

IV. 有意性の検定

IV-1. 検定の作業

有意性・有意差の検定にはさまざまなものがありますが、検定しようとするデータのタイプの違いと検定の目的の違いが、様々な統計検定が生まれる理由です。すべての検定を説明することは不可能ですが、データや分析の目的によって、分析方法をタイプ分けしておくことは、適切に検定法を選択するために必要でしょう。よくあるのは、データについて、パラメトリックなデータか、ノンパラメトリックなデータかという分け方です。しかし、この考え方は多少観念的です。正直に告白すると、筆者にはパラメトリック、ノンパラメトリックという分け方の境界が何なのかよくわかりません。計量的 (quantitative) なデータはパラメトリック、形質的 (qualitative) なデータだとノンパラメトリックと完全に二分的に分けることはできません。確かに、性別とか、国籍とか名義的変数で表現されるデータもありますが、多くのデータは、計量的な性質と形質的な性質を両方もっています。このことについては、データとして表現するために使われる尺度の水準の問題として、対応分析のところでもう一度改めて触れます。性別のような名義変数も、男性は男性性1、女性性0、女性は女性性1 男性性0と表現すれば、数値的に表現できます。そんなものは定量性がないということではありますが、無理に理屈をこねれば、男性度1の物を男性、男性度0を女性としたと言い訳することはできないわけでもありません。パラメトリックな統計解析とは、データに母集団分布が確率的に分布していることが仮定できて、その分布を決める母数を求めることによって、確率が計算できると考えられ場合に用いる検定法です。ノンパラメトリックとは、そのような分布をあらかじめ仮定できない場合に用います。少し、実践的に考えると、少数例の実験データのように、等分散性が確認できない場合には、パラメトリックなデータでも、ノンパラメトリックの分析方法を考えるとというやり方もあります。カイ二乗検定は、期待値からの差の分布であるカイ二乗分布を前提にしていますから、パラメトリックな分析なのですが、名義変数で表現されたグループ間の違いの検定に使われます。F検定や student の t 検定などは、そのままでは形質的なデータに使えませんが、例えば、アンケートなどの選択肢などが、程度を表す順序データであれば、データの分布の形を確かめて、単峰形に分布するのであれば、そのままでも、あるいは、プロビット変換やロジット変換あるいは何らかの関数を使って変換を行って、正規性と等分散性を作り出してパラメトリックな手法で分析することも考えられます。著者が若かったころには、コンピュータを使った統計解析のソフトウェアはなかったので、様々な変形をして、無理やり F 検定や t 検定をしました。おそらく今は、そういう無理なことをしないで、ノンパラメトリックな分析法のソフトウェアを使って統計解析が行われているものと思います。

大切なことは分布が正規的で標本集団間に等分散性があるかないかを確認することです。厳密に正規分布して等分散性が保証されているデータなど現実にはめったにないと言われ

れば、確かにその通りなのですが、ここでは、正規性や等分散性が明瞭に否定されていないければよいぐらいにゆるく考えてください。データが得られたら、まず、記述統計的にそれぞれの集団ごとに平均値と分散を求め、全体としてデータの頻度分布を描いてみるべきです。それによって、データについて様々な情報が得られます。習慣化しておきましょう。場合によっては、データの数少なすぎて等分散性や分布の形がわからない場合もあります。そういう場合は、データが持っている性格や過去の知見を参考に分析方法を選択します。もし、それが有効で必要であれば、なんらかの関数でデータを変形して正規化して t 検定 F 検定などのパラメトリックな分析をすることもあり得ます。一般的に t 検定は一对の集団の平均値の差の有意性に関する検定に用いられます。分子の自由度を 1 として 2 つのグループ間の差の F 検定で有意差を検定することもできますが、2 グループの差の有意性の検定は t 検定で行うことが一般的です。それは t 検定の方がわかりやすいからです。3 つ以上のグループ間の差の検定は F 検定で行います。3 つ以上のグループの場合、その中から 2 つずつ選んで、t 検定を行うというのも考えられます。しかし、初めからそれを行うことは薦められません。統計検定を繰り返して行くと、統計的なエラーの可能性が増加するからです。たとえば、一つの組み合わせについて、それが同じであるという帰無仮説が正しい可能性が $19/20$ のであれば、そうでないことが起こる可能性は $p = 0.05$ です。つまり、そのようなことが起これば、帰無仮説を否定して、それは同じではないと判断します。しかし、これを 2 回繰り返すと、2 つの組み合わせとも帰無仮説が正しいという可能性は $\left(\frac{19}{20}\right)^2 = 0.9025$ になるので、二のうちどれかで帰無仮説が正しくないという確率は $p = 1 - 0.9025 = 0.0975$ となります。いくら固い馬券でも続けて勝負して勝ち続けることはできないということと同じです。もし、5 つのグループがあると、そこから 2 つ取り出す組み合わせの総数は

$${}_{10}C_2 = 10 \times 9 / 2 = 45$$

ですので、それらの組み合わせのすべてで差がないという可能性は

$$\left(\frac{19}{20}\right)^{10} = 0.598734$$

となります。もともと、帰無仮説（すべてのグループ間で差がない）を否定できる可能性が 40% ぐらいしかないのです。あまり意味のある検定とは言えません。その場合、もともと有意差のレベルを $p = 0.005$ 程度に厳しく設定することも考えられます。こういう補正をボンフェローニ補正と言います。そのような補正をして計算することが適切かどうかはわかりません。すべての帰無仮説が否定されなければ全体として差がないと判断できないというのは、あまりに極端な立場でしょう。この場合、まず F 検定でグループ間の平均的な分散がランダムな分散や組み合わせによる分散（交互作用）に比べて、十分に大きいかどうかを議論すべきでしょう。個別の組み合わせについての検定は、そのあとどこに差があるのかを調べるものと考えて良いでしょう。

下図にパラメトリックな分析について、分析法の選択のフローチャートを示しました。

