。だが、さっきも言ったように、データ数が少ないので、ここでは正規性のチェックはせずに、Wilcoxon検定を行うことにするよ。



まず、Wilcoxonの順位和検定から

```
> wilcox.test(V,W)
Wilcoxon rank sum test
data: V and W
W = 30, p-value = 0.06494
alternative hypothesis: true location shift is not equal to 0
```

つぎに、同じデータでt-検定をすると、

```
> t.test(V,W)
Welch Two Sample t-test
data: V and W
t = 2.2557, df = 9.976, p-value = 0.04778
alternative hypothesis: true difference in means is not equal to 0
```

Aさん：有意水準 α を0.95とすると、t-検定だと有意差あり、Wilcoxonの順位和検定では有意差なしになってしまいますね。私たちの研究では、データ数が少ないのはいつものことですし、t-検定でいいのかWilcoxon検定にしなければならないのか、どう見分ければいいんですか。

X教授：難しい質問だが、統計学的に判断するより、科学的に判断すべき問題だと思うよ。類似のデータを扱った研究がいくつもあると思うから、その論文を読んで、データにどんな特徴があるのかを考えるといいと思う。単に、論文でこう検定しているからではなく、正規分布と考えても矛盾がないのかといった点に注目してデータを見るといいと思うね。

B君：こういう意味でも、関連する論文を読むことが大切なんですね。でもRの入力の仕方は覚えたし、自分のデータで試せそうです。

Aさん：ソフトの使い方だけ覚えて、統計ができる気になるなんて。本当に、教えてもらった方法を理解できたんですか？

B君：それ、どう言う意味？

X教授：A born fool is never cured. Throw in the towel. でもいいかな。

B君：……。

X教授：ソフトの使い方を習得すると計算はできるが、結果の解釈はできないね。統計では、計算結果の解釈が重要だと何度も言ってきたと思うが。

Aさん：先輩、今年も反省ばかりですね。

B君：申し訳ありません。

条件付き確率

X教授：統計処理法の理解も大事だが、統計的な感覚を持っていることも大事だ。2人の統計的なセンスを試すために、次の問題を考えてみようか。

【演習問題】

Zさんが、画期的な細菌Jの簡易検出法を開発した。細菌Jは、毒性が強いため注意が必要な細菌であるが、これまで感度のいい検出法がなかったので、検査場でサンプリングしたDNAをPCRで検査して、細菌Jの有無の判定を行ってきた。

Zさんの方法では、細菌Jが一つでもいた場合、98%の確率で検出が可能である。細菌Jがまったく存在しない場合、間違って存在するとしてしまう確率は5%である。実際に、細菌Jが検査する場所に存在する確率は経験的に4%であるとされている。

ある場所を、Zさんの方法で検査したところ、細菌Jを検出した。どう考えればいいか。

B君：98%って非常に感度が高い方法で見つかったのですから、すぐに、消毒したほうがいいんじゃないですか？

Aさん：待ってください。もっと慎重に考えないと。

X教授：これから会議なんだ。次回までの宿題にしておこうか。

この連載を読んでいただいている皆さんは、どう考えられるでしょうか。AさんとB君と一緒に考えてみてください。

参考文献

1) 内田英明ら：生物工学, 85, 75 (2007).

以下の本も、統計的な感覚とはどのようなものかを知るのに役に立つと思うので、機会があれば手にとって見ていただきたい。

2) ダレル・ハフ著、高木秀玄訳：統計でウソをつく法、講談社ブルーバックス (1968).

3) 神永正博：ウソを見破る統計学、講談社ブルーバックス (2011).

(【第9回】は95巻8号に掲載予定です)