

こころの 測り方

サンプルサイズの決め方

専修大学人間科学部心理学科 教授

大久保街亜（おおくほ まちあ）

参加者や被験体の数、つまりサンプルサイズを皆さんはどのように決めていますか？ サンプルサイズは、実証科学である心理学に必ずついて回る問題です。最近、重要性が再認識されています。実際、サンプルサイズを事前に正確に決めることは、研究の信頼性、妥当性、そして再現可能性を担保するために欠かせません。そして、サンプルサイズを事前に決めず、なんとなくデータを取り足す、あるいは、途中で取るのをやめることは、研究における不正行為として認識されるようになってきました。この認識がないなら、あなたは知らないうちに不正を行っているかもしれません。

メンデルの法則

サンプルサイズと不正行為についてある逸話があります。メンデルの法則は、中学校の理科で教えられる有名な法則です。メンデルは修道院で日々エンドウをまき、膨大なデータを集めてこの法則を導きました。ご存知の方もいるかもしれませんが、メンデルのデータには様々な疑問や批判があり、不正が疑われてきたのです。

例えば、現代統計学の父、R・A・フィッシャーは、メンデルのデータを再分析し、奇跡に近いレベルでデータが理論的な期待値に一致することを指摘しました。この結果が偶然に得られるチャンスが1/30000と著しく低いことから、データが故意に操作された可能性を指摘しました（Fisher, 1936）。

このような奇跡のデータが得られたのでしょうか。現代では、メンデルは自分の考える法則に一致

するデータが得られたところでエンドウ豆を数えるのをやめたと言われています。なるほど、それなら期待値とぴったり一致するはずです。メンデルは不正をする意図もなかったでしょうし、当時はそのような認識もありませんでした。しかし、現代でこのような行為は、データの加工と同じ不正行為になるのです。

Simmonsたちのコンピュータ・シミュレーション

データを分析して有意差が得られなかったとき、参加者を足した経験を持つ人は多いと思います。John, Loewenstein, & Prelec (2012) によれば、およそ7割の研究者にこの経験があるそうです。しかも、これを問題だと認識している人は、調査当時、ほとんどいませんでした。読者の中にも、同じように何が問題なのか判然としない人もいるでしょう。

なんとなくデータを取り足したり、取るのをやめたりすることが、なぜ不正につながるのでしょうか？ Simmonsたちは、そのようなことをしてきた人々を震え上がらせるコンピュータ・シミュレーションを発表しました。彼らはデータセットからランダムにデータを取り出して、二つの条件に振り分けました。ランダムに取り出しているのに、当然、二つの条件に差はありません。しかし、差がないはずのデータセットでもデータの取得ごとに（例、一人の参加者のデータが得られたごとに）検定を行うと、22%のケースで有意な結果が得られたのです（Simmons, Nelson & Simonsohn,

2011）。これは全く差がないデータでも、検定を繰り返せば、有意な結果が得られてしまうことを示しています。

検定を繰り返してはならない。これは推測統計の基本です。検定を繰り返せば、実質的な有意水準が上昇し、本当は差がないのに差があるという確率が増加します。結果として、誤った検定結果が導かれてしまうのです。

なんとなくデータを取り足すような研究における問題のある習慣はQuestionable Research Practices, 頭文字をとって、QRPsと呼ばれます。このような習慣は改めなくてははいけません。QRPsについて日本語で読めるものでは平石・池田（2015）があるので、気になる方はそちらをご覧ください。

サンプルサイズを事前に正確に決める。データの取り足しや、途中でやめることを防ぐには、これが重要です。事前に何人の参加者のデータを取るのか正確に決まっているなら、データの取り足しや途中での打ち切りは生じません。

大きすぎるサンプルサイズ・ 小さすぎるサンプルサイズ

「大は小を兼ねる」ということわざがあります。しかし、心理学では、単純にサンプルサイズが大きければよいわけではありません。まず、データをとるための実験や調査は、参加者の負担になります。過度に多くデータを集めることは参加者に不要な負担をおわせていることになり、倫理的に問題があります。また、研究者も無駄な時間や労力を割くことになり

ます。加えて、帰無仮説検定を行う場合、サンプルサイズが大きいと取るに足らない差や効果が有意になってしまいます。これは帰無仮説検定が、サンプルサイズが大きいほど有意になりやすい性質を持っているからです。例えば、相関係数で $r = .01$ という、実質、無相関に近いようなデータについて考えてみましょう。もし、サンプルサイズが50、例えば、50人の参加者からデータを取ったのなら、この相関は有意になりません ($p = .95$)。ところが、同じ $r = .01$ でもサンプルサイズが50000なら、 $p < .05$ で有意になるのです。もちろんこのような場合でも、相関係数の値そのものに注目すれば、実質的に無相関であることがわかります。しかし、帰無仮説検定をすると、有意か否かだけが一人歩きして、効果の大きさが無視されることがしばしばあります。そして、実質は無相関にもかかわらず、有意な相関があったと検定の結果だけが、喧伝されることがあるのです。

小さすぎるサンプルサイズにも問題があります。帰無仮説検定が、サンプルサイズが大きいほど有意になりやすい性質がある以上、サンプルサイズが少なくと有意である結果を、有意でないと判断することになってしまいます。結果として、重要な知見を発見しそこなうことになるかもしれません。大きすぎたり、小さすぎたりするサンプルサイズの問題については、大久保・岡田(2012)に詳しく書きました。そちらもご覧ください。

検定力を使った決め方

サンプルサイズを事前に正確に決めると、不正を防ぐことにつながります。では、どのように決めたらいいのでしょうか？適切なサンプルサイズは、調べたい現象

や効果の大きさによって異なります。強い効果なら、比較的小さなサンプルサイズでも良いですが、弱い効果ならサンプルサイズを大きくしなくてはなりません。

サンプルサイズの決め方にはいくつか種類があります。帰無仮説検定を行う場合、心理学では、検定力を基準にすることが一般的です。検定力とは、本当に差があるとき、きちんと差があると判断できる確率です。ですから、検定力に基づくサンプルサイズは、本当に差があるときに、きちんと差があると判断できるだけの数となります。

ここで検定力について復習をしておきましょう。統計的な判断には、2種類の誤りが生ずる可能性があります。ひとつは、第1種の誤りで、本当は差がないのにあると判断する誤りです。 α と呼びます。帰無仮説検定ではいわゆる p 値がそれにあたります。もうひとつは、第2種の誤りで、本当は差があるのに差がないと判断する誤りです。 β と呼びます。検定力は、この第2種の誤りと強く関係し、定義としては、全体の確率から第2種の誤りの確率を引いたもの、つまり、 $1 - \beta$ となります。

検定力に基づくサンプルサイズの決定では、検定力が.80になるようにすることが一般的です。これは本当に差があるなら8割は検出できるサイズということです。検定力、有意水準(通常は.05)、効果量、サンプルサイズは互いに影響をし、これら四つのうち三つが決まると残りの一つは自動的に決まります。ですから、サンプルサイズを決めるには、検定力、有意水準、効果量を先に決めなくてはなりません。検定力と有意水準は慣習に従ってそれぞれ.80と.05に設定することが多いです。効果量は先行研究などを参考におお

よその値を決めます。これらを元に計算を行うと適切なサンプルサイズを求められます。計算は、G*PowerやRなどのフリーソフトを使うと簡単です。ただし、測定する変数や検定の手法などによって求め方の詳細は異なります。実施にあたっては村井・橋本(2017)や大久保・岡田(2012)を参考にしてください。これらの書籍を参考にサンプルサイズを事前に正確に決め、研究の信頼性、妥当性、そして再現可能性を高めていきましょう。

文 献

- Fisher, R. A. (1936) Has Mendel's work been rediscovered? *Annals of Science*, 1, 115-137.
- 平石界・池田功毅(2015) 心理学な心理学研究: Questionable Research Practice. 心理学ワールド, 68, 5-8.
- John, L. K., Loewenstein, G., & Prelec, D. (2012) Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, 23, 524-532.
- 村井潤一郎・橋本貴充(2017)『心理学のためのサンプルサイズ設計入門』講談社
- 大久保街亜・岡田謙介(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
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22, 1359-1366.

Profile — 大久保街亜

2002年、東京大学大学院人文社会科学系研究科博士課程修了。2014年より現職。専門は認知心理学。著書は『伝えるための心理統計』(共著、勁草書房)など。