2022 年度春学期　統計学　第１５回

分布についての仮説を検証する — 仮説検定（２）

「統計学」最終回は，前回から続いて仮説検定（検定）を説明します。検定による結論がどんなことを言っていて，検定はどんなときにするものなのかを，よく理解してください。

両側検定

では，この例題で，設問を「薬 A を与えた場合と薬 B を与えた場合で，検査の数値に違いがあると言えるでしょうか」に変えてみます。最初の設問との違いは，最初の設問が「薬 B での数値のほうが高くなる」かどうかを知りたいのに対して，今度は「薬 B での数値のほうが，薬 A での数値よりも，高いか低いか，いずれにしても本質的な差がある」かどうかを知りたいわけです。

この場合，帰無仮説はさきほどと同じ「$\mu = 0$」ですが，それが棄却されて得られる対立仮説は，「μ はもっと大きい」だけではなく，「μ はもっと小さい」という場合も考える必要があります。つまり，対立仮説は「$\mu \neq 0$」となります。

帰無仮説「$\mu = 0$」が棄却された時に採択される対立仮説が「μ はもっと大きい」と「μ はもっと小さい」の両方ですから，この検定では，帰無仮説が正しいとしたときに，t 統計量が「大きすぎる」ときも「小さすぎる」ときも，帰無仮説を棄却します。したがって，この検定の棄却域は，図 2 のように，確率密度関数のグラフの両側にあります。有意水準を，前の例と同じく 5%とすると，t 統計量が棄却域に入る確率は両側合わせて 5%で，左右それぞれ 2.5%ずつとなります。つまり，帰無仮説が正しいとするとき，

> t 統計量が，上側 2.5%点より大きいか，または，下側 2.5%点より小さい

ときに，帰無仮説を棄却します。このやり方の検定を，**両側検定**といいます。

帰無仮説「$\mu = 0$」が正しいとするとき，t 統計量の値は，前の片側検定の例と同じく，+2.121 です。一方，上側 2.5%点 $t_{0.025}(9) = +2.2622$，下側 2.5%点は -2.2622 ですから，t 統計量は上側 2.5%点と下側 2.5%点の間にあり，帰無仮説「$\mu = 0$」は棄却されません。したがって，対立仮説「$\mu \neq 0$」は採択されず，「薬 A を与えた場合と薬 B を与えた場合で，検査の数値に違いがある」とはいえない，ということになります。

帰無仮説が棄却されないときは

「帰無仮説が棄却されない」ときは，その理由は「帰無仮説が正しい ($\mu = 0$) とするとき，いま得られているような t 統計量が得られる確率は，非常に小さい（5%）とまではいえない」ということになります。したがって，

> 「帰無仮説が間違っているかどうかはわからない」「対立仮説が採択できるかどうかはわからない」

という結論を導かなくてはなりません。今回の例でいえば，帰無仮説が棄却されなかった場合は，「$\mu = 0$ でないとはいえない」「$\mu = 0$ でないとまで断言する自信はない」という結論になります。

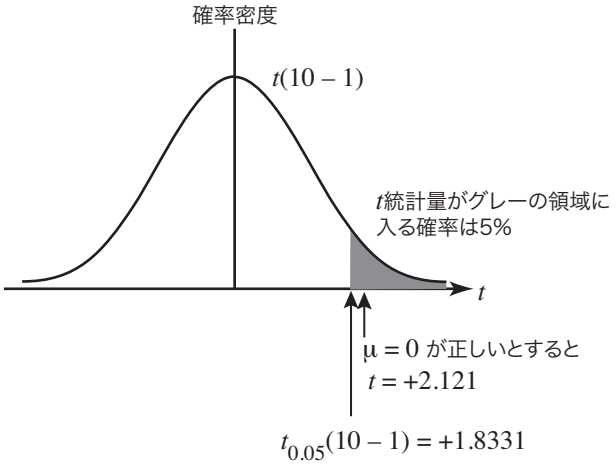

図 1: 片側検定（第 14 回の図を再掲）.

注意しなければならないのは，あくまで，「いま得られているような t 統計量が得られる確率は，非常に小さいとまではいえない」のであって，「確率が大きい」のではない，ということです。したがって，帰無仮説が棄却されなかったときに，「帰無仮説が正しい」「対立仮説は間違っている」という結論が得られるわけではありません。今回の例でも，「$\mu = 0$ である」「薬 A を与えた場合と薬 B を与えた場合で，検査の数値に違いはない」などと答えてはいけません。つまり，

帰無仮説を棄却しない
＝　× 帰無仮説を採択する
　　◯ 対立仮説を採択するべきかどうか断言できない

ということです。なお，「帰無仮説を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**といいます[1]。

有意水準について

ここまでの例で，有意水準は 5%としていました。有意水準の値は，5%あるいは 1%がよく用いられます。

有意水準の値は，検定をする人の「大胆さ・慎重さ」の程度を表しています。

有意水準が大きい (5%) ときは，帰無仮説が仮に正しいとしても，いま起きている現実（t 統計量の値が $+2.121$）が起きる確率が 5% と「そこそこ小さな確率」であれば，「そんなことが起きるはずがない，帰無仮説は間違っている」と結論します。はっきり物を言う態度ではありますが，帰無仮説が実は正しいときでも「間違っている」と断言してしまう可能性があります．大胆ですが，勇み足も多い，というわけです。

有意水準が小さい (1%) ときは，いま起きているような現実が起きる確率が，1%と相当小さくないと，「まあそんなことも起きるかもしれない，帰無仮説は間違っているとは言い切れない」となり，結論を出しません。慎重ですが，煮え切らない態度ということになります。

[1]第２種の誤りを，俗に「ぼんやり者の誤り」といいます。第２種の誤りの確率をしばしば β で表すことにかけています。

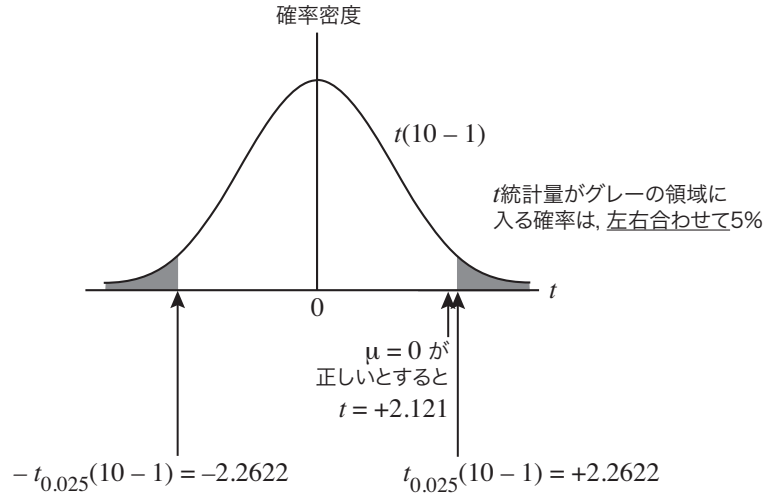

図 2: 両側検定.

片側検定と両側検定

どうもおかしい…

ところで，ここまで見てきた「片側検定」と「両側検定」，おかしくないでしょうか？ 有意水準が 5% なのは同じなのに，

- 「薬 B での数値のほうが高い」かと聞かれたら，「高いと言える」と答える（片側検定）
- 「薬 B での数値と薬 A での数値に違いがある」かと聞かれたら，「違いがあるとは言えない」と答える（両側検定）

という結論になりました。「高いと言える」のに「違いがあるとは言えない」というのも変な話です。

実はおかしくない

これは，おかしくないのです。片側検定と両側検定では，「調べている」こと，つまり検定している内容が違うのです。

検定とは，帰無仮説での想定（例題では，$\mu = 0$，すなわち「薬 B での数値と薬 A の数値に本質的な差はない」）が，現実にデータを調べた結果（つまり標本，あるいは標本から求めた標本平均などの値）と食い違っているかどうかを検査しています。そしてそのような食い違いが，確率 5%でしか起こらないような，つまり偶然とは言えない（有意な）食い違いのとき，帰無仮説での想定は誤りとして，帰無仮説を棄却します。

両側検定の場合

両側検定は，帰無仮説が標本と食い違っているかどうかだけを検査しています。ですから，帰無仮説での想定が，標本に比べて，大きい方に食い違っていても，小さい方に食い違っていても，帰無仮説を

棄却します。今回の例でいえば，帰無仮説でいう「μ=0」が，標本平均に比べて大きすぎても小さすぎても，帰無仮説を棄却します。すなわち，「標本について，薬 B での数値が，薬 A での数値よりも著しく大きい（例えば，標本平均が +10）」場合でも，「著しく小さい（例えば，標本平均が −10）」場合でも，いずれも「薬 B での数値と薬 A の数値に本質的な差はない」という帰無仮説を棄却します。

片側検定の場合

これに対して, 片側検定は, 帰無仮説が標本に比べて, 大きすぎるか, または小さすぎるかのどちらか一方を検査します。ですから，帰無仮説での想定が，標本に比べて「ある一方向に」食い違っているときだけ帰無仮説を棄却します。例題では，対立仮説が「$\mu > 0$」，すなわち「薬 B での数値のほうが，薬 A での数値より高い」という片側検定をしています。つまり，帰無仮説の「$\mu = 0$」が，標本平均（例題では +2）に比べて小さすぎると言えるかどうかだけを検査していますから，帰無仮説の「$\mu = 0$」が標本平均に比べて小さすぎるときだけ，帰無仮説を棄却します。

片側検定は，調べていないことは「見逃す」

では，もし例題で，標本平均が例えば「−10」，つまり標本について「薬 B での数値が，薬 A の数値よりも著しく小さ」くて，帰無仮説の「$\mu = 0$」という想定が標本平均に比べて大きすぎる時はどうなるのでしょうか？

両側検定では，この場合も帰無仮説を棄却します。しかし，対立仮説が「$\mu > 0$」という片側検定では，帰無仮説を棄却しません。この場合も，帰無仮説の「$\mu = 0$」が標本平均と大きく食い違っているにかかわらず，片側検定はそれを見逃し，「帰無仮説を棄却できない」と答えてしまいます。それは，「$\mu = 0$ は標本平均に比べて大きすぎるかどうか」は，この片側検定では検査の対象ではないからです。たとえこれが，「$\mu = 1000$」つまり「薬 B での数値が，薬 A の数値よりも，本質的に 1000 高い」などというとんでもない帰無仮説でも，「そんなことは今検査していることではない」といって棄却しないのです。

同じ「有意水準 5%の検定」でも

片側検定では，棄却域は図 1 のように「片側で確率 5%」に対応する部分です。一方，両側検定では，棄却域は図 2 のように「両側それぞれに確率 2.5%」に対応する部分です。同じ「有意水準 5%の検定」でも，片側だけをみれば，片側検定のほうが棄却域が広くなります。そのために，今回の例題のような，片側検定と両側検定の結果が一見矛盾するようなことがおきます。

片側検定の例題で，「薬 B での数値は，薬 A での数値よりも高い」かどうかを検定しました。この検定をするのは，あらかじめ「薬 B での数値は，薬 A での数値よりも高いだろう」という目論見があるからです。仮に，逆の「薬 B での数値が，薬 A での数値よりも低い」という結果になっても，それは見逃してもかまわないので，そのぶん「大胆」な検定を行います。

一方，両側検定の場合は，「薬 B での数値は，薬 A での数値よりも高い」「薬 B での数値が，薬 A での数値よりも低い」のどちらの場合も見逃してはいけないので，同じ有意水準でも「慎重」な検定となります。

くじびきを例にして考えてみる

片側検定と両側検定の違いを，「くじびき」を例にして考えてみましょう。くじをひくほうの立場からすると，「当たり確率は 50%」と称するくじが「10 回ひいて全部はずれ」れば不満です。しかし，「10 回ひいて全部当たり」の時は，「当たり確率は 50%」というのは正しくないような気はしますが，得をした

のですから，別に不満は持ちません。

一方で，賞品を出すほうの立場に立てば，逆に「10 回ひいて全部当たり」の時は賞品を皆持っていかれて不満ですが，「10 回ひいて全部はずれ」でも，客に「残念でしたね」というだけで，とくに不満は持ちません。

こういうふうに，「当たる確率は 50%」という帰無仮説と現実の当たり数を比べて，現実の当たりが「少なすぎる」という不満，あるいは「多すぎる」という不満の，どちらかだけを検査するのが片側検定です。

ところが，このくじびきを主催している商店街の商店会長からすると，「あそこのくじびきは何かおかしい」という噂が流れると困ります。ですから，現実の当たりが「少なすぎる」ときも「多すぎる」ときも不満です。この両方の不満をとりあげるのが両側検定で，つまり「くじびきが双方にとって公正かどうか」を問題にすることになります。

大事なのは，「どちらの検定をするかは，検定の目的に沿って，データを調べる前に決める」ことです。データを見てから，帰無仮説が棄却されそうな検定を選んではいけません。それは，アンフェアなやりかたです。

検定はどんなときにするものなのか

有意水準 5%の検定では，帰無仮説が仮に正しいとするとき，確率 5%でしか起きないはずのことが起きていることになってしまうのなら，帰無仮説を棄却します。

しかし，「確率 5%でしか起きないはずのこと」は，言い換えれば確率 5%で起きるのであって，確率ゼロではありませんから，それが偶然起きることはあるはずです。ですから，例えばここまでの例題で，母平均 μ が本当に 0 である，つまり「母集団について，薬 A と薬 B での検査結果の差が平均 0 である」という帰無仮説が正しいときでも，得られた標本が偶然母平均から大きくはずれていて，その結果帰無仮説を偶然棄却してしまうことが，確率 5%で起きます。これは間違った判断ですが，このような間違いをする確率が 5%であるわけです。このような間違いを**第 1 種の誤り**といいます[2]。つまり，

> 帰無仮説が本当に正しいとしても，有意水準 5%の仮説検定を何度も行うと，そのうち 5%の場合では第 1 種の誤りを犯して棄却し，採択すべきでない対立仮説を採択してしまう

ことになります。

ですから，同じ現象について何度もデータを集めて，同じ帰無仮説について検定を繰り返し，たまに対立仮説が採択されても，直ちに「帰無仮説は間違っている」とはいえません。例えば，「血液型と性格に関係はない」という帰無仮説について何度もデータを集めて検定を行い，たまに「血液型と性格に関係がある」という結論が出ても，直ちに「やっぱり血液型と性格に関係がある」ということにはなりません。何度も検定を行うと，帰無仮説が間違っていない場合でも，たまに対立仮説が採択されるのは，むしろ自然なことです。血液型と性格の問題でいえば，ごくたまに「血液型と性格に関係がある」という結論が出る程度であれば，「血液型と性格に関係があるとは今のところ言えない」というのが，科学的態度です。

[2]第 1 種の誤りを，俗に「あわて者の誤り」といいます。第 1 種の誤りの確率（＝有意水準）をしばしば α で表すことにかけています。

では，検定の結論は結局何を言っているのでしょうか？ それは，

> 私は，帰無仮説は間違いだ，と判断する。
> ただし，私は 100 回中 5 回はウソを言う（第 1 種の誤りを犯す）人間である。
> 私が今回，本当のことを言っているのか，ウソを言っているのか，それは誰にもわからない。

というのと同じことです。

この程度のことしか言っていないのに，検定にはどういう意味があるのでしょうか？ それは，検定とは，小さなサイズの標本しか調べられず，しかもそれを 1 度だけしか調べられないときに，「それだけのデータからでも十分な確信をもって述べられる疑いだけを述べる」方法ということなのです。何度も検定できるほどデータを集められるのなら，検定を用いるのは不適切です[3]。

今日の演習問題

片側検定と両側検定の 2 種類の検定がありますが，どういう問題のときにどちらを選べばよいのでしょうか。100 文字以内で簡潔に答えてください。

[3]今回例に使った「薬の効果」の問題では，実際の研究では，「薬 A と B でどのくらいの差（効果量）があれば効果があるといえるか」を考えて，標本サイズを決める手続きがあります。