統計的仮説検定：資料

芳賀昭弘 [*]

1　統計的仮説検定の基本

仮説を予め設定し、その仮説が正しいと仮定した上で、それに従う母集団から、実際に観察された標本が抽出される確率を求め、その確率の値によりその仮説が正しかったかどうかを判断する方法を統計的仮説検定という（例として図 1 参照）。

手順	やること
1	母集団に関する帰無仮説 [*a] と対立仮説 [*b] の設定。問題の設定の仕方により両側検定 [*c]or 片側検定 [*d] が決まる。
2	検定統計量（z 値、t 値、χ^2 値、F 値など）を選ぶ（検定方法を選ぶ）
3	有意水準 α の値を決める
4	データの検定統計量の実現値を求める
5	検定統計量の実現値が棄却域に入れば帰無仮説を棄却して対立仮説を採択する。

a: 「差がない」「効果がない」という仮説を設けること。H_0 という記号を使う。
b: 帰無仮説とは逆（「差がある」「効果がある」）の仮説。H_1 という記号を使う。
c: 差があるかどうかを知ることが重要であり、どちら（どれ）が「大きい」or「小さい」のかは気にしない（図 2 参照）。
d: ある集団より別の集団の方が「大きい」or「小さい」ことを知りたい場合（図 2 参照）。

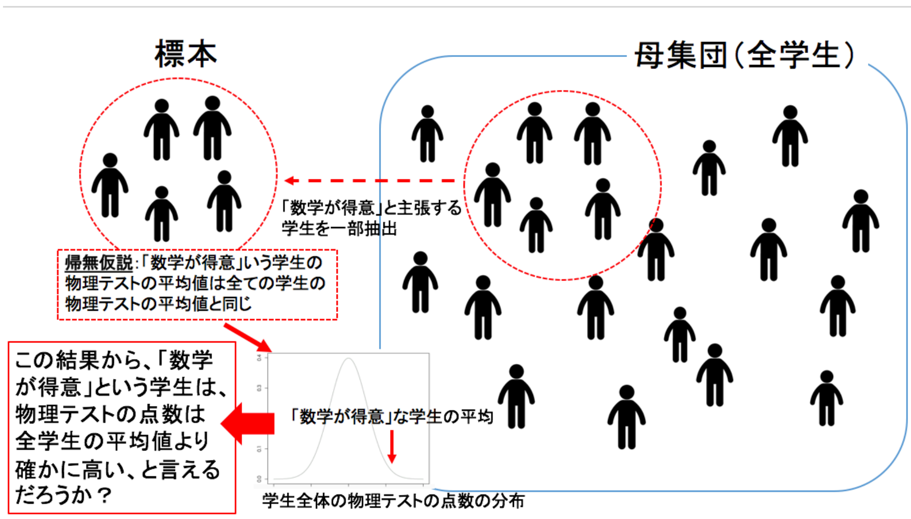

図 1　統計的仮説検定の考え方

* Electronic address: haga@tokushima-u.ac.jp

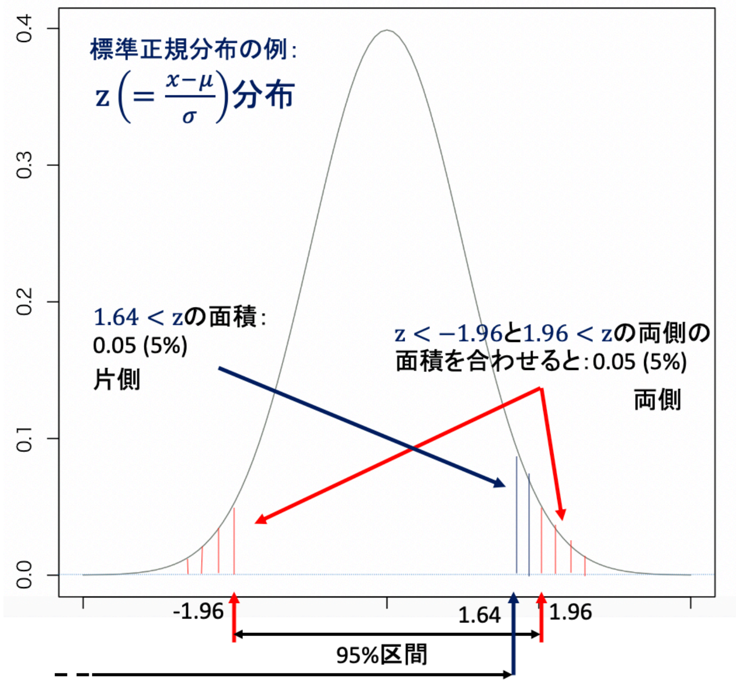

図 2　両側・片側検定における 95% 信頼区間

1.1　標準正規分布を用いた検定

母集団の性質（平均値や分散）がわかっている場合もしくはサンプル数が十分に大きい場合に利用できる検定。

例題 1 ：0 から 10000 番まで番号がついた病院アンケートを、0~5000 番までは女性、5001~10000 番を男性の患者に配布した。100 枚のアンケートを回収したときの番号の平均値が 4135 であった。回収されたアンケートは、ランダム・サンプルされたものと結論することができるか？ → 母集団 0~10000 を全て使って解析可能（平均値や分散も求めることができる）なので、母集団の性質がわかっている場合に利用できる検定を使う。

（手順）

帰無仮説 H_0：回収アンケート数に男女間の差はない（対立仮説 H_1：回収アンケート数に男女間の差がある）

検定統計量：z 値（標準正規分布の変数)。ここで $z = \frac{(x-\mu)}{\sigma/\sqrt{n}}$ で、μ、σ はそれぞれ母集団の平均値と標準偏差。n はサンプル数。

有意水準：両側 5% とする

（考え方）

母集団の平均 μ は 5000、標準偏差 σ は $\sqrt{(1/10001)\sum_{i=0}^{10000}(x_i-\mu)^2} = 2887$ なので、4135 に対する z 値は $z = \frac{(x-\mu)}{\sigma/\sqrt{n}} \sim\ -2.996$ である。標準正規分布の場合、95% 信頼区間は $(-1.96, 1.96)$ なのでそれから大きく外れている。偶然では起こりにくい結果が得られたということ。

理解を深めるため、同じことを前回やったように平均値を推定することで確かめてみよう。回収した 100 枚のアンケートが標本。この 1 回の標本抽出で得た値が 4135 であり、この値から期待される真の平均値の 95% 信

頼区間は、$\bar{x} - 1.96\sigma/\sqrt{n} \leq \mu \leq \bar{x} + 1.96\sigma/\sqrt{n}$ より $3569.1 \leq \mu \leq 4700.9$ となる。実際には $\mu = 5000$ なので、z 値で得られた結論と同じくやはり大きく外れていることが確かめられる。

今回得られた 100 枚のアンケート番号の平均値 4135 というのは非常に起こりにくいということがわかりました。では、どのくらいの確率で起こり得る値なのでしょう？それは z 値が-2.996 という値から教科書の付録にある標準正規分布表から調べることができます。$z \sim\ 3$ のところを見てみましょう。上側確率で 0.0013 となっていますので両側確率で 0.0026 ということです。つまり、平均値 4135 を得る確率（正確に言うと、5000 から 865 以上離れた平均値を得る確率）は 0.3% もない、非常に稀な結果だったということが定量的に示されました。

p 値が____________であるので､統計検定量は有意水準______であり､帰無仮説は______________________________。

0∼5000 番までは女性、5001∼10000 番を男性の患者に配布したので、若い番号に偏りがあったということは女性の方がアンケートに答えてくれた、ということです。そして、以上の統計的な定量値から、それが統計的に有意に示されたことになります。

1.2　t 分布を用いた検定

母集団の分散がわからない場合に利用する検定。

例題 2：ある日、ある酪農家が搾乳中のホルスタイン 5 頭の乳量を調べたら、一頭あたりの平均乳量は 22.1 リットル、標本標準偏差は 6.5 リットルでした。一方、これまでのデータの平均値は 25.8 リットルでした。この日のホルスタイン 1 頭あたりの乳量は平均値から有意にずれていると言えるでしょうか？

（手順）

帰無仮説 H_0：この日のホルスタイン 1 頭あたりの乳量は平均値と差がない（対立仮説 H_1：差がある）

検定統計量：t 値（t 分布の変数)。

有意水準：両側 5% とする

（考え方）

今、母集団の標準偏差がわかりません。また標本も 5 つなので標本標準偏差に母集団の標準偏差を代用するわけにもいきません。そういう場合には t 分布を使います。ちなみに t 分布は z 分布の標準偏差に曖昧さがある場合に拡張した分布です。

自由度 $n-1$ の t 分布は

$$t = \frac{x - \mu}{s/\sqrt{n-1}}$$

であり（μ は母集団の平均、s は標本標準偏差)、上記のホルスタインのデータでは $t = \frac{22.1-25.8}{6.5/\sqrt{5-1}} = -1.14$ となります。教科書の付録の t 分布表を見ると、df：自由度=4 で 1.14 となるところは上側確率で $p > 0.1$ のところのようです（p が 0.1 のところで $t = 1.533$ なので、$t = 1.14$ となる p はもっと大きいはず)。ということは、今回の乳量は平均より少ないようでも、普段からそのくらいの増減があるあまり珍しくない結果だった、ということがわかりますね。

p 値が____________であるので､統計検定量は有意水準______であり､帰無仮説は______________________________。

1.3 t 分布を用いた検定２：相関係数

２つのデータに相関があるかどうかを調べるときに利用する検定。

例題３：密度がわかっている 7 つの物質を CT で撮影したところ、CT 値（画像上での濃淡を表す信号値）は次のようなデータとなった。

密度 0.26, 0.52, 0.93, 1.0, 1.25, 1.46, 1.92

CT 値 -817, -522, -90, 0, 491, 741, 1547

密度と CT 値に相関があると言えるか？

（手順）

帰無仮説 $H_0: \rho = 0$、すなわち母相関はない（対立仮説 $H_1: \rho \neq 0$、すなわち母相関がある）

検定統計量：t 値。なぜなら r を相関係数とすると、$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ は自由度 $n-2$ の t 分布となる。

有意水準：両側 5% とする

相関係数 $r = 0.993$ である（計算してみましょう）。データ数 $n = 7$ のときの t 値は 18.3。教科書の付録の t 分布表を見ると、df : 自由度=5 で上側確率の p 値が 0.001 となるところでさえ $t = 5.893$ です。それより大きいので、p 値は相当小さいことがわかります。

p 値が__________であるので、統計検定量は有意水準_____であり、帰無仮説は____________________。

ちなみに相関係数の検定では、p 値が小さいということはそれだけ相関が高いということを示しています。なぜそうなのか、また、なぜ t 分布で相関係数の検定ができるのか、などの詳細を知りたいという人は、言ってください。厳密に証明する資料を配布します（初等数学で証明できますが、少し大変です）。

1.4 χ^2 検定（カイ二乗検定）

好きか嫌いか、などのような 2 つもしくはそれ以上の質的変数が独立かどうかを調べるために利用する検定。

例題４：治療 A と治療 B の効果を調べたところ、次のような結果が得られた。

	治療 A	治療 B	計
有効（P）	8	13	21
無効（N）	2	7	9
計	10	20	30

治療 A と治療 B に差があると言えるか？

（手順）

帰無仮説 H_0：治療 A と治療 B は独立、すなわち差がない（対立仮説 H_1：治療 A と治療 B に関連あり、すなわち差がある）

検定統計量：χ^2 値。$\chi^2 = \sum_i (O_i - E_i)^2/E_i$ で、O_i は観測値、E_i は期待値。自由度は 1。

有意水準：片側 5% とする

なお、期待値は次のようになる：

	治療 A	治療 B	計
有効（P）	$10 * 21/30 = 7$	$20 * 21/30 = 14$	21
無効（N）	$10 * 9/30 = 3$	$20 * 9/30 = 6$	9
計	10	20	30

$\chi^2 = (8-7)^2/7 + (13-14)^2/14 + (2-3)^2/3 + (7-6)^2/6 = 0.714$ この実現値は珍しい値でしょうか？教科書の付録の χ^2 分布表を見ると、自由度 1 では $\chi^2 = 0.714$ を与える確率は p 値が 0.1 以上になるようです。

p 値が____________であるので、統計検定量は有意水準______であり、帰無仮説は______________________________。

一見、差がありそうですが、この結果だけでは差があるとは言えないようです。なお、χ^2 分布についても厳密に証明できます。

1.5　F 検定

２つのグループの分散が等しいかどうかを調べるために利用する検定。

例題５：教科書第 1 章のキュウリの収量のデータを使って、栽培法 A と栽培法 B のキュウリの収量の母分散が等しいかどうかを検定せよ。

（手順）
帰無仮説 H_0：母分散は等しい（対立仮説 H_1：母分散は異なる）
検定統計量：F 値（F 分布の実現値）。$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ で計算する。ここで $\hat{\sigma}^2$ は不偏分散であり、分子が分母より大きいように計算すること。
有意水準：片側 5% とする

栽培法 A の不偏分散は 183227、栽培法 B の不偏分散は 133515.4 である。よって $F = 1.372$ となる。この実現値が珍しいかどうかを教科書の付録 F 分布表から確かめてみよう。データ数はともに 15 なので自由度はともに 14 となる。F 分布表の上側 5% の値を見ると、信頼区間 95% を与える範囲がおおよそ 2.5 以下だとわかります。実現値はそれより随分値が小さいので全く珍しくないありふれた結果が得られたことになります。つまり、栽培法 A と栽培法 B のキュウリの収量の母分散は等しいと考えて良さそうです。

2　用語についての補足

第 1 種の誤り（過誤）$\cdots$ 帰無仮説が実際には真であるのに棄却してしまう過誤（偽陽性）

第 2 種の誤り（過誤）$\cdots$ 対立仮説が実際には真であるのに帰無仮説を採用してしまう過誤（偽陰性）

有意水準 $\cdots$ 片側（上側 or 下側）、両側以上の検定量が発生する確率（パーセント）のことで、5% や 1% がよく使われる（物理学の世界では $3\sigma(\sim 0.13\%)$ で Evidence レベル、$5\sigma(\sim 0.00003\%)$ で Discovery レベル、数学の世界では証明されなければ、どんなに確率が低い事象があったとしても無意味（実際に何千万とい

う事象から正しいと考えられた「予想」が証明によって覆されることがあるということは、数学史が証明している）と言われているように、分野によって有意水準の値や考え方は異なる)。

p 値 $\cdots$ 統計的仮説検定において、帰無仮説の元で検定統計量がその値となる確率（上の有意水準における 5% や 1%）のこと。p 値が小さいほど、検定統計量がその値となることはあまり起こりえないことを意味する。

3　演習問題

1. ある血液の μL 辺りの白血球数を 5 回測定したところ、
   6623, 6887, 7054, 6289, 7003
   という観測値が得られた。真の白血球数 [個 $/\mu L$] の 95% 信頼区間及び 99% 信頼区間を求めよ。
   ヒント：
   自由度 $n-1$ の t 分布は

$$t = \frac{x-\mu}{s/\sqrt{n-1}}$$

   である（μ は母集団の平均、s は標本標準偏差)。
   95% 信頼区間を与える t 値（上側）を p95level、99% 信頼区間を与える t 値（上側）を p99level と書くとするとしたがって、

$$-\mathrm{p95level} < \frac{x-\mu}{s/\sqrt{n-1}} < \mathrm{p95level}$$
$$-\mathrm{p99level} < \frac{x-\mu}{s/\sqrt{n-1}} < \mathrm{p99level}$$

   となる x が白血球数 [個 $/\mu L$] の 95% と 99% の信頼区間。p95level と p99level はデータ数によって変わることに注意。

2. 過去のデータから、あるテストの平均値は 60 の正規分布に従うことがわかっている。今回 20 名でそのテストを実施したところ、平均値が 65、標本分散が 8 であった。この 20 名は、過去の人たちと同じ母集団からの無作為標本と考えてもよいか？
   (手順)
   帰無仮説 H_0：無作為標本と考えて良い。
   検定統計量：t 値（母集団の標準偏差が不明)。
   有意水準：両側 5% とする。
   平均値が 65、標本分散が 8、標本数 $n=20$ に対する t 値を求める。自由度____________の t 分布なので、t 値に対する確率 p 値を求める。
   p 値が____________であるので、統計検定量は有意水準______であり、帰無仮説は______________________________。