

検定: 統計学の考え方

1 はじめに

2015年1月14日の講義で、私は超能力の実験を実施しました¹。私には、指先で色を感じる能力があるように思うのです。そこで、実験をすることにしました。

実験には5枚の札を使いました。札の片面は白、もう片面は黒です。札の表面はプラスチックで被われるようにしたので、感触では区別がつかないようにになっています。これを、私が目でみてわからないように黒い袋に入れて、黒い面を上にしながらか取り出すことができれば、指で色(少なくとも明るさ)を判定できると考えることができます。実験の結果、5枚中4枚は黒を上にして取り出すことができました。しかし、1枚だけは白が上でした。これは、学生の皆さんの前でやった実験ですから、当日の受講生は証言してくれると思います。

2 超能力を証明するために

2.1 実験結果の解釈

さて、ここまで書いたような実験結果から、私には常人には無いと思われるような超能力があると、判断できるか考えてみましょう。目で見ていても、うっかり表裏を間違えて取り出すこともあるくらいです。黒い面を出すことが100%できなくてもいいでしょう。それに、普通はこのようなことが全くできないと思われるのに、5枚中4枚もうまくいったので、私は指先で黒と白を判別できると考えていいのではないのでしょうか。

これに対して、もちろん、反論がありえます。それは偶然そうっただけなので、これでは指先で色を判別したことにはならないという主張です。

さて、どうしたものでしょうか。どちらの考え方も、もっともであるように思われます。せっかく実験したのに、結局は何も結論が得られないような実験をしてしまったのでしょうか。私の超能力について何もいうことができないのでしょうか。

2.2 統計学の考え方-検定-

そんなはずはありません。せっかくデータを出したのだから、このデータを基にして議論することができるはず。そのための、最初のステップとして、私が超能力を持っていることを証明す

¹ ちなみに、目以外にも、人間の脳には光を感じる組織があると聞いたことがあります。少なくとも進化の段階をさかのぼると、光を感じる組織だったものが脳にはあります(松果体というようです)。もしかしたら、人間には常人でも指先(あるいは皮膚)で光を感じるができるのではないかと、という議論もあります。つまり、超能力ではなく、常人が備えている能力であるということです。ここでは、超能力かどうか、全員ができるのかどうか、などの点は議論しません。ここで議論したいのは、私が、指先で、光(あるいは明るさや色)を感じるができるかどうか、について議論することです。以下ではわかりやすさのために、超能力の実験、と表現することにします。

る、という発想を、一度、捨ててしまいましょう。逆に、私は超能力が無い人間であるという仮説を立てて議論することを試みます。

このような仮説は、私が超能力を持っている場合には、否定されるべき仮説です。否定されることを想定している仮説なので帰無仮説と名づけられています。

さて、超能力がないとしたら、黒い面を上にして札を取り出すのも、白い面を上にして札を取り出すのも、同じ確率で起こるはずで、つまり、どちらも確率 0.5 (50%) で発生する出来事です。確率が決定 (計算) できるような出来事は、統計学では事象といいます。1枚1枚について、確率が決まっているとすれば、5枚中4枚が黒となる事象の確率も計算することができます。その計算方法は決して難しくありません。しかし、ここでは、その計算をすることが目的ではないので、途中は省略します。結果は約 15.6% です。つまり、超能力が全く無かったとしても、100回中 15~16回は発生するような事象であった、あるいは、7回に1回 (確率約 14.3%) 以上発生するような事象であった、ということです。このようなことを考えると、超能力がないのに5枚中4枚が黒になるのは、起こりにくいことであることがわかります。しかし、滅多に起こらないこと、ほとんどありえないこと、ではなく、わりと起こりうる事である、とも言えます。

結局、断言はできません。しかし、ここまでくると、話は具体的になりました。少なくとも、議論しやすい数字になったのは間違いありません。そして、ここから先は、人間が作る基準によって、判断が分かれることとなります。その基準とは、通常、2通りです。発生確率 5%(20回に1回)以下の事象が起こったのかどうか、あるいは、発生確率 1%(100回に1回)以下の事象が起こったのかどうか、です。5%以下とか1%以下とかが、「滅多に起こることではないことが起こった」と判断する基準になっているのです。

ここから先の議論のパターンは決まっていますので、それにのっとって、以下、議論していきます。まず、今回の実験結果を考えてみましょう。5枚について実験して、4枚の黒が表になる確率は、4枚のときと5枚の時を合わせると、約 18.8%です。これは、1%どころか、5%を越えており、先ほどの説明した通常の基準からすると、「帰無仮説滅 (超能力は無い) が正しいとした場合に、滅多に起こらないようなことが起きた、とは言えない」となります。これを、帰無仮説は、危険率 (あるいは有意水準) 5%で棄却されなかったと表現します。5%以下の確率で起こるようなことが起こったんだったらともかく、今回の事象はもっと起こることだから、偶然で起こったと考えても説明できるよね、という話です。上の表現はこういうことを言っています。

ここで、気をつけたいことがあります。帰無仮説が否定されなかった = 帰無仮説が正しかった、ではありません。帰無仮説 (超能力がない) を否定する (超能力があると認める) ができなかつただけです。帰無仮説は、議論の都合上、作った仮説であることを忘れてはいけません。今回の場合、この帰無仮説に積極的な意味はありません。

次に、もしも、5枚中5枚を黒を表にして出すことができたとして考えてみましょう。帰無仮説に基づく確率は3%しかありません。ということは、5%を基準に考えれば、滅多に起きないことが起きた、ということになります。これを帰無仮説は、危険率 (あるいは有意水準) 5%で棄却されたと表現します。5%以下の確率で偶然そうなったと考えることもできけど、それは考えにくいから、帰無仮説 (超能力はない) は否定されたと考えよう、という話です。

ここでも気をつけたいことがあります。まず、危険率という言葉が表すように、たまたま、滅多

に発生しない事象が現れたのだ、という可能性もあることです。これは、しっかりと、肝に命じておく必要があります。

もう一つは、危険率(あるいは有意水準)を1%に設定した場合との兼ね合いです。今回の場合、1%に設定すると帰無仮説は棄却されない(否定されない)ことになります。基準が変われば、話は変わってきますから注意して下さい。通常は、危険率5%を使って帰無仮説を棄却し、もしも、危険率1%でも帰無仮説が棄却できるなら、それを用いる、という態度でいいと思います。危険率が低いほど、滅多に起こらない事象が発生したことになるので、もしも帰無仮説が棄却できるなら説得力が増します。

以上のような手続きは検定(統計検定)と呼ばれています。また、ここで説明したような考え方や表現は、私たちが流用すべきパターンであることを強調しておきます。微妙な問題を扱う場合に、どのように扱ったらいいのか、先人が悩みながら努力して作り上げたパターンですから、ありがたく使わせてもらいましょう。このパターンを使って表現することで、他の人に正確に伝わるようになります。また、他の人から伝えられた場合に、きちんと受け止められるように、意味を理解しておきましょう。

2.3 統計学の考え方-推定-

統計学で、もう一つ大切な考え方は推定の考え方です。

もう一度、今回の超能力実験について考えてみましょう。超能力があったとしても、完全に目で見たように判別できるとは思えませんから、例えば、「80%の確率で判別できる」とか、そういった話になることが、あらかじめ予想できます。その確率はどのように求めることができるでしょうか。

今回は5枚とりだして黒が上であるような札が何枚あるかを調べました。これを何度も繰り返せば、真の確率に近づいていくことが予想されます。無限回行えば、真の値に到達するでしょう(というより、無限回行った時の値を真の値と言わないで、何を真の値と言うべきでしょうか)。ところが、実際には私たちは無限回の実験を行うことは決してできません。それは、つまり、真の確率を求めることができないということです。

検定の時と同じように、ここで虚しさを感じてあきらめてしまっってはいけません。詳細は述べませんが、ここでも統計学が情報を与えてくれます。統計学の(区間)推定という考え方は、真の値がどの範囲にあるかについて教えてくれます。適当な仮定の下で、例えば、95%の確率でこの範囲(これを信頼区間といいます。)に真の値がある、といった形で情報を与えてくれます。検定の時と同じように、ある一定の確率で、というところが重要です。また、その意味がわかれば、95%の信頼区間よりも、99%の信頼区間の方が広いことはすぐにわかるでしょう。

今回のような実験の場合、95%の信頼区間で、私が黒の面を出せる真の確率は約28~99%の中にあると計算できます。検定の際の帰無仮説である確率50%がこの範囲の中に入っていることに注目しましょう。今回の実験結果は、黒い面を出せる確率50%で考えても、起こりうる(5%以上の確率で生じる)ものと考えられるのです。

ただし、今回の実験は、明らかにデータ数が少なすぎるように思われます。データ数が少なすぎるかどうかを判定する方法も、区間推定で考えるべき論点です。

2.4 より有意義な値を得るために

私の超能力の実験に関連して、5枚中4枚以上が黒い面になる確率は、超能力が無い(同じ確率で白と黒が表になる)と仮定すると、18.8%程度であると説明しました。この実験を10枚で行ったとしましょう。その時に、8枚以上が黒い面を上にして取り出せたとしたら、同じように考えた確率はいくつになるでしょうか。5枚中4枚も、10枚中8枚も、同じ割合(8割)ですから、同じ値18.8%になるでしょうか。これも計算できるので、その計算結果だけ説明しましょう。答えは、約5.5%です。

一般的に、ある事象が発生する確率は、回数を繰り返せば繰り返すほど、真の値に近づきます。今の場合、超能力が無いと仮定していますから、白と黒が半々になるような状態に近づくことになります。半々から外れた状態は、生じにくくなるわけです。帰無仮説は、超能力が無いということでした。そこで、もしも本当に超能力があるのであれば、回数を増やせば増やすほど、帰無仮説と現実との違いがはっきりしてくるはずで

当たり前の話ですね。でも、しっかりとここで確認しておきましょう。

3 まとめ

理科の各分野や、社会科学の各分野では、「その原因はこれだ」と断言できる研究は少なくなっています。そのような研究は、既に行われてしまっていることが多いからです。また、コンピュータで多くのデータを処理できるようになった現代では、データを駆使して物事の間を議論するような研究が増えてからでもあります。

データを用いたこのような研究の場合の多くは、白黒はっきりさせるようなことが困難です。例えば、今回の超能力実験がそうです。そしてその時に、統計学の考え方が大活躍します。

今回は、統計学そのものではなく、統計学の考え方について説明しました。ある意味、統計学の入り口でもっともわかりにくいところだと思うからです。今回の学習でその発想を理解してもらえるといいと思います。