

統計入門

[第6回]

平均値と比率に関する検定方法

高木 廣文* 新田 裕史** 中井 里史***

I. 2群の母平均値の差の検定

ここでは、二つのグループ、例えば患者群と健康者群で、母平均値に差があるのかという比較を行う場合の方法について説明する。

1. 母集団の分散が等しい場合

ある感染症の患者を無作為に2群に分け、一方に新薬Aを他方に旧薬Bを投与したとする。標本数は各50例、症状消失までの平均日数と標準偏差(不偏分散の平方根)は、表1のようになっているとしよう。この成績から、両薬の効果に差があるといえるだろうか。

“2群の母平均値の差の検定”は、2群の母集団での分散(母分散)が“等しい場合”と“等しくない場合”の2通りに分けて考える必要がある。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する。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、後述する。

2群、A、Bの標本数を m 、 n とし、A群での平均値と不偏分散を $\bar{x}$ と s_A^2 とし、同様にB群での平均値を $\bar{y}$ 、不偏分散を s_B^2 としよう。

2群で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ればよい。

$$t_0 = \frac{|\bar{x} - \bar{y}|}{\sqrt{S_T \left(\frac{1}{m} + \frac{1}{n} \right)}}$$

を求める。ただし、

* 文部省統計数理研究所

[〒106 東京都港区南麻布 4-6-7]

** 国立環境研究所

*** 東京大学医学部疫学

表1 2群の平均値と標準偏差

	例数	平均値	標準偏差
A群	50	3.1日	1.2日
B群	50	5.3日	0.8日

$$S_T = \frac{(m-1)s_A^2 + (n-1)s_B^2}{m+n-2}$$

である。 t_0 は t 分布と呼ばれる分布に従うことがわかっている。したがって、適当な t 分布表より両側 2α 点(通常5%か1%)の値 t_α を求め、 t_0 と比較する。ただし、 t 分布は自由度によってその分布が異なっているので、自由度を求めておく必要がある。この場合、自由度は $m+n-2$ である。

検定は、 $t_0 \geq t_{0.025}$ ならば、5%水準で2グループ間の母平均値には有意差があるものとし、そうでなければ、母平均値に差があるとはいえなくなる。

実際に計算してみると、

$$t_0 = \frac{|3.1 - 5.3|}{\sqrt{\frac{49 \times 1.2^2 + 49 \times 0.8^2}{50 + 50 - 2} \times \left(\frac{1}{50} + \frac{1}{50} \right)}} = 10.786$$

であり、自由度は $(50+50-2)=98$ となる。 t 分布表から自由度98の両側5%点、1%点、0.1%点を求めると、1.986、2.629、3.397である(数値表に対応する自由度が掲載されていない場合、その自由度より小さいもので最も近い自由度に対する数値を用いればよい)。これらの数値と t_0 の値を比較すると、すべての数値で t_0 の方が大き

い。したがって、有意水準 0.1% で母平均値に有意差が認められる。

上記の検定結果から、A 薬は旧薬 B に比べ、症状消失に関して優れた効果を持つものと考えてよいだろう。

2 群の母平均値の差の $100(1-2\alpha)$ 信頼区間は、

$$(x-y) \pm t_{\alpha} \sqrt{S_T \left(\frac{1}{m} + \frac{1}{n} \right)}$$

により、求められる。

例題の場合、母平均値の差の 95% 信頼区間は、1.795~2.605 となる。

2. 母集団の分散が等しくない場合

前述の方法は、2 群の母集団での分散（母分散）が等しい場合のものであった。実際には、母集団での分散が等しいか否かは明らかでないことが多いだろう。したがって、まず母集団での母分散が等しいか否かを検定し、等しいと仮定できれば前述の方法を用い、そうでなければ後述の方法を用いる必要がある。

二つのグループ A と B の母分散を σ_A^2 と σ_B^2 としよう。前述した方法は、二つの母分散が“ $\sigma_A^2 = \sigma_B^2$ ”という場合に、二つの母平均値が等しいかを検定する方法であった。この方法を用いるためには、あらかじめ母分散が等しいことを、“等分散の検定”と呼ばれる方法を用いて調べておけばよい。

仮説“ $\sigma_A^2 = \sigma_B^2$ ”を検定するためには、

$$F_0 = \frac{s_A^2}{s_B^2}$$

を求めればよい。ただし、 $F_0 \geq 1$ になるように分子と分母は適宜入れかえる。

F_0 は、自由度 $(m-1, n-1)$ の F 分布と呼ばれる分布に従うことがわかっている。有意水準 5% で検定を行う場合、二つの自由度に対応する F 分布の上側 2.5 点 $F_{0.025}$ を数値表から求め、 $F_0 > F_{0.025}$ ならば、5% 水準で有意差ありとし、二つの母分散は異なるものとする。その場合は、以下に説明する“ウェルチの方法”により母平均値の差の検定を行う必要がある。また、有意でなければ、二つの母分散は等しいものと仮定し、前述の方法を適用すればよい。“ウェルチの方法”は、

二つのグループの母分散について等分散の前提なしに、母平均値が等しいという仮説を検定するものである。

$$t_0 = \frac{|x-y|}{\sqrt{\frac{s_A^2}{m} + \frac{s_B^2}{n}}}$$

を求める。 t_0 は t 分布に従うが、その自由度は先ほどの場合と異なり、

$$\text{自由度} = \frac{\left(\frac{s_A^2}{m} + \frac{s_B^2}{n} \right)^2}{\frac{s_A^4}{m^2(m-1)} + \frac{s_B^4}{n^2(n-1)}}$$

となる。ここで、自由度は整数になるとは限らない。自由度の値が 10 未満の場合、特別な表として“小数自由度の t 分布表”が必要となる。また、自由度が 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全く同一に行えばよい。なお、二つのグループの標本数が等しい場合、 t_0 の値は等分散を仮定した場合でも、仮定しない場合のウェルチの方法による場合でも、その値は完全に一致する。

例題について、まず“等分散の検定”を行う。

$$F_0 = \frac{1.2^2}{0.8^2} = 2.25$$

となる。 F_0 は自由度 (49, 49) の F 分布に従うことになるので、対応する F 分布の上側 2.5% 点を数表より求めると約 1.76 である。 F_0 の値はこれより大きいので、両側 5% 水準で等分散の仮説は棄却される。すなわち、2 群の母分散には有意差がある。この結果は、A 薬と B 薬の効果（症状消失までの日数）の分布のばらつき具合が異なることを示している。したがって、この結果のみからでも 2 薬剤の効果は異なると結論できるが、分布の代表値である平均値についての検定は以下のようなになる。

ウェルチの検定を行うと、

$$t_0 = \frac{|3.1-5.3|}{\sqrt{\frac{1.2^2}{50} + \frac{0.8^2}{50}}} = 10.786$$

となる（等分散の場合と数値が一致）。この t_0 の

自由度は,

$$\text{自由度} = \frac{\left(\frac{1.2^2}{50} + \frac{0.8^2}{50} \right)^2}{\frac{1.2^4}{50^2 \times (50-1)} + \frac{0.8^4}{50^2 \times (50-1)}} = 85.4$$

となる。自由度85の t 分布表の両側 0.1% 点を数表から求めると、約3.411である。 t_0 はこれより大きいので、等分散を仮定した場合と同様に、高度に有意差が認められる。これらの結果から、A薬はB薬に比べ、平均的な効果は優れているが、効果の現れ方について、より個体差が大きいものと考えられる。

2群の母平均値の差 $100(1-2\alpha)$ の信頼区間は,

$$(x-y) \pm t_{\alpha} \sqrt{\frac{s_A^2}{m} + \frac{s_B^2}{n}}$$

により求められる。母平均値の差の95%信頼区間は、自由度85の t 分布の両側 5% 点は約 1.990 なので、1.794~2.606となる。

II. 2群の母比率の差と4分表の検定

ある疾患の患者に、無作為にA薬とB薬を各15例ずつ投与し、症状の改善状況を調べたところ、表2のような結果が得られた。A薬とB薬に効果の差があるといえるだろうか。表2のように2変数のどちらも2カテゴリからなるクロス表は、“4分表”と呼ばれる。

A薬での改善の比率 $p_A = 11/15 = 0.733$ であり、B薬では $p_B = 0.4$ であった。2群の母比率に差があるかを検定すればよい。正規分布を用いる方法もあるが、ここでは、計算が簡単なカイ2乗分布を用いる方法を説明する。

検定のための統計量は、表2の記号を用いると、

$$\chi_0^2 = \frac{N(|ad-bc|-N/2)^2}{(a+b) \times (c+d) \times (a+c) \times (b+d)}$$

である。ここで、 $N=a+b+c+d$ は全標本数である。

この χ_0^2 は自由度1のカイ2乗分布に従うので、上側5%点の値 3.84、もしくは1%点の値

表2 A薬とプラセボの投与結果

	改善	不変	計
A薬	11(a)	4(b)	15
B薬	6(c)	9(d)	15
計	17	13	30

6.63と比べ、 χ_0^2 の方が大きい場合に有意とすればよい。なお、式の分子中の“ $-N/2$ ”は、“連続性の補正 (イエーツの補正)”と呼ばれるもので、人数のように整数値しかとらない (離散的) データから得た統計量が、連続的なカイ2乗分布に対して、あてはまりがより良くなるようにするためのものである。ただし、連続性の補正は、

$$|ad-bc| > N/4$$

の場合に行い、そうでない場合には、 $\chi_0^2=0$ とする。

表2について計算を行うと、

$$\chi_0^2 = \frac{30(|11 \times 9 - 4 \times 6| - 30/2)^2}{15 \times 15 \times 17 \times 13} = 2.172$$

となる。自由度1のカイ2乗分布の上側5%点の値3.84と比較すると、 χ_0^2 の値は小さいので有意とはいえない。したがって、A薬とB薬による症状改善の効果に差があるとはいえないことになる。

2群の母比率の差の $100(1-2\alpha)$ 信頼区間は、A群、B群の標本数を m, n とし、標本比率を p_A, p_B としたとき、

$$(p_A - p_B) \pm z_{\alpha} \sqrt{\frac{p_A(1-p_A)}{m} + \frac{p_B(1-p_B)}{n}}$$

により求められる。ここで z_{α} は標準正規分布の上側 $\alpha\%$ 点であり、95%信頼区間を求めるのなら、1.96を用いればよい。例題では、両薬の効果の母比率の差の95%信頼区間は、 $-0.001 \sim 0.667$ となる。

ところで、4分表の場合には、カイ2乗分布を用いるのではなく、2変数に関連がないという仮定のもとで、クロス表が偶然できる確率を直接的に求める方法がある。その方法は“フィッシャーの直接確率 (Fisher's exact test)”とか“フィッシャーの正確な検定”とか呼ばれている。

フィッシャーの直接確率は、偶然に四つのセルの度数が表2のように a, b, c, d となる確率、およびそれ以上に特定方向に偏った状態の確率を求めるものである。A薬で改善の人数が11人となる確率 p_0 は、

$$p_0 = \frac{15! \times 15! \times 17! \times 13!}{30! \times 11! \times 4! \times 6! \times 9!} = 0.057046$$

となる。ここで、 $15! = 15 \times 14 \times \dots \times 1$ 、のように階乗を計算する。

検定のためには、表2の“計”の欄の数値を変えずに、対角成分の積の大きいセルの数値を一つずつ多くし、逆に積の小さいセルの数値は一つずつ減らしていき、各クロス表の確率を全て求める必要がある。そのような状態は、表には示していないがA薬で“不変”のセルが3, 2, 1, 0となる四つのクロス表がある。これら、残り四つのクロス表の確率を求めるのは、以下のようにすればよい。

A薬で不変のセルの度数が3の場合の確率 p_1 は、

$$p_1 = \frac{4 \times 6}{12 \times 10} \times p_0 = 0.011409$$

となり、度数が2の場合の確率 p_2 は、

$$p_2 = \frac{3 \times 5}{13 \times 11} \times p_1 = 0.001197$$

となる。また、度数が1の場合と0の場合の確率 p_3 と p_4 は、

$$p_3 = \frac{2 \times 4}{14 \times 12} \times p_2 = 0.00005699$$

および、

$$p_4 = \frac{1 \times 3}{15 \times 13} \times p_3 = 0.000000088$$

のように、順にその前の状態の確率を用いて計算できる。

全体の確率の和 p を求めると、

$$p = p_0 + p_1 + p_2 + p_3 + p_4 = 0.0697$$

となる。これが、“フィッシャーの直接確率（正確な確率）”と呼ばれるものである。ただし、上

表3 A薬とB薬の症状改善について

	著効	有効	不変	悪性	計
A薬	8	10	6	1	25
B薬	3	9	10	4	26
計	11	19	16	5	51

記の確率 p は一方方向の偏りのみを考えて計算されており、カイ2乗検定と対応させるためには、逆方向への確率も計算する必要がある。すなわち、上記の手頭とは逆に、 a の数値を1ずつ減らしてゆき、連続性の補正をしないカイ2乗値（分子で $N/2$ を引かないもの）が大きくなる4分表の正確な確率を全て求める必要がある。しかし例のように二つのグループの標本数が等しい場合には、 p の値を2倍したものをを用いればよい。 $2p = 0.139$ であり、有意確率は13.9%となり、5%水準に達しないことがわかる。したがって、A薬の効果があるとはいえないことになる。すでに述べたカイ2乗検定で求めた χ_0^2 の有意確率を正確に求めると14.1%であり、フィッシャーの直接確率と大差ない。

4分表の検定としてフィッシャーの直接確率を用いるか、カイ2乗検定を用いるかという問題があるが、できる限りフィッシャーの直接確率を求めるべきであろう。しかし、標本数が大きい場合には、例でみたように両者の結果はほぼ一致するので、カイ2乗検定を用いることができる。

III. 一般のクロス表の検定

表3は、2種の薬剤A, Bをある病気の患者に無作為に投与し、その効果を調べたものである。このような場合、どのようにすれば、二つの薬剤の効果の差を検定できるだろうか。

クロス表の検定は、二つに大別して考えることができる。すなわち、クロス表の二つの変数の一方がグループを示すもので、他方の変数の分布にグループ間で差があるか否かを検定する場合は、“一様性の検定”と呼ばれる。また、2変数間の関連の有無を検定する場合は、“独立性の検定”と呼ばれる。しかし、これらの検定を行うための実際の計算手順は、まったく同一であり、以下に

説明する方法により検定することができる。

2変数に関してクロス表が作成されているものとし、一方の変数 X は K 個のカテゴリからなり、他方の変数 Y は L 個のカテゴリからなるものとする。したがって、クロス表は $(K \times L)$ 個のマスをもち、各マス目は“セル (cell)”と呼ばれる。表3の場合は、 2×4 のクロス表ということになる。

二つの変数の組み合わせについて、 X について X_i で、かつ Y について Y_j というセルの度数を n_{ij} と表わすことにする。また、一方の変数のみに注目して、変数 X のカテゴリ X_i の度数の合計を $n_{i\cdot}$ とし、変数 Y のカテゴリ Y_j の度数の合計を $n_{\cdot j}$ としよう。また、全体の標本数を N とする。

仮説を“2変数間に関連なし”とした場合、2変数 X, Y の独立性の検定を行うためには、統計量として、

$$\begin{aligned}\chi^2_0 &= N \cdot \left(\frac{n_{11}^2}{n_{1\cdot} \cdot n_{\cdot 1}} + \frac{n_{12}^2}{n_{1\cdot} \cdot n_{\cdot 2}} + \right. \\ &\quad \left. \dots + \frac{n_{KL}^2}{n_{K\cdot} \cdot n_{\cdot L}} - 1 \right) \\ &= N \cdot \left(\sum_{i=1}^K \sum_{j=1}^L \frac{n_{ij}^2}{n_{i\cdot} \cdot n_{\cdot j}} - 1 \right)\end{aligned}$$

を求める。 χ^2_0 は、カイ2乗分布と呼ばれる分布になることがわかっている。

表3の場合には、

$$\begin{aligned}\chi^2_0 &= 51 \times \left(\frac{8^2}{25 \times 11} + \frac{10^2}{25 \times 19} + \frac{6^2}{25 \times 16} + \right. \\ &\quad \left. \frac{1^2}{25 \times 5} + \frac{3^2}{26 \times 11} + \frac{9^2}{26 \times 19} + \right. \\ &\quad \left. \frac{10^2}{26 \times 16} + \frac{4^2}{26 \times 5} - 1 \right) \\ &= 5.108\end{aligned}$$

となる。カイ2乗分布は、その自由度により分布の形状が変わってくる。クロス表の自由度は、“計”の欄の数値を変えずに、セルの数値を最大何個自由に求められるかという、その個数と一致する。したがって、自由度は、 $(K-1) \times (L-1)$ となる。例題では、 $(2-1) \times (4-1) = 3$ となる。検定の方法は、 χ^2_0 の値がカイ2乗分布の上側5%（もしくは1%）点の値より大きければ、有意差があるとし、小さければ差が認められないとすればよい。自由度3のカイ2乗分布の上側5%点の値は7.815であり、求めた χ^2_0 の値はこれより小さいので、有意ではない。したがって、A薬とB薬の効果に差があるとはいえないことになる。

なお、一方のカテゴリに順序がある場合には、より効率のよい方法があるが、その点については次回以降に解説する予定である。

* * *