

2 グループの平均値を比べる

高木 廣文

統計学講座

2 グループの平均値を比べる

高木 廣文

はじめに

多くの研究では、患者と健常者、喫煙者と非喫煙者のように、ある特性により調査対象者を分類し、その分類されたグループごとに同一項目の測定値の平均値を比較することがよく行われる。身長や体重などのような測定値の場合、測定値から平均値を計算し、それを用いて各グループの比較を行うのが普通だろう。場合によっては、収集したデータそのままの値ではなく、順序や大小関係をもとにした比較も行われるが、その方法については別の機会に紹介する。ここでは、主に2つのグループの平均値を比べるための方法について解説を行う。

2 グループの平均値の比較の方法

2 グループの平均値の比較という場合、この「平均値」という言葉は、実際には各グループが代表している母集団での平均値である「母平均値」を示していることを、まず断っておく必要があるだろう。データが無作為に収集された場合、偶然の影響により、求められた平均値（標本平均値）が、2つのグループで完全に一致する場合よりも、一方が大きく他方が小さいという、不一致の場合の方が多いに違いない。したがって、単に標本から得られた平均値を見比べて、「患者グループで大きかった」といった結論をすぐに導くことはできない。

統計学的に母集団での2グループの平均値を比

較するためには、

- (1) 母平均値の信頼区間を求めて比較する
- (2) 帰無仮説「2グループの母平均値は等しい」を検定する

などの方法に大別できよう。

一般によく用いられる方法は(2)であり、2グループの母平均値に有意差が認められた、などとその結果を表現する。しかし、「有意差」の有無で2グループ間の違いを表現する前に、各グループの母平均値がどのくらいの範囲にあるのかを求めておくことは、情報量の点からいえば、単純な検定よりもずっと優れたものである。この点から、まず母平均値の信頼区間の求め方について解説する。

母平均値の信頼区間

母平均値の信頼区間を求めるために、まず「データの標準化」と正規分布について解説する。その後、データの標準化に不偏分散を用いた場合について、t分布との関係を説明し、最後に母平均値の信頼区間の計算方法について解説し、簡単な例について実際の計算の過程を示す。

1) 標準化と正規分布

身長ならば170cm、体重ならば70kgのように、各変数の単位や平均値は異なっているのが当たり前であるが、一般的な方法を考える場合、これは極めて都合が悪いことになる。そこで、データの平均が0、分散が1になるようにデータの変換を行うことがある。この操作は「標準化」と呼ばれる。母集団からデータ x が無作為に抽出されており、かつそのデータ x が平均 μ （みゅう）、分散 σ^2 （しぐま2乗）の正規分布から抽出されたものであれば、

$$(1) z = \frac{x - \mu}{\sigma}$$

と変換すると、変換後のデータ z は、平均 0、分散 1 の正規分布（標準正規分布）に従うことになる。標準正規分布は、図 1 に示したような正規分布の中で、平均が 0、標準偏差が 1 の場合の分布であり、平均 0 を中心としたピークが 1 つという、左右対称の一峰性の分布である。標準正規分布を用いることで、確率計算が容易になるというメリットがある。例えば、よく知られているように、仮に $z = 1.96$ であれば、それは上位 2.5% の中に入る大きさのデータであることがわかる。

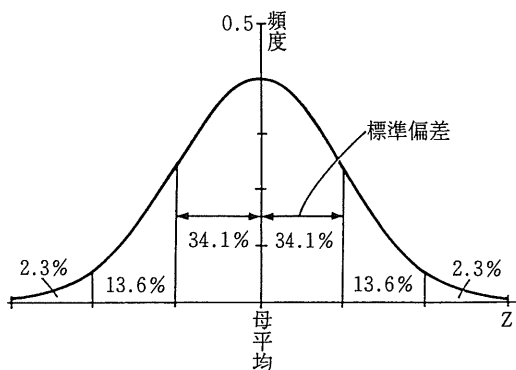


図 1 正規分布

この標準化の考え方を、データの平均値に応用してみよう。いま、母集団での分布が平均 μ 、分散 σ^2 の正規分布に従う分布から、 n 個のデータ $x_1, x_2, \dots, x_n$ が無作為に抽出されたものとしよう。前回簡単に説明したように、データの平均値 x_M の期待値は母平均値 μ に一致する。また、平均値の分布の分散は、 σ^2/n となる（ $\sigma/\sqrt{n}$ は平均値の「標準誤差」と呼ばれる）ことが知られている。これらのことから、

$$(2) z = \frac{x_M - \mu}{(\sigma/\sqrt{n})}$$

と変換すると、 z は平均 0、分散 1 の標準正規分布に従うことになる。

標準正規分布の -1.96 から 1.96 の値が 95% の範囲を与えるので、 $-1.96 < z < 1.96$ の不等式を解

くことにより、母平均 μ の 95% 信頼区間を求めることができることになる。

実際に不等式を解くと、

$$(3) x_M - 1.96 \times (\sigma/\sqrt{n}) < \mu < x_M + 1.96 \times (\sigma/\sqrt{n})$$

となる。母集団の分散 σ^2 の値がわかっているのならば、(3)式により母平均値の 95% 信頼区間の計算が行える。

2) 不偏分散と t 分布

上記(3)で求めた方法は、残念ながら実際的ではない。なぜならば、母平均値 μ がわからないのに、その分散がわかっている場合というのは、普通はあり得ないからである。

母分散の推定値としては、不偏分散 S^2 がよく用いられている。不偏分散は、

$$(4) S^2 = \frac{1}{n-1} \sum_i (x_i - x_M)^2$$

により求められる。ここで、 $\sum_i$ はすべてのデータについて和を求めることを示している。

(2)式の σ のかわりに、不偏分散の平方根 S （標準偏差）を用いた場合、

$$(5) t = \frac{x_M - \mu}{(S/\sqrt{n})}$$

により求められる t は、どのような分布に従うかを知る必要がある。この問題については、すでに Student（本名 W. G. Gosset）により解答が与えられている。すなわち、上記の t は、自由度 $n-1$ の t 分布に従う。 t 分布は、その発見者の名前から「Student の t 分布」とも呼ばれている。 t 分布は標準正規分布に形状が極めて近く、やや分布の裾が長い³⁾、平均 0 を中心に左右対称の分布である。自由度が大きくなるにつれ（これは標本数が大きくなることを意味する）、標準正規分布に近づき、理論上は自由度が無限大で一致するのである⁴⁾、実際上は自由度が 200 程度でほぼ一致する。

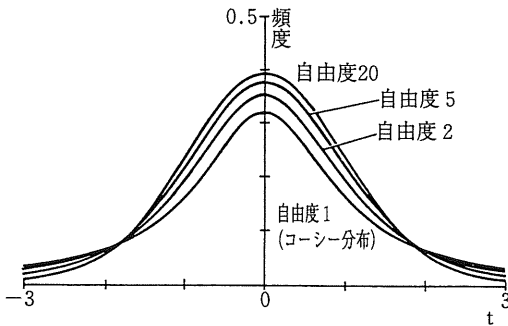


図2 t分布の自由度と形状

ここで、「自由度」という言葉が登場するが、不偏分散を計算する場合の(4)式に出てくる「 $n-1$ 」に一致するものである。(5)式のSは母集団での σ の推定値であるが、この計算には、 n 個のデータ $x_1, x_2, \dots, x_n$ と、平均値 x_M とが用いられている。ところで、平均値 x_M は母平均値の推定値として n 個のデータ $x_1, x_2, \dots, x_n$ から求められている。もともと n 個のデータしかなかったものが、平均値 x_M を求めたため、これらのデータを使用して計算される不偏分散は、 $n-1$ 個のデータを用いた計算を行った段階で、 n 個のデータの合計は平均値 x_M の n 倍になるために、最後の n 番目のデータがどのような値になるかは決まってしまう。このような状況を、データの自由度は $n-1$ であるといい、不偏分散は各データと平均値の偏差平方和をその自由度で割ったものとなっている。t分布は(5)式の分母にあるSの自由度と同じ自由度を持つことになる。

3) 母平均値の信頼区間の計算

母平均値の95%信頼区間を求めるには、t分布の上側2.5%点の値が必要である。これは、自由度 ν (にゆう) により値が異なるので、ここでは、 $t_{0.025}(\nu)$ と書くことにする。

(3)式の標準正規分布の場合の母平均値の95%信頼区間を求める式では、1.96を用いているが、1.96を $t_{0.025}(\nu)$ とし、 σ をSにかえると、

$$(6) \quad x_M - t_{0.025}(\nu) \times (S/\sqrt{n}) < \mu < x_M + t_{0.025}(\nu) \times (S/\sqrt{n})$$

となる。99%信頼区間を求めたい場合には、t分布の上側2.5%点の値 $t_{0.025}(\nu)$ のかわりに、上側0.5%点の値 $t_{0.005}(\nu)$ を用いればよい。

実際の計算では、t分布の上側2.5%点の数値などが必要であるが、これは統計学の教科書の巻末などに掲載されているものを用いればよい。最近では、パソコンで簡単に計算できるようなソフトもある。また、著者のホームページ (<http://www.ism.ac.jp/~takagi/>) でも簡単に計算が行えるようになっており、必ずしもt分布表は必要ではない。

4) 健常群と動脈硬化群の脈波速度の比較

ここでは、健常群25名と動脈硬化群20名について、超音波パルスドプラ法により上肢動脈(手関節部)における平均脈波速度(Mean Velocity, MV)の比較を行おう(表1)。

表1 健常群と動脈硬化群の平均脈波速度 (cm/sec)

グループ	人数	平均値 (標準偏差)
動脈硬化群	20	15.6 (4.9)
健常群	25	23.2 (8.6)

グループごとに母平均値の95%信頼区間を求めよう。

まず、動脈硬化群では、 $n=20$ であるから、t分布の自由度は19となる。t分布表の自由度19の上側2.5%点 $t_{0.025}(19)=2.093$ 、0.5%点 $t_{0.005}(19)=2.861$ となり、これらの数値を(6)式に代入すればよい。

動脈硬化群の母平均値の95%信頼区間は、

$$2.093 \times 4.9 / \sqrt{20} = 2.293$$

となるので、 15.6 ± 2.293 を計算すると、

$$13.3 < \text{動脈硬化群の母平均値} < 17.9$$

と求められる。同様な計算により、99%信頼区間は、

$$12.5 < \text{動脈硬化群の母平均値} < 18.7$$

と求められる。

健常群では $n=25$ であり、 $t_{0.025}(24)=2.064$ 、0.5%点 $t_{0.005}(24)=2.797$ であることに注意して計算すると、

95%信頼区間：19.7<健常群の母平均値<26.8

99%信頼区間：18.4<健常群の母平均値<28.0
となる。

上記の結果を比べると、母平均値の95%信頼区間は2グループで重なりはない。しかし、99%信頼区間では、健常群の信頼区間の下限が動脈硬化群の信頼区間の上限よりも小さくなっており、重なりが認められた。この結果からすると、95%の信頼性で2グループのMVの母平均値には違いがあり、健常群で平均脈波測定は大きいことがいえることになる。

母平均値の差の検定

ここでは、統計学的仮説検定法による母平均値の比較について解説する。すなわち、帰無仮説「2つのグループの母平均値は等しい」が成り立つと仮定した場合、2つの標本から得られた平均値の差が、偶然得られる確率を計算し、その確率（p値）がある設定した値（通常は0.05、5%）より小さいか否かにより、検定を行なう。p値が大きければ、偶然によっても観察されたような平均値の差が生じると判断できる。逆に、p値が小さければ、そのような偶然がたまたま生じたと考えるのは困難であり、その計算のもととなった仮説が誤りであると考え、その帰無仮説を棄却する。その結果、2つのグループには母平均値に差があることになる。

上記の「2群の母平均値の差の検定」は、理論上は2グループの母分散が「等しい場合」と「等しくない場合」に分けて考える必要がある。ここでは、まず母分散が等しいものと仮定できる場合について説明し、母分散が等しいことを仮定しない場合、もしくは等分散の検定（後述）で有意差が認められた場合については、その後で説明する。

1) 母分散が等しい場合の2つの母平均値の差の検定

2グループA、Bの標本数をm、nとし、ある変数の母集団での分布が正規分布に従い、そこから無作為にデータが抽出されたものとする。

A群での平均値を x_M 、不偏分散を S_x^2 、B群での平均値を y_M 、不偏分散を S_y^2 とする。2グルー

プの母分散が等しいと仮定できる場合、以下の統計量を求める。すなわち、2つの平均値の差と、その不偏分散を用いて、

$$(7) \quad t_0 = \frac{|x_M - y_M|}{\sqrt{\left[\frac{(m-1)S_x^2 + (n-1)S_y^2}{m+n-2} \times (1/m + 1/n) \right]}}$$

を求める。上記の t_0 は自由度 $m+n-2$ のt分布に従う。したがって、t分布表より両側 α 点（通常5%か1%）の値 $t_{\alpha/2}(m+n-2)$ を求め、 t_0 と比較する。

$$t_0 \geq t_{\alpha/2}(m+n-2)$$

ならば、有意水準 α で2グループの母平均値間に有意差があるものとし、そうでなければ、2群の母平均値に差があるとはいえないことになる。

前述した健常群と動脈硬化群の平均脈波速度（表1）の母平均値の差の検定を行ってみよう。
(7)式にそれぞれ数値を代入すると

$$\begin{aligned} t_0 &= \frac{|15.6 - 23.2|}{\sqrt{\left[\frac{(20-1) \times 4.9^2 + (25-1) \times 8.6^2}{20+25-2} \times (1/20 + 1/25) \right]}} \\ &= \frac{7.6}{\sqrt{\left[\frac{19 \times 4.9^2 + 24 \times 8.6^2}{43} \times (1/20 + 1/25) \right]}} \\ &= \frac{7.6}{\sqrt{4.67}} = 3.517 \end{aligned}$$

となる。この数値と自由度43のt分布の上側2.5点2.017、0.5%点2.695と比べると、 t_0 の方が大きいので、2グループには1%水準で有意差が認められることになる。正確なp値を求めてみると、 $p=0.001$ であり、実際には0.1%水準で有意ということになる。

この結果は、信頼区間による方法が1%水準で母平均値が異なるといえなかったのに比べて、極めて有意となっており、両者に矛盾があるように思えるかもしれない。この原因は、母平均値の信頼区間を求める場合に用いる標本数に比べて、上記のt検定で用いる標本数が約2倍になっていることによる。一般に検定の効率は標本数が多いほどよく、ほんのわずかの差でも、標本数が数千、数万になると検出してしまうことになる。本例の

ように小標本の場合、異なる方法による分析結果が、有意性などで異なってしまうことがよくある。母平均値の差の検定は、2つの母平均値が等しいか否かを述べるにすぎない。これに比べ、信頼区間などを求める統計学的推定は、母平均値などの存在範囲を示すので、情報としてはずっと有用である。

2) 母分散が等しくない場合の2つの母平均値の差の検定

2つの母分散について等分散性を仮定せずに、帰無仮説「2グループの母平均値は等しい」を検定する方法として、一般によく用いられている「ウェルチの方法」を紹介する。

等分散の前提なしに、上記の帰無仮説を検定するためには、

$$(8) \quad t_0 = \frac{|x_M - y_M|}{\sqrt{\left[\frac{S_x^2}{m} + \frac{S_y^2}{n} \right]}}$$

を求める。上記の t_0 は t 分布に従うが、その自由度 ν は、

$$(9) \quad \nu = \frac{\left[\frac{S_x^2}{m} + \frac{S_y^2}{n} \right]^2}{\frac{S_x^4}{m^2(m-1)} + \frac{S_y^4}{n^2(n-1)}}$$

となる。ここで、自由度 ν は必ずしも整数になるとは限らない。検定のためには、 ν の値が10未満の場合、特別な表として「小数自由度の t 分布表」が必要である。ただし、 ν が10以上ならば、小数部を切り捨て、整数部のみを自由度として採用すればよい。検定の方法は、等分散が仮定できる場合とまったく同様である。なお、標本数 m, n が等しい場合、(8)式と(7)式による t_0 の値は完全に一致する。

ここでも、表1のデータについて検定を行ってみよう。

$$t_0 = \frac{|15.6 - 23.2|}{\sqrt{\left[\frac{4.9^2}{20} + \frac{8.6^2}{25} \right]}}$$

$$= \frac{7.6}{\sqrt{4.1589}} = 3.727$$

となる。また、自由度 ν は、

$$\begin{aligned} \nu &= \frac{\left[\frac{4.9^2}{20} + \frac{8.6^2}{25} \right]^2}{\frac{4.9^4}{20^2(20-1)} + \frac{8.6^4}{25^2(25-1)}} \\ &= \frac{4.1589^2}{0.4405} = 39.3 \end{aligned}$$

となる。 t_0 と自由度39の t 分布の上側2.5点2.023、0.5%点2.708と比べると、 t_0 の方が大きいので、2グループには1%水準で有意差が認められることになる。正確な p 値を求めてみると、 $p=0.0006$ であり、実際には0.1%水準で有意ということになる。

この結果は、等分散を仮定した場合の t 検定とはほぼ同一のものであった。ただし、標本数が小さい場合などは、必ずしもいつも結果が同じようになるとは限らないので注意が必要である。

3) 等分散の検定

すでに述べたように、2グループの母平均値の差の検定の方法は、母分散が等しい場合と等しくない場合では、計算方法が異なっていた。2グループの不偏分散を眺めただけでは、どちらの方法を使えばよいのかは、必ずしも明らかではない。一つの方法としては、母分散が等しいことを仮定しないウェルチの方法を常に用いることである。または、以下に説明する方法で、母分散が等しいか否かの検定（等分散の検定）を行い、その結果をもとにどちらの方法にするかを決めるという分析手順が考えられる。

帰無仮説「2グループの母分散は等しい」を検定するためには、

$$(10) \quad F_0 = \frac{S_x^2}{S_y^2}$$

を求める。ただし、 S_x^2, S_y^2 は各グループの不偏分散である。

上記の F_0 は自由度 $(m-1, n-1)$ の F 分布に、

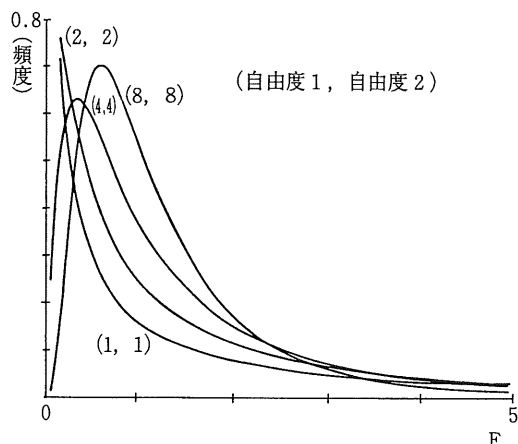


図3 F分布の自由度と形状

また $1/F_0$ は自由度 $(n-1, m-1)$ の F 分布にそれぞれ従う。なお、F 分布は、2つの分散比に関する分布であり、自由度が2つあり、その自由度により分布の形状は大きく異なっている。

検定を両側 α の確率で行なう場合、自由度 (ν_1, ν_2) の F 分布の上側 $\alpha/2$ 点を $F_{\alpha/2}(\nu_1, \nu_2)$ のように表すと、

$$F_0 > F_{\alpha/2}(m-1, n-1)$$

$$1/F_0 > F_{\alpha/2}(n-1, m-1)$$

ならば仮説を棄却すればよい。その場合には、(8)式によりウェルチの方法で検定を行ない、そうでなければ、仮説を採択し、等分散を仮定して、(7)式により検定すればよい。ただし、(10)式により F_0 を計算する場合、 $F_0 > 1$ になるように、分子と分母を設定して、一方のみの条件を満たすか否かで検定することが多い。

健常群と動脈硬化群の平均脈波速度の分散が等しいといえるかを検定しよう。分子の方が大きくするようにして計算すると、

$$F_0 = \frac{8.6^2}{4.9^2} = 3.080$$

となる。 F_0 は自由度 $(24, 19)$ の F 分布に従うので、その上側2.5点2.344 (実際には数値表に掲載されていないことがある。ここでは、第1自由度が15の場合が2.437と20の場合が2.327であり、途中を線形補間して求めた) と比べると F_0 の方が大きい。したがって、2グループ母分散は等しくないものと考えられる。この結果から、母平均値の差の検定にはウェルチの方法を用いた方がよいことになる。

ところで、上記の手順でF検定とt検定を行うと、それぞれの検定の有意水準を α とすると、全体の有意水準は 2α となってしまう。したがって、全体の有意水準を α とするためには、「等分散の検定」も「母平均値の差の検定」も各水準を $\alpha/2$ とする必要があるのではないかと考えられる。もしもそうだとすると、全体の有意水準を5%にしたい場合には、 F_0 の比較にはF分布の上側1.25%点を用い、 t_0 の比較にはt分布の両側2.5%点を用いなければならなくなる。このように有意水準を設定することは、等分散が仮定されやすくなり、また母平均値の差が有意になりにくくなることを意味している。ただし、実際にはこのように検定の回数を考慮して、分析全体の有意水準を設定することは減多にない。このような「多重比較」の問題は、3グループ以上の母平均値の差の検定を、2グループの各組み合わせについて行うような場合に、より重要な意味を持ってくる。この点については、次回以降の適当な箇所でも触れることにしよう。

参考文献

- 1) 三谷哲也, 土居忠文, 石川満子, 西原えり子, 雑賀光一, 山崎文靖, 影岡武士: 動脈硬化における骨動脈血行動態の検索—超音波パルスドプラ法による血流パラメータと血流パターンおよびスコア解析—, 超音波検査技術 1997; 22: 264-270.