

分布についての仮説を検証する — 仮説検定 (1)

「統計学」の最後は、**仮説検定** (あるいは**検定**) という考え方について説明します。これは、前回までに説明した「区間推定」と同じような考え方をを用いて、例えば「母平均は 0 でない」「母平均は 0 より大きい」といった、母集団分布についての仮説が、適切かどうかを推測する方法です。

検定の考え方

次の例を考えてみましょう。

店員が「確率 50% で当たる」と宣伝しているくじがあるとします。ところが、あなたがこのくじを 10 回引いてみたところ、1 回も当たりませんでした。

店員は「運が悪かったねー」と言っていますが、あなたはどうも納得がいきません。「『確率 50% で当たる』という宣伝はウソじゃないの?」と思います。さて、店員かあなたか、どちらが正しいのでしょうか?

店員の言っていることが正しいかどうかは、くじ箱を開けて中のくじを全部調べれば、確実にわかります。もちろん、そんなことはふつうはできません。しかし、そのようにして調べない限り、店員がウソをついているのか、それともあなたの運がものすごく悪いのか、結論は出せません。そこで、次のように考えてみます。

店員の宣伝では、1 回のくじ引きで、当たりもはずれも確率は $\frac{1}{2}$ だと言っています。そこで、1 回目のくじ引きと 2 回目のくじ引きが独立とみなせるのであれば、2 回続けてはずれる確率は、それぞれではずれる確率の積で、 $\frac{1}{2} \times \frac{1}{2}$ となります。

同じように考えると、「くじを 10 回引いて 1 回も当たらない」確率は、 $\left(\frac{1}{2}\right)^{10}$ すなわち $\frac{1}{1024}$ ということとなります。つまり、店員の「確率 50% で当たる」という宣伝を信じるのであれば、「くじを 10 回引いて 1 回も当たらない」という結果になる確率は $\frac{1}{1024}$ ということとなります。

確率とは、「すべての可能性のうち、どの結果になりやすいか」の割合を表すものです。ということは、「店員の宣伝を正しいと信じる」ことは、「10 回のくじ引きの結果のすべての可能性のうち、 $\frac{1}{1024}$ という小さな確率でしか起きないことが、たまたま今、目の前で起きている」という考えを受け入れることとなります。そんな無理のある考えを受け入れるよりも、「『確率 50% で当たる』という宣伝のほうが間違っている」と考えるほうが自然ではないでしょうか?

こういう論理で、「『確率 50% で当たる』という宣伝は間違っている」という判断を下すのが、検定の考え方です。検定の論理は基本的にはこれだけで、問題によって異なるのは、「小さな確率でしか起きないことが、今たまたま目の前で起きているなどという考えは、受け入れられない」という考えを導くための、確率の計算のしかたです。次節では、もうすこし実際的な例で、検定の使い方と確率の計算のしかたをみてみましょう。

t 分布と検定

次のような問題を考えてみます。

10 人の実験協力者に、薬 A を与えた場合と薬 B を与えた場合とで、それぞれある検査を行うと、その結果の数値は次の表の通りとなりました。このとき、薬 B は、薬 A よりも、検査の数値を高くする働きがあるといえるでしょうか。

実験協力者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66

検査結果の数値自体は、人によって大きく違います。ここでいう「薬 B のほうが高くなる」というのは、それぞれの実験協力者において、数値がどう変化しているかを問題にしています。そこで、各実験協力者について、数値の差（薬 B での数値引く薬 A での数値）を求めてみます。

実験協力者番号	1	2	3	4	5	6	7	8	9	10
薬 A	60	65	50	70	80	40	30	80	50	60
薬 B	64	63	48	75	83	38	32	83	53	66
差	4	-2	-2	5	3	-2	2	3	3	6

差を見ても、実験協力者によって正だったり負だったり、ばらつきがあります。つまり、薬 B での数値のほうが高くなっている実験協力者もいれば、逆に低くなっている実験協力者もいます。ここで問題にしているのは、差の平均値をもって「薬 B のほうが高くなる」といえるかどうかです。

差の平均値は +2 で、薬 B のほうが平均的には数値が高いことを示していますから、「薬 B のほうが高くなっている」とかといえば、なっていることはなっています。問題は、その差が、偶然のために生じたものではなく、本質的な差であるかどうかです。本質的かどうかなど、どうすればわかるのでしょうか。

この疑問に対して、次のように考えます。

1. 「母集団（ここでは、世界のすべての患者）について『薬 A と薬 B での差』を求めると、平均は 0 になる」と仮説を設定する。つまり、「本質的な差はない」という仮説を設定する。
2. 実験協力者は、母集団から無作為抽出された、10 人からなる標本と考える。
3. 実験協力者 10 人について、「薬 A と薬 B での差」の平均値を求める。
4. 実験協力者 10 人について求められた「薬 A と薬 B での差」が、「本質的な差はない」はずの母集団から標本を無作為抽出したときに、偶然生じる確率を求める。
5. その確率が小さければ、「こんな差が偶然生じるとは思わない」と考える。すなわち、「本質的な差はない」という当初の仮説は誤り、と結論する。

このような考えで、「母集団に関する仮説が正しいとすると、いま標本について得られている結果は、偶然生じたとは考えにくい」→「母集団に関する仮説は間違っている、と考える」という推論を、**仮説検定（検定）**とといいます。次節以降で、最初にあげた例題について、どのように検定を行うかを考えます。

片側検定

では、例題の、薬Bでの結果のほうが、薬Aでの結果よりも「本質的に」高いといえるか、を考えます。ここで、母集団全体での「薬Aと薬Bでの差」は、平均 μ の正規分布にしたがうと考えます。そうすると、前回の「 t 分布にもとづく区間推定」で説明したように、

- 標本サイズを n (例題では 10)
- 標本平均を \bar{X} (例題では、10 人の実験協力者における差の平均値で、+2)
- 不偏分散を s^2 (例題では、10 人の実験協力者についての不偏分散で、計算すると 8.89)

とすると、 t 統計量

$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \quad (1)$$

は、自由度 $n - 1$ の t 分布にしたがいます。

さて、前節の「検定の考え方」の第 1 項にある「母集団について『薬 A と薬 B での差』を求めると、平均は 0 になる」という仮説を考えます。この仮説は、つまり「 $\mu = 0$ 」ということになります。

この仮説が正しいとすると $\mu = 0$ で、また例題では $n = 10$ 、 $\bar{X} = +2$ 、 $s^2 = 8.89$ です。これらの数値を (1) 式に代入すると、 t 統計量の値は $t = +2.121$ となります。

一方、例題では、 t 統計量は自由度 $(10 - 1) = 9$ の t 分布にしたがいます。このとき、 t 統計量が上側 5% 点よりも大きくなる確率は 5% で、自由度 9 のときの t 上側 5% 点、すなわち $t_{0.05}(9)$ は、数表から +1.8331 となります。これらを、 t 分布の確率密度関数のグラフを使って図示したのが、図 1 です。グラフのグレーの部分の面積が 5% で、 t 統計量が上側 5% 点よりも大きくなる確率が 5% であることを示しています。

この図からわかることは、「 $\mu = 0$ 」という仮説が正しいとき、 t 統計量は「確率 5% でしか起きないほど大きな値」になる、ということです。つまり、「母集団について『薬 A と薬 B での差』を求めると、平均は 0 になる」という仮説が正しいとすると、 t 統計量が偶然このような大きな値になる確率は 5% しかなく、偶然とは考えにくい、ということになります。したがって、この仮説は正しくなくて、いまの標本についての差の平均 (+2) は「本質的な差」と考える方が自然、ということです。

では、この仮説が間違いであれば、かわりにどんな結論が得られるのでしょうか。(1) 式を見ると、 μ が 0 でなくもっと大きければ、 t 統計量はもっと小さくなり、図 1 のグレーの領域から外れます。つまり、「確率 5% でしか起きないほど大きな値」ではなくなります。 μ が 0 でなくもっと大きい、すなわち「 $\mu > 0$ 」とは、「薬 B での結果のほうが数値が高い」ということです。つまり、「薬 B での結果のほうが数値が高い、と考える方が自然だ」というのが、この例題に対する答えとなります。

検定の言葉

以上のように、この例題に検定を使って答えました。ただ、検定には独特の用語があり、統計学の教科書ではその用語が使われています。ここでは、例題を使って、用語について説明します。

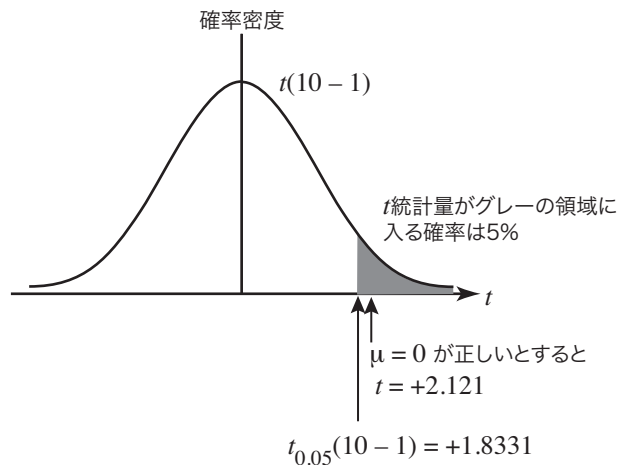


図 1: t 統計量.

例題で、「 $\mu = 0$ 」という仮説は「間違っている」と判断されました。このときの「 $\mu = 0$ 」という仮説を**帰無仮説**といい、 $H_0: \mu = 0$ と表します¹。また、帰無仮説を「間違っている」とした判断を、**帰無仮説を棄却する**といいます。さらに、帰無仮説を棄却した結果、正しいと判断した「 $\mu > 0$ 」という仮説を**対立仮説**といい、 $H_1: \mu > 0$ と表します。この判断を、**対立仮説を採択する**といいます。

上の推論では、「確率5%でしか起きないことが、偶然起きていると考えるのは不合理」と考えています。つまり、5%の確率でしか起きないことが起きたということを説明する時、「偶然起きた」という説明ではなく、帰無仮説が間違っているという「必然」によって起きた、という説明のほうが合理的だ、と考えているのです。偶然ではなく必然的に何かが起きることを「**有意である**」といい、この「5%」を**有意水準**といいます。

例題では、帰無仮説が正しいとするとき、「 t 統計量が $-t_{0.05}(10-1)$ 以上である」ならば帰無仮説を棄却する、という推論をしました。つまり、図1のように、「帰無仮説が正しいとするとき、 t 統計量があそこに入ったら、帰無仮説を棄却する」という区間（グレーの部分）が、 t 分布の確率密度関数で片側（右側）にあります。その意味で、今回のやりかたの検定を**片側検定**といいます。

なお、上の「帰無仮説が正しいとするとき、 t 統計量があそこに入ったら、帰無仮説を棄却する」という区間のことを**棄却域**といい、棄却域を表すのに用いる統計量（ここでは t 統計量）を**検定統計量**といいます。また、検定統計量の値が棄却域に入ること、**棄却域に落ちる**という表現をします。

今日の演習問題

仮説検定とは、どういう状況で何をすることかを、100文字以内で簡潔に教えてください。

¹ H は、hypothesis (仮説) という英語の頭文字です。