

検定と検出力

検定は、くじの当たりはずれや、無作為抽出による標本など、偶然によって得られるデータをもとにした推論ですから、結論を誤る可能性がつねにあります。例えば、本当に確率 50% で当たるくじであっても、それを 20 本ひいたときの当たりの数が、偶然非常に少なくなることはあります。そのときは、「『確率 50% で当たる』というのはウソ」という、間違った結論が出てしまいます。

実は、検定については、「検定は、データの数が少ないときに、そんなときでもいえる結論を出すもの」ということが言われています。今回の講義では、それがどういう意味なのかを、「第 1 種と第 2 種の誤り」「検出力」という考え方を示しながら説明します。

第 1 種の誤り

「有意水準 5% の検定」の論理を、くじびきの例を使ってひとことで説明すると、下のようになります。

あるくじが、確率 50% で当たる、と仮に信じる。このとき、ある本数だけくじをひいたときの当たりの本数が少なく、その本数以下しか当たらない確率が 5% 以下ならば、「確率 50% で当たる」と信じたのは間違いだった、と結論する。

さて、ここで、「実はそのくじは、本当に確率 50% で当たるものだった」としましょう。そのときでも、当たり本数が偶然少なかったために、棄却してしまう可能性はあります。その確率はいくらでしょうか？

上に書いたように、有意水準 5% の検定で棄却するのは、「そのくじが確率 50% で当たると信じたとして、当たり本数が確率 5% でしか起きないくらい少ないとき」です。それならば、本当に確率 50% で当たるくじでも、それほど当たりが少なくなる確率は 5% であるはずですが、

このように「本当は帰無仮説が正しいのに、棄却してはいけないのに、誤って棄却してしまう」ことを、**第 1 種の誤り**とといいます。第 1 種の誤りをおかす確率 (α で表します) は、上の通り、有意水準と同じです。有意水準 5% の検定は「帰無仮説が実は正しいとしたときに、5% の確率でしか起きないはずのことが、偶然いま起きている」ならば帰無仮説を棄却します。ですから、帰無仮説が本当に正しいときも、5% の確率で起きることがいま起きれば、帰無仮説を棄却してしまうわけです。

では、逆に、「当たる確率が 50% よりも小さいので、帰無仮説を棄却しなければならないのに、棄却せずに見逃してしまう」という誤りを考えましょう。当たる確率が 50% より小さくても、当たり本数が偶然多かったために、棄却しない可能性があります。このような「対立仮説が正しく、帰無仮説は棄却すべきなのに、棄却しない」という誤りを**第 2 種の誤り**とといいます。その確率 (β で表します) を考えてみましょう。

検定の考えを説明するのに、この講義の前半で用いたのは、

あるくじは、50% の確率で当たるとします。このくじを 20 本ひくと、6 本当たって 14 本ははずれでした。このくじは、本当に 50% の確率で当たるのでしょうか。当たる確率をもっと小さいのではないのでしょうか？

という問題でした。このとき、当たり本数を S 、当選確率を p 、くじの本数を n とすると、

$$Z = \frac{S - np}{\sqrt{np(1-p)}} \quad (1)$$

は標準正規分布 $N(0, 1)$ にしたがうことを説明しました。今後の説明のために、当たり本数 S のかわりに、当たりの割合 S/n を考え、これを \hat{p} とおくことにします。(1) 式の分母分子を n で割ると

$$Z = \frac{\frac{S}{n} - p}{\frac{\sqrt{np(1-p)}}{n}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \quad (2)$$

となります。

何度も述べていますが、有意水準 5% の検定では、

仮に帰無仮説が正しい ($p = 0.5$) とするとき、当たり割合 (\hat{p}) が、確率 5% でしかおきないくらい小さいならば、帰無仮説を棄却する

というルールで、帰無仮説を棄却します。一方、標準正規分布の数表によると、 $Z \leq -1.64$ となる確率が 5% です。したがって、上のルールを言い換えると、

仮に帰無仮説が正しい ($p = 0.5$) とするとき、当たり割合 (\hat{p}) が、 $Z \leq -1.64$ となるくらい小さいならば、帰無仮説を棄却する

こととなります。この問題では、くじを 20 本ひいていますから ($n = 20$)、(2) 式に、 $p = 0.5, n = 20$ を代入すると、 $Z \leq -1.64$ となるのは、 $\hat{p} \leq 0.316$ のときであることがわかります。つまり、この問題では、

くじを 20 本ひいてみて、当たり割合 \hat{p} が 0.316 以下ならば、帰無仮説を棄却する

というルールで、帰無仮説を棄却します。

さて、帰無仮説ではなく対立仮説のほうが正しい場合に、この検定で帰無仮説を棄却する、すなわち「当たりの割合 \hat{p} が 0.316 以下となる」確率を調べましょう。

ここでの対立仮説は、「当選確率をもっと小さい」、すなわち $p < 0.5$ です。ひとくちに $p < 0.5$ といってもいろいろな場合がありますから、次の 2 つの場合を考えてみましょう。

1. 本当の当たり確率が、 $p = 0.5$ よりも十分小さい場合

例えば、 $p = 0.2$ である場合を考えてみましょう。(2) 式に $p = 0.2$ と $n = 20$ を代入すると

$$Z = \frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{20}}} \quad (3)$$

となります。

今回の検定では、「 $\hat{p} \leq 0.316$ ならば帰無仮説を棄却する」というルールで、帰無仮説を棄却するかどうかを決めています。 $\hat{p} = 0.316$ のとき、(3) 式に $\hat{p} = 0.316$ を代入すると、 $Z = +1.30$ となります。

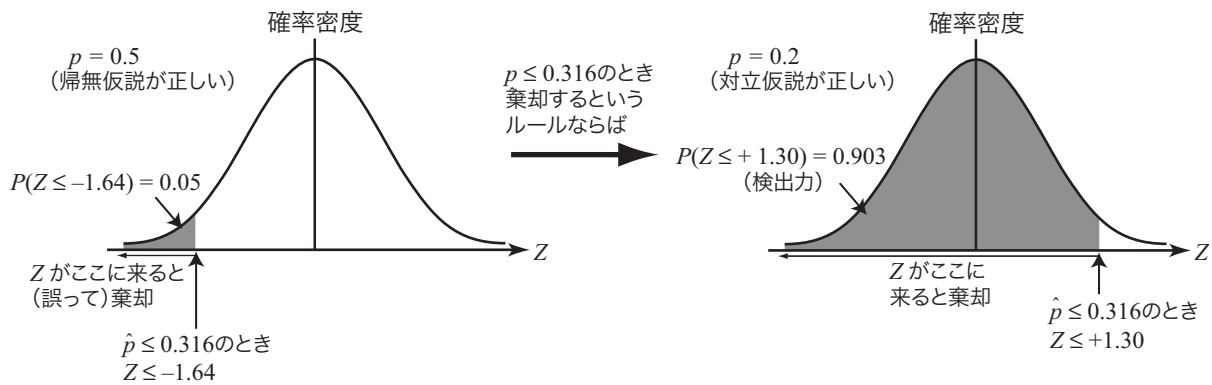


図 1: 検出力

つまり、「 $p = 0.2$ のときは、 $Z \leq +1.30$ ならば帰無仮説を棄却する」ということがわかります。標準正規分布の数表から、そうなる確率は $P(Z \leq +1.30) = 1 - P(Z \geq +1.30) = 1 - 0.0968 = 0.903$ であることがわかります。つまり、「対立仮説が正しく、 $p = 0.2$ であるとき、正しく帰無仮説を棄却する確率」は 0.903 です。

この値を**検出力**（または**検定力**）といいます。また、帰無仮説を棄却しない確率、すなわち第 2 種の誤りの確率は $1 - 0.903$ となります。

2. 本当の当たり確率が、 $p = 0.5$ に近い場合

今度は、 $p = 0.49$ である場合、すなわち「 $p < 0.5$ という対立仮説は確かに正しいが、帰無仮説との食い違いはわずかである場合」を考えましょう。

このときも、上と同様に、(2) 式に $p = 0.49, n = 20$ 、それに $\hat{p} = 0.316$ を代入すると、 $Z = -1.55$ となります。

つまり、 $p = 0.49$ のときは、「 $Z \leq -1.55$ ならば帰無仮説を棄却する」ということになり、その確率は $P(Z \leq -1.55) = 0.0606$ となります。つまり、 $p = 0.49$ のときの検出力、すなわち「 $p = 0.49$ であるとき、正しく帰無仮説を棄却する確率」は 0.0606 です。

これらのことからわかるように、検出力は、同じように「対立仮説が正しい」場合でも、 p の値に依存して変化します。

これは当然といえば当然で、対立仮説が $p = 0.2$ のときのように、「当たり確率が帰無仮説 ($p = 0.5$) よりも十分小さい」ときは、自信をもって「帰無仮説は誤りである」といえます。このときは、帰無仮説を棄却できるほど当たり本数が少なくなる可能性が高いため、検出力は高く、第 2 種の誤りの確率は小さくなります。

一方、同じように「対立仮説が正しい」場合でも、 $p = 0.49$ のように帰無仮説 ($p = 0.5$) に非常に近い状態のときは、「帰無仮説は誤りである」とはつきりとは言いにくいわけです。このときには、当たり本数が偶然多くなって帰無仮説を棄却できない可能性も高いため、検出力は低く、第 2 種の誤りの確率は大きくなります。

標本サイズと検出力

ここまでの話では、標本サイズ（くじの本数） n は、 $n = 20$ で一定としていました。では、標本サイズが変化すると、検出力はどうかわるのでしょうか。

そこで、くじの本数をずっと多くして、 $n = 500$ の場合を考えてみましょう。もう一度、前節の最初にもどります。今の問題では、

仮に帰無仮説が正しい($p = 0.5$)とするとき、当たり割合(\hat{p})が、 $Z \leq -1.64$ となるくらい小さいならば、帰無仮説を棄却する

というルールで、帰無仮説を棄却します。

今度はくじを500本ひいていますから、(2)式に $p = 0.5, n = 500$ を代入すると、 $Z \leq -1.64$ となるのは、今度は $\hat{p} \leq 0.463$ のときであることがわかります。

つまり、今度は、

くじを500本ひいてみて、当たり割合 \hat{p} が0.463以下ならば、帰無仮説を棄却する

というルールで、帰無仮説を棄却します。

くじを20本ひいたときは、「当たり割合 \hat{p} が0.316以下ならば、帰無仮説を棄却する」というルールでした。このことは、くじが500本の場合には、20本の場合に比べて、当たりくじの割合が0.5により近い場合でも棄却することを示しています。

つまり、くじの本数が少なければ、当たりくじの割合が0.5よりもとても小さいときだけ棄却します。しかし、くじの本数が多くなると、本当に確率50%で当たるのなら、当たりくじの割合もほとんど50%になるはずですから、当たりくじの割合が0.5よりも少し小さいだけでも、「確率50%で当たる」という帰無仮説を棄却します。

このとき、さきほどと同様に、対立仮説のほうが正しいけれども帰無仮説に十分近い、 $p = 0.49$ である場合を考えてみましょう。(2)式に $p = 0.49, n = 500$ を代入すると

$$Z = \frac{\hat{p} - 0.49}{\sqrt{\frac{0.49(1-0.49)}{500}}} \quad (4)$$

となります。今度は「 $\hat{p} \leq 0.463$ ならば帰無仮説を棄却する」というルールです。そこで、(4)式に $\hat{p} = 0.463$ を代入すると、 $Z = -1.21$ となります。つまり、 $p = 0.49, n = 500$ のときは、「 $Z \leq -1.21$ ならば帰無仮説を棄却する」となり、その確率 $P(Z \leq -1.21) = 0.113$ となります。

つまり、ひいたくじの本数 $n = 500$ のときは、 $p = 0.49$ のときの検出力、すなわち「 $p = 0.49$ であるとき、正しく帰無仮説を棄却する確率」は0.113となります。

この結果を、 $n = 20$ の場合と比べてみましょう。「 $p = 0.49$ であるとき、正しく帰無仮説を棄却する確率」は、 $n = 20$ のときは0.0606、 $n = 500$ のときは0.113です。

このように、本当の当たり確率 p が同じ、すなわち対立仮説の正しさの程度（帰無仮説からの離れ方）が同じであっても、**くじの本数（標本サイズ） n が大きくなると、検出力は大きくなります。**

つまり、ひいてみたくじの本数が少ない（標本サイズが小さい）ときは、当たり確率 p の値が帰無仮説（ここでは $p = 0.5$ ）から十分に大きくかけ離れていないと、帰無仮説を棄却し対立仮説（ここでは $p < 0.5$ ）を採択する（つまり対立仮説の正しさを支持する）可能性はほとんどありません。

一方、くじの本数が多くなると、当たり確率の値がほんの少しでも帰無仮説と異なっていれば、帰無仮説を棄却し、「当たり確率をもっと小さい」という対立仮説を採択する可能性が大きくなります。

何かおかしいと思いませんか

検定では、

帰無仮説を棄却→対立仮説は正しい（積極的結論）

帰無仮説を棄却できない→対立仮説は正しい、とはいいい切れない（消極的結論）

という論理になっています。つまり、帰無仮説は『内心』では棄却されることが期待されています。目論見通り棄却されると、『対立仮説を採択する』という積極的な結論が得られます。

ところが、前節の説明では「当たり確率 p が、少しでも帰無仮説での値（ $p = 0.5$ ）より小さければ、ひいてみたくじの本数が多くなると、帰無仮説が棄却される確率は大きくなる」といっています。このことは、「くじをいくらでも多くひけるのなら、当たり確率 p が少しでも 0.5 より小さければ、対立仮説を採択することができる」ことを意味しています。

つまり、「対立仮説は $p < 0.5$ である」としたとき、たとえ $p = 0.49999$ であっても、くじをいくらでも多くひけるのなら、ほぼ確実に対立仮説を採択し「 $p < 0.5$ である」「『50%当たるというのはウソだ』」と言える、ということになります。つまり、「くじの本数を増やすことによっていくらでも思うような結論が出せる」ということになってしまいます。

くじの本数は、ひく人が自由には決められないかもしれませんが、第Ⅷ部で説明した標本調査における仮説検定の場合は、くじの本数に相当するのは標本サイズです。それは、調査予算が許すかぎり、調査する人が自由に決められます。つまり、「調査する人の都合で、思うような結論が出せる」ことになってしまいます。これはおかしくないでしょうか？

このことは、「対立仮説が正しい」とは何を意味しているのか、という問題をよく考えなければならぬということの意味しています。「対立仮説は $p < 0.5$ である」としたとき、数学としては、 p が 0.5 よりも少しでも小さければ、対立仮説は正しいと言えます。しかし、現実の問題では、 p が 0.5 からどのくらい離れていれば、確かに対立仮説は正しいといえるのかを考える必要があります。

$p = 0.49$ のとき、対立仮説は正しいのでしょうか？ $p = 0.49999$ なら、対立仮説は正しいと本当にいえるのでしょうか？ それを決めるのは統計学ではありません。「 p がどのくらい帰無仮説からかけ離れていれば、確かに帰無仮説が間違っていると言えるのか」を決めるのは、いま統計学を適用しようとしている対象の性質の問題です。

ですから、ある問題で仮説検定を行ない、「有意水準 5% で帰無仮説が棄却された」とだけ答えても、本当は不十分です。くじの数や標本サイズと、今日の例なら p の各値に対応する検出力の値の関係¹を示し、「この問題において、帰無仮説を棄却するのに十分なほどの、 p の帰無仮説からの離れ方はどのくら

¹これをグラフにしたものを「検出力曲線」といいます。

いで、そのときの検出力はいくらか」を示さなければなりません。少なくとも、検定の結果に標本サイズが示されていないと、読者が自分で検出力を求めることもできないので、検定の結果には疑いがもたれることとなります。

とくに、標本サイズが極端に大きいと、対立仮説の帰無仮説からの隔たりがわずかでも検出力が高くなりすぎ、帰無仮説との隔たりが実質的に意味のないほどわずかでも帰無仮説を棄却してしまうので、無意味な結論を導いてしまうおそれがあります。これが、最初に述べた「そもそも検定は、少ししかデータを集められないときに、それでも十分に言えることを言う手法である」ということなのです。

このことは、区間推定の場合を考えても理解することができます。第9回で説明した母平均の区間推定で、標本サイズが十分大きければ、信頼区間は十分狭くなり、ほぼ確実に母平均を推定することができます。今回の例で、当たり確率を推定すると考えても、500回もくじをひけば、かなり確実に当たり確率を推定することができます。すなわち、当たり割合 \hat{p} と当たり確率 p は、事実上ほぼ等しいと考えてさしつかえありません。それならば、その推定された当たり確率と、帰無仮説で述べている当たり確率との隔たりとが、実質的な意味のある差であるかどうかを考えればいいのであって、こんな場合に検定をする必要性は薄いといえます。