60分でわかる仮説検定

第2版　2002.05.20.

小野　滋

ガソリン燃費問題

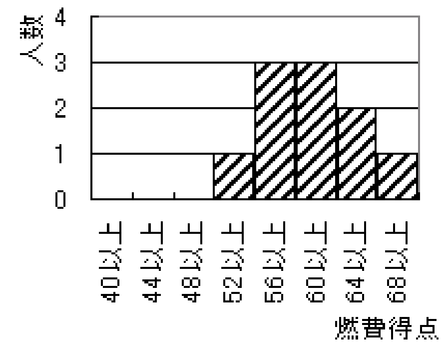
とあるタクシー会社では，ガソリンA,Bの燃費を比較するために，20人の運転手を選び，ガソリンA,Bのどちらかを一定期間利用させて，燃費をしらべた。結果は下のようになった。ガソリンA,Bの間には，燃費の差があるといえるだろうか？

ガソリンA

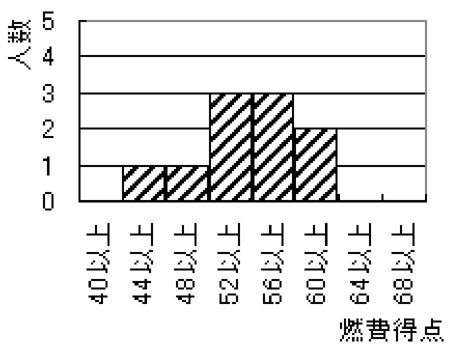
運転手	燃費得点
鈴木	61
佐藤	60
福田	56
三木	63
中曽根	56
田中	64
小泉	59
海部	69
吉田	52
森	64
平均	$\overline{x_A} = 60.4$
標準偏差	$s_A = 4.67$
人数	$n_A = 10$

ガソリンB

運転手	燃費得点
土井	55
不破	54
鳩山	47
小沢	59
村山	51
志井	61
辻元	57
羽田	54
管	62
神崎	58
平均	$\overline{x_B} = 55.8$
標準偏差	$s_B = 4.35$
人数	$n_B = 10$



ガソリンAの燃費データ



ガソリンBの燃費データ

キーワード
母集団, 標本, 標本のサイズ, 不偏性, 独立性, 無作為抽出

1.1 母集団と標本

“ガソリン燃費問題” で私たちに求められているのは,「データをもとに推測する」ことである。つまり，問題になっているのは，いま手元にあるデータの性質 (ガソリン A の燃費得点は平均 60.4 点だった，などなど) そのものではない。我が社で今後，ガソリン A を採用したら，はたして燃費はどうなるだろうか? B を採用したらどうなるだろうか? — という，現時点ではわからないことについての推測を求められているのである。

このように，実験研究では，ふつう，手元にあるデータそのものには本質的な関心がないことに注意してほしい。私たちは，手元にあるデータを通じて，より大きなデータ集合 (ガソリン A ないし B を採用した場合の，将来の燃費) の性質を推測しようとしているのである。このとき，

- 関心を持っているデータ全体を**母集団** と呼ぶ。
- 手元にあるデータの集まりを**標本** と呼ぶ。標本のなかに含まれているデータの数を，**標本のサイズ**と呼ぶ。

標本はいわば，母集団から取り出した (抽出した) データである，と考えることができる。

1.2 母集団の性質

母集団は，そのすべての値を測定することが原理的には可能な場合もあれば，そもそも不可能な場合もある。そこで，次のように考えよう。

- 母集団のなかには，無限個の値が含まれている。
- 私たちにはわからないが，それらの値についてヒストグラムを描くこともできるし，平均や標準偏差も計算できる。

母集団の平均を μ (「ミュー」), 母集団の標準偏差を σ (「シグマ」) と呼ぶことにする。この μ や σ の値については，私たちは知ることができないが，神様は知っている，と考えるわけである。

1.3 望ましい標本とは

標本の性質をもとに，母集団の性質を推測するためには，標本は次の性質をもっている必要がある。

- 不偏性 ... 母集団から偏りなく抽出されていること
- 独立性 ... 個々のデータが，互いに影響を及ぼしていないこと

これらの性質をそなえた標本のことを，いっぱんに**無作為標本**と呼ぶ。つまり，良い標本とは，母集団からランダムに抽出した (**無作為抽出**した) かのような性質をもっている標本だ，というわけである。

統計的推測の手法は，その種類をとわず，データが母集団からの無作為標本であることを前提としている。

“ガソリン燃費問題” について，つぎのように考えることにしよう。

- 「ガソリン A を採用した場合の将来の燃費」の集合を，母集団 A と呼ぶことにする。
- 母集団 A には，無限個の値が含まれている。その分布についてはわからないが，その平均値を μ_A, その標準偏差を σ_A と呼ぶことにする。
- 手元にあるデータ $61, 60, \cdots 64$ は，母集団 A からの無作為標本である，と考えることにする。

母集団 A　　　　　　　　　　　　　　標本 A

平均 $\mu_A =?$　　　　　　　　　　平均 $\bar{X}_A = 60.4$

標準偏差 $\sigma_A =?$　　　　　　　標準偏差 $s_A = 4.67$

ガソリン B についても同様に考える。

キーワード

統計的仮説検定, 帰無仮説 (H_0), **対立仮説** (H_1) **検定統計量, 棄却域, 臨界値, 正規分布, 自由度,** t **分布, 有意水準, 帰無仮説の棄却**

統計的推測の手法にはいろいろなものがある。以下では，**統計的仮説検定**という考え方を紹介しよう。

広末涼子問題

広末涼子らしき女性が街を歩いていた。本物か，そっくりな別人かを知りたくて，すれ違いざまに手元をのぞき込んでみると，J フォンの携帯を持っていた。そこでつぎのように考えた。「広末涼子ならば，J フォンの携帯は持っていない。しかし，あの女性は J フォンの携帯を持っていた。だから，あの女性は広末涼子ではない。」

このような論理的手続きは，私たちの日常においてもなじみ深いものである。整理してみよう。

Step 1. まず，あの女性は広末涼子である，と仮定する。
Step 2. さて，携帯に着目すると，あの女性は J フォンの携帯を持っていた。
Step 3. もし Step 1. の仮定が正しければ，Step 2. のような事柄が起こるはずはない。
Step 4. だから，Step 1. の仮定が間違っている，すなわち彼女は広末涼子ではない，と判断できる。

ここで注目すべきなのは，まず仮説を設け，それに反する証拠を挙げる，というアイデアである。この仮説は，分析者の予想や主張とは必ずしも一致しないことに注意してほしい。上の例でいえば，分析者は「あの女性は広末涼子でない」という結論を導くために，いったん「あの女性は広末涼子だ」という仮説を設けているにすぎない。

それでは，ガソリンの燃費の問題について考えてみよう。つぎのような手順を踏めば，ガソリン A とガソリン B のあいだに燃費の差がある，と判断することができる。

Step 1. ガソリン A と B のあいだに燃費の差がない，と仮定してみよう。
Step 2. さて，ガソリン A と B の燃費データのあいだには，×××のちがいがあった。
Step 3. もし Step 1. の仮定が正しいならば (ガソリン A と B のあいだに燃費の差がないならば)，Step 2. のような事柄が起こるはずはない。
Step 4. だから，Step 1. の仮定が間違っている，すなわちガソリン A とガソリン B のあいだには燃費の差がある，と判断できる。

統計的仮説検定とは，確率論の知見を生かして，上の論理的手続きを洗練した手法である。それでは，この 4 つのステップについて，順に検討していこう。

2.1 Step 1. 帰無仮説を設定する

> Step 1. ガソリンAとガソリンBのあいだに燃費の差がない，と仮定してみよう。

この「ガソリンAとBのあいだに燃費の差がない」という仮説を，**帰無仮説**と呼ぶ(H_0と略記する)。また，帰無仮説の反対側の仮説(「ガソリンAとBのあいだに燃費の差がある」)を，**対立仮説**と呼ぶ(H_1と略記する)。

どのような帰無仮説を設定するかは，分析者の予想や主張とは無関係に，ただテクニカルな都合によって決まることに注意してほしい[1]。

> → 帰無仮説を「ガソリンAとガソリンBのあいだに燃費の差がない」とする。

2.2 Step 2. 検定統計量の値を求める

> Step 2. さて，ガソリンAとBの燃費データのあいだには，×××のちがいがあった。

このステップは，手元のデータについて，＜帰無仮説が正しいときには起こり得ないような＞なんらかの特徴を指摘するステップである。“広末涼子問題” の場合では，謎の女性がJフォンの携帯を持っている，という指摘がこれに相当する。

“ガソリン燃費問題” についていえば，帰無仮説が真のときと偽のときとで，値が大きく変わるような，なんらかの変量をとりあげて，その値を指摘できると都合がよい。たとえばガソリンAの燃費データの平均$\bar{X}_A$と，Bの燃費データの平均$\bar{X}_B$の差，$\bar{X}_A - \bar{X}_B$あたりが，その候補になるだろう。

- もし帰無仮説が正しければ，$\bar{X}_A - \bar{X}_B$は0に近くなるはずである。
- もし帰無仮説が正しくないならば，$\bar{X}_A - \bar{X}_B$は0から離れるはずである。

だから，手元のデータから得られた$\bar{X}_A - \bar{X}_B$が，もし0から遠く離れていたならば，それを根拠に，帰無仮説は正しくはなさそうだ，すなわちガソリンAとBの間には燃費の差がありそうだ，と主張できる。

> → 例題のデータでは，$\bar{X}_A - \bar{X}_B = 4.6$ であった。

実際には，$\bar{X}_A - \bar{X}_B$ではなく，それを修正した，次の変量をもちいるのが一般的である(その理由はあとで述べる)。

$$t = \frac{\bar{X_A} - \bar{X_B}}{\sqrt{\frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}}$$

このtの値も，もし帰無仮説が正しければ0に近くなるはずであり，もし帰無仮説が正しくないならば，0から離れるはずである。

> → 例題のデータでは，$t = 2.16$であった。

なお，このステップでもちいる変量を**検定統計量**と呼ぶ。ここでは，tが検定統計量である。

---

[1] 例題の場合でいえば，もし「ガソリンAとBのあいだに燃費の差がある」という帰無仮説を設定してしまうと，Step 3.で行き詰まってしまう。

2.3 Step 3. 帰無仮説のもとでの検定統計量の分布を求める

Step 3. もしガソリンAとBのあいだに燃費の差がないならば，このような事柄が起こるはずはない。

このステップでは,「もし帰無仮説が正しいならば，検定統計量の値がこんな値になるはずがない」と主張しなければならない。

このステップはなかなか難しい。“広末涼子問題” の場合ならば，

謎の女性はJフォンの携帯電話を持っていた。帰無仮説(「彼女は広末涼子である」)がもし真ならば，これはとうてい起こりそうにないことだ。

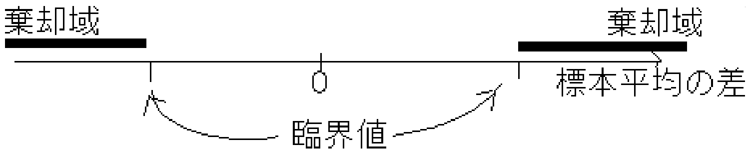
という理屈を，誰もが納得してくれるだろう。しかし，“ガソリン燃費問題” の場合，

$\bar{X}_A - \bar{X}_B = 4.6$であり，この値はあまりにも0から離れている。帰無仮説(「ガソリンAとBのあいだに燃費の差がない」)がもし真ならば，これはとうてい起こりそうにないことだ。

と主張しても，はたしてそれだけで，納得してもらえるだろうか？

つきつめていえば，$\bar{X}_A - \bar{X}_B$の値がどれほど0から離れていても，帰無仮説をくつがえす確実な証拠にはならない。しかし常識的には，$\bar{X}_A - \bar{X}_B$の値が0からある程度離れていたら，帰無仮説は正しくない，と判断するのが妥当だろう。つまり，ここで問題になっているのは，この「ある程度離れている」がどこからなのか，そのボーダーラインをどこに設定するか，ということなのである。

このボーダーラインのことを**臨界値**と呼び，$\bar{X}_A - \bar{X}_B$がこのボーダーラインを超える範囲のことを**棄却域**と呼ぶ。

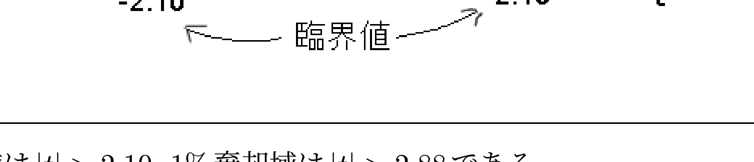


それでは，このボーダーラインをどのようにして決めるのか?

結論から先に述べよう。

- $\bar{X}_A - \bar{X}_B$ **ではなく，**t**を用いる**。
- ± 2.10 **を臨界値とし**，その外側を棄却域とする。これを5% 臨界値・5% 棄却域という。

ここで，この5% という値を**有意水準**と呼ぶ。



→ 5% 棄却域は$|t| > 2.10$, 1% 棄却域は$|t| > 2.88$である。

したがって，t**の値が臨界値を超え，棄却域に含まれている場合には，**H_0**が正しくない，と判断する**ことになる。では，臨界値と棄却域をこのように決める理由を説明しよう。

2.3.1 H_0 のもとでの $\bar{X}_A - \bar{X}_B$ の分布

ここで確率論の知見が登場する。

確率論の知見 (1)

仮に，“ガソリン燃費問題” の実験を，たとえば 10000 回，繰り返すことができたとしよう。すると，10000 回目の実験が終了した時点で，$\bar{X}_A - \bar{X}_B$ の値が 10000 個，手元に蓄積されていることになる。

さて，**ほんとうは帰無仮説が真である** (ガソリン間に燃費の差はない) 場合について考えよう。その場合，この 10000 個の $\bar{X}_A - \bar{X}_B$ は，0 に近いものが多く，0 から遠いものは少ないはずである。ヒストグラムを描けば，0 を中心とした山形になるはずである。

いま，つぎの前提条件が満たされているとする。

a. 母集団の分布が**正規分布**であること。つまり，母集団のすべての値をもしヒストグラムに描くことができたなら，そのかたちが正規分布に近いものであること。
b. 母集団の標準偏差 σ_A と σ_B が等しいこと。

このとき，$\bar{X}_A - \bar{X}_B$ のヒストグラムの形状について，数学的に予想することができる。

さて，“ガソリン燃費問題” のデータをみてみよう。2 つのヒストグラムは,

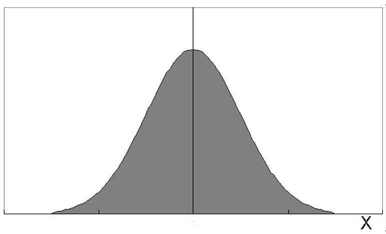
a. それぞれ，左右対称な山形をしており,
b. 標準偏差 s_A,s_B の間に大きな差はない。

この 2 つのヒストグラムは，2 つの母集団の分布を反映しているはずである。これらの特徴からみるかぎり，上の前提条件 a,b が満たされていると考えてよさそうだ。

ちょっと寄り道: 正規分布とは

正規分布とは，右図のように，平均を中心に左右対称な山形を示す確率分布である。その分布曲線は下の式で与えられる。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < \infty)$$



たとえば「性別・年齢が同じ人々の身長の分布」「パチンコ玉の直径の分布」など，自然界のさまざまな分布が，正規分布に近い分布を示すことが知られている。だから，“ガソリン燃費問題” でも，母集団における燃費の分布が正規分布にしたがう，と考えることには，さほど無理がない。

2.3.2 H_0 のもとでの t の分布

したがって，帰無仮説が真のときに，$\bar{X}_A - \bar{X}_B$ がどのような分布を示すかは，数学的に予想できることになる。もっとも，この数学的予想について知るためには，いちいち数学者に問い合わせる必要がある。というのは，標本のサイズ n_A, b_B や標本の標準偏差 s_A, s_B によって，この予想は変わってくるからだ。

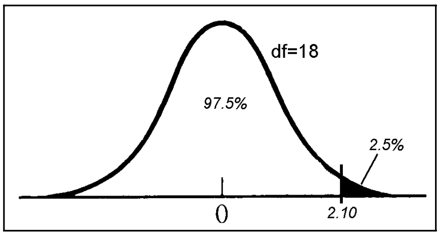
そこで，もっと簡単な方法が用意されている。$\bar{X}_A - \bar{X}_B$ のかわりに，$\bar{X}_A - \bar{X}_B$ を標本のサイズ n_A, n_B や標準偏差 s_A, s_B の組み合わせで割った変量 t を用いるのである。10000 個の t 値についてヒストグラムを描けば，それは $\bar{X}_A - \bar{X}_B$ と同じく，0 を中心とした山形になるはずである。

確率論の知見 (2) ←(1) のつづき

$\bar{X}_A - \bar{X}_B$ だけでなく，t のヒストグラムの形状についても，数学的に予想することができる。それは,「**自由度 $n_A + n_B - 2$ の t 分布**」と呼ばれる形状になる。

t 分布は，0 を中心に左右対称な，なだらかな山形を示す分布である。その形状は t 分布表に記載されており，数学者に問い合わせなくても，その性質を簡単に調べることができる[2]。

t 分布表の，自由度 18, $P = 0.025$ の欄をみてみよう。2.10 と書かれている。これは，t の値が 2.10 以上になる確率が 2.5% である，ということを意味している。

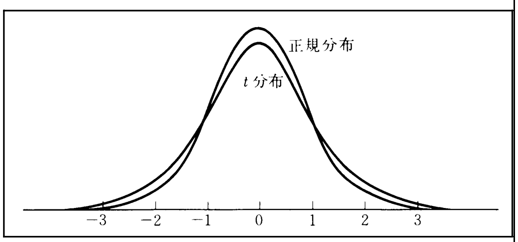


つまり，帰無仮説が真のとき (燃費の差がないとき)，例題の実験を 10000 回繰り返して，t の値を 10000 個，手に入れたならば，そのうち 250 個くらいは，(燃費の差がないにもかかわらず，) 2.10 以上の値になるだろう — ということである。

この項で述べたことはすべて，**帰無仮説が真である場合** (ほんとうは燃費の差がない場合) に限られていることに注意してほしい。**帰無仮説が偽である場合** (ほんとうは燃費の差がある場合)，$\bar{X}_A - \bar{X}_B$ が (そして t が) どのような分布を示すかは，予想がつかない[3]。

ちょっと寄り道: t 分布とは

t 分布とは，右図のように，0 を中心に左右対称な山形を示す確率分布である。標準正規分布 (平均 0, 分散 1 の正規分布) を，すこし下に押しつぶしたようなかたちをしている。**自由度 (df)** が大きくなるほど，正規分布に近づく。



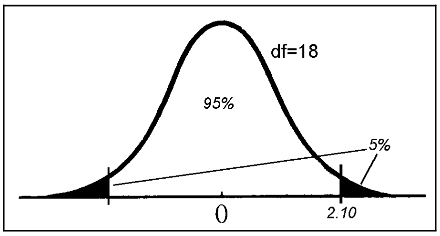
[2] なお，ここでなぜ $n_A + n_B - 2$ が自由度と呼ばれるのか，疑問に思われるだろうが，その説明は省略する。あまり気にしなくて良い。

[3] これが，ステップ 1 でいったん「燃費の差はない」という帰無仮説を設定した理由である。

2.3.3　有意水準とは

さて，問題に戻ろう。「5% 臨界値が ±2.10」というのは，どういうことなのか？

実は，この ±2.10 という値は,「自由度 18 の t 分布を両側から塗りつぶしていき，全体の 5% を塗りつぶしたところでストップした位置」なのである。



ふたたび，**帰無仮説 H_0 が真である場合** (ほんとうは燃費の差がない場合) に限定して考えよう。
例題の実験を 10000 回繰り返し，10000 個の t 値を得て，そのたびに t 値を棄却域と照らし合わせると，10000 回のうち 500 回くらいは，t 値が棄却域に含まれることになる。この 500 回の実験では，H_0 が正しくない，と判断することになる。
しかし，H_0 はほんとうは真なのだから，この判断は結果的には間違っている。ここで，この種の間違いを犯す確率は, 10000 回のうち 500 回くらい，すなわち 5% である。

このように，有意水準とは,「帰無仮説がほんとうは真であるにも関わらず，帰無仮説が正しくないと判断してしまう確率」に相当する。わたしたちは，この確率が 5% になるように，臨界値を ±2.10 に設定したわけである。

それでは，有意水準はなぜ 5% なのか? 4% でも 3% でもよいのではないか? — 原理的にはそのとおりである。分析者は有意水準を自由に設定してよい。しかし実際には，有意水準をどのように設定するかは，慣習的に決まっている。心理学では，5%, 1% の 2 種類が用いられることが多い。

2.4　Step 4. 帰無仮説が棄却されるかどうか判断する

Step 4. だから，ガソリン A とガソリン B のあいだには燃費の差がある，と判断できる。

データから得られた t 値が，臨界値を超え，棄却域に含まれていた場合には,

> この t 値はあまりにも 0 から離れている。帰無仮説 (「ガソリン A と B のあいだに燃費の差がない」) がもし真ならば，これはとうてい起こりそうにないことだ。

と判断することになる。これを，帰無仮説の**棄却**と呼ぶ。

帰無仮説の棄却は，対立仮説を支持する証拠となる。つまり,

> ガソリン A と B のあいだには，燃費の差がある。

と主張する際に，その有力な根拠となる。

→ t 値は 5% 棄却域に含まれている。すなわち，帰無仮説は有意水準 5% で棄却された。ガソリン A と B のあいだには，燃費の差があるようだ。

キーワード
仮説検定の非対称性，タイプIのエラー，タイプIIのエラー

Step 4. で述べたように，帰無仮説が棄却されたことは，対立仮説を支持する証拠になる。“ガソリン燃費問題” の場合でいえば，t 値が0から離れていることは,「燃費の差がある」ことの証拠になる。

それでは，帰無仮説が棄却されなかったことは，帰無仮説を支持する証拠になるだろうか? “ガソリン燃費問題” でいえば，もし t 値が0に近かったら，それは「燃費の差がない」ことの証拠になるだろうか?

結論からいえば，帰無仮説が棄却されなかった，という出来事からは，なにも得られない。つまり，**帰無仮説が棄却されなかったら，沈黙すべきである。**

これを**仮説検定の非対称性**と呼ぶ。以下では，その理由を説明しよう。

3.1 説明1: 仮説を確証するのはむずかしい

まず，おおざっぱな説明を試みよう。

一般的にいって，ある仮説を反証するのは簡単である。その仮説と矛盾する証拠を，ひとつ挙げるだけでよいからだ。しかし，ある仮説を確認する方向の証拠を挙げるのは，なかなか難しい。

“広末涼子問題” について考えよう。もし，謎の女性が，ドコモの携帯を持っていたら，どうなっていただろうか。4つのステップを順に追ってみよう。
Step 1. まず，あの女性は広末涼子である，と仮定する。
Step 2. さて，携帯に着目すると，あの女性はドコモの携帯を持っていた。
Step 3. もし彼女が広末涼子ならば，上のような事柄は...起きても不思議ではない。
Step 4. だから ...?
このように，Step 4.では，謎の女性が広末涼子だとも，そうでないとも主張できない。

このように，帰無仮説と矛盾しない特徴を指摘しても，帰無仮説を確証することにはならない。つまり，帰無仮説が棄却されなかったとしても，それは帰無仮説を支持する証拠にはならないのである。

3.2 説明2: タイプIIエラーが制御されていない

以上の説明に納得し，かつ，ややこしい話が嫌いな人は，この項は読み飛ばしてよい。ここでは，もうすこし厳密な説明を試みる。

結局のところ，帰無仮説は真か偽かのどちらかである (どちらなのか，を知ることはできないが)。また，仮説検定の結果，帰無仮説は棄却されるかされないかのどちらかである。だから，私たちのおちいる状況は，次の4通りであることになる。

- ほんとうは帰無仮説は真。帰無仮説は棄却されなかった。→ ○
- ほんとうは帰無仮説は真。帰無仮説は棄却された。→ × (**タイプIのエラー**という)
- ほんとうは帰無仮説は偽。帰無仮説は棄却されなかった。→ × (**タイプIIのエラー**という)
- ほんとうは帰無仮説は偽。帰無仮説は棄却された。 → ○

表にすると，下の図のようになる。

	H_0 は真	H_0 は偽
H_0 を棄却しない	○	× **タイプIIのエラー**
H_0 を棄却する	× **タイプIのエラー**	○

まず，H_0 が真のとき(表の左列) について考えよう。このとき，それと知らずに，運悪く H_0 を棄却してしまう場合 (ほんとうはガソリンの燃費に差がないのに，t 値がたまたま0から離れていたせいで，帰無仮説「燃費に差がない」を棄却してしまう場合) が，**タイプIのエラー**である。すでに述べたように，統計的仮説検定とは，このタイプIのエラーの確率が一定の値 (有意水準) になるように棄却域を設定する，というアイデアを採用した技法であった。すなわち，タイプIのエラーの確率は，私たちの手で制御されている。

次に，H_0 が偽のとき(表の右列) について考えよう。それにもかかわらず，運悪く H_0 が棄却されない場合 (ほんとうはガソリンの燃費に差があるのに，t 値がたまたま0に近かったせいで，帰無仮説「燃費に差がない」が棄却されなかった場合) が，**タイプIIのエラー**である。タイプIIのエラーが起こる確率は，わからない[4]。つまり，タイプIIのエラーは，私たちの手で制御されていないのである。

さて，H_0 が棄却されたとき(表の下行)，私たちはそれを根拠に，H_0 の反対側の主張を支持することができる。もちろん，この推測はまちがっているかもしれない (実は H_0 は真かもしれない)。しかし，その間違い (タイプIのエラー) を犯す確率を，私たちは私たちの手で制御している。

いっぽう，H_0 が棄却されなかったとき(表の上行)，それを根拠に H_0 を支持するのは，非常に危険である。当然，その推測はまちがっているかもしれない (実は H_0 は偽かもしれない)。問題は，その間違い (タイプIIのエラー) を犯す確率が，非常に高いかもしれない，ということだ。だから，H_0 が棄却されなかったときには，そこから積極的な主張を引き出さず，単に沈黙するのが正しい。

[4] タイプIIのエラーが起こる確率は，数学的に予想できるのだが，詳しくは述べない。たいていの研究では，その確率はかなり高いと考えて良い。

ここまでで紹介してきたのは，仮説検定の手法のなかのひとつで，一般に「t検定」と呼ばれている手法である。

それでは，手元の問題に，仮説検定の技法を適用する手順をまとめておこう。

データ解析の教科書には，いろいろな検定手法が紹介されている。たとえばt検定については，次の情報が載っているだろう。

通称	t **検定** (対応のないt検定)
データの形	独立な2群から得られた量的データ。 $\begin{array}{r\|c} \multicolumn{2}{c}{\text{A群}} \\ \hline & x_{A1} \\ & x_{A2} \\ & \vdots \\ & x_{n_A} \\ \hline \text{平均} & \bar{x}_A \\ \text{標準偏差} & s_A \end{array} \qquad \begin{array}{r\|c} \multicolumn{2}{c}{\text{B群}} \\ \hline & x_{B1} \\ & x_{B2} \\ & \vdots \\ & x_{n_B} \\ \hline \text{平均} & \bar{x}_B \\ \text{標準偏差} & s_B \end{array}$
前提条件	2つの母集団が，それぞれ正規分布にしたがっており (正規性)，分散が等しい (等分散性)。
帰無仮説	2つの母集団の平均μ_A, μ_Bが等しい ($H_0 : \mu_A = \mu_B$)。
検定統計量	$t = \dfrac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2}(\frac{1}{n_A} + \frac{1}{n_B})}}$ は，H_0のもとで自由度$n_A + n_B - 2$のt分布に従う。

問題を与えられたとき，まず必要なのは，その問題に，教科書のなかのどの手法が適用できるか，に気づくことである。

> 準備 1. “ガソリン燃費問題” には,「(対応のない) t検定」が適用できそうだ。

つぎに必要なのは，手元の問題が,「(対応のない) t検定」の前提条件に沿っているかどうか，の判断である[5]。

> 準備 2. 標本のヒストグラムをみるかぎり，母集団の正規性・等分散性という仮定には，無理がなさそうだ。

前提条件が満たされているならば，仮説検定の4つのステップに従って考える。

> Step 1. 帰無仮説H_0を「ガソリンの燃費に差はない」とする。
> Step 2. 検定統計量の値は$t = 2.16$である。
> Step 3. H_0のもとで，検定統計量tは自由度18のt分布に従う。5% 棄却域は$|t| > 2.10$である。
> Step 4. 帰無仮説は有意水準5% で棄却される。したがって,「ガソリンの燃費には差がある」と判断できる。

[5] データが母集団からの無作為標本であることは，ここに書くまでもない大前提である。

5.1 統計的推測のしくみについて

Q1. 手元にあるデータ (標本) が，関心を持っているデータ全体 (母集団) から無作為抽出されたものであることが，統計的推測の前提になる，ということはわかりました。それでは，いま私が持っているデータは，母集団からの無作為標本だ，といってよいでしょうか?

A. 現実の研究においては，手元のデータが母集団からの無作為標本だ，ということを保証する方法は，なにもありません。むしろ，母集団は「手元のデータの性質を一般化できる範囲」のことだ，と考えた方が，わかりやすいでしょう。

“ガソリン燃費問題” について考えてみましょう。実験をおこなった間，たまたま快晴が続いていた，とします。すると,「この実験の結果は，雨や雪の日には適用できないかもしれない」という疑いがわきます (晴れの日にはガソリン A のほうが燃費がよかったけれども，雨や雪の日には，ガソリン B のほうが燃費がよいかもしれません)。もっともな疑問ですが，この疑問をもたらしてくれたのは，統計学の知識ではなく，当該の問題についての常識と考察です。

このように，母集団についてどのように考えるべきかは，もはや統計学的な問題ではありません。(cf. [佐伯・松原:2000] 3 章)

Q2. 無作為割り当てとはなんですか?

A. 実験研究においては，データが無作為標本だ，という前提に並んで，もうひとつ，大事な前提があります。それは，データが条件の各水準に，無作為に割り当てられている，という前提です。

“ガソリン燃費問題” でいえば，たとえばベテランの運転手を A 群に，初心者の運転手を B 群に割り当てたとしましょう。このとき，A 群と B 群の燃費に差がみられても，その差はガソリンの差によるのか，それとも運転スキルの差によるのか，区別することができません。したがって，実験研究においては，データが**無作為抽出**されていることと**無作為割り当て**されていることの 2 点が重要だ，といえます。(cf. [佐伯・松原:2000] 3 章)

5.2 仮説検定について

Q3. 帰無仮説は，分析者が主張したいことの逆の仮説になる，と聞いたことがあります。ほんとうですか。

A. 必ずしもそうなるわけではないのですが，そうなることが多いのは事実です。その理由は，仮説検定の非対称性にあります。

“ガソリン燃費問題” について考えてみましょう。いま分析者が「ガソリン A, B には燃費の差がない」という方向で論述を進めている，とします。しかし，分析者が内心どのように考えていようが，H_0 は「ガソリン A, B には燃費の差がない」となります (そうでないと，t 検定の手法が適用できません)。すると，

- H_0 が棄却された場合，分析者の論述を否定する証拠となり
- H_0 が棄却されない場合，そのことは特に意味を持たない

ので，どちらにしても，分析者の論述を支持する結果にはなりません。ですから，分析者が「ガソリン A, B には燃費の差がない」という方向での論述を進めている場合には，この仮説検定はあまり適切な手法ではありません。

このような事情のせいで，帰無仮説は分析者の主張の反対側の仮説となることが多くなります。しかし,「帰無仮説をどのように設定したらよいか」という問題は,「自分がどの方向の主張をしたいか」とは切り離して，テクニカルな問題としてとらえてください。

Q4. それでは，たとえば「2 つのガソリンの間に燃費の差がない」と主張したい場合には，どのような実験をすればよいのでしょうか。

A. これは難しい問題です。分析者が示したい主張が，仮説検定論における帰無仮説の側になってしまっているからです。

この場合は，“ガソリン燃費問題” のような実験研究をおこなうこと自体が，分析者にとって適切な研究手法ではありません。この種の実験をいくら積み重ねても,「2 つのガソリンの間に燃費の差がない」という証拠を得ることは難しいからです。それでは，どのような研究をすればよいのか? という問いには，簡単には答を出せません。

“ガソリン燃費問題” のようなタイプの実験の結果から,「2 つのガソリンの間に燃費の差がない」ということをあえて示したいのならば,

- 標本のサイズ n_A, n_B を大きくし，タイプ II のエラーの確率が十分に小さいことを示す
- 母平均の差 $\mu_A - \mu_B$ についての**信頼区間**を求め，それが 0 を含んでいることを示す
- 一般的な仮説検定ではなく，まったく異なる統計的推測手法を用いる
- 統計的推測をあきらめ，図と論述のみで読み手に納得してもらえるよう努力する

といった方法が考えられるかもしれませんが，困難な路線であることはたしかです。(cf. [佐伯・松原:2000] 4 章, [繁桝ほか:1999] p.30)

Q5. 有意水準とは,「判断が誤っている確率」と考えてよいでしょうか。

A. たとえば "ガソリン燃費問題" では,，帰無仮説は有意水準 5% で棄却され,「燃費に差がある」という判断が得られました。この判断は，5% の確率で間違っている，といってよいでしょうか。

そうではありません。下の表をもう一度みてください。

	H_0 は真	H_0 は偽
H_0 を棄却しない	○ $(1-\alpha)$	× (β) **タイプ II のエラー**
H_0 を棄却する	× (α) **タイプ I のエラー**	○ $(1-\beta)$

H_0 を棄却したとき，その判断が誤っている確率は，＜表の下の行の 2 マス (α と $1-\beta$) を 100% としたときの，左側のマスの割合＞に相当します。

いっぽう，有意水準とは,「H_0 が真のときに H_0 を棄却してしまう確率」です。すなわち，＜左表の上下 2 マスを 100% としたときの，下のマスの割合 (α) ＞に相当します。

Q6. 「標本のサイズが小さいと，帰無仮説はたいてい棄却されない。標本のサイズが大きいと，帰無仮説はたいてい棄却される。」と聞きました。ほんとうですか。

A. まず,「標本のサイズが小さいと，帰無仮説はたいてい棄却されない」というのは事実です。このことは筋が通っています。つまり，標本のサイズが小さいときには，母集団の性質の推測が困難ですから，母集団についての判断を下すのはなるべく避けるべきです。

問題は,「標本のサイズが大きいと，帰無仮説はたいてい棄却される」という点です。これも事実です。このことは，検定統計量の式をよくみるとわかります。t 検定の場合には，n_A, n_B が極端に大きければ，t 値は大きくなる傾向があります。

はたしてこのことは，筋が通っているといえるでしょうか？どんな対立仮説であれ，データさえ増やせば支持される，ということにならないでしょうか？

残念ながら，仮説検定にはそのような弱点があります。このことは，帰無仮説と対立仮説が，もともと対称的でない，ということに由来しています。たとえば "ガソリン燃費問題" でいえば，対立仮説「燃費に差がある」には，その差の程度に無限の可能性があるのに対して，帰無仮説「燃費に差がない」は非常に限定された仮説です。もともとこの 2 つを比較すること自体がフェアでない，という見方もできます。

以上の理由により，仮説検定は，大量のデータの解析にはあまり適していません。(cf. [繁桝ほか:1999] p.75)

Q7. 仮説検定の結果を文章で述べるときには，どのようにすればよいでしょうか。

A. (1) データの記述統計量, (2) 検定手法の種類，(3) 検定統計量の値，(4) 自由度，(5) 有意水準，(6) 棄却の有無，について要領よく述べてください。

以下に，"ガソリン燃費問題" の結果について報告した文章の例を挙げます。6 つの項目に下線をつけてあります。

> ...被験者20名を10名づつ2群にわけ，一方の群にはガソリンAを，他方の群にはガソリンBを使用させ，燃費得点を得た。平均(SD)はそれぞれ60.4点(4.7点), 55.8点(4.4点)であり，t検定の結果，有意な差が示された($t = 2.16$, $df = 18$, $p < .05$)。

帰無仮説が5%有意水準で棄却されたときには,「$p < .05$」と書く習慣があります。

Q8. p値とはなんですか。

A. 仮説検定の結果を報告する場合，p値と呼ばれる数値を付け加えることがあります。“ガソリン燃費問題” の結果でいえば，以下のようになります。

> ...であり，t検定の結果，有意な差が示された ($t = 2.16$, $df = 18$, $p = .044 < .05$)。

下線部がp値です。コンピュータ・ソフトウェアをつかって仮説検定をおこなうと，たいていp値を出力してくれるので，その値をそのまま書く，と考えれば良いと思います。

p値の意味についてはここでは触れませんが，p値には「有意水準を××%としたとき，$p <$××%ならば，H_0は棄却される」という性質があります。上の場合では，p値は5%より小，1%よりも大です。これをみるだけで，帰無仮説は有意水準5%では棄却できるけど，1%では棄却できないんだな，ということがわかります。

なお，p値にはそれ以上の意味はありません。p値が小さければ小さいほど，2群の差が大きい，といった捉え方は避けてください。(cf. [繁桝ほか:1999] p.34)

Q9. “ガソリン燃費問題” で，分析者がほんとうに知りたいのは,「2種類のガソリンの燃費に差があるかどうか」ではなくて,「2種類のガソリンの燃費にどの程度の差があるか」なのではないのでしょうか。だから，t検定は適切な手法ではないと思うのですが。

A. この意見にも一理あります。これはたいへん難しい問題で，容易に答えが出せません。ここでは,「t検定では，2つのガソリンの燃費にどの程度の差があるかはわからない」ということを確認しておきたいと思います。

5.3 t 検定について

Q10. 母集団の正規性や等分散性が仮定できない場合は，どうすればよいでしょうか。

A. 標本のヒストグラムが，左右対称になっておらず，極端にゆがんでいるような場合には，母集団の正規性という前提が疑われます。また，標本の標準偏差 s_A と s_B が大きく異なっている場合には，母集団の等分散性という前提が疑われます。

こうした場合には，

- **ウェルチ検定**などの近似的検定手法を用いる
- **変数の変換**をおこなって，分布の正規化・等質化をはかる
- データを量的でなく質的なものとみなして，質的データのための検定手法 (**ノンパラメトリック検定**) を用いる

といった方法があります。それぞれについての詳細は参考文献を参照してください (たとえば，[森・吉田:1990])。

なお，標本のサイズ n_A, n_B が十分大きいときには，母集団の正規性と等分散性のうち，どちらかが欠けていても，t 検定をおこなってかまわない，ということがわかっています。

Q11. 2 群の標本のサイズは等しくなければならないでしょうか?

A. 2 群の標本のサイズ n_A, n_B は異なっていてもかまいません。ただし，n_A, n_B が極端に異なるデータについての分析にあたっては，t 検定をおこなう以前の問題として，なぜそのようにサイズがちがうのか，よく考察しておく必要があります。(cf. [繁桝ほか:1999] p.26)

Q12. 自由度が大きすぎて，t 分布表から棄却域を調べられません。どうすればよいでしょうか。

A. 正規分布表をつかってください。自由度が大きくなると，t 分布は次第に標準正規分布 (平均 0, 分散 1 の正規分布) に近づきます。(cf. [繁桝ほか:1999] p.27)

Q13. “ガソリン燃費問題” において，第 3 のガソリン C についても同様のデータが得られていた，とします。このとき，(1)A と B を比較する t 検定，(2)B と C を比較する t 検定，(3)A と C を比較する t 検定，の 3 回の t 検定をおこなってよいでしょうか?

A 群	
	x_{A1}
	x_{A2}
	$\vdots$
	x_{n_A}
平均	$\bar{x_A}$
標準偏差	s_A

B 群	
	x_{B1}
	x_{B2}
	$\vdots$
	x_{n_B}
平均	$\bar{x_B}$
標準偏差	s_B

C 群	
	x_{C1}
	x_{C2}
	$\vdots$
	x_{n_C}
平均	$\bar{x_C}$
標準偏差	s_C

A. これは典型的な誤りです。ひとことでいえば，**同じデータをつかって，複数回検定をおこなってはいけません**。この点についての説明は，参考文献を参照してください。([永田・吉田:1997] にわかりやすい説明があります。)

このように，3つ以上の水準がある要因の影響について検討する場合には，**分散分析**や**多重比較**という方法を用います。参考文献を参照してください (たとえば，[森・吉田:1990])。

Q14. 対応のある t 検定とはなんですか?

A. 例として，次の問題を考えてみましょう。

ガソリン燃費問題 II

とあるタクシー会社では，ガソリン A,B の燃費を比較するために，10 人の運転手を選び，それぞれの被験者に，ガソリン A,B の両方を，一定期間利用させて，燃費をしらべた。結果は下のようになった。ガソリン A,B の間には，燃費の差があるといえるだろうか？

	A 使用時		B 使用時
鈴木	61	—	55
佐藤	60	—	54
⋮	⋮		⋮
森	64	—	58
平均	50.4		55.8
SD	4.7		4.4

上のデータでは，たとえば鈴木さんの A 使用時燃費得点 (61) と，B 使用時燃費得点 (55) を，線で結ぶことができます。このような 2 組のデータを，**対応がある**データ，と呼びます。

データに対応があるときには，t 検定量として

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2 + s_B^2 - 2 \times r \times s_A \times s_B}{n-1}}}$$

を用いて検定をおこないます (r は 2 組のデータから得られる相関係数です)。H_0 のもとで，この t は自由度 $n-1$ の t 分布に従います。この手法を，**対応のある t 検定**と呼びます。

参考文献

- 森敏昭・吉田寿夫 (編著) 「心理学のためのデータ解析テクニカルブック」, 北大路書房, 1990.
- 永田靖・吉田道弘「統計的多重比較法の基礎」, サイエンティスト社, 1997.
- 繁桝算男・柳井晴夫・森敏昭 (編著) 「Q&A で知る統計データ解析 DOs and DON'Ts」, サイエンス社, 1999.
- 佐伯胖・松原望 (編著) 「実践としての統計学」, 東京大学出版会, 2000.

おわり