2007 年度後期　統計データ解析Ｂ　第８回

分布の平均を一部から推測する **(3)** － 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」「母平均は 100 より大きい」といった，母集団分布についての仮説の真偽を推測する方法です．

両側検定

次の区間推定の問題を考えてみます．

> ある実験で，水の沸点を９回測定して，「100.0　100.1　101.0　99.3　97.8　100.2　98.5　100.1　101.0」（℃）という値を得ました．物質の沸点の測定結果が，真の沸点を平均とする正規分布にしたがうとするとき，真の沸点を信頼係数 95%で区間推定してください．

この問題の考え方は，次のようになります．

９回の測定結果は，正規分布にしたがう母集団からのサイズ 9 の標本と考えることができます．そこで，標本平均を $\bar{X}$，不偏分散を s^2 とすると，$\bar{X} = 99.78$，$s^2 = 1.149$ です．標本サイズを $n(= 9)$ とし，真の沸点を μ とすると，

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

すなわち t 統計量は，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，

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

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

上の (2) 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%で当たっている，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，当たっている確率が 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．

> ある実験で，水の沸点を９回測定して，「100.0　100.1　101.0　99.3　97.8　100.2　98.5　100.1　101.0」（℃）という値を得ました．物質の沸点の測定結果が，真の沸点を平均とする正規分布にしたがうとします．このとき，「真の沸点（母平均）は 101 ℃ではない」という「仮説」を考えると，この仮説は当たっているといえるでしょうか？

上で述べたように，今得られている標本については，$n = 9, \bar{X} = 99.78$，$s^2 = 1.149$ です．ここで，仮に「真の沸点（母平均）は 101 ℃である（$\mu = 101$）」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，$t = -3.41$ となります．

一方，$t_{0.025}(9-1) = 2.306$ です．したがって，仮に $\mu = 101$ 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，当たっている確率が 5%でしかないはずの記述が当たっていることになります．

つまり，

「t 統計量が $-t_{0.025}(9-1)$ 以下かもしくは $t_{0.025}(9-1)$ 以上である」という記述は，当たっている確率が 5%でしかない
→仮に「$\mu = 101$ 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 $t = -3.41$ で，一方 $t_{0.025}(9-1) = 2.306$ であるから，
　「t 統計量が $-t_{0.025}(9-1)$ 以下かもしくは $t_{0.025}(9-1)$ 以上である」
　という記述が正しいことになる
→当たっている確率が 5%でしかないはずの記述が，
　いま偶然当たっていると考えざるを得ない
→ 5%の確率でしか起きないはずのことが起きているのに，「偶然起きている」と考えるのは不合理なので，
　「$\mu = 101$ である」という仮説は間違っていると判断する
→「$\mu = 101$ 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．つまり，「母平均は 101 ℃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 101 ℃であるという可能性は考えにくい」と言っているのです．

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 101 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 101 である」という仮説を**帰無仮説**といい，$H_0 : \mu = 101$ と表します．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 101 ℃ではない」という仮説を**対立仮説**といい，$H_1 : \mu \neq 101$ 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%の確率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．

本当は，5%の確率でしか起きないことでも，偶然起きることは，5%の確率であるはずです．ですから，例えば母平均が本当に 101 である，つまり帰無仮説が正しいときでも，信頼区間が偶然大きく母平均からはずれて，帰無仮説を偶然棄却してしまうことが，5%の確率であります．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．また，有意水準はここまで 5%としてきましたが，これは 5%がよく用いられるからというだけで，本当は第 1 種の誤りによる損失の期待値を見積って決めます．これは，区間推定のときの信頼係数の決め方と同じです．

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 分布のヒストグラム（確率密度関数）で，左右両側にあります．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．

棄却されないときは

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[1]．

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？ 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 ($\mu = 101$) 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，「$\mu = 101$ でないとはいえない」つまり，「目論見ははずれた．$\mu = 101$ 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，「いま起きている現実が起きる確率は，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，「$\mu = 101$ 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．

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，すなわち

$$P\left(-t_{0.025}(n-1) \leq \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \leq t_{0.025}(n-1)\right) = 0.95$$

は，図 1(a) のようにヒストグラムにおいて左右対称になっていました．

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

[1] 帰無仮説を「無に帰す仮説」とよぶのはその意味です．

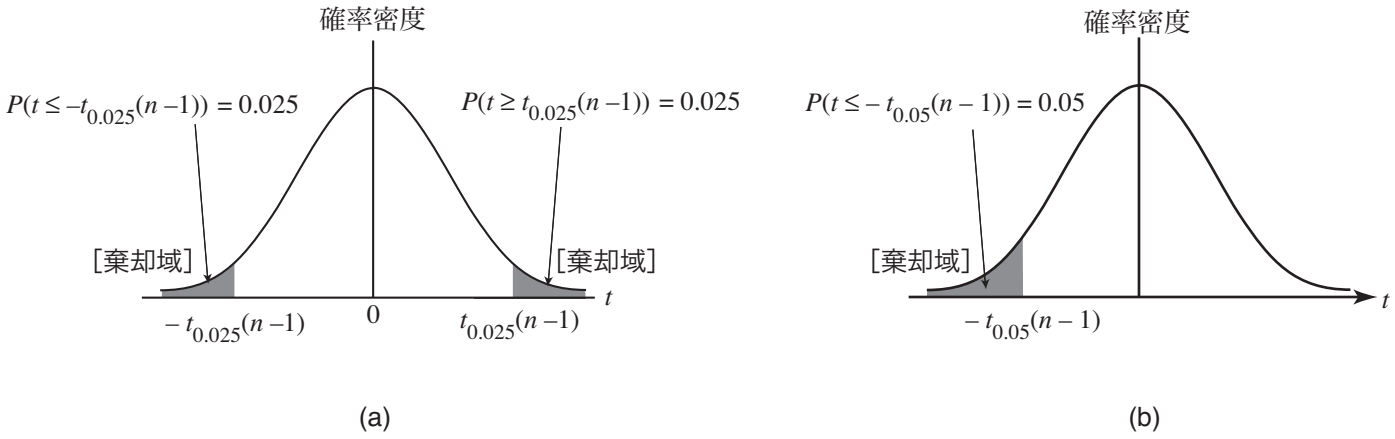

図 1: 検定の棄却域．(a) 両側検定，(b) 片側検定

95%である区間」です．

$$P\left(-t_{0.05}(n-1) \le \frac{\bar{X}-\mu}{\sqrt{s^2/n}}\right) = 0.95 \tag{3}$$

この場合をヒストグラムで表すと，図 1(b) 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．

(3) 式は，「t 統計量が $-t_{0.05}(n-1)$ 以下である」という記述が当たっている確率は，5%でしかないということを述べています．

ここで，上の両側検定の例と同じ問題例を考えて，同様に「真の沸点 (母平均) は 101 ℃である ($\mu = 101$)」という帰無仮説が正しいとしましょう．そうすると，上の例と同じく，t 統計量は $t = -3.41$ となります．

一方，$t_{0.05}(9-1) = 1.860$ です．したがって，仮に $\mu = 101$ 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，当たっている確率が 5%でしかないはずの記述が当たっていることになります．

つまり，

「t 統計量が $-t_{0.05}(9-1)$ 以下である」という記述は，当たっている確率が 5%でしかない
→仮に「$\mu = 101$ である」という帰無仮説が正しいとすると，
そのとき t 統計量は $t = -3.41$ で，一方 $t_{0.05}(9-1) = 1.860$ であるから，
「t 統計量が $-t_{0.05}(9-1)$ 以下である」という記述が正しいことになる
→「当たっている確率が 5%でしかないはずの記述が，
いま偶然当たっていると考えざるを得ない
→ 5%の確率でしか起きないはずことが，いま偶然起きていると考えるのは不合理なので，
「$\mu = 101$ である」という帰無仮説は間違っていると判断する（帰無仮説を棄却する）

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 $-t_{0.05}(9-1)$ 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 (1) 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 $\mu = 101$ 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 101 よりももっと小さい」，すなわち $H_1 : \mu < 101$ 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？

ここまでの話によると，両側検定は「μ は 101 ℃ でない」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 101 ℃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 $\mu = 101$）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「$\mu < 101$」という片側検定なら，帰無仮説の「$\mu = 101$」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「$\mu = 101$」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「$\mu = 101$」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「$\mu < 101$」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「$\mu = 101$」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，「$\mu = 101$ 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．

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．