

仮説検定の限界を乗り越えるためのベイズ統計学

岡田謙介（専修大学）

1. 仮説検定とその限界

t 検定や F 検定のような統計的仮説検定（statistical hypothesis testing）は、Fisher および Neyman, Pearson らが確立した、データに基づいて仮説を評価するための体系的な方法論である。

たとえば 3 群の母平均 μ_1, μ_2, μ_3 について検討する仮説検定では、次のような帰無仮説 H_0 と対立仮説 H_1 を設定する。

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_1: \text{not } H_0$$

ここで、 H_0 のもとで既知の分布にしたがう検定統計量を考え、その実現値をデータから算出する。そして、 H_0 が真である場合にめったに（確率 α 以下でしか）生じないような極端な実現値が得られたならば、 H_0 を棄却し、 H_1 を採択する。これが仮説検定の基本的な考え方である。

仮説検定はもっとも広く普及した統計学的な仮説評価の方法と言って過言ではない。科学やビジネスの現場において、仮説検定が広く活用されていることは論を俟たない。しかしながら、その役割は本来の含意を大きく超えてしまっている現状があり、研究者の間では仮説検定への過度な依存に対する危機感が高まっている。

仮説検定への代表的な批判のうちいくつかを紹介する。多くの応用場面において、検定によって分析者が主張したいことは H_1 である。しかしながら、仮説検定では H_1 を評価することはせず、 H_1 が真である状況について考えることもしない。行うのは H_0 が真であると仮定した場合の検定統計量についての検討であり、 H_0 が棄却されることをもって H_1 が支持されるという、持って回った意思決定となる。このようなロジックのため、仮説検定では H_0 を棄却することはできても、積極的に支持することはできない。さらに、 H_0 は各母数が完全に等しいといった極端な仮説であり、現実場面において厳密に正しいことはまずありえない。それにもかかわらず、仮説検定では一般に標本サイズを大きくすれば検定力が大きくなり、 H_0 が棄却されやすくなる。しかし、標本サイズに頼ってだされた有意差が、果たして現実的に意味のある差なのかは疑わしい。

2. ベイズ統計学

ベイズ統計学と伝統的・頻度論的な統計学の違いは、データと母数のどちらを定数とみなし、どちらを確率変数とみなすかにある。頻度論においては、母数は定数であり、データは母集団からランダムに抽出された確率変数であると考えられる。一方、ベイズ統計学においては、データはすでに分析者の手元にあるので定数であり、母数は未知で不確実であるから確率変数と考える。ここが両者の本質的な違いであり、ほかにしばしば指摘される違

いはここから導かれるものと考えると理解しやすい。

H_0 や H_1 といった仮説は母数に対する仮説である。その母数は頻度論では定数、ベイズでは確率変数である。したがって、頻度論に基づく限り、「 H_1 が真である確率」を考えることはできないが、ベイズではそれができる。一般にデータを D 、仮説を H で表すとき、頻度論的な仮説検定では $P(D|H)$ を考えるが、ベイズ統計学の枠組みでは $P(H|D)$ を考えることができるのである。本来評価を行いたいのは仮説 H の適切さであるから、その確率 $P(H|D)$ を求められるベイズ統計学の枠組みは魅力的である。

3. 情報仮説の評価

仮説検定の限界をためのひとつの方法として、情報仮説 (informative hypothesis) の評価と呼ばれる方法論が提案されている。この枠組みでは、たとえば次のような2つの仮説を扱う。

$$H_i: \mu_1 < \mu_2 < \mu_3$$

$$H_a: \mu_1, \mu_2, \mu_3$$

H_i が情報仮説と呼ばれるものであり、ここでは1群の母平均 μ_1 よりも2群の母平均 μ_2 の方が大きく、さらにそれよりも3群の母平均 μ_3 が大きいという分析者の仮説を表している。情報仮説では、このように分析者の仮説を等号および不等号を用いて明示的に表現する。一方、 H_a は母数に制約をおかない仮説であり、対立仮説と意味的には同じである。

ここで H_a と比較した H_i の相対的なよさを表すベイズファクターを求めることにより、分析者の持っている仮説を定量的に評価することができる。 H_i と H_a とを比較するベイズファクターとは、その2つの仮説のもとでの周辺尤度の比である。これが、情報仮説の評価の基本的な考え方である。なお H_i として帰無仮説に相当する仮説を評価することもできる。

こうした方法は、仮説検定と比較していくつか理論的に有利な点がある。まず、情報仮説の評価では無制約の仮説 H_a と比較した分析者の仮説 H_i の相対的なよさを問題にしており、2つの仮説を「平等に」扱っている。この点で、帰無仮説 H_0 のもとでの検定統計量の挙動のみを問題とし、対立仮説 H_1 が真の状況は考えない仮説検定よりも優れていると考えられる。また、ベイズファクターでモデルを評価することにより、データの仮説を支持する証拠の大きさを定量化し、それに基づいた判断を下すことができる。

本セミナーではこのベイズ統計学の考え方をを用いた情報仮説の評価というアプローチについて、考え方や実際の分析方法、ソフトウェア、その限界、関連するアプローチなどを紹介し論じる。

参考文献

Hojtink, H. (2011). Informative Hypotheses: Theory and practice for behavioral and social scientists. Champan & Hall/CRC.

大久保・岡田 (2012). 伝えるための心理統計：効果量・信頼区間・検定力. 勁草書房.