

信頼区間と検定の関係は？

高木 廣文

新潟大学医学部保健学科教授

1. はじめに

これまで何度も述べているように、統計学が多くの分野で多くの研究者たちに使用されている理由は、少ない標本データでより大きな母集団についてのことが言えるからである。もう少しいえば、小さな標本データを統計学的解析、すなわち検定や推定の様々な方法を用いて、ずっと大きな母集団の特性について、言及することができるからである。

もしも、新しい薬の効果を調べるために、わざわざその病気の患者全員について調査研究をしなければならないのなら、新薬など全く開発することができないのではなかろうか。

統計学は記述統計学と推測統計学に大きく2分類されるが、推測統計学の方法を使うことで、科学的な方法として少ないデータから一般的な結論を得ることが可能になる。

ただし、どのようなデータでも、分析方法さえ良ければよいというわけではない。統計学的推測の前提条件は、標本の無作為抽出や無作為割付などの、無作為ということが必要である。この点に関しては、すでに説明した。

また、推測統計学の中心である検定や推定の方法についても、何度も触れてきた。今回は、統計学的推定のうち、信頼区間について、少し詳しく説明したいと思う。

2. 統計学的な推論について

研究のための標本が正しく選ばれた場合（多くの場合、「無作為」に選ぶ必要がある）、統計学の手法により、母集団に関する推論を行うことができる。母集団に関する統計学的推論は、「推定 estimation」と「検定 test」に分けられるが、まず推定の考え方を簡単に以下に説明しよう。

(1) 統計学的推定とは

例えば、20歳の日本人男性1000人を無作為に抽出し、その身長を測定したところ、平均値が170cmだったとしよう。我々は、この結果から、20歳の日本人男性の平均身長は170cmであると主張することができる。すなわち、なにも20歳の男性全員の身長を測定する必要はないことになる。これが、統計学が示した結論であり、「推定」と呼ばれる方法である。すなわち、推定とは、少数の標本データからその背後にある母集団全体の特性について言及するための方法ということになる。

上記の例では、統計学的推定の方法として、20歳日本人男性の身長の平均値を170cmという、一つの数値で推定を行った。このような推定方法は、推定値として数直線上のある一点を用いて行うため、「点推定 point estimation」と呼ばれている。

このように、標本データから求めた平均値は、「データが母集団から無作為（ランダム）に集められた」場合には、そのまま母集団のある量的な特性の母平均値の点推定値として、使用することができる。

同様に、20歳の男性日本人1000人の喫煙状況を調べたところ、喫煙者が40%いたとしよう。この標本での喫煙割合も、無作為に標本が選ばれていれば、母集団である20歳の日本人男性での喫煙の母割合は40%である、と推定することができる。このように、標本での割合もまた、そのまま母集団での

ある質的な特性の母割合の点推定値として用いることができる。

このように、標本がランダムに選ばれたのであれば、その標本から収集されたデータの平均値や割合は、そのまま母集団の推定値として扱ってよいことになる。当然ではあるが、この前提となる「ランダム」に標本が選ばれていなければ、この推論はできないことになる。現実問題では、医学領域での標本は滅多にランダムに集められることはないので、その推定値はバイアス（前回の記事参照）が含まれる可能性が高い。

(2) 区間推定の考え方と信頼区間

ところで、点推定は1つの数値を用いて、母集団の特性値を推測するため、必ずしも母集団の特性値である母数と完全に一致するものではない。というよりも、かならず誤差があるので、幾分は母数からはずれた数値である可能性がある。そこで、一つの数値ではなく、その前後にある程度の幅を設定した区間を考え、その区間の中に推定したい母数が含まれるようにできれば、その方が推定方法としては確実なものになるだろう。

このように、ある確からしさのもとで一定の区間を設定し、その区間内に母平均値などの母数が存在することを示す方法は、「**区間推定** interval estimation」と呼ばれている推定方法である。区間推定によって求められた母数が存在すると推定される区間は、「**信頼区間** confidence interval」と呼ばれている。

信頼区間の「信頼」とは、この場合、母集団のある特性値である母数の存在する可能性（確率）を示すものである。その確率の大きさに応じて、母平均値の「95%信頼区間」などと呼ばれている。

このように、「信頼」という用語のもつ意味は、確率的な意味である。すなわち、95%信頼区間であれば、同一の調査を100回行った場合、同一の計算方法で母数の区間推定を行ったとき、求めた信頼区間の中に、母数が実際に95回は確実に入るということを、「95%」により表わしている。

逆にいえば、95%信頼区間では、たまたま母数とその信頼区間に入らない場合も、5%はあるということである。このような推定の誤りのリスクを小さくしたいのならば、例えば100回中99回は母数とその区間に入るように信頼性を高めるためには、「99%信頼区間」を求めればよいことになる。

3. 統計学的検定について

実際のデータ解析では、推測統計学の方法でよく使われているのは、母数の信頼区間を推定するのではなく、多くの検定方法であろう。

統計学的検定も推定と同様に、標本のデータを用いて、より一般的に母集団の各種の特性について言及するための方法である。しかし、推定が直接、母集団の特性値を求めるのに比べて、検定では間接的な方法を用いている。すなわち、「帰無仮説」と呼ばれる母集団についての仮説を設定し、それが成り立つかどうかを、データを用いて調べるという手順をとる。例えば、喫煙と血管拡張率に関して、「喫煙者は非喫煙者に比べて、血管拡張率が低い」という作業仮説を立てて調査研究を行ったとしよう。

このような場合、調査対象者を喫煙群と非喫煙群に分け、血管拡張率の計測の結果の比較を行うことになるだろう。その結果、調査データでは確かに、喫煙群で血管拡張率が低値であったとしよう。しかし、2つしかグループがなければ、完全に2つの平均値が一致する場合よりも、偶然の影響により、どちらかが高く、どちらかがより低くなり、見た目の差が生じるのが普通だろう。したがって、その偶然性による影響の有無を統計学的に検討する必要がでてくる。

この偶然性を確率として計算するために、統計学的仮説検定法では、作業仮説とは異なり、「喫煙者と非喫煙者の血管拡張率の母平均値は等しい」というような帰無仮説を立てる。そして、帰無仮説が正しいと仮定した場合に、調査データで生じた2グループ間の差が生じる確率（有意確率、 p 値）を計算する。求められた確率が大きければ、2つのグループ間の差は偶然によっても起こり得ることになり、必ずしも2グループ間に真の差があるとはいえないことになる。

その逆に、求めた確率が小さければ、そのように滅多に起こらないことが、たまたま自分の研究で起こったと考えるのは無理があるということになる。したがって、その確率計算の前提となった仮説に誤があると考えねばならないだろう。すなわち、帰無仮説は否定され（「棄却する」という）、2つのグループでの母平均値には差（有意差）があると言えることになる。これが、統計的な仮説検定の考え方である。

上記のように統計学的仮説検定における仮説は、棄却されるために存在する。このために、帰無仮説（無に帰す仮説）と呼ばれるのである。かりに求めた確率が大きいため、帰無仮説を棄却することができないのならば、その検定はほとんど意味のないものとなる。何故ならば、帰無仮説を棄却できないということは、その帰無仮説を積極的に支持する根拠を与えた訳ではなく、そのような状況も確率的にはありえるということを示したに過ぎないだけだからである。

4. 信頼区間と仮説検定の関係について

喫煙者と非喫煙者の血管拡張率を比較する場合、前述のように帰無仮説「喫煙者と非喫煙者の血管拡張率の母平均値は等しい」を検定する方法以外にも、いくつかの方法が考えられるだろう。帰無仮説の検定と同様の内容の推定を行うことができる。例えば、

(1) 喫煙者と非喫煙者の血管拡張率の母平均値の95%信頼区間を、それぞれ求める。

それぞれのグループの信頼区間内には、母集団の母平均値が含まれるはずなので、2グループの95%信頼区間に重なっている区間があれば、それは2つのグループの母平均値が等しい場合もあり得ることを示している。すなわち、2つのグループの信頼区間が重なっている場合には、2つの母平均値には差がない可能性があり、有意差は認められないことになる。

逆に、2グループの95%信頼区間に重なる区間がないのであれば、2つの母平均値は異なっていると考えてよいことになる。すなわち、5%のリスクのもとで、2つのグループの血管拡張率には有意差があると考えてよいことになる。

もう一つの方法として、

(2) 喫煙者と非喫煙者の血管拡張率の母平均値の差の95%信頼区間を求める。

2つのグループの母平均値の信頼区間を個々に求めるのではなく、2つの母平均値の差について、95%信頼区間などを求めることが考えられる。求めた信頼区間に0が含まれている場合、差が0の場合もあり得ることを示している。すなわち、2つのグループの母平均値が等しい場合もあることを示しており、有意差はないものと考えられる。

もしも、95%信頼区間に0が含まれないのであれば、2つの母平均値には5%のリスクのもとで有意差があることになる。

このように、信頼区間を求めることは、統計学的検定を行うことと同等の結果を得ることができる。さらにいえば、検定では、母数の推定に関して、ほとんど情報は得られない点から、どちらかといえば、95%信頼区間を求める方がより多くの情報を得られることから、推奨できる方法といえよう。

なお、今回は説明のために、解説が平均値を例としているが、相関係数、オッズ比などでも同様の議論が成り立つことは、言うまでもないことだろう。

参考文献

1) 松元香緒里, 坂口勝哉, 大串正道: 反応性充血試験を用いた血管拡張反応の超音波による検討; 喫煙者, 非喫煙者及び喫煙中止者を対象として, 超音波検査技術, 23(2), pp175, 1998。