

特集「色への興味を研究に：研究計画から論文投稿まで」 実験計画をたてる

The way to sublimate your interest in color to academic research: From designing research to submitting paper

仮説検証のための実験計画

Experimental design for testing statistical hypothesis

池田 尊司
Takashi Ikeda金沢大学子どもこのころの発達研究センター
Research Center for Child Mental Development, Kanazawa University

キーワード：実験計画, 仮説, 変数, 統計的仮説検定, サンプルサイズ

Keywords : experimental design, statistical hypothesis, variable, statistical hypothesis testing, sample size

1. はじめに

実証とは、これまでの経験を経て生じてくる発想に対して、データによる裏付けを得て整合性を評価することである。直観と実際のデータには乖離があることは珍しくなく、性格特性と血液型の関係のように、研究の結果からは繰り返し否定されている事柄も、直観的に受け入れられているという現実もあるため、実証によって現実を正確にとらえ、合意形成を行う重要性は高い。

実証研究のアプローチには、大まかに調査型のもの、実験型のものがある。どちらのアプローチも事前に何らかの仮説を設定してデータ取得計画を立て、データが仮説を支持するか否かを判定するという過程については同じである。調査型研究では、仮説を検証するために適した既存のデータを分析することで主張を裏付ける。または、仮説を明確に設定せずに既存データからどのようなことがいえるのかを分析する探索的な調査も行われており、予備調査としての性質も持っている。実験型研究では、仮説に対する反証に対応できるように積極的に実験的操作を行い、その後の分析まで含めた実験計画 (experimental design) に基づいたデータ取得を行い、結論を導き出す。本稿では、調査型研究における計画についても「実験」計画と呼ぶこととする。

2. リサーチクエスションと先行研究

まずは、仮説 (statistical hypothesis) の設定を含めた実験計画の作成、そして仮説検証の過程について、事例を通して概観してみたい。ここでは、2005年に Nature 誌に掲載された「Red enhances human performance in contests」という研究¹⁾を題材に、実験計画を定めて実証するまでの一連の過程を追いかけてみる。タイトルが示すとおり、これは「赤には人間

のパフォーマンスを促進させる効果がある」という因果関係があるかどうかの実証を目的とした研究である。

研究計画を立てるための第一歩が先行研究の調査である。これまでに行われてきた研究がどこまで明らかにしてきたのかを知ることで、リサーチクエスションに挑む上での出発点を定めることができる。上記の論文の導入部分では、動物の世界では赤色がテストステロンに裏打ちされた個体の強さと関連していることや、人間は怒りによって顔が紅潮することを示した先行研究が複数引用されている。これは、生物における闘争状態では赤い色彩をまとった個体が相対的に優位な立場になりやすいことを主張するためのものであり、人間を含む生物にとって、赤い色彩には相対的に優位な力関係を構築する効果があるという可能性を紡ぎ出すための糸口となる。また、先行研究を調べることで、研究のゴールに向かうためには何が足りないのかも見えてくる。引用された先行研究からは、自然に発生する赤への色彩変化が生物学的な優位な立場と関連していることまではわかったが、強い個体が赤くなりやすいといった逆の関係性を持っている可能性や、さらには単なる偶然という可能性もある。これらを否定し、主張の正当性を高めていくために必要な作業が、適切な仮説の設定と確率計算に基づく実証である。

3. 実験計画の作成 — 変数の決定と仮説の設定

先行研究で得られている結果をまとめると、「赤」と「優位な力関係」に関連性はあるものの、因果関係を特定する論拠としては用いるには不十分であることがわかった。つまり、赤の促進効果というリサーチクエスションに答えるためには証拠が弱く、まだ保留せざるを得ない段階であるといえる。赤がパフォーマンスを促進させる原因であることを積極的に主張するならば、人為的に赤い彩色を施すことでパフォーマンスが

促進されるのかどうかを調べればよい。これが研究における当面の目的となる。

さて、「赤には人間のパフォーマンスを促進させる効果がある」という当初の仮説をブラッシュアップするにあたり、まずは原因とその結果に分解して考えてみる。原因にあたるものは「赤」であり、その影響を受ける結果にあたるものは「パフォーマンス」である。統計学の用語では、原因に相当する箇所は「独立変数」、結果に相当する箇所は「従属変数」と呼ばれている。これらの変数(variable)は、仮説に沿わないデータが偶然発生する確率を計算することになるため、何らかの数値的な表現にできることが必須である。

まずは独立変数について検討すると、赤がパフォーマンスを促進させると主張したいのであれば、赤以外の色よりも促進効果が強いことを示せばよい。少なくとも「赤」か「赤以外」の2つの状態を設定し、得られる結果が異なるということを実証すればよい。ここで考慮しておくべき点は、「赤以外」の定義によって主張できる範囲が変わってくることである。例えば、赤以外の色として灰色を選択した場合、灰色と比べて赤は効果があるという主張は可能だが、緑や紫といった他の色よりも促進効果があるかどうかについての言及はできない。比較対象となる色を増やすことで主張できる範囲は広がるが、どこかの比較で偶然主張が通ってしまう危険性も増えてしまう。このことはサイコロを1回振って1の目が出る確率よりも、10回振ったうち少なくとも1回は1の目が出る確率のほうが高いことからわかる。そのため、計画全体において得られた結果が偶然得られたものと結論づけられてしまうおそれを回避するため、多重比較補正などの統計的手法を援用して妥当性を担保することとなる。

次に、従属変数となる「パフォーマンス」であるが、促進効果があったと主張するためには、他の変数と同様にデータとして提出する測定値の性質を慎重に吟味する必要がある。シンプルな測定値としては跳躍の高さや遠投の距離を従属変数とすることもできるが、生理学的にはホルモンの分泌量、神経科学的には一次運動野の興奮性なども候補となるだろう。しかし、「パフォーマンス」という漠然とした抽象概念全体を直接的に定量化できる便利な測定方法は存在せず、どのような測定方法を選んだとしても何らかの未解決点が残ることには留意しておくべきである。そのため、具体的な測定値から主張できる範囲に仮説を絞り込み、実証可能なものにする作業が必要となる。

独立変数以外にも従属変数に対して影響力を持ちう

る変数のことを「剰余変数」と呼ぶ。本来は独立変数と従属変数の間に関連性があったとしても剰余変数が打ち消してしまったり、逆に独立変数と従属変数が無関連であっても関連性がみられたかのような結果が得られたりすることがある。例えば、同じ実験参加者に対してシューズの色を変えて長距離走のタイムを比較するという実験状況を想定すると、必ず赤条件と赤以外条件の順序で実施していたときには、疲労の影響を無視できない。ここでは疲労が剰余変数となるため、統制のために何らかの措置を講じる必要がある。先に実験者が赤条件で走る群と赤以外条件で走る群が同数となるように振り分けて影響を相殺するカウンターバランスや、データに対する主観的バイアスを排除するために実験者・参加者以外の第三者が割り付けを行うランダム化比較試験(RCT: randomized controlled trial)などがよく用いられる。データ取得前に統制が困難である場合は、剰余変数も定量化して統計解析に組み入れ、剰余変数の影響力を推定して除外することもある。

独立変数と従属変数が満たすべき条件、そして考慮すべき剰余変数について狙いが定まったら、データの取得前に統計的仮説検定(statistical hypothesis testing)のための具体的な手法と必要なサンプルサイズ(sample size)が決定される。t検定や分散分析といった検定手法は、確率計算のために様々な前提が設けられているため、これらの前提に沿ったデータを投入する必要がある。研究で陥りがちな罠として、データ取得が終わった後で検定手法を選ぶというものがあるが、せっかく取得したデータが前提を満たしていないためにどの手法も使えない、または誤った手法を採用してしまう事態を招く危険性がある。統計の専門家に相談する場合、遅くともデータ取得前までには持ちかけておくことをお勧めする。

この論文では調査研究アプローチに沿ってデータを取得している。調査対象に選ばれたのは、2004年に行われたアテネオリンピックのボクシング・テコンドー・レスリング(グレコローマン・フリースタイル)競技における勝率である。これらの競技では、赤または青のユニフォームを着て試合を行い、勝敗が必ず決まるルールがある。ユニフォームの色(赤または青)を独立変数に、試合の勝率を従属変数に設定すると、「アテネオリンピックでの4競技においては、ユニフォームの色によって試合に勝つ確率が変化する」という仮説が最終的に生成される。独立変数の違い、すなわち赤ユニフォームの場合と青ユニフォームの場合で、勝率

が偶然起こりうる範囲を超えて変化していたことを実証すれば、この研究についての結論が得られることとなる。このような独立変数・従属変数の関係に対して有効な統計的検定手法はカイ二乗検定であるため、これを採用している。なお、ここで生成された仮説には問題点があるが、それについては後述する。

そして、特に実験や介入を伴う研究に必要なデータを取得する時には、倫理的配慮を十分に行わなくてはならない。実験参加者の安全や健康が少しでも脅かされる可能性がある場合や、結果の報告およびデータ保管に際しての個人情報の取り扱いには特段の注意が必要である。近年では、各研究機関に設置された倫理委員会に研究計画書を事前に提出し、審査を受けることが研究実施の要件となってきた。論文投稿時にも倫理審査に関する情報を記載するよう求められるケースが増えているため、対応できる体制を整えておくことが重要である。また、人体に対する侵襲を伴うおそれのある研究は、臨床研究法に則った審査が可能な審査機関に諮る必要がある。

本節の最後に、用語について補足しておく。(a)独立変数・(b)従属変数・(c)剰余変数に相当する用語は研究分野や書籍によって呼び方が様々であるが、基本的にはこの3変数のいずれかを指していることを念頭に置くと、分類の一助となるであろう。高野・岡他²⁾は、因果関係を想定しない観察的な調査研究においては、(a)予測変数・(b)基準変数・(c)共変数という呼び分けを推奨している。また、南風原³⁾は同様の文脈において、(a)説明変数・(b)目的変数・(c)共変量という呼び方があることも紹介している。また、(b)グループには応答変数・反応変数・結果変数・外的基準・被説明変数、(c)グループには交絡変数・外的変数という呼称もあるが、指し示す対象を把握しておけば適切に読み替えることができるであろう。

4. 仮説検証による実証

仮説を設定してデータの取得が終わったら、いよいよデータを分析して仮説を検証する段階へと移行する。ここでは、どのようにすれば「色の効果がある」という仮説を採択することができるのかについて解説する。

アテネオリンピックのデータの特徴は、試合の勝敗は必ず決まることと、赤または青のユニフォームは各選手にランダムに割り当てられることにある。そのため、独立変数である色の効果がないと仮定される場合には、取得するデータが増えるに従って赤および青ユニフォームの選手の勝率は5割に収束していくことが

期待される。このように、主張したい仮説(対立仮説: alternative hypothesis)を否定する仮説のことを、帰無仮説(null hypothesis)と呼ぶ。

仮説検証は、帰無仮説を棄却することで対立仮説を採用するという一見遠回りにも思える手続きを通して行われる。具体的には、まず帰無仮説が正しいと仮定した確率分布のもとで検定対象のデータが得られる確率(p 値)を計算する。これが一定の基準値以下の低い確率でしか起こりえない状態であったとき、帰無仮説を棄却して対立仮説を採択する。これでようやく統計的に「有意」な結果が得られたと報告できることになる。この基準値は有意水準と呼ばれ、社会科学やその近接領域では慣習的に5%が用いられている。つまり、これほど低い確率でしか起こりえないことが起こっていたので、単なる偶然として片付けてしまう合理性がない、という判断を行うのである。

論文で検証すべき帰無仮説は、色の効果がない、すなわち赤ユニフォームの選手と青ユニフォームの選手が勝った試合数に偏りがないというものである。この帰無仮説に沿った勝率の確率分布を計算で求めることができる。具体的には、色の効果がなければ勝率が5割となる確率が最大となり、勝率が5割から遠ざかるに従って起こる確率は低くなっていくことが予測されるのである。実際のデータをグラフから読み取ると、赤ユニフォームの選手の勝率は約55%であった。帰無仮説が正しいという仮定のもとでこの勝率が得られる確率である p 値は4.1%であることが文中に示されている。この値は有意水準の5%よりも低いため、帰無仮説を棄却して対立仮説を採択するという判断が下される。このような手続きを経て、色の効果が有意であったという結論を導くことができるのである。

仮説検証では、帰無仮説を棄却することで対立仮説を採択するという手続きを経ているため、 p 値が有意水準を超えていて有意と判定されなかった場合の解釈には注意が必要である。帰無仮説が棄却されなかった場合でも、色の効果がなかった(勝率差がゼロであった)ということの意味するわけではない。データから計算された差は偶然起こりうる範囲の内側にあり、統計的に意味のある差ではないという判断をしたにすぎないため、効果がないという強い主張はそぐわない。このような枠組みの中では、 p 値が有意水準を超えたことは同等性の証明にはなりえないため、帰無仮説の設定および検証の際には注意が必要である。

また、 p 値の計算においては、実験参加者数に代表されるサンプルサイズが影響することにも気をつけな

くてはならない。サンプルサイズが大きくなるに従って p 値は小さくなる傾向を持つため、先ほどの帰無仮説が棄却されなかった場合 ($p \geq .05$) を考えてみると、本当にわずかな勝率上昇の効果しか得られなかったのか、単にサンプルサイズが不足していたのかの区別ができないのである。反対に、莫大な参加者が参加した研究においては、ごくわずかな差であっても有意であるという判断が下りやすくなるのである。つまり、 p 値は有意であるかどうかの判定にのみ用いられ、効果の量または強さの指標ではないことに気をつけなくてはならない。

それでは、「意味のある」差かどうかは何が決めるのだろうか。勝率が5%変動することに意味があるかどうかは、研究目的次第である。もっと強力な促進作用が得られなくてはならないのか、逆に競り合っている状況下で1%の促進効果であっても重要となるのかによって、効果の大きさの意味合いが変わってくるのである。そのため、研究の仮説や検定手法が決まった段階で、どのくらいのサンプルサイズがあれば研究にとって意味のある差を検出できるようになるのかという基準を定めることも、実験計画を作成するうえで欠かせない項目となる。適切なサンプルサイズ設計がなされた計画に沿ってデータを取ることで、参加者が少ないために起こる誤った帰無仮説の棄却や、闇雲に多いデータ取得によって参加者を疲弊させたり研究予算を圧迫したりしてしまう恐れを減らすことができるのである。

5. 仮説の再構成

論文に話題を戻すと、色の効果について実証した後、対戦相手との実力差で分割したデータについて分析を行っている。その結果、実力差がある試合では色の効果はなく、実力が伯仲しているときにのみ色の効果がみられたことが述べられている。

上記の研究結果からは、ユニフォームの色によって勝率が変化することが実証されたといえるが、まだ多くの但し書きが残された状態であるともいえる。但し書きの一つは、いずれかの選手は赤、もう一方の選手は青ユニフォームを着用する条件下においては、赤条件と青条件の勝率が連動してしまうことに起因する。勝率の変化は赤による促進か、青による抑制か、またはその両方かということを検証できる仮説ではなかったため、勝率の違いは赤の促進効果以外の理由によっても説明できてしまうのである。もう一度アテネオリンピックのデータを調査対象とすると決めた段階で生

成された仮説を確認すると、帰無仮説は「色の違いは勝率に影響しない」というものであり、この帰無仮説を棄却することで、ユニフォームの色の違いが勝率を変化させたことしか実証できておらず、赤または青といった具体的な色の効果には言及できない仮説設定であったことに注意したい。より確実に赤の効果を検証するためには、「赤ユニフォームを着ることで勝率が有意に上昇する」という仮説を設定し、これに沿ったデータ取得を行う必要がある。

そして、サッカーの国際トーナメント戦 Euro 2004 を題材に分析を行っている。このデータの特徴は、赤以外に青と白のユニフォームがあることと、団体競技であること、そして得点数という対戦相手と連動しない従属変数を選んでいることにある。分析の結果からは赤ユニフォームを着用したチームがより多くのゴールを決めていたことが示され、赤の促進効果について積極的に主張できるようになったことと、競技が変わっても成立することが新たにわかった点である。

しかし、これで十分な証拠が揃ったかという点、そうとも言い切れない。アテネオリンピックや Euro 2004 以外でのデータセットでも結果を再現できるかどうかなど、今回の実証が例外的である可能性を打ち消すことはできていない。さらに、試合中には勝敗が決するまでに色の効果による選手の身体的変化や心理的变化が伴っていたはずであり、なぜ勝率が上がったのかというメカニズムに迫ることも、解決すべき問題として残っている。

リサーチクエスチョンを解明する過程において、ごく少数の実験結果だけで結論に到達できることはまれであり、先行研究とオリジナルデータから得られた結果を数多く蓄積していくことによって完成する。特定の色が人間のパフォーマンスに及ぼす影響を実証するためには、依然として綿密な実験計画に基づいた新たな調査と実験データ収集が必要となりそうである。

6. 参考文献

- 1) Hill, R. A.; Barton, R. A. Red enhances human performance in contests. *Nature*. 2005, vol. 435, no. 7040, p. 293, doi: 10.1038/435293a.
- 2) 高野陽太郎, 岡隆 編. 心理学研究法: 心を見つめる科学のまなざし. 補訂版, 有斐閣, 2017, 372p.
- 3) 南風原朝和. 心理統計学の基礎 - 統合的理解のために, 有斐閣, 2002, 394p.