

超心理学研究の最終結論

・ [Daryl Bem Proved ESP Is Real](#) Which means science is broken. MAY 17, 2017 , By Daniel Engber , Slate

タイトルを直訳すると、「ダリル・ベムが ESP が本当だったと証明 - つまり科学が壊れているということ -」みたいな感じ。ただし、ESP が実在したという結論ではない。ベムの実験手法は科学的に正しいとされていたのだが、その方法そのものが間違っていたと言うこと。つまり、科学（統計的評価法）は壊れていたなので修正を余儀なくされている。当然超心理学も科学を前提としているのなら、根底から見直されなければいけない。これまでの手法はもはや使えないので、その意味で今までの超心理学は終了なのである。

この記事を一言で要約すると、「とんでもない結果が出てしまったとき、その結果に飛びつくのではなく、その実験方法に間違いがないか、きちんと確かめるべき」という基本的な考えを述べているだけ（これが意外と非常に難しい...）。たとえ、それが標準的な科学的実験方法だったとしても。おそらくこれで超心理学はとどめを刺されるだろう。しかし、諦めの悪い連中はどこにでもいるので、研究はまだまだ続くと考えられる。超心理学研究はもう 30 年くらい前からゾンビ化していると思っておいたほうがいい。

結局、ベムら超心理学者がこぞって多用した統計的な手法に問題があった。懐疑論者がず～～～っと指摘してきたように、統計的有意性だけでは意味がないということがようやく認知されてきたのである。これまでも、統計評価を多用する分野の研究（たとえば心理学）は再現性が低い（再現性の危機）ことが指摘されており、バイアスの影響が懸念されていた。事ここに及んで、権威ある心理学者であるベムが、荒唐無稽な超能力の実験をすると、無視できない統計的有意性が検出され、論文が権威ある学術誌に受理されてしまった。その結果、やっと多くの専門家が、「今までの標準的な統計的手法になにかおかしい点があるのではないか？」と気付いたということである。これはひとつのパラダイムシフトだと思う。p 値検定のような統計的評価は所詮人間が決めた基準である。きちんと実際の効果の大きさを測量して、その原因を論理的に説明できたとき、初めてそれは科学となる。

なぜ統計的有意性にバイアスが容易に入り込むかということ、「後付けの理由」の効果が大きい。つまり、実験終了後に実験データをどう評価するか決めると、恣意性が入り込みやすい（p-ハッキング）ということ。ところが、心理学分野では、こうした後付けの評価が 2010 年頃まで普通に行われていたというのである。超心理学は、統計を多用する心理学の弱点を巧みに突いて、その主張を正当化してきたのである。

今後は、実験を行う前に、どういう実験を行い、それをどう解析するか、宣言しておいてから実験を行う「open science」が主流の時代が来るかも知れない。

この結論がもたらしたもう一つの衝撃は、レイ・ハイマンとチャールズ・ホノートンが 1986 年に発表した共同コミュニケが、その後、30 年近くにわたりまったく遵守されていなかったということである。要するに、強制力のない規律は無意味だったのである。「人間の良心」に依拠するシステムは簡単に崩壊するということ。

これまでの研究方法が間違っていたとしても、それを修正すれば心理学の研究は続けられる。しかし、超心理学は超能力がなければ成立しないので、今後は非常に苦しい状況となるであろう。

そもそも統計学とは、数が限られたサンプルから全体のだいたいの傾向を推測する手法であり、ごく微弱な効果が本当にあるかどうか判定できるような手法ではない。

以下の記事も参照。

- ・ [P-Hacher Confessions: Daryl Bem and Me](#) Stuart Vyse, June 13, 2017, Skeptical Inquirer, 25-27

これは、当時の心理学研究者はベムも含めた皆が当たり前のように p-ハッキングを行っていたという告白の記事である。

この記事の出版を受けてスーザン・ブラックモアが、ダリル・ベムに対してさらに批判的な記事を書いている。

- ・ [Daryl Bem and Psi in the Ganzfield](#) Susan Blackmore, Skeptical Inquirer Volume 42.1, January/February 2018

p 値の誤用 (p 値だけで判断してはいけない)

超心理学に限らず、一般に p 値の誤用や誤解釈が多すぎるので、アメリカ統計学会 ASA は 2016 年に以下のような声明を発表することとなった。

- ・ "[The ASA's Statement on p-Values: Context, Process, and Purpose](#)", Ronald L. Wasserstein & Nicole A. Lazar, Pages 129-133, Accepted author version posted online: 07 Mar 2016, Published online: 09 Jun 2016

ここでは、p 値の使用について、以下のような原則を挙げている。

- 1 . データが指定された統計モデルとどの程度相容れないか、p 値で示すことができる。
- 2 . 研究されている仮説が真実なのか、データがランダムチャンスのみによって生じたのか、その可能性を p 値で計ることはできない。
- 3 . p 値が特定の閾値を越えたかどうかだけに基づいて、科学的な結論、ビジネスまたは政策上の決定を下してはいけない。
- 4 . 適正な推論は、完全な報告と透明性を必要とする。
- 5 . p 値または統計的有意性だけでは、効果の大きさまたは結果の重要性を測ることはできない。
- 6 . p 値だけでは、モデルや仮説について、良い程度の証拠を提供しない。

これは今さらながらごく当たり前のことを言っているだけなのだが、要するに、p 値だけでは、超能力があるかどうかなど判断できないし、その効果の大きさを測ることもできないということだ

- [False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant](#) (pdf) Joseph P. Simmons, Leif D. Nelson, and Uri Simonsohn, Psychological Science, 22(11) 1359-1366

以下の文献では、2千人以上の心理学者を調査したところ、半数以上が問題のある研究手法を実践していたという結論になっている。

- [Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling](#) Leslie K. John, George Loewenstein, and Drazen Prelec, Psychological Science, 23(5) 524-532

p 値判定の問題は、例えば以下のリンクを参照。

- [72人の統計学者が連名で p 値の閾値を 0.05 から 0.005 にしようと言っている。](#) 2017年8月6日(日)、Fallen Physicist, Rising Engineer
- [Not Even Scientists Can Easily Explain P-values](#) By Christie Aschwanden, FiveThirtyEight

しかし、p 値の閾値を変えたところで根本的な解決にはならないだろう。バイアスを排除する実験設計が重要。p 値が特定の閾値を越えたかどうかだけに基づいて、科学的な結論を下してはいけない。

- [Scientists rise up against statistical significance](#) Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects., Nature 567, 305-307 (2019)

2019年3月には、「統計学的有意性」という概念自体を放棄しようというコメント(800人以上の科学者が署名)がネイチャー誌上で発表された。