EBMを見抜くための統計学キーワード
逆説のｐ値

高木　廣文
新潟大学医学部保健学科　教授

１．ｐ値とは何か？

データを集めて統計学的分析を行うと、いろいろな計算結果が統計ソフトから出力される。その中で最も気になる数値は、なんといっても「ｐ値」であろう。もしもｐ値が0.04ならば、５％より小さいので「統計学的に有意」となり、大喜びである。しかし、ｐ値が0.06ならば６％となり、「有意でない」ので地団駄踏んで悔しがることだろう。わずか0.01、１％の違いで、天国と地獄ほども差がついてしまう。

このように、統計学の中にあって重要な決定権をもつｐ値であるが、そもそもｐ値とは何を表しているのだろうか。また、何で５％がそんなに重要なのだろうか。

ｐ値とは、そもそも″ｐ-value″という英語を、ｐはそのままで、″value″を「値」とした、変な訳語である。それでは、ｐは何かというと″probability″のｐであり、「確率」ということになる。したがって、そのまま全てを訳すのならば「確率値」ということになるだろう。しかし、確率値という訳はほとんど使われていないようである。

ｐ値という用語が、これほど多用されるようになる前は、もともと「危険率」とか「有意確率」という統計学用語が使用されていたのである（今でも使用されているのだが）。

従来から、統計学的検定では、ｚ値、ｔ値、Ｆ値などという統計量を使用していたので、同様な呼称として、ｐ値という用語が用いられるようになったのではなかろうか。いずれにせよ、ｐ値は有意確率と同一の意味であり、このことから明らかなように、統計学的検定の有意性を示すための確率値である。今回は、統計学的仮説検定とこのｐ値について、間単に説明しよう。

２．統計学的仮説検定法の考え方

調査研究などで統計学を使用する目的は、収集したデータから得られた調査結果を普遍化するためであろう。すなわち、統計学を用いれば、母集団のほんの一部である標本から得られたデータを分析することで、その結果を母集団全体の特性として、一般化して考えることができる。統計学的推定や検定は、標本から母集団への一般化のために存在するのである。

統計学的推定の考え方は、日常的な意味での「推定」と大差ないので、理解が難しいということはない。しかし、統計学的検定の考え方は、かなり普通とは異なっている。

統計学的検定も推定と同様に，標本データを用いて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。ただし，推定と異なる点は，「帰無仮説 null hypothesis」と呼ばれる母集団についての仮説を用いることである。

すなわち、ある帰無仮説が正しいと仮定した場合、実際に問題となっている特性値が、偶然性のみから標本データとして得られる確率を計算し、その仮説が本当に成り立つのかどうかを調べるという手順をとる。

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研究を行ったとする。

健常者を対象にして、喫煙習慣と血管拡張率の計測値を調べたとしよう。分析の結果，調査データでは確かに，喫煙者グループの血管拡張率が、非喫煙者グループの平均値に比べて低値であったとしよう。

このような場合、研究前に考えていた作業仮説通りだからといって、仮説が証明されたと考えるのは早計である。

なぜならば、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，どちらかで数値が大きく，どちらかで小さくなるのが普通だからである。したがって、観察された結果が、思った通りだからといって、それでよしとするわけには行かない。

この偶然性を評価するのが、統計学的仮説検定法である。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」という仮説を立てる。

次に帰無仮説が正しいと仮定した場合，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、偶然生じる確率を計算する。実は、ｐ値とはこの偶然性についての確率のことである。

求めた確率（ｐ値）が大きい場合，母集団では差がなくても、偶然によって観察された２グループ間の差は観測できることになる。したがって、データでは２グループ間に平均値の差があるように見えても、母集団で本当に差があるとはいえないことになる。

逆に，ｐ値が小さい場合，偶然そのような結果が観察されるのは稀なことになる。したがって、自分の行った研究でたまたま稀な現象が観察されたと考えるのは無理がある。それでは何がおかしいかというと、求めたｐ値が小さすぎる点である。そうすると、そのｐ値の計算のもとになった帰無仮説がおかしいということになるだろう。すなわち、そもそもの前提となった帰無仮説が誤っていたと考える必要がある。

この結果，帰無仮説は否定され（「棄却する reject」という），「２グループの母平均値の差は０ではない」ことになる。すなわち、喫煙者グループと非喫煙グループでの、血管拡張率の母平均値には差（有意差）があることになる。

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。もしも、求めたｐ値が大きく、仮説が棄却されなければ，帰無仮説が証明されたのではなく、統計学検定は意味がなくなってしまう点にも注意が必要である。何故ならば，帰無仮説が棄却できないということは，その仮説を積極的に支持する根拠を与えた訳ではなく，偶然によってそのような状況が観察されることもありえるということを，単に示すに過ぎないからである。

統計学的検定で帰無仮説を棄却する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。しかし、この基準は絶対的なものではない。大体なぜ５％なのかというのは、根拠のある数値ではない。単に我々がまあ小さい数値であると感じているだけである。手足20本の指のうちの僅か１本程度の割合ならば小さいに違いない、という感覚である。したがって、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%を採用してもよい。実際には、慣用的に５％が使われているので、これに反対するのはなかなか困難であるが、状況によって適宜判断を自分自身で下してもよいものである。もっとも、そのような場合には、他人が見てある程度は納得する値でなくてはならないので、それほど極端に甘い基準（例えば20％）や、厳しい基準（例えば、0.01％）などはなかなか採用できないだろう。

現在では，検定結果を示す場合，以前よく見かけたように＊印などを単に付けるのではなく，帰無仮説が正しい場合の偶然性の確率であるｐ値の数値を示すのが一般的である。

３．ｐ値のもつ意味と標本数

ここまでで明らかになったように、ｐ値が示しているのは、帰無仮説が正しい場合に得られるデータに関する偶然性の確率の大きさである。したがって、ｐ値が小さければ、確かに帰無仮説を棄却でき、「統計学的に有意」であることを言うことができるが、その他の情報に関しては何も言うことができない点には注意が必要である。

２グループの比較では、ある変数についての帰無仮説は「２グループの母平均値は等しい」であり、その否定は「２グループの母平均値は等しくない」ということである。また、食塩の摂取量と血圧のような相関関係についてであれば、帰無仮説は「２変数間の母相関係数は０」であり、その否定は「２変数間の母相関係数は０でない」ということになる。

したがって、ｐ値が小さいからといって、２グループ間の母平均値の差が大きいとか、母相関係数が大きいとかを示している訳ではない。

この種の誤解は、かなりよく見かけるものである。しかし、ｐ値自体は偶然性を示す確率であり、単に帰無仮説を否定する場合に、その決定がどれくらい妥当かという価値判断に関係するだけである。すなわち、５％よりは１％、１％よりは0.1%、0.1%よりは0.01%の方が、一般には偶然による確率が低いのだから、帰無仮説を棄却してもよいだろうと言えるだけである。

実は重要な点として、統計学的検定において、帰無仮説の棄却のためのｐ値は、標本数と大きく関係していることがあげられる。

例えば、健常者での食塩摂取量と血圧の相関が0.1である場合、よく使われる「無相関の検定」を行うと、100例の場合、ｐ値は0.322となり、５％よりもずっと大きく、有意ではない。しかし、1000例の場合には、ｐ値は0.0015となり、１％水準でも有意である。更に言えば、かりに相関係数が0.01であっても、40000例あればｐ値は0.045となり、５％水準で有意となる。

ここでは特に数式等は示さないが、現在では、様々な解析用のソフトがあるので簡単に調べられる。相関係数が0.01であるというのは、確かに０ではない。したがって、帰無仮説「母相関係数は０」は標本数が多ければ棄却することができる。しかし、このような場合、統計学的有意性にどのような実質的な意味があるのだろうか。

４．やはりｐ値か？

実質的な２グループ間での差や相関係数の大きさという点は、考慮すべき重要な問題である。ｐ値だけでは、帰無仮説を棄却できるか否かの情報しか得られないことに注意すべきである。

しかしながら、一方では、とにかく統計的に有意にしたい場合には、標本数を大きくすればよいという戦略も考えられることになる。実際、ある研究で必要とされる標本数とは、ｐ値が５％より小さくなるような標本数のことである。したがって、この種の考え方自体は、すでに研究デザインを考える場合の標本数の設定として、一般に用いられている。

現実問題として、ｐ値による統計学的検定が単に帰無仮説の採択／棄却のみに関係するとすれば、検定以外に何を行えばよいのだろうか。

それは、統計学のもう一つの方法論である「推定」を行うことである。かりに相関係数に関して情報が必要なのであれば、「母相関係数の95%信頼区間」などを求めればよい。信頼区間を求めることにより、実際的な母集団での特性値の範囲を知ることができる。差がないかどうか、とか相関が０かどうかというよりも、ずっと実質的に有用な情報を入手できるはずである。

単純ではあるが、統計学では検定と推定が両輪となって、多くの研究を支えているのである。