2011 年度前期　推測統計学への招待　第１０回

水の沸点は本当に 100 ℃か？—t 分布と両側検定

第８回の講義で，t 分布を用いた区間推定について説明しました。これは，母集団のデータが正規分布にしたがって分布するとし，さらに母分散がわからない時に，標本から不偏分散を計算して，母平均の区間推定を行なうものでした。今回は，区間推定から導かれる，「両側検定」というもうひとつの検定を，同じく t 分布を使って説明します。

区間推定について，ちょっと復習—「95%信頼区間」の真の意味は？

ここで，第７回の講義で説明した「母平均の 95%信頼区間」すなわち「母平均が入っている確率が 95%であるような区間」の意味を，もう一度思い出してみましょう。

ここで，例えば「10 人からなる標本を取り出して標本平均を求め，95%信頼区間を求める」という計算を，何度も行なったとしましょう。そうすると，標本平均の値は，標本をとりだして計算するごとに違うでしょうから，求められる信頼区間も毎回違います。このとき，ランダムなのは標本平均のほうですから，標本をとり出して計算するごとに変化するのは，母平均ではなく，それを囲んでいる「区間」のほうです。

「95%信頼区間」の意味は，何回も標本を取り出しては計算し信頼区間を求めると，

> 100 回あたり 95 回は
> 確かにその信頼区間の中に母平均が入っている「本当の」信頼区間が求められるが
> 残り 5 回は
> その信頼区間の中に母平均が入っていない「ウソの」信頼区間が求められてしまう

ということです。

現実には，標本を取り出して計算するのは 1 回だけです。「たまたま」その時に取り出された標本を用いて，標本平均を計算して求めた信頼区間が，母平均を本当に含んでいる「本当の」信頼区間か，それとも含んでいない「ウソの」信頼区間か，それはわかりません。

「信頼係数 95%」というのは，「この計算は，100 回中 95 回も『本当の』信頼区間を求められる方法だから，その方法で今求めた信頼区間も，『本当』であると信じてください」と言っているにすぎません。

区間推定から導かれる，もうひとつの検定

さて，次の区間推定の問題を考えます。

> 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。

この問題の解答を，第１１回の講義で説明した方法で計算すると，母平均（真の沸点）の 95%信頼区間は [98.96, 100.6]（℃）となります。

ということは，すなわち

> 母平均の信頼区間を，第１１回の講義で説明した方法で [98.96, 100.6] と推測すると，その推測法が「ウソ」の信頼区間を答える確率は 5%である。

ということになります。

「98.96 から 100.6」という推測がはずれたとすると，そのとき母平均は「98.96 以下，または 100.6 以上」だということになります。よって，

> 母平均の信頼区間を上に述べた方法で「98.96 以下，または 100.6 以上」と推測すると，その推測法が「本当」の信頼区間を答える確率は 5%でしかない。

ということが言えます。

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」という仮説は当たっているといえるでしょうか？

これを上の話の進め方から次のように考えてみましょう。すると，

> 母平均の信頼区間を「98.96 以下，または 100.6 以上」と推測すると，その推測法が「本当」の信頼区間を答える確率は 5%でしかない
> →「母平均は 101 である」という仮説が正しいとすると，
> 　101 は「98.96 以下，または 100.6 以上」に入っているから，
> 　「5%の確率でしか当たらないような推測法で答えた結果が当たっている」と
> 　考えざるを得ない
> → 5%の確率でしか起こらないことが，現に起こっていると考えるのは不合理なので，
> 　「母平均は 101 である」という仮説は間違っていると判断する
> →「母平均は 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。

この推論は，第 6 回で説明した検定とは考え方が違っていますが，これも仮説検定のひとつです。上の例では，「母平均は 101 である」が帰無仮説で（$H_0 : \mu = 101$），それが棄却されています。前回の例と異なるのは，帰無仮説が棄却されたときに採択される対立仮説です。前回は「母平均は 100 ℃より小さい」という対立仮説（$H_1 : \mu < 100$）が採択されていたのに対して，今回は「母平均は 101 ℃ではない」（$H_1 : \mu \neq 101$）となっています。

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。また，上の推論で「5%の確率でしか起こらないことが現に起こっていると考えるのは不合理」と考えているので，有意水準は 5%です。

ところで，今回の検定では，帰無仮説での μ の値が「98.96 以下，または 100.6 以上」のとき帰無仮説を棄却する，という推論をしました。つまり，ヒストグラム（確率密度関数）のうえで，棄却域が大小

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。これに対して前回までに説明した検定では，棄却域が大小どちらか片側にありますから，**片側検定**といいます。

両側検定の問題の解き方

ここまでの説明にもとづいて，さきほどの両側検定の問題の解き方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」と言えるかを，有意水準 5%で答えてください。

9 回の測定結果は，正規分布にしたがう母集団からの 9 個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」（上側 2.5 パーセント点）とすると，

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

となります。

区間推定の時には，ここから μ の信頼区間を求めました。しかし，両側検定のときはその必要はありません。(2) 式から，「(1) 式の t が，$-t_{0.025}(n-1)$ 以下または $t_{0.025}(n-1)$ 以上である確率は 0.05」であることがわかります。ですから，帰無仮説を $\mu = 101$ として，これが正しいとしたときに「(1) 式の t が，$-t_{0.025}(n-1)$ 以下または $t_{0.025}(n-1)$ 以上」であれば，帰無仮説を棄却し，$\mu \neq 101$ という対立仮説を採択します。

$\mu = 101$ とするとき，他の数値と合わせて (1) 式に代入すると，$t = -3.41$ で，一方 $t_{0.025}(9-1) = 2.306$ ですから，「(1) 式の t が，$-t_{0.025}(n-1)$ 以下または $t_{0.025}(n-1)$ 以上」ということになり，帰無仮説を棄却します。したがって，「有意水準 5%で，真の沸点（母平均）は 101 ℃ではないと言える」という答えになります。

どちらの検定を選ぶか？

ここまでの話によると，両側検定では「μ は 101 ℃ でない」「当たり確率は 0.5 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 100 ℃より小さい」「当たり確率は 0.5 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 $\mu = 100$）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，標本から推測される値に比べて，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「$\mu < 100$」という片側検定なら，帰無仮説の「$\mu = 100$」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「$\mu = 100$」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「$\mu = 100$」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「$\mu < 100$」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「$\mu = 100$」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，「$\mu = 100$ 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。

講義の前半で扱った「くじびきの問題」を例と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」の時は不満ですが，「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するものの，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。

このように，「当たる確率は 50%」という帰無仮説と，現実の当たりとを比べて，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，どちらかだけをとりあげ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。

今日の演習

ある土砂の汚染の度合を調べるために，対象の土砂をよく混ぜてから 1 グラムずつ 10 個の土の標本を抽出し，各々汚染物質の量を調べました。その結果，汚染物質の量（ミリグラム）は次の通りでした。

1.2　0.5　0.3　0.1　1.0　1.3　1.2　1.0　0.1　0.5

1. 「この土砂の 1 グラムあたりの汚染物質の量は，1.1 ミリグラムではないと言えるか」という質問に，講義で説明した知識を用いて答えるために，必要な仮定を 2 つあげてください。
2. 1. の仮定が正しいとして，この質問に統計学的に答えてください。