

推定と検定について

高木 廣文* 新田 裕史** 中井 里史***

はじめに

これまでの連載で、統計学でよく用いられる用語やその概念、基本的な統計量と分布に関して説明を行った。その内容は、主に記述的な方法と呼ばれるものが中心であった。

実際に臨床研究を行う場合には、いわゆる“推測統計学”と呼ばれる方法が、きわめて重要になってくる。推測統計学とは、標本データから母集団に関する各種の特性について、推論を行う方法のことである。すなわち、研究者自身が持っているデータから、より一般的な結論を導き出す統計学的方法といえる。

ここでは、その基礎となる“推定”と“検定”についての説明を行う。

I. 推定とは何か

10例の標本から収縮期血圧のデータを取り、その平均値が 130 mmHg だとしても、それがそのまま母集団の平均（母平均）と一致するわけではない。図1で示したように、標本の取り方は無数にあるのだから、標本の取り方が間違っただけであれば、もはや母集団については何もいえないことになる。また、標本が正しく選ばれた場合、すなわち無作為抽出された場合であっても、やはり標本の取り方は無数にあり、平均値が 130 mmHg のこともあれば、135 mmHg のこともあるだろ

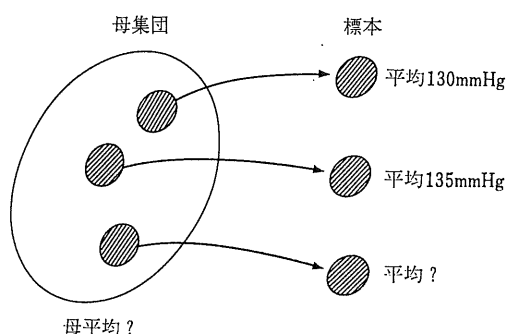


図1 母集団と標本の関係

う。それでは、どうすれば標本のデータから、母集団の平均値，“母平均値”を求めることができるのだろうか。

標本が正しく選ばれた場合、推測統計学の手法により、標本データから母集団についての推定を行うことができる。統計学で用いる“推定”という用語は、少数の標本から得られた限られたデータから、母集団での各種特性を求めるという普遍化、一般化の方法といえる。

統計学では、推定の方法は大きく2分類して考えることができる。すなわち、“点推定(point estimation)”と“区間推定(interval estimation)”に分類される。

1. 点推定とは

“点推定”とは、母平均値などを一つの数値を用いて推定するものである。例えば、全国の新生児科医 100 人に対して無作為調査を行い、治療経験のある児の最低体重を調べたとしよう。その結果、100 人の医師の回答した平均値は 450 g であったとしよう。この結果を一般化して、“新生児

* 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

*** 東京大学医学部保健学科学疫学

科医が治療経験した児の平均最低体重の母平均値は450gである”というような場合が、母平均値の点推定にあたる。よく用いられる平均値や比率に関して、“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比率は、母集団の母平均値と母比率の点推定値となる”ということが知られている。

2. 区間推定と信頼区間

ところで、点推定は一つの数値しか用いないため、必ずしも母集団の母数（母平均値や母比率など）と一致するものではない。そこで、適当な幅の区間を設定し、その範囲内に母平均値などの母数が存在することを示す方法も考えられる。そのような方法が“区間推定”と呼ばれるものである。

問題は、どのようにその幅を設定するかである。大きな区間をとれば、必ず母数が入るはずであるが、不必要に大きくとっては実用的ではない。あまり小さくしても、母数が入らないこともありえる。そこで、統計的に信頼される程度を考え、その中に母数が存在すると推定される区間を設定すればよい。そのような区間は“信頼区間(confidence interval)”と呼ばれる。さらに、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。ここで、“95%”という数値の示す意味は、同一の方法で100回調査を行い、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表現している。逆にいえば、5%は間違える可能性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”などを求めればよい。

先の例では、“新生児科医が治療経験した児の平均最低体重の母平均値の95%信頼区間は420～480gである”といった表現により、母平均値の存在する区間を推定することになる。

II. 検定とは何か

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、推定と異なり“帰無仮説(null hypothesis)”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後述するように、研究上の作業仮説とは異なるものである。

1. 統計学的仮説検定と帰無仮説

例として、妊婦の喫煙に関して、“喫煙者は非喫煙者に比べて、未熟児出産率が高率である”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、妊婦を喫煙群と非喫煙群に分け、各グループでの未熟児の割合を調べることになる。調査データでは確かに、喫煙群で未熟児の割合が高率であったとしよう。しかし、二つしかグループがなければ、完全に二つの割合が一致する場合よりも、偶然の影響により、どちらかが高率となる可能性の方がありそうである。

統計学的な仮説検定法では、“喫煙者と非喫煙者の未熟児出産の母比率（母集団での割合）は等しい”という仮説を立てる。

次に、この仮説が正しい場合に、調査データで認められるような2グループ間の比率の差が偶然生じる可能性（確率）を計算する。これは“有意確率P”と呼ばれることがある。実際の計算の方法は、どのような仮説を設定するかに依存している。一般的な手順としては、ある統計量をデータから計算し、その統計量が従う分布（t分布、 χ^2 分布、F分布など）から、二つのグループで観察された以上の差が偶然に起きる確率を求めることになる。

もしも、計算した有意確率Pが大きければ、二つのグループでの差は偶然によるものともいえることになり、それ以上のことは何もいえなくなる。逆に、確率が小さければ、そのように稀なことが、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、そもそも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(reject)”という）、二つのグループでの母比率には差があるといえることになる。

こ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し、そのために“帰無仮説”と呼ばれるのである。かりに仮説が棄却されなければ、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

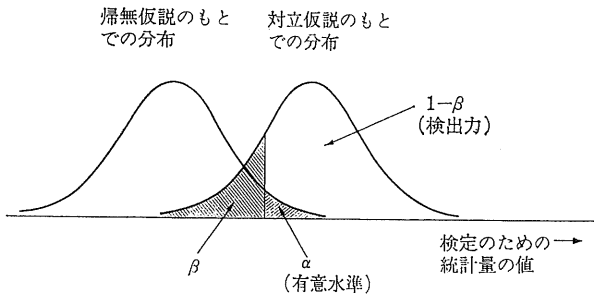


図2 有意水準と検出力

支持する根拠を与えたわけではなく、そのような状況が偶然に起こり得るということを示したに過ぎないからである。

ところで、帰無仮説を棄却するか否かという、検定の目安となる基準は“有意水準(significance level)”と呼ばれ、慣習的に5%か1%がよく用いられている。したがって、有意確率 P が5%以下ならば、通常は“有意差が認められた”といった表現で、検定の結果を示すことになる。しかし、これ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。この点から、“5%水準で有意差が認められた”といった検定結果の提示方法は、あまり適切なものではない。可能ならば、“有意確率は3.5%であり、有意差があると考えられる”といった形式を用いて、正確な有意確率を提示することが望まれる。有意水準は確定的なものではないことに注意して欲しい。

2. 検定における二つの誤り

有意水準はすでに述べたように、“仮説が正しい場合に、誤って仮説を棄却する誤り”の判断基準となる確率を示すものであり、“第1種の誤り(過誤)”とも呼ばれる。逆に、“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)”も存在し、“第2種の誤り(過誤)”と呼ばれている。

ところで帰無仮説を“喫煙者と非喫煙者の未熟児出産の母比率は等しい”とした場合、“喫煙者の未熟児出産の母比率は、非喫煙者の母比率より大きい”という仮説も考えられる。後者は帰無仮

説に対して“対立仮説”と呼ばれるものである。図2は帰無仮説と対立仮説の関係を図示したものである。

図2で、第1種の誤りは α (有意水準)、第2種の誤りは β という記号で表わされている。有意水準 α は一般には、すでに述べたように5%か1%を用いることが多い。ところで、図からわかるように、帰無仮説のもとでの統計量の分布と、対立仮説のもとでの分布がそれほど離れていないと、第2種の誤りは大きくなる。言い換えると、2グループの母比率などにわずかの差しかない場合には、帰無仮説を棄却することが難しくなる。この点から、 $[1-\beta]$ は“仮説が誤っている場合に、仮説を棄却する”確率を示すもので、この値が小さいと検定の結果はあまり信頼できないことになるだろう。仮説検定では $[1-\beta]$ は“検出力(power of a test)”とも呼ばれ、検出力が大きいほど検定の効率がよいことになる。

3. 片側検定と両側検定

前節では対立仮説として“喫煙者の未熟児出産の母比率は、非喫煙者の母比率より大きい”(対立仮説1)を考えたが、“喫煙者の未熟児出産の母比率は、非喫煙者の母比率より小さい”(対立仮説2)という仮説を考えることもできる。

図2の状況は、帰無仮説と対立仮説1に関する状況を示すものと考えてよい。このような場合、有意水準 α を用いて行う検定は“片側検定”と呼ばれる。帰無仮説に対立仮説2を対応させて考える場合も、上記と同様であり、やはり片側検定になることは、容易に理解できるだろう。

実際問題では、二つの対立仮説のうちどちらが適切なものか、事前に判明していないことが多いだろう。調査データの結果をみて、自分の判断でどちらかの対立仮説を採用するのは“後知恵”と呼ばれ、絶対に行ってはならない方法の一つである。

それでは、どのようにすればよいのだろうか。通常は、帰無仮説を否定した、“喫煙者と非喫煙者の未熟児出産の母比率は等しくない(差がある)”という仮説を対立仮説にすればよい。すぐにわかるように、この対立仮説は上記の対立仮説

1と2を合わせたものである。したがって、有意水準として帰無仮説に基づく分布の片側の部分を考えるのではなく、両側の部分を考えることになる。この意味から“両側検定”と呼ばれている。図2のような場合には、有意確率 P を2倍した値を、新たに有意確率 P として、その値が5%以下か否かで検定を行うことが多い。ただし、この場合でも、正確な有意確率を提示する方が、より望ましいのはいうまでもないだろう。

実際にどのような対立仮説を考えるかは、その分野での既知の情報がどれだけ使用可能かによ

る。図2のような対立仮説の分布を求めることも、二つの母比率の差がどの程度あるかが、あらかじめわかっているか、仮定できなければならないので、見た目ほど単純な計算によるものではない。したがって、事前情報がないかぎり、検定には両側検定を用いるべきだろう。

なお、ここでは検定の例として2グループの母比率の差の検定を考えたが、実際には多種多様な検定方法があり、次回以降に使用上の注意なども含めて、順に解説することにする。