

統計講座 (第3回)

3 グループ以上の平均値を比べる

高木 廣文

はじめに

前回は、患者と健常者、喫煙者と非喫煙者のように、2グループの母集団でのある特性の平均値(母平均値)に差があるかを検討する方法について説明した。実際には、喫煙に関してでも、現喫煙者(現在喫煙中)、既喫煙者(以前は吸っていたが現在は止めている)、非喫煙者(全く喫煙したことがない)のように、3グループに分けて研究することが多い。そのような場合、どのような統計的方法を用いて解析を行えばよいだろうか。

統計学的に母集団での3グループ以上の母平均値を比較するためには、

- (1) 各グループの母平均値の信頼区間を求めて比較する。
- (2) 帰無仮説「3グループの母平均値は等しい」を検定する。

という方法がまず考えられる。

一般によく用いられる方法は(2)であり、その方法は「一元配置分散分析 One-way layout Analysis of Variance (One-way ANOVA)」と呼ばれる。一元配置分散分析の結果は、多群の母平均値に一つでも他と異なるものがあるかを検定するものであり、どの2グループ間に差があるのかは、明らかにできない。したがって、一元配置分散分析の結果が統計学的に有意、すなわち少なくとも1つ以上のグループの母平均値が他と異なることが明らかになった場合、どのグループ間の母平均値に

差があるのかを検定する必要がある。この方法は、多重比較 multiple comparison の問題として知られているが、基本的には2グループ間の母平均値の差の検定を行うか、上記(1)のように信頼区間を求めることになる。ただし、同時に複数個の検定や推定を行うために、設定した検定の有意水準(例えば5%)を確保するために、特別な方法を用いる必要がある。ここでは、まず一元配置分散分析について説明し、次に多重比較の方法について説明する。

一元配置分散分析の考え方

グループが多数ある場合、どのようにすれば各グループの母平均値に差があるかが分かるだろうか。ここでは、データのばらつき具合(「変動 variation」と呼ばれる)に注目する。

変動は、各データと平均値の差(「偏差 deviation」と呼ばれる)の平方和として定義される。すなわち、これを標本数で割れば分散となり、その自由度で割れば不偏分散が求められる。

グループを考えずに、全体の平均値(全平均)と全データの偏差平方和は、「全変動」と呼ばれる。すなわち、

- (1) 全変動 T = 各データの全平均からの偏差平方和

により求められる。

一方、各グループ内でも同様に、そのグループの平均値からの偏差平方和を考えることができる。このグループごとに求めた偏差平方和を合計したものは、「級内変動(誤差変動)」と呼ばれる。すなわち、

- (2) 級内変動 W = 各グループ内でのグループ平均からの偏差平方和の合計

により求められる。

これらの2つの変動の他に、グループの平均的な離れ具合を表す「級間変動」を考えることができる。級間変動は、各グループのデータがすべてそのグループの平均値に集まっているとした場合の、全平均値との偏差平方和の合計によって求められる。すなわち、

- (3) 級間変動 B = [グループの人数 $\times$ (グループ平均 - 全平均)²] の合計

によって求められる。

これらの全変動、級内変動、級間変動の間には、

- (4) 全変動 = 級内変動 + 級間変動

の関係がある。

級間変動がグループ間の離れ具合を示し、一方、級内変動がグループ内のばらつきを示している。

帰無仮説「すべてのグループの母平均値は等しい」を検定するには、級内変動と級間変動をそれぞれの自由度で割り、不偏分散を求め、誤差よりもグループ間の離れ方の方が大きいかな否かを検討すればよい。

全変動の自由度は、通常の不偏分散を求める場合と同様に、「全標本数 - 1」である。誤差変動は、グループ数だけの平均値を推定しているの、その自由度は「全標本数 - グループ数」となる。自由度についても(4)式が成り立つので、級間変動の自由度は「グループ数 - 1」となる。これらの自由度を用いて、不偏分散は、

- (5) 級間分散 V_W = 級間変動 / [グループ数 - 1],
誤差分散 V_E = 誤差変動 / [全標本数 - グループ数]

により求められる。

前回説明した等分散の検定と同じように、分散の比を計算する。すなわち、

- (6) F_0 = 級間分散 / 誤差分散

を求める。上記の F_0 は、自由度 (グループ数 - 1, 全標本数 - グループ数) の F 分布にしたがう。

求めた F_0 の上側確率 (p 値) を適当なプログラムで計算し、その値が 0.05 (5%) より小さければ、仮説を棄却し、少なくとも1つのグループの母平均値が他のグループの母平均値と異なると考えてよいことになる。一方、 p 値が 0.05 より大きければ、仮説を棄却することはできないので、各グループの母平均値に差があるとはいえない。

上記の一元配置分散分析の計算結果を、表1のように「分散分析表」として示すのが普通である。

表1 一元配置分散分析表

変動	平方和	自由度	不偏分散	F 値	p 値
級間	W	グループ数 - 1	V_W	F_0	(確率)
級内	B	全標本数 - グループ数	V_E		
全体	T	全標本数 - 1			

脈波速度の3グループでの比較

前回例題に用いた健常群25名と動脈硬化群20名の超音波パルスドプラ法による上肢動脈 (手関節部) における平均脈波速度 (Mean Velocity, MV) のデータに、表2のように運動実施群23例のデータが追加された場合を考えてみよう (架空例)。なお、表2の標準偏差は不偏分散の平方根を示すものとする。

表2 3グループの平均脈波速度 (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)
運動実施群	23	25.4 (8.7)

表2から、全平均値 = $(20 \times 15.6 + 25 \times 23.2 + 23 \times 25.4) / 68 = 21.709$ 、となる。まず、級内変動 W を求める。級内変動 (誤差変動) は各グループの、不偏分散の人数 - 1 倍の合計になるので、
 $W = (20 - 1) \times 4.9^2 + (25 - 1) \times 8.6^2 + (23 - 1) \times 8.7^2 = 3896.41$

となる。一方、級間変動 B は、

$$B = 20 \times (15.6 - 21.709)^2 + 25 \times (23.2 - 21.709)^2$$

$$+23 \times (25.4 - 21.709)^2 = 1115.3147$$

となる。級内変動と級間変動から全変動 T は、

$$T = B + W = 3896.41 + 1115.3147 = 5011.7247$$

と求められる。

これらの変動から、それぞれ不偏分散を求める。級間変動の自由度は $3 - 1 = 2$ 、級内変動の自由度は $68 - 3 = 65$ 、全体の自由度は $68 - 1 = 67$ 、となる。これらの自由度で各変動を割ると、

$$\text{級間分散 } V_W = 1115.3147 / 2 = 557.657,$$

$$\text{級内分散 } V_E = 3896.41 / 65 = 59.945$$

となる。また、

$$\text{全分散} = 5011.7247 / 67 = 74.802$$

と求められる。

「3 グループの母平均値はすべて等しい」という帰無仮説を検定するための検定統計量は、

$$F_0 = V_W / V_E = 557.657 / 59.945 = 9.303$$

と求められる。 F_0 は自由度(2,65)のF分布にしたがうので、その上側確率(p値)を求めると、 $p = 0.0027$ となり、高度に有意な結果であることを示している。これらの結果をまとめると表3のようになる。

表3 平均脈波速度の一元配置分散分析表

変動	平方和	自由度	不偏分散	F値	p値
級間	3896.4	2	557.7	9.303	0.0027
級内	1115.3	65	59.9		
全体	5011.7	67	74.8		

一元配置分散分析の結果は、表3に示したようにp値が通常の検定基準である5%よりもはるかに小さいため、「有意差あり」と判定される。しかし、この有意差がどのグループ間に存在するのかは、このままでは明らかではない。そのため、一元配置分散分析のあとは、2グループごとに母平均値の差の検定を行うのが通常の手順となっている。その場合、いくつも検定を繰り返すため、有意水準を一定に保つための工夫が必要になる。

多重比較の方法

一元配置分散分析の結果が有意な場合、各2グループ間の母平均値の差について検定を行うこと

で、どのグループ間に有意差があるかを調べる方法、いわゆる「対比較(対比)」について、いくつかの方法を説明しよう。

1) ボンフェロニの方法

いま、グループ数が g であるとする。この場合、各2グループ間の母平均値の差を、全ての組合せについて検定すると、 g グループから2グループを選ぶ組合せの個数は、 $g(g-1)/2$ 個ある。したがって、各検定の有意水準を α とすれば、検定全体での有意水準はその約 $g(g-1)/2$ 倍となってしまう。例えば、グループ数が5の場合、組合せの個数は10通りあり、各検定の有意水準を5%としても、全体では、 $1 - (1 - 0.05)^{10} = 0.40$ 、すなわち40%程度の確率となり、極めて大きな有意水準を設定することになってしまう。

このような弊害を避けるための一つの方法として、各検定の有意水準を組合せの個数で割ったものとすればよい。すなわち、全体の検定の有意水準を α としたければ、個々の検定の有意水準を $2\alpha / \{g(g-1)\}$ とすればよいことになる。この方法は「ボンフェロニ Bonferroni の方法」と呼ばれる方法である。この場合、前回説明した2群の母平均値の差の検定(t検定)を、そのまま用いることもあるが、計算式の分母に、一元配置分散分析で求めた誤差分散(級内分散) V_E を用いるのが一般的である。すなわち、2グループA、Bの標本数を n_A 、 n_B とし、平均値を M_A 、 M_B としたとき、

$$(7) \quad t_0 = \frac{|M_A - M_B|}{\sqrt{[V_E \times (1/n_A + 1/n_B)]}}$$

を求める。この t_0 は、自由度(全標本数-グループ数)のt分布にしたがうものとして検定を行えばよい。通常の2群の母平均値の差の検定のt値の自由度は「2つの標本数の和-2」であるが、ここでは、差の分散の推定値として誤差分散を用いているので、自由度が異なっているのに注意が必要である。

実際にこの基準を用いると、グループ数が大きくなるにつれ、個々の検定の有意水準は極めて小さくなり、検定結果は有意となりにくくなる。したがって、このような有意水準の設定は厳し過ぎ

るといえる。

実際に、平均脈波測度について、動脈硬化群 A、健常群 B、運動実施群 C として、(7)式を用いて検定を行ってみよう。全体の検定水準を 5% とすると、A-B、A-C、B-C の 3 通りの組み合わせがあるので、 $5/3=1.7\%$ (0.017) 以下の場合、各検定を有意とすればよい。

誤差分散 $V_E=59.945$ なので、動脈硬化群 A と健常群 B の組み合わせについては、

$$t_{AB} = \frac{|15.6 - 23.2|}{\sqrt{[59.945 \times (1/20 + 1/25)]}} = 3.272$$

となる。この t_{AB} は自由度が $68-3=65$ の t 分布にしたがうので、有意確率を求めると、 $p=0.0017$ 、となり、2 グループの母平均値に有意差があるものと認められる。同様に、動脈硬化群 A と運動実施群 C については、

$$t_{AC} = \frac{|15.6 - 25.4|}{\sqrt{[59.945 \times (1/20 + 1/23)]}} = 4.140$$

となる。自由度 65 の t 分布から有意確率を求めると、 $p=0.0001$ 、となり有意差ありと考えてよい。健常群 B と運動実施群 C については、

$$t_{BC} = \frac{|23.2 - 25.4|}{\sqrt{[59.945 \times (1/25 + 1/23)]}} = 0.983$$

となる。自由度 65 の t 分布から有意確率を求めると、 $p=0.3290$ 、となり有意差は認められない。

これらの結果から、動脈硬化群の母平均値が、健常群と運動実施群の 2 群とは異なっており、健常群と運動実施群の母平均値には有意差が認められないことになる。

2) シェッフエの方法

場合によっては、いくつかのグループの母平均に差がなく、それらをまとめて他の複数のグループと比較したいこともある。そのような場合、いくつかの平均値の線形結合 (重み付けして、加算したもの) により、新たに適当な統計量を求め、それを用いて比較することもできる。そのような方法は「線形比較」と呼ばれている。ここではそのような方法のうち「シェッフエ (Schéffe) の方法」を説明する。

いま、 g 個のグループでの変数 X の平均値をそれぞれ、 $M_1, M_2, \dots, M_g$ とする。各平均値にそれぞれ重みとして、 $w_1, w_2, \dots, w_g$ を掛け、その合計を計算すると、

$$(8) \quad M_A = w_1 M_1 + w_2 M_2 + \dots + w_g M_g$$

とする。ここで、重みの合計は 0 になるように定めるのが普通である。

(8)式のように平均値に重み付けした合計点 M_A の期待値 $E(M_A)$ は、

$$(9) \quad E(M_A) = w_1 \mu_1 + w_2 \mu_2 + \dots + w_g \mu_g = \mu_A$$

となる。ここで、 $\mu_1, \mu_2, \dots, \mu_g$ は各グループの母平均値を示す。

また、分散 $V(M_A)$ は、

$$(10) \quad V(M_A) = \sigma^2 \left[\frac{w_1^2}{n_1} + \dots + \frac{w_g^2}{n_g} \right]$$

となる。ここで、 $n_1, \dots, n_g$ は各グループの標本数、 σ^2 は母誤差分散であり、この値は一元配置分散分析から得られる誤差分散 V_E により推定する。したがって、この分散の推定値 V_A は、

$$(11) \quad V_A = V_E \left[\frac{w_1^2}{n_1} + \dots + \frac{w_g^2}{n_g} \right]$$

により求められる。

この結果を用いて、重み付け合計得点 μ_A の $100(1-\alpha)\%$ 信頼区間を求めることができる。全標本数を N とし、自由度 $(g-1, N-g)$ の F 分布の上側 α 点の値を $F_\alpha (K-1, N-g)$ とし、

$$(12) \quad d = \sqrt{[(g-1) \cdot F_\alpha (g-1, N-g)]}$$

とする。この d を用いて、 μ_A の信頼区間の上側の点を μ_U 、下側の点を μ_L とした場合、

$$(13) \quad \begin{aligned} \mu_U &= M_A + d \cdot \sqrt{V_A}, \\ \mu_L &= M_A - d \cdot \sqrt{V_A} \end{aligned}$$

により求められる。

求めた信頼区間 (μ_L, μ_U) の間に 0 を含まなければ、有意水準 α でグループ群の間に差があることになる。

実際の計算では、一方のグループに含まれるグループ数を G_A 、他方を G_B とした場合、比較に用いないグループの平均値にかけの重みは 0 とし、

前者の重み $= 1/G_A$ 、後者の重み $= -1/G_B$ のようにグループ内では同一の重みに、全体では重みの合計が 0 になるように設定すればよい。

2 グループの母平均値の検定を行う場合には、一方に 1 を、他方に -1 の重みを付ければよい。また、(13) のように信頼区間を推定するのではなく、同一の結果を与える検定統計量は、

$$(14) \quad F_0 = M_A^2 / [(g-1) V_A]$$

となる。この F_0 は自由度 $(g-1, N-g)$ の F 分布にしたがうので、その上側確率 (p 値) が、0.05 (5%) 以下ならば、有意差ありとすればよい。

平均脈波測度について、(14) 式を用いて検定を行ってみよう。誤差分散 $V_E = 59.945$ なので、動脈硬化群 A と健常群 B の組み合わせについては、

$$F_{AB} = \frac{(15.6 - 23.2)^2}{(3-1) \times 59.945 \times (1/20 + 1/25)} = 5.353$$

となる。この F_{AB} は自由度 (2, 65) の F 分布にしたがうので、有意確率を求めると、 $p = 0.0070$ 、となり、2 グループの母平均値に有意差があるものと認められる。同様に、動脈硬化群 A と運動実施群 C については、

$$F_{AC} = \frac{(15.6 - 25.4)^2}{(3-1) \times 59.945 \times (1/20 + 1/23)} = 8.570$$

となる。自由度 (2, 65) の F 分布から有意確率を求めると、 $p = 0.0006$ 、となり有意差ありとなる。

健常群 B と運動実施群 C については、

$$F_{BC} = \frac{(23.2 - 25.4)^2}{(3-1) \times 59.945 \times (1/25 + 1/23)} = 0.484$$

となる。自由度 (2, 65) の F 分布から有意確率を求めると、 $p = 0.6188$ 、となり有意差は認められない。これらの結果は、ボンフェロニの方法による結果と同一である。

3) テューキー法

全てのグループの標本数 n が等しい場合、より検定効率のよい方法を用いることができる。い

ま、グループの中で最大の平均値と最小の平均値の差を、範囲 R とした場合、これを誤差分散 V_E で標準化した統計量を、

$$(15) \quad q = \frac{R}{\sqrt{[V_E/n]}}$$

とする。この q は「スチューデント化された範囲」と呼ばれる。

実際の検定では、標本数が大きく異ならなければ、近似的に本方法を適用することができる。いま、2 グループ A, B の標本数を n_A, n_B とし、平均値を M_A, M_B としたとき、

$$(16) \quad q_0 = \frac{|M_A - M_B|}{\sqrt{[V_E \times (1/n_A + 1/n_B)/2]}}$$

を求める。この q_0 を、スチューデント化された範囲の数表の自由度 (グループ数, 全標本数 - グループ数) の 5% 点、1% 点の値と比較し、 q_0 が大きい場合に有意とする。検定には特殊な数表が必要であり、ここでは掲載していないが、例えば文献 3 などを参照されたい。

平均脈波測度について、(16) 式を用いて検定を行ってみよう。動脈硬化群 A と健常群 B については、

$$q_{AB} = \frac{|15.6 - 23.2|}{\sqrt{[59.945 \times (1/20 + 1/25)/2]}} = 4.627$$

となる。この q_{AB} と自由度 (3, 65) のスチューデント化された範囲の 5% 点 $q_{0.05}$ が手持ちの数表に掲載されていないので、自由度 (3, 60) の 5% 点 3.399、1% 点 4.282 と比べると、 q_{AB} の方が大きいので、1% 水準で有意差ありと考えてよいことになる。このように、第 2 自由度に一致した値が数表に掲載されていない場合には、より小さい自由度で代用するか、線形補間を行う必要がある。実際には数値の差は極めて小さいので、判定が極めて微妙な場合でなければ、線形補間をする必要はない。

同様に、動脈硬化群 A と運動実施群 C については、

$$q_{AC} = \frac{|15.6 - 25.4|}{\sqrt{[59.945 \times (1/20 + 1/23)/2]}} = 5.855$$

となる。前述の結果と同様に、1%水準で有意差ありと考えてよい。

健常群 B と運動実施群 C については、

$$q_{BC} = \frac{|23.2 - 25.4|}{\sqrt{[59.945 \times (1/25 + 1/23)/2]}} = 1.391$$

となり有意差は認められない。

4) ダネット法

すべてのグループ間で比較を行うのではなく、特定のグループを基準として、そのグループと他のグループの比較のみを行う場合がある。このような比較の方法は、ダネット Dunnett 法と呼ばれている。基本的には、すべてのグループで標本数が等しい場合の多重比較の方法であるが、標本数に大差なければ、すでに示した(7)式を用いて計算を行えばよい。ただし、求めた統計量は t 分布を仮定するのではなく、ダネット法による比較のための数表に掲載された自由度(グループ数-1, 全標本数-グループ数)の数値と比較する。数値表の数値よりも大きければ、基準グループと比較グループに有意差があることになる。数値表はここには掲載できないが、文献 3 などを参照して欲しい。

ボンフェロニ法の説明で求めた数値を再掲すると、動脈硬化群 A を基準群とした場合、健常群 B とは、 $d_{AB} = 3.272$ 、運動実施群 C とは、 $d_{AC} = 4.140$ 、であった(誤解を避けるため、t でなく d とした)。ここでは、健常群と運動実施群の比較は行わないものとする。自由度(2, 65)の 5%点、1%点の数値はないので、自由度(2, 60)の 5%点 2.27、1%点 2.90 と比較すると、いずれも d_{AB} 、 d_{AC} の方が大きいので、1%水準で有意差が認められた。

おわりに

今回は、3 グループ以上の母平均値の差の検定

の方法として、一元配置分散分析の方法を解説した。さらに、一元配置分散分析では明らかにすることができない 2 グループ間の母平均値の差の検定について、いくつかの多重比較の方法を解説した。実際の生データから計算を行うのは結構大変である。また、検定によっては特別な数表が必要となる。この目的のために、私のホームページで計算ができるようにした。インターネットが利用できるのならば、簡単に使用できる (URL: <http://www.ism.ac.jp/~takagi/>)。

今回解説した 4 つの多重比較の方法のうち、どれを用いればよいのか、迷うかもしれない。ボンフェロニの方法は、検定基準が厳しいことが知られている。また、シェッフェの方法は 2 群の比較のみならず、2 群対 1 群のような比較も考慮しているので、やはり検定が有意になりにくい。チューキーの方法は、2 群間の比較のみなので前述の 2 方法よりは有意になりやすいが、基本的にグループでの標本数は等しいことが前提なので、極端に標本数が異なる場合は、不適当な方法となる。ダネット法はすべての 2 群間で検定を行うのではなく、基準群とのみの比較を行う場合に適用できる。このように、各検定方法には特徴があるので、何でもやってみて有意な結果がでた方法を採用するというのは感心できない。重要な意志決定に関しては、有意に出にくい方法を用いるなど、目的に応じた使い分けが必要である。

参考文献

- 1) 高木廣文：2 グループの平均値を比べる，超音波検査技術 23(5)：400-405, 1998
- 2) 三谷哲也，土居忠文，石川満子ほか：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 22：264-270, 1997
- 3) 柳井晴夫，高木廣文編著：新版看護学全書 統計学，メヂカルフレンド社（東京），1995