2018 年度春学期　統計学　第１４回

分布についての仮説を検証する — 仮説検定

今回は，**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用いて，例えば「母平均は 100 である」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。

両側検定

前回とりあげた，t 分布を用いた区間推定の例題を，もう一度見てみます。

> ある試験の点数の分布は正規分布であるとします。この試験の受験者から，10 人からなる標本を無作為抽出して，この 10 人の点数を平均したところ 50 点で，またこの 10 人の点数の不偏分散が 5^2 でした。このとき，受験者全体の平均点の 95%信頼区間を求めてください。

この問題の考え方は，次のようなものでした。標本平均を $\overline{X}$，不偏分散を s^2 とすると，$\overline{X} = 50$，$s^2 = 25$ です。標本サイズを $n(= 10)$ とし，受験者全体の平均点を μ とすると，t 統計量

$$t = \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}} \tag{1}$$

は，自由度 $n-1$ の t 分布にしたがいます。そこで，$t_{0.025}(n-1)$ を「自由度 $n-1$ の t 分布において，t 統計量が $t_{0.025}(n-1)$ 以上である確率が 0.025 になるような t の値」（2.5 パーセント点）とすると，

$$P\left(-t_{0.025}(n-1) \leqq \frac{\overline{X} - \mu}{\sqrt{\frac{s^2}{n}}} \leqq t_{0.025}(n-1)\right) = 0.95 \tag{2}$$

となります。この問題のように区間推定を行う時には，ここから μ の信頼区間を求めます。

上の式では，

> t 統計量が $-t_{0.025}(n-1)$ と $t_{0.025}(n-1)$ の間に入っているという記述は，確率 95%で的中している

ということを述べています。ということはすなわち，

> 「t 統計量が $-t_{0.025}(n-1)$ 以下かもしくは $t_{0.025}(n-1)$ 以上である」という記述は，的中している確率が 5%でしかない

ということになります。

では，ここで，次の問題例を考えてみましょう。

ある試験の点数の分布は正規分布であるとします。この試験の受験者から，10 人からなる標本を無作為抽出して，この 10 人の点数を平均したところ 50 点で，またこの 10 人の点数の不偏分散が 5^2 でした。このとき，「受験者全体の平均点は 54 点」という「仮説」を考えると，この仮説は的中しているといえるでしょうか？

この問題でいま得られている標本については，標本平均 $\overline{X} = 50$，不偏分散 $s^2 = 25$，標本サイズ $n = 10$ です。ここで，仮に

「受験者全体の平均点は 54 点である（$\mu = 54$）」

という仮説が正しいとしましょう。そうすると，これらの数値を (1) 式に代入して t 統計量を求めると，$t = -2.53$ となります。

一方，t 分布表から，$t_{0.025}(10-1) = 2.2622$ であることがわかります。したがって，仮に $\mu = 54$ が正しいとすると，

t 統計量が $-t_{0.025}(n-1)$ 以下かもしくは $t_{0.025}(n-1)$ 以上である

という，的中している確率が 5%でしかないはずの記述が的中していることになります。

ここまでのことをまとめると，下のような推論ができます。

1. 「t 統計量が $-t_{0.025}(10-1)$ 以下かもしくは $t_{0.025}(10-1)$ 以上である」という記述は，的中している確率が 5%でしかない

2. 仮に「$\mu = 54$ である」という仮説が正しいとすると，そのとき t 統計量は $t = -2.53$ で，一方 $t_{0.025}(10-1) = 2.262$ であるから，

   「t 統計量が $-t_{0.025}(10-1)$ 以下かもしくは $t_{0.025}(10-1)$ 以上である」

   という記述は正しいことになる

3. つまり，的中している確率が 5%でしかないはずの記述が，いま偶然的中していると考えざるをえない

4. 「確率 5%でしか起きないはずのことが，いま偶然起きている」とわざわざ考えるのは不合理なので，「$\mu = 54$ である」という仮説は間違っていると判断する

5. 「$\mu = 54$ ではない」という仮説が正しいと判断する

つまり，「母平均は 54 点よりもずっと大きいかずっと小さいかのどちらかであって，とにかく 54 点であるという可能性は考えにくい」と言っているのです。

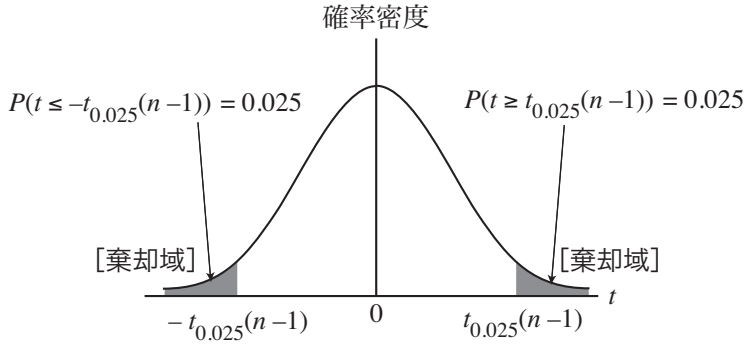

図 1: 両側検定の棄却域

このような推論のしかたを**仮説検定**といいます。上の例で,「母平均は 54 である」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「母平均は 54 である」という仮説を**帰無仮説**といい，$H_0 : \mu = 54$ 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「母平均は 54 ではない」という仮説を**対立仮説**といい，$H_1 : \mu \neq 54$ 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,「5%の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。

上の推論では，帰無仮説が正しいとするとき,「t 統計量が $-t_{0.025}(n-1)$ 以下かもしくは $t_{0.025}(n-1)$ 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,「帰無仮説が正しいとするとき，t 統計量がここに入ったら，帰無仮説を棄却する」という区間が，t 分布の確率密度関数で，左右両側にあります（図 1)。その意味で，今回のやりかたの検定を**両側検定**といいます。

なお，上の「帰無仮説が正しいとするとき，t 統計量がここに入ったら，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を，**棄却域に落ちる**という表現をします。

片側検定

両側検定は区間推定をもとにしたものなので，(2) 式で表される，t 統計量が入っている確率が 95%である区間，再掲すると

$$P\left(-t_{0.025}(n-1) \leqq \frac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}} \leqq t_{0.025}(n-1)\right) = 0.95 \tag{3}$$

は，図 1 のように，確率密度関数のグラフにおいて左右対称になっていました。

[1] H は，hypothesis（仮説）という英語の頭文字です。

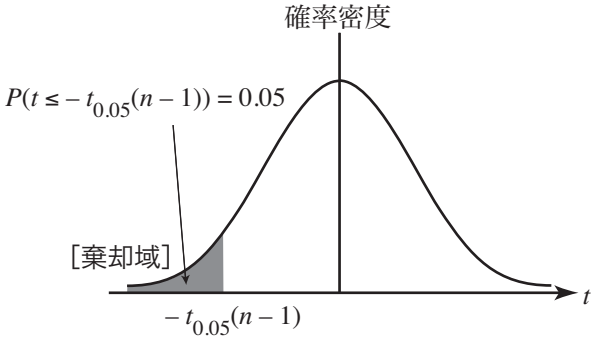

図 2: 片側検定の棄却域

しかし，「t 統計量が入っている確率が 95%である区間」を求めるだけなら，別に左右対称でなければならない理由はありません。ですから，次の式で表される区間も，やはり「t 統計量が入っている確率が 95%である区間」です。

$$P\left(-t_{0.05}(n-1) \leqq \frac{\overline{X}-\mu}{\sqrt{\frac{s^2}{n}}}\right) = 0.95 \tag{4}$$

この場合を確率密度関数のグラフで表すと，図 2 のようになります。この区間を使った検定を考えてみましょう。

上の (4) 式は，「t 統計量が $-t_{0.05}(n-1)$ 以下である」という記述が的中している確率は，5%でしかないということを述べています。

ここで，両側検定の時と同じ問題例を考えて，同様に「母平均は 54 点である（$\mu = 54$)」という帰無仮説が正しいとしましょう。そうすると，さきほどの例と同じく，t 統計量は $t = -2.53$ となります。

一方，$t_{0.05}(10-1) = 1.8331$ です。したがって，仮に $\mu = 54$ という帰無仮説が正しいとすると，「t 統計量が $-t_{0.05}(n-1)$ 以下である」という，的中している確率が 5%でしかないはずの記述が的中していることになります。

つまり，

1. 「t 統計量が $-t_{0.05}(10-1)$ 以下である」という記述は，的中している確率が 5%でしかない

2. 仮に「$\mu = 54$ である」という仮説が正しいとすると，そのとき t 統計量は $t = -2.53$ で，一方 $t_{0.05}(10-1) = 1.8331$ であるから，

   「t 統計量が $-t_{0.05}(10-1)$ 以下である」

   という記述は正しいことになる

3. つまり，的中している確率が 5%でしかないはずの記述が，いま偶然的中していると考えざるをえない

4. 「確率 5%のでしか起きないはずのことが，いま偶然起きている」とわざわざ考えるのは不合理なので，「$\mu = 54$ である」という仮説は間違っていると判断する

5. 「$\mu = 54$ ではない」という仮説が正しいと判断する

という推論ができます。

ここまでは，両側検定の場合と同じです。違うのは，帰無仮説が棄却された結果，導かれる対立仮説です。

上の推論で，帰無仮説を棄却した理由は，t 統計量が $-t_{0.05}(10-1)$ 以下であったためです。それならば，t 統計量がもう少し大きければ，帰無仮説は棄却されません。t 統計量は，この節の最初に再掲したように (3) 式の形をしていますから，t 統計量を大きくするには，帰無仮説で述べられている $\mu = 54$ を，もっと小さくすればよいわけです。

つまり，この推論では，帰無仮説が棄却されることによって，「μ は 54 よりももっと小さい」，すなわち $H_1 : \mu < 54$ という対立仮説が採択されます。この検定は，帰無仮説がヒストグラムの片側にありますから，**片側検定**といいます。

どちらの検定を選ぶか？

ここまでの説明によると，両側検定は「μ は 54 点ではない」という形の対立仮説しか得られないのに対して，片側検定では「μ は 54 点より小さい」といった，より詳しい対立仮説を求めているように思われます。ですから，片側検定の方がより優れた検定のように感じられるかもしれません。

しかし，それは誤りです。片側検定と両側検定とでは，検査している内容が違うのです。

検定とは，帰無仮説で想定しているパラメータの値（例えば $\mu = 54$）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，帰無仮説で想定しているパラメータの値は誤りとして，帰無仮説を棄却します。

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，帰無仮説で想定しているパラメータの値が，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 の値が，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。

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，つまり標本に比べて「ある方向に」食い違っているかどうかを検査します。ですから，帰無仮説で想定しているパラメータの値が，標本に比べて「ある方向に」食い違っているときだけ帰無仮説を棄却します。例えば，対立仮説が「$\mu < 54$」という片側検定なら，帰無仮説の「$\mu = 54$」は大きすぎると言えるかどうかだけを検査していますから，帰無仮説でいう「$\mu = 54$」という値が標本平均に比べて大きすぎるときだけ，帰無仮説を棄却します。

では，帰無仮説でいう「$\mu = 54$」という値が，標本平均に比べて小さ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「$\mu < 54$」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，帰無仮説でいう「$\mu = 54$」という値が標本に比べて食い違っているにかかわらず，片側検定はそれを見逃し，「対立仮説が採択できるかどうかはわからない」と答えてしまいます。それは，「$\mu = 54$ は標本平均に比べて小さすぎるかどうか」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「母平均は 0 である」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。

この違いを，「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。

棄却されないときは

さて，ここまでに述べたように，検定では，「内心では」帰無仮説が棄却されて，対立仮説が採択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」という結論が得られるわけです。帰無仮説を「無に帰す仮説」と名前でよぶのは，その意味合いがあります。

では，帰無仮説が棄却されない場合は，どういう結論になるのでしょうか？ そこで，今回の両側検定の例で，有意水準を 1%にしてみましょう。この場合，

> 「t 統計量が $-t_{0.005}(n-1)$ 以下かもしくは $t_{0.005}(n-1)$ 以上である」という記述は，的中している確率が 1%でしかない

ということから出発します。$n = 10$ なので，数表で $t_{0.005}(9)$ を調べると，$t_{0.005}(9) = 3.2498$ であることがわかります。一方，先に計算したように，「母平均は 54 である」という仮説が正しいとすると，t 統計量は $t = -2.53$ です。したがって，次のような推論をすることになります。

1. 「t 統計量が $-t_{0.005}(10-1)$ 以下かもしくは $t_{0.005}(10-1)$ 以上である」という記述は，的中して

いる確率が1%でしかない

2. 仮に「$\mu = 54$ である」という仮説が正しいとすると，そのとき t 統計量は $t = -2.53$ で，一方 $t_{0.005}(10-1) = 3.2498$ であるから，

「t 統計量が $-t_{0.025}(10-1)$ 以下かもしくは $t_{0.025}(10-1)$ 以上である」

という記述は正しいとはいえない

3. つまり,「的中している確率が1%でしかないはずの記述が, いま偶然的中している」とまではいえない

4. 「$\mu = 54$ である」という仮説は間違っているとは言い切れない

という，なんとも煮え切らない結論となります。検定の言葉づかいでは，「帰無仮説は棄却されない」となります。

つまり，帰無仮説が棄却されなかったとすれば，その理由は「帰無仮説が正しい ($\mu = 54$) 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，「$\mu = 54$ でないとはいえない」つまり，「目論見ははずれた。$\mu = 54$ でないとまで断言する自信はない」という結論になるのです。

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,「$\mu = 54$ である」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[2]。

有意水準について

ここまでの例では，有意水準が5%のときは「帰無仮説を棄却する」と結論され，有意水準が1%のときは「帰無仮説を棄却しない」という結論になりました。しかし，帰無仮説の内容や標本平均・不偏分散・標本サイズは同じで，有意水準は勝手に決めたのに，こんなに違った結論になってもよいのでしょうか？

[2]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。

これについては，「検定とはそういうものだ」ということを，よく理解しなければなりません。有意水準は，検定をする人の「大胆さ・慎重さ」の程度を表しているのです。

有意水準が大きい (5%) ときは，帰無仮説が仮に正しいとしたときに，いま起きている現実（t 統計量の値が-2.53）が起きる確率が 5% 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。

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，母平均が本当に 54 である，つまり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[3]。

> **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。

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。**

[3]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。

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，少ない数のデータしか調べられず，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。

今日の演習

下の各質問に，各々 100 文字以内で簡潔に答えてください。

1. 仮説検定とは，どういう場合に何をすることですか。
2. 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。