

# メタ分析 p 値とその挙動について

鹿児島大学 理工学研究科 杉本知之

## はじめに

p 値は有用な統計的測度であるが、一般に誤用され誤解されている。このことが p 値の利用を非推奨とする科学雑誌、科学者、統計家を生み出してきたが、そこでの議論は p 値が最初に紹介されたときから本質的に変わっていない(Wasserstern & Lazor, 2016)。統計的な差の有意性を評価することにおいて、p 値に完全にとって代わることができる指標を見つけることは困難であるが、ほとんどの場面で、何らかの代替指標、もしくは併用指標を用いることは、確かに重要である。ただし、このような他の指標を検討する目的のために、統計が用いられる場面の特徴の中で、p 値がどのような挙動をするのかを十分把握しておくことも必要である。簡単のため、よくデザインされ比較可能性が達成された状況を考える。ここで得られたデータに対し、連続な検定統計量に基づく p 値は、標本サイズに関係なく、帰無仮説のもとで(0,1)上の一様分布に従うが、対立仮説のもとではそうではない。対立仮説のもとでの p 値の分布は、一般に複雑であるが、標本サイズや対立仮説のパラメータによって指定される関数によって表現される(Hung et al., 1997)。さらに、メタ分析手法を用いて、複数の研究の結果をまとめることを考えれば、ここで得られる p 値の分布は、さらに複雑になる。例えば、ある標本サイズ  $N$  をもつ単一試験(研究)での p 値の分布と、複数の試験結果(研究)を集めてメタ分析手法に基づく p 値の分布は、統合した標本サイズが同じ  $N$  であったとしても、一般に異なるであろう。因みに、有意水準と検出力をある一定の水準に保ってデザインする場合では単一試験よりも複数の試験の方が必要標本サイズは大きくなる(上坂, 2003)。一方で、単一試験から得られる p 値と、その試験をいくつかの標本にランダムに分割し再びそれらを統合して得られるメタ分析 p 値は、試験間変動もないため、互いにほとんど同じ挙動を示すという興味深い特徴がある。

本報告では、p 値を巡る最近の議論を契機として、メタ分析 p 値の挙動に焦点をあてて、その挙動が、標本サイズや対立仮説のパラメータの水準に加えて、統合する研究の個数、それらの試験間変動の大きさの違いで、どのように変化するかを調べる。それらの結果に基づいて p 値の特徴、意義、その代替などについて議論する。

## 参考文献

- Wasserstern, R. L. & Lazor, N. A. (2016). The ASA's statement on p-values: context, process and purpose. *The American Statistician* **70**(2), 129-133.
- Hung, H.M.J., O'Neill, R.T. Peter, B. & Kohne, K. (1997). The behavior of the P-value when the alternative hypothesis is true. *Biometrics* **53**(1), 11-22.
- 上坂浩之(2003). 検証試験における有意水準と検証試験の個数について. *計量生物学* **24**, Special Issue, 10-16.