いくつかのものを統計的に比較してみよう

筑波大学　数学系　　青嶋　誠

1. はじめに

皆さんは学校で**統計**を勉強しているかも知れませんね。高校までの範囲では、正規分布の母平均に対する信頼区間を作ることがゴールになっています。よく読むと、母平均の信頼区間を作るために、どうやら、母分散の値は分かっていることを仮定しています。母分散の値が分からない場合、信頼区間はどうやって作ったらよいのでしょうか？また、２つの母平均の差に対して信頼区間を作りたい場合、どうしたらよいのでしょう？２つのクラスのデキを平均点で比較する場合が、これに当たります。３つ以上の母平均を、それぞれ対にして比較したい場合は、どうすればよいのでしょう？単純に２つの母平均の差の信頼区間を繰り返して作ればよいのでしょうか？この辺の議論に、数学が使われるのです。

教科書にも書いてあるように、統計的に推定を行うとは、

母集団を設定し、母集団から無作為に取り出された標本の値に基づいて
母集団を特定する母数を推定すること

です。**統計的推定**とは、まさに、一部から全体を推定する理論と方法なのです。皆さんの中には、一部などとケチなこと言わずに全部を調査してしまえばいいじゃないか、と思う人がいるかも知れませんね。でもね、日本人の成人男子の平均身長を知ろうとばかりに、長万部から渋谷から那覇からと、二十歳の男性を調査期間内にくまなく測定することなんて、できるだろうか！

統計学という学問は、数学のひとつの分野として厳密な理論をもつ一方、方法としての使用がそれほど困難でないため、応用範囲が極めて広いという特徴をもっています。ときとして、この学問的性質が災いし、解釈や使い方に誤解や誤用が発生し易いことも、特徴であるかも知れません。
本講義では、方法論を闇雲に与えることなく、出来る限り理論の背景を重視しながら、高校から大学そして大学院・社会人のレベルまでを、ひとつの物語でお話したいと思います。

2. 母平均の区間推定

母集団の未知母数を、標本の値から、精度まで考慮にいれて推定をしようというのが**区間推定**である。正規分布の母平均の区間推定を考えよう。正規分布は記号で $N(\mu, \sigma^2)$ と表記される。正規分布のカッコの中の最初の母数は分布（グラフ）の母平均（重心）を表し、その値が μ（「ミュー」と読む）という訳だ。ここでは、μ の値が具体的にいくつか分からない状況を考えていて、標本から区間推定で μ の値を推測しようとしている。一方、カッコの中の後ろの母数 σ^2（「シグマ自乗」と読む）は正規分布の分散を表し、この値はちょうど正規分布の重心から変曲点までの長さの 2 乗に等しい。高校の教科書で仮定されるように母分散 σ^2 の値が分かっているといいのだけれど、実際にデータを解析してみようとすると、なかなかそうもいかない。母平均 μ の推測をするときに、困った事に、母分散 σ^2 の値も分からないことが多いんだ。そこで、高校の教科書で勉強する母分散 σ^2 の値が分かっている場合の母平均 μ の区間推定を少しずつ修正して、母分散 σ^2 の値が分からない場合の母平均 μ の区間推定を作ることにしよう。

母集団から無作為に抽出された大きさ n の標本 $(X_1, ..., X_n)$ をもとに、母平均 μ を推定するために**標本平均** $\overline{X} = \sum_{i=1}^{n} X_i/n$ を計算する。標本平均の計算値は母平均の周りで変動し、その変動の様子について、高校の教科書から

$$\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \tag{2.1}$$

の分布が平均 0 で分散 1 の正規分布 $N(0, 1)$ になることが知られている。正規分布 $N(0, 1)$ はあらゆる分布の基本となる分布で、**標準正規分布**と呼ばれている。ところで (2.1) 式は、大きさ n の標本を無作為に抽出する実験を何回も繰り返して、毎回計算される $\overline{X}$ の度数分布を（少し工夫して）グラフに描くと、母平均 μ を中心として左右対称で変曲点までの長さが $\sqrt{\sigma^2/n}$ であるような正規分布になる、ということを意味している。そこで、

上記の実験を 100 回繰り返して母平均 μ に対する 100 個の信頼区間を作ったとき、
そのうち 95 個の信頼区間は母平均 μ の値を含む

ことを区間推定の精度として要求しよう。この信頼区間の精度のことを、**信頼度** 95% と言うんだ。教科書の最後のページにある標準正規分布 $N(0, 1)$ の数表は、

$$P\left\{\left|\frac{\overline{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right| \leq z\right\} = .95 \tag{2.2}$$

なる z の値が $z = 1.96$ であることを教えてくれる。(2.2) 式による確率的保証から、母平均 μ に対する信頼度 95% の信頼区間が

$$\overline{X} - 1.96\sqrt{\frac{\sigma^2}{n}} \le \mu \le \overline{X} + 1.96\sqrt{\frac{\sigma^2}{n}} \tag{2.3}$$

で与えられる。ところが、母分散 σ^2 の値が分からないので、(2.3) 式の上限の値も下限の値も具体的には計算できず、このままでは役に立たない。そこで、先ほど抽出した標本を利用して σ^2 の値も推定して、(2.3) 式に推定値を放り込むことを考えよう。母分散の推定として、

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \tag{2.4}$$

という計算式が知られている。この式は、大学の統計学の授業で最初に学ぶ大切な式で、**標本分散**と呼ばれるものなんだ。標本分散 S^2 の分布や推定としての性質は、そのとき詳しく学ぶことにしよう。ところで、母分散の推定に、何でまたこのような式を考えるのだろう？実は、きちんとした理由があるんだ。(2.4) 式の分子にも分母にも。その辺の物語は、またの機会にお話しよう。ともかくも、(2.3) 式の中の母分散 σ^2 を標本分散 S^2 で推定してしまおう。すると、母平均 μ の信頼区間は

$$\overline{X} - 1.96\sqrt{\frac{S^2}{n}} \le \mu \le \overline{X} + 1.96\sqrt{\frac{S^2}{n}} \tag{2.5}$$

で与えられる。しかし、確率変動する標本分散 S^2 を単純に放り込んでしまって、はたして、(2.5) 式の信頼区間は精度の要求「信頼度 95%」を満たすのだろうか？もともと信頼区間の「1.96」という数値は、(2.1) 式と (2.2) 式とから、標準正規分布 $N(0,1)$ の数表を拝借して得られた数値であった。母分散 σ^2 を 標本分散 S^2 で推定する場合、本来、(2.1) 式は

$$\frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}} \tag{2.6}$$

と書き変わるべきである。実は、この分布は、標準正規分布 $N(0,1)$ を上から少しつぶして、その分だけ裾が重くなった形状の、正規分布とは別の分布になることを大学で勉強する。この分布は、(**自由度** $n-1$ の）t **分布** と呼ばれているんだ。すると、(2.2) 式は、

$$P\left\{\left|\frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{n}}}\right| \le t\right\} = .95 \tag{2.7}$$

なる t を求める式に書き変わり、この t 値については、今度は大学で学ぶ教科書の最後のページにある t 分布の数表を引くことになるんだよ。この t 値は自由度（ここでは $n-1$）

と呼ばれる値に依存して決まり、例えば標本の大きさが $n = 10$ のとき、$t = 2.26$ が数表から得られる。すると、最終的に、母平均 μ に対する信頼度 95% の信頼区間は

$$\overline{X} - 2.26\sqrt{\frac{S^2}{10}} \leq \mu \leq \overline{X} + 2.26\sqrt{\frac{S^2}{10}} \tag{2.8}$$

で与えられることになる。(2.5) 式と比べて、区間の幅がほんの少し広がっていることに気が付くね。実は、(2.5) 式のままだと、

$$P\left\{\left|\frac{\overline{X} - \mu}{\sqrt{\frac{S^2}{10}}}\right| \leq 1.96\right\} = .918 \tag{2.9}$$

となることが自由度 9 $(= n - 1)$ の t 分布の確率から計算できて、信頼区間の精度が信頼度 95% からやや落ちることが分かるんだ。

3. 2つの母平均の差の区間推定

母集団を2つにして、2つの正規分布の母平均の大小関係を比較することを考えよう。2節の話を応用しよう。2つの母集団の分布を記号で $N(\mu_i, \sigma_i^2),\ i = 1, 2$ と表記しておく。2つの母集団を母平均 μ_1 と μ_2 の大小関係で比較したいのだが、これらの値が分からないので標本から推定する必要がある。ここでの推測の目的は母平均の大小関係を調査することにあり、2節にあったような母平均そのものの値を知ることが目的ではない。その意味で、母平均の差 $\mu_1 - \mu_2$ が推定のターゲットになる。母平均の差 $\mu_1 - \mu_2$ について区間推定を行えば、

差の信頼区間の下限が正であれば第1母集団の母平均が第2母集団の母平均よりも
大きいと統計的に判断でき、逆に、上限が負であれば第1母集団の母平均が
第2母集団の母平均よりも小さいと統計的に判断できる

という訳だ。もしも、差の信頼区間が原点をまたぐ場合、2つの母集団の間には統計的に見て母平均に有為な差は検出されなかった、ということになる。これらの結論の精度は、前節と同じ「信頼度 95%」に設定しておく。

話を簡単にしよう。

2つの母集団の母分散は等しい $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$

ことを仮定する。実は、2つの母分散が異なる場合の母平均の差の推測は **ベーレンス・フィッシャー問題** と呼ばれていて、統計学の歴史的な問題のひとつなんだ。この問題を

解くいくつかのアプローチが、世界中の研究者によって精力的に開発された。この辺の研究開発の物語は、別の機会にお話しよう。ともあれ、共通の分散 σ^2 の値は、2節と同様に、信頼区間を作る過程の中で必要に迫られた時に推定することにしよう。

それでは、2節で述べた区間推定の手順を真似て、母平均の差 $\mu_1-\mu_2$ の信頼区間を作っていこう。母集団分布 $N(\mu_1, \sigma^2)$ をもつ第1母集団から大きさ n_1 の標本 $(X_{11}, ..., X_{1n_1})$ が無作為に抽出され、一方、母集団分布 $N(\mu_2, \sigma^2)$ をもつ第2母集団からは大きさ n_2 の標本 $(X_{21}, ..., X_{2n_2})$ が無作為に抽出されたとする。2つの母平均の差 $\mu_1 - \mu_2$ を推定するために、それぞれの母集団から計算される標本平均 $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i,\ i = 1, 2$ の差 $\overline{X}_1 - \overline{X}_2$ を計算しておこう。標本平均の差 $\overline{X}_1 - \overline{X}_2$ の確率変動については、

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2}} \tag{3.1}$$

の分布が標準正規分布 $N(0, 1)$ になることを大学で学ぶ。信頼度 95% の信頼区間を作るために、標準正規分布 $N(0, 1)$ の数表から (2.2) 式と同じ $z = 1.96$ を用いる。母平均の差 $\mu_1 - \mu_2$ に対する信頼度 95% の信頼区間は、値の分からない共通の母分散 σ^2 を含んだ式で、

$$\overline{X}_1 - \overline{X}_2 - 1.96\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2} \leq \mu_1 - \mu_2 \leq \overline{X}_1 - \overline{X}_2 + 1.96\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2} \tag{3.2}$$

のように与えられる。ここで、先ほど抽出した標本を利用して σ^2 の値を推定しよう。母分散 σ^2 の値は両方の母集団で共通なので、第1母集団から抽出された標本 $(X_{11}, ..., X_{1n_1})$ と第2母集団から抽出された標本 $(X_{21}, ..., X_{2n_2})$ との両方が、母分散 σ^2 の情報を握っていることになる。これを使わない手はない。そこで、(2.4) 式で導入した推定方法で両方の母集団から標本分散 $S_i^2 = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2/(n_i - 1),\ i = 1, 2$ を計算し、これらを併合して母分散 σ^2 の推定に次の式を考える。

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \tag{3.3}$$

(3.3) 式で定義される S^2 が、分布の導出や推定などで都合の好い格好をしていることも、大学の統計学の授業で学ぶ事柄のひとつなんだ。2節の最後で経験したことから類推すると、(3.2) 式で与えられる信頼区間の母分散 σ^2 を確率変動する S^2 で単純に置き換えただけでは、区間推定に要求される「信頼度 95%」は確保できそうにない。そこで、(3.1) 式の代わりに、今度は

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)S^2}} \tag{3.4}$$

の分布を知る必要が出てくる。これも大学で詳しく学ぶ事柄なんだけど、実はこれも、(2.6) 式と同様に、標準正規分布 $N(0,1)$ を上から少しつぶした「t 分布」になるんだ。2 節でも少し触れたけど、t 分布は「自由度」と呼ばれる母数に依存して形状が決まる。自由度は S^2 の分母の値になっていて、これは標本数に関係している。今の場合、(3.3) 式の分母から、t 分布の自由度は n_1+n_2-2 になる。ちなみに標本数が大きくなるにつれ S^2 は母分散 σ^2 の値をより上手く推定するので、t 分布の形状は次第に中心が吊り上がり、しまいには標準正規分布 $N(0,1)$ のグラフとぴったり重なるんだ。

ともあれ、

$$P\left\{\left|\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S^2}}\right|\leq t\right\}=.95 \tag{3.5}$$

なる t 値を t 分布の数表から引こう。第 1 母集団の標本数が $n_1=10$ で 第 2 母集団の標本数が $n_2=15$ のとき、自由度は $n_1+n_2-2=23$ となり、このとき t 値は $t=2.07$ となる。最終的に、母平均 $\mu_1-\mu_2$ に対する信頼度 95% の信頼区間は、

$$\overline{X}_1-\overline{X}_2-2.07\sqrt{\left(\frac{1}{10}+\frac{1}{15}\right)S^2}\leq\mu_1-\mu_2\leq\overline{X}_1-\overline{X}_2+2.07\sqrt{\left(\frac{1}{10}+\frac{1}{15}\right)S^2} \tag{3.6}$$

で与えられることになる。

4. ３つ以上の母平均の差の区間推定

母集団が 3 つ以上ある場合、それらの正規分布の母平均の大小関係を**同時に**比較したい。ここで注意しなければならないことは、

同時に複数個の信頼区間が成立する確率は、個々の信頼区間を 3 節の方法で個別に扱った場合に設定される信頼度 95% よりも小さくなってしまう

という点だ。従って、3 節で作った信頼区間を母平均の全ての対にそのまま使う訳にはいかず、t 値を「やや大きめ」に設定する必要がある。どの程度に「やや大きめ」にするかの議論に、統計学の理論である**数理統計学**が登場する。

問題を明確にしよう。いま、$k\ (\geq 3)$ 個の正規分布に従う母集団 $N(\mu_i,\ \sigma^2),\ i=1,...,k$ があり、これらの母平均の大小関係を全ての対で比較したい。母分散については、3 節と同様に全ての母集団で共通の値をもつことを仮定している。全ての母平均の差 $\mu_i-\mu_j,\ 1\leq i<j\leq k$ を 95% の信頼度で同時に比較することが、ここでの問題である。

各母集団 $N(\mu_i, \sigma^2)$ $(i = 1, ..., k)$ から大きさ n_i の標本 $(X_{i1}, ..., X_{in_i})$ が無作為に抽出されたところから話を始めよう。それぞれの母集団で標本平均 $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i,\ i = 1, ..., k$ と標本分散 $S_i^2 = \sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2/(n_i - 1),\ i = 1, ..., k$ を計算しておこう。標本平均は全て対にして、$\overline{X}_i - \overline{X}_j,\ 1 \le i < j \le k$ で母平均の差 $\mu_i - \mu_j,\ 1 \le i < j \le k$ を推定する。標本分散の方は併合して、(3.3) 式を一般化した次の式で共通の母分散 σ^2 を推定する。

$$S^2 = \frac{\sum_{i=1}^{k}(n_i - 1)S_i^2}{\sum_{i=1}^{k} n_i - k} \tag{4.1}$$

３節までの議論から類推すれば、どうやらここでは、それぞれの分布が自由度 $\sum_{i=1}^{k} n_i - k$ の t 分布になる $k(k-1)/2$ 個の

$$T_{ij} = \frac{(\overline{X}_i - \overline{X}_j) - (\mu_i - \mu_j)}{\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_j}\right) S^2}}, \quad 1 \le i < j \le k \tag{4.2}$$

からなる関数の分布を議論する必要がありそうだ。具体的には、区間推定の精度に対して要求されている「信頼度 95%」を確保するために、$k(k-1)/2$ 個の不等式が同時に成り立つ確率が

$$P\left\{|T_{ij}| \le t,\ 1 \le i < j \le k\right\} = .95 \tag{4.3}$$

なる t 値を求めなければならず、そのためには、とどのつまり

$$\max_{1 \le i < j \le k} |T_{ij}| \tag{4.4}$$

の分布が必要になる。

実は、この分布の理論は大変に難しい！

大学院レベルになるので、長い話は省略して、結果だけを紹介しておこう。分布に関する込み入った議論の末、(4.3) 式は以下のように積分で表現される。

$$\int_0^\infty \int_{-\infty}^\infty \sum_{j=1}^{k} \prod_{\substack{i=1 \\ i \ne j}}^{k} \left\{ \Phi\left(\frac{\sqrt{\lambda_{ij}}x}{\sqrt{1 - \lambda_{ij}}}\right) - \Phi\left(\frac{\sqrt{\lambda_{ij}}x - ty}{\sqrt{1 - \lambda_{ij}}}\right) \right\} \phi(x)dx\, g(y)dy = .95 \tag{4.5}$$

ここで、$\lambda_{ij} = n_i/(n_i + n_j)$ である。また、$\Phi(\cdot)$ は標準正規分布 $N(0, 1)$ の確率を表す関数、$\phi(x)$ は $N(0, 1)$ の分布を表す関数で、$g(y)$ は自由度 $\sum_{i=1}^{k} n_i - k$ に依存して決まる $\sqrt{S^2/\sigma^2}$ の分布を表す関数なんだ。この t 値については、大学の教科書をいくら捜しても数表が与えられていない。数表を与えるには、計算機をフルに使って、個々のケースに合わせて数値計算をしなくてはならない。計算すべき標本数 $(n_1, ..., n_k)$ の (比の) パターン

があまりに多過ぎて、困りものなんだ。実は、５０年ほど前にテューキーという統計家が、t 値の簡便な代用値を見つけた。(4.5) の方程式よりもはるかに実用に便利な

$$k\int_0^{\infty}\int_{-\infty}^{\infty}\left\{\Phi(x)-\Phi(x-\sqrt{2}ty)\right\}^{k-1}\phi(x)dx\,g(y)dy=.95 \tag{4.6}$$

といった、標本数 $(n_1,...,n_k)$ の和 $\sum_{i=1}^{k}n_i$ だけの関数で表現される方程式を解くことで、t 値の代用とすることを提唱したんだ。方程式 (4.6) は、方程式 (4.5) での $n_1=...=n_k$ とおいた特別な場合なんだけど、テューキーの偉いところは、

方程式 (4.6) で求められる値を方程式 (4.5) の t 値の代用として用いれば、そのとき

$$P\left\{|T_{ij}|\leq t,\ 1\leq i<j\leq k\right\}\geq .95 \tag{4.7}$$

なる不等式が成立する

ことを予想したところなんだ。つまり、

計算が面倒な方程式 (4.5) の t 値を、簡便な方程式 (4.6) の解で代用しても、
区間推定は信頼度 95% の精度を少なくとも保証する

と主張できるだろう、という訳だ。この予想は **テューキー予想** と呼ばれ、これを解決するために多くの研究者がその証明を試みた。それはもう膨大な文献があるんだ。

1984 年、ついに、この予想は完全に解決された。

ヘイターという研究者が、予想が正しいことを複雑な数学を使って証明したんだ。ここに至るまでの歴史物語は、また皆さんに会ったときにでもゆっくり話すことにしよう。私の研究室も含め世界中で、今では、予想の一般化を検証することが研究されている。研究の矛先は、時代に合わせた具体性をもって、さらなる抽象化へと移行しているんだ。

少し気の早い話をしたかも知れないが、ともあれ、方程式 (4.6) を数値計算して t 値の代用値を具体的に求めてみよう。母集団の個数が３つで、標本数がそれぞれ $(n_1,n_2,n_3)=(8,10,15)$ のとき、(4.6) 式を解いて $t=2.47$ が得られる。ちなみに、面倒でも方程式 (4.5) を解いて得られる正確な t 値は $t=2.46$ となり、(4.6) 式から得られる代用値が若干大きめに設定されていることが分かる。正確な t 値は、３節の方法で任意の２つの母平均の比較を個別に行うとしたときの t 値 である $t_{12}=2.12,\ t_{13}=2.08,\ t_{23}=2.07$ の、どれよりも大きい。このことからも、母集団の個数が３つ以上の場合、２つの母平均の差の区間推定を単純に繰り返す方法では、推測に十分な精度が得られないことが分かる。最終的に、

全ての母平均の差 $\mu_i - \mu_j,\ 1 \le i < j \le 3$ に対する 95% **以上**の信頼度をもつ**同時信頼区間**は、

$$\overline{X}_1 - \overline{X}_2 - 2.47\sqrt{\left(\frac{1}{8} + \frac{1}{10}\right) S^2} \le \mu_1 - \mu_2 \le \overline{X}_1 - \overline{X}_2 + 2.47\sqrt{\left(\frac{1}{8} + \frac{1}{10}\right) S^2}$$
$$\overline{X}_1 - \overline{X}_3 - 2.47\sqrt{\left(\frac{1}{8} + \frac{1}{15}\right) S^2} \le \mu_1 - \mu_3 \le \overline{X}_1 - \overline{X}_3 + 2.47\sqrt{\left(\frac{1}{8} + \frac{1}{15}\right) S^2}$$
$$\overline{X}_2 - \overline{X}_3 - 2.47\sqrt{\left(\frac{1}{10} + \frac{1}{15}\right) S^2} \le \mu_2 - \mu_3 \le \overline{X}_2 - \overline{X}_3 + 2.47\sqrt{\left(\frac{1}{10} + \frac{1}{15}\right) S^2}$$

で与えられることになる。

「いくつかのものを統計的に比較」する話は、まだまだつきないけれど、続きはまた別の機会にお話することにしましょう。