

基礎講座

統計学の基本的な考え方

高木 廣文**はじめに**

超音波検査という私にとってはあまり馴染みのない分野で、これから統計学の解説を行うことになった。できるだけ平易に分かりやすくという気持ちで、実際の研究にどのように統計学が関わるのかを解説する予定である。

私がこれまで主に関係してきた公衆衛生学の分野では、超音波を用いた測定というと、皮脂厚の測定ぐらいしか思いつかない。したがって、具体的な解析例は、本誌や研究発表会での報告から、適当にアイデアをいただくつもりでいる。内容をよく理解しない、おかしな例題にならないか心配である。もしも、適切ではないと思われる例題がある場合には、その解説の真意だけを読み取っていただきたいと考えている。

統計学といっても、基礎的な分布に関する理論、データを記述的に表現する方法、標本抽出法、実験計画法、基礎的な検定と推定に関する方法、複数の変数を一度に扱う多変量解析など、どこに焦点を絞るかで、その内容もかなり異なってくる。

ここでは、実際の臨床研究に役立つような統計学に関する解説を中心に行う予定である。まず、はじめに統計学の基本的な考え方について解説を行う。次回以降、順によく用いる検定方法などの解説を行うつもりである。

統計と統計学の違い

「統計」をとるといい方がある。例えば、

ある病院での外来患者数の診療科別の「統計」をとる、というような使われ方をするようである。このような場合、統計という意味は、「調査をして、診療科ごとの人数を集計する」ということである。人口動態統計などは、死亡などに関する集計結果であり、まさにこのような使われ方である。

歴史的には、世界中のどこでも国が誕生すると徴兵、徴税などのために自国民の人数調査が必ず行われており、「統計」をとることが行われてきた。しかし、このようにデータを収集し、数をかぞえてまとめることを、現代の「統計学」と考えていては、大きな間違いを冒すことになる。

確かに、データを収集し数をかぞえることは、統計学の基本である。様々な表やグラフにまとめることも同様に、統計学の基本であることは確かである。しかし、統計学の果たす重要な役割は、データをまとめたり図示したりすることにあるのではなく、データから得られた結果を、一般化、普遍化することにある。TVの視聴率調査で、日本のすべてのTVに調査用の機器が設置されている訳ではない。数千万台の内、わずかに数百台～数千台程度の調査結果から、各番組の視聴率が計算されている。全体のほんの一部を調べた結果から、あたかも全体を調べたような結論を導いても文句がでないのは、統計学のおかげである。

この統計学の特質を理解するために、調査の基本的な考えである標本と母集団の関係を、次に説明する。

母集団と標本

動脈硬化のリスク因子である喫煙の影響を検出するため、反応性充血前後の血管拡張反応を超音波で測定する場合を考えてみよう。日本人の健常

者を対象として喫煙の影響を調べるのであれば、概念上は「日本人の健常者」が、調査対象ということになる。このような概念上の対象集団を、統計学では「母集団：population」と呼んでいる。実際の調査対象となる集団もまた「母集団」と呼ばれるが、「ユニバース：universe」と呼んで区別している。実際に調査を行う場合、ユニバースの構成員全員を対象とする「全数調査(悉皆調査)」を行うのは、国勢調査のような国の行政機関が行う場合を除けば、極めてまれといえる。普通は、予算や人材の問題から、ユニバースのほんの一部が調査対象として選ばれ、データが収集される。実際の調査対象は「標本：sample」と呼ばれている。統計学の方法は、この標本から母集団に関する様々な推論を行うための手段を与えるものといえる。問題は、どんな標本でもよいという訳ではないことである。例えば、ある病院の外来患者から適当に対象を選んで、喫煙の調査と血管拡張反応を超音波測定したデータから得た分析結果は、そのまま一般の健常者に適応するのは、かなり困難が伴うだろう。

それでは、どのようにデータを集めれば、統計学の要求にあうのだろうか。

統計学の要求するデータの集め方

喫煙の心疾患への影響を調べるのに、全世界中の喫煙者而非喫煙者の比較を行わねばならないのならば、医学研究は不可能に違いない。実際には、喫煙の影響を調べるにも、ある薬剤の効果を判定するにも、全患者数に比べれば、遥かに少数の対象者を追跡することで、様々な研究が行われている。それではなぜ、少数の対象者のデータだけでよいのだろうか。

統計学の多くの教科書では、標本(調査対象)の選び方について、多くのページを割いている。すなわち、単純無作為抽出法、系統抽出法、多段階抽出法、層別抽出法などの標本抽出法について必ず記述があるはずである。調査対象を選ぶのにも、まずどの方法がどのような長短があるのか知らないで、対象を選ぶことさえ困難である。

多くのページを割いているからには、標本の選び方は重要に違いない。それでは、このような標

本抽出法の何が重要なのであろうか。

それぞれの標本抽出法の手順は異なっても、「標本はランダムに選ぶ」という点が重要ある。このランダムは「無作為」と訳されており、「作為」的でないという意味を示すものである。

よく行われるように、臨床研究では自分が勤務している施設の患者を研究対象とする場合が多いだろう。このような調査方法は、研究者の意識の中では「作為」的ではないかもしれない。しかし、たまたま自分の周辺にいる人達から対象者を選んでおり、分類すれば「機縁法」という作為的な方法であり、「無計画抽出法」と呼ばれる「悪い」標本調査方法の一つである。

自分の周囲にたまたまいる人や知人、友人、親類縁者などを選んだのでは、統計学の要求するような無作為標本抽出法になっていないことになる。アルバイトの学生や応募者を調査対象とした場合なども、同様に無計画抽出法の一つである。

無作為抽出という用語は、すでに述べたように英語の random sampling の訳であるが、この言葉を適切に表現するのは容易ではない。例えば、正しく作られたサイコロでは、1から6までのどの目がでるかはランダムである、宝くじで各賞が当たる可能性はランダムに決まる、など使用できる。すなわち、「ランダム」とは人為的、作為的な操作を排除して、くじ引きやサイコロ投げのような要領で、対象者を選ぶことを示す用語であるといえる。それでは、ランダムに対象を選ぶと何がよいのであろうか。それは、無作為に対象を選ぶことにより、調査対象者に選ばれる可能性を、ある集団内でその各構成員について、誰でも等しくなるようにできるからである。これは「等可能性の原理」と呼ばれる標本抽出法の中心的な原則である。ランダムに選ばれた標本から収集されたデータは、偏りがないものと期待でき、その対象者が所属する集団(母集団)を「代表」するものとして扱うことができる。

このように、研究対象の選び方で最も重要な点は、調査対象が母集団の代表性をもつことである。母集団を代表しないようなデータをいくら集めても、科学的な研究に必要な普遍的な推論は不可能である。

臨床研究での対象の選び方

無作為標本抽出法は、母集団の特性を推定し、研究結果を一般化するためには不可欠な方法論であり、そのために統計学の教科書ではよく説明されている。しかし、臨床研究ではランダムに対象を選ぶことは不可能であろう。

それでは、心疾患のリスク因子の研究のために、どのような調査対象で研究が行えるだろうか。人が病気になるか死亡したりするのに、普通は作為はないと考えてよいだろう。また、喫煙や飲酒などの状態については、とくに強制していなければ、自然のままなので無作為であると考えてよいだろう。したがって、喫煙の有無などによる心疾患罹患や死亡の頻度などの比較を行うことは可能だろう。これはデータの「比較可能性」と呼ぶべき性質であり、収集したデータがもつべき特性の一つといえる。

よく行われる研究形態で、小規模であるが効率のよい方法がある。患者を研究の対象とするのなら、大学病院などでは比較的容易に対象者の選択は行えるだろう。十分なインフォームドコンセントを行ったのち、患者から研究に関する情報の収集を行えばよい。問題は、患者だけのデータではケース・スタディや実態調査的な結果しか得ることはできないことである。そのような研究では、特定の病院などの患者を対象としているため、一般集団を代表するものとは考えにくく、対象の選択に偏り（バイアス）があるとみなされる。バイアスのあるデータからは、どのような統計学的な解析を行っても、母集団の特性値について推定を行うことはできない。

しかし、母集団の様々な特性値の推定はできなくても、比較すべき適当な集団を設定することにより、効率のよい比較研究を行うことは可能である。

比較研究では、患者に対して「コントロール」、「対照群」と呼ばれる健常者のグループを設定する必要がある。2つのグループで比較したい特性、例えば喫煙、非喫煙の影響を、心疾患群とコントロール群で比較したいのなら、理論上はそれ以外の特性をすべて2グループで一致させる必要が

ある。これは「マッチング」と呼ばれる方法であるが、現実には同一人以外の人間の特性をことごとく一致させることは不可能である。通常は、各患者に対して、コントロールとして性、年齢、職業、居住地などを一致した健常者を選べばよい方であろう。

このような「ケース・コントロール研究」で一番の問題は、コントロールには何の疾患も有しない健常者が最良なのだが、病院のデータでは外傷や風邪などの、比較したい疾患とは無関係と考えられる患者データを収集することしかできない場合が多いことである。このような場合には、やはりデータにその病院特有の偏りが生じる可能性が常にあるので、分析にはその点の注意が必要である。

コントロールさえ適切に選ぶことができれば、ケース・コントロール研究は、2グループ間の比較に重点を置く、比較可能性を高めた優れた研究方法となる。

データの比較可能性を高める方法として、公衆衛生分野でよく行われている健康教育などの例がある。ある地域の老人に健康教育を目的とした講演会などを実施し、その教育講演の効果を調べるとしよう。このような場合、教育講演の効果を検討するために、どのようなデータが研究に用いることができるのだろうか。通常は、その会の前後で同一の項目について調査を行い、その間の測定値の差を調べるという方法を用いるのが普通である。この場合、対象者は最低でも2回の同一の測定を行うことになる。このタイプの研究では、自分自身をケースとコントロールにした調査研究となる。すなわち、教育前の対象者はコントロールであり、教育後の対象者はケースとなる。講習会などの前後で得られるデータによって、2つのグループを作ると、各対象者については、ある教育を受けた以外では時間経過を除けば、すべての特性が、1対1に完全にマッチしていることになる。したがって、極めて比較可能性の高いデータが収集できることになる。ただし、実際にはどのような教育効果を測定すればよいかという大問題がある。知識や健康意識変化を調べたい場合には、その目的に応じて設定した項目からなる質問紙を

用いることが多い。血圧や体重のように直接測定できる場合もある。重要な点は、このように各対象がそれ自身がケースとコントロールを兼ねるために、統計学的方法で教育効果の有無を調べるための検定を行うことができるということである。このような研究では、どのようにその対象者を選んだかは、検定などを行う際には大きな問題にはならないことである。ただし、それらの対象者が女性のみであれば、結果を一般化する場合、女性に限られるし、ある地域の50歳代と60歳代の住民のみであれば、結果の解釈もその地域の50歳代、60歳代に限る必要がある。研究結果を普遍化したい場合には、研究対象として、広範囲な年齢層などで同様なデータを収集し、同様な結果を得ることができることを示さねばならない。統計学的方法が適用できるかと、その結果をどの程度一般化できるかは、必ずしも一致する訳ではない。

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、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」と「検定」に分けて考える。ここでは、まず推定の考え方を説明し、ついで検定の基本的な考え方を解説する。

1) 統計学的推定について

統計学的推定の方法は、「点推定」と「区間推定」にさらに分類される。

点推定とは、図1に示したように母平均値などを1つの数値を用いて推定するものである。たとえば、一定時間安静仰臥位後、まず上腕動脈径を計測し、その後左前腕部を250mmHgで4.5分間駆血し、駆血解除後1分後に再度上腕動脈径を計測し、その拡張率を調べた結果、喫煙者の平均値は2.7%、非喫煙者では7.7%であったとしよう（松元ほか、1998）。標本が無作為に選ばれている場合、標本から得た平均値と割合（「比率」と呼ばれることが多いが、これは間違った用語であり、できるだけ使用しない）は、母集団の母平均値と母割合の点推定値となることが知られている。したがって、データがランダムに選ばれたのであれば、母集団の推定値も同様な値と考えてよいことになる。

ところで、点推定は1つの数値しか用いないため、必ずしも母集団の母数と一致するものではない。そこで、ある統計学的な確からしさ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が考えられる。そのような方法が「区間推定」と呼ばれる推定的方式である。母数が存在すると推定された区間は「信頼区間」と呼ばれ、その信頼性の程度をつけて「母平均値の95%信頼区間」などと呼ばれる。この信

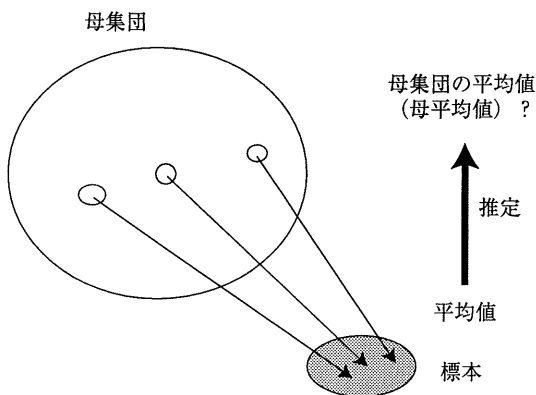


図1 推定の考え方

頼性の意味は、95%信頼区間であれば、100回の同一の調査を行い、同一の計算方法を用いれば、95回はこの信頼区間の中に母平均値が入るということを、「95%」により表現している。逆にいえば、5%は信頼区間に入らない場合もあるということであり、これを小さくしたければ「99%信頼区間」などを求めることになる。先の例では、「喫煙者の血管拡張率は、2.0~3.4%である」といった表現により、95%信頼区間を表すことになる。

2) 統計学的検定について

統計学的検定も推定と同様に、標本についてではなく、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、推定とは異なり、「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。ただし、検定のための仮説は、研究上の作業仮説とは異なる点に注意が必要である。

例として、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を行ったとしよう。

このような場合、対象者を喫煙群と非喫煙群に分け、各グループでの血管拡張率の計測を行うこ

とになる。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、どちらかが低くなるのが普通であろう。

統計学的な仮説検定法では、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」という仮説を立てる。この仮説が正しいとした場合に、調査データで認められるような2グループ間の平均値の差が生じる可能性(確率)を計算する。計算し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グループ間に差があるとはいえないことになる。逆に、確率が小さければ、そのようにまれなことが、たまたま自分の研究で起こったと考えるのは無理があり、そもその前提となる仮説が誤っていたと考えねばならない。したがって、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。上記の検定の手順を、図2に示しておいた。

このように統計学的仮説検定における仮説は、

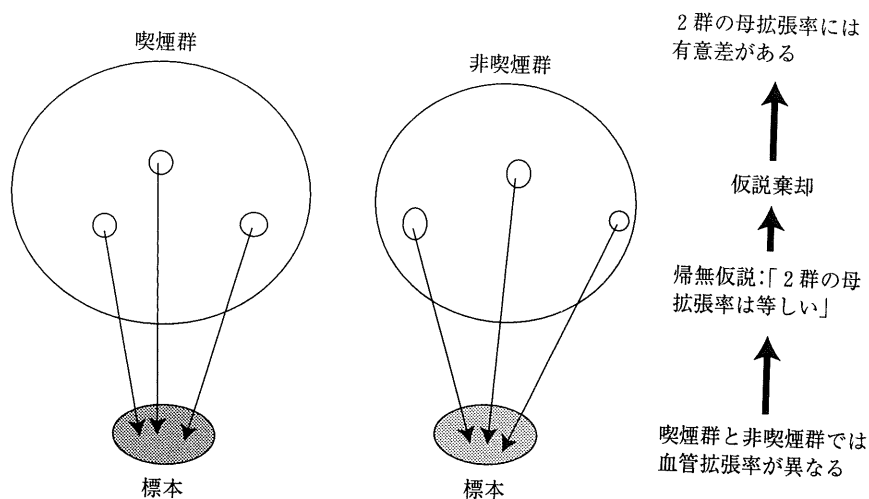


図2 検定の考え方

棄却されるために存在し、そのために帰無仮説と呼ばれるのである。かりに棄却されなければ、検定は無意味となる。何故ならば、仮説が棄却できないということは、仮説を積極的に支持する根拠を与えた訳ではなく、そのような状況もありえるということを示すに過ぎないからである。

統計学的検定のための目安となる確率の大きさは、通常「有意水準」と呼ばれ、慣習的に5%か1%がよく用いられている。この基準は絶対的なものではなく、状況によっては、10%程度の甘い基準でよいかもしれないし、重要な意志決定に関わる場合には、0.1%を採用するなど、状況によって適宜判断を自分自身で下す必要がある。現在では、検定の結果を示す場合、有意水準との比較のみで*などを付けるのではなく、確率そのものを「p値」として示すのが一般的である。

3) 統計学的検定における2つの誤り

すでに述べたように、帰無仮説を棄却するか否かという、検定の目安となる確率の大きさは有意水準と呼ばれ、「仮説が正しい場合に、誤って仮説を棄却する誤り」を冒す確率を示すものである。これは、「第1種の誤り（過誤）type I error」と呼ばれることがある。

一方、「仮説が誤っているのに、仮説を採択してしまう誤り」（仮説を棄却できない誤り）があり、これは「第2種の誤り（過誤）type II error」と呼ばれている。

一般に、第1種の誤り（有意水準）は α （アルファ）、第2種の誤りは β （ベータ）という記号

で表わされることが多い。有意水準 α は、慣習的に5%か1%を用いることが多い。第2種の誤り β に対して、 $1 - \beta$ は「帰無仮説が誤っている場合に、帰無仮説を棄却する」確率を示すので、「検定の検出力 power of a test」と呼ばれている。検定の検出力は、研究計画を立案する場合、研究に必要な標本数を設定するのに必要なものであり、通常は80%程度に設定することが多いが、詳細については次回以降、適当な箇所ですべて詳述したい。

おわりに

今回は、統計学の基本的な考え方を中心に解説を行い、併せて簡単な統計学用語の説明を行った。

今後は、適当な例題を用いて、統計学の検定方法を中心に解説を行う予定である。最後に、調査や統計学の基礎的な考え方を理解するための、参考図書を示しておいた²⁻⁵⁾。

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 1998: 23: pp175.
- 2) 林知己夫監修, 多賀保志編集: 調査とサンプリング, 同文書院, 東京, 1985.
- 3) 高木廣文, 佐伯圭一郎, 中井里史: HALBAU によるデータ解析入門, 現代数学社, 京都, 1989.
- 4) 高木廣文, 三宅由子: よくわかる医療・看護のための統計学入門, メディカ出版, 大阪, 1991.
- 5) 駒澤勉, 高木廣文, 佐藤俊哉: ヘルスサイエンスのための統計科学, 医歯薬出版, 東京, 1996.