Japanese Psychological Review
2018, Vol. 61, No. 1, 142–146

心理統計の新しい展開と今後の統計教育

南風原　朝　和
東京大学

New developments in psychological statistics and future statistics education

Tomokazu HAEBARA
The University of Tokyo

In this article, I have selected and will comment on three topics from the papers in this special issue: power analysis, Bayesian hypothesis testing, and Bayesian statistical modeling. The major problem in power analysis is the quantitative evaluation of effect size, and this problem is shared in the estimation of effect size. Bayesian hypothesis testing is theoretically superior to classical null hypothesis testing. However, setting prior distributions properly continues to be difficult to overcome. Bayesian statistical modeling changes the role of statistics in psychological research and is promising in many aspects. However, the ease of computation can mask the necessity of careful thinking in modeling, which includes setting prior distributions. I also will discuss future statistics education in the era of rapid developments in methodology, as documented in this special issue.

Key words: power analysis, effect size, Bayesian hypothesis testing, Bayes factor, Bayesian statistical modeling
キーワード：検定力分析，効果量，ベイズ的仮説検定，ベイズファクター，ベイズ統計モデリング

1. 統計革命？

本特集号の編集者の一人である岡田謙介氏が，2012年に共著で出版された大久保・岡田（2012）は，その副題を「効果量・信頼区間・検定力」として，それらを活用することを「統計改革」と呼んでいる。その内容は，従来，帰無仮説の検定一辺倒であったことの限界を認識し，それを克服しようという試みであった。それから6年後のいま，「統計革命」という，より強い表現のテーマの特集号が組まれたが，全部で9論文のうち，効果量，信頼区間，検定力を正面から取り上げたものは村井・橋本（2018）の1編のみで，他の論文では，これらの言葉が本文に一度も，あるいはほとんど出てこない。短い期間でのこの変化は，ある意味，革命的ではあるかもしれない。また，その「統計改革」で批判の対象となっている帰無仮説の検定は，1940年頃から短期間のうちに心理学に一気に普及し，「推測革命」（inference revolution）と呼ばれていたこと（Gigerenzer & Murray, 1987；南風原，2014a）を考えると，感慨深いものがある。

「統計改革」については，筆者も比較的早くから発言してきた（南風原，1991）。（本特集号の巻頭言（三浦・岡田・清水，2018）に豊田（2017）がテキストに「有意性検定からの脱却」という副題をつけたことが書かれているが，私のその報告のタイトルは「有意性検定からの脱却は可能か」であった。）そして，上述の大久保・岡田（2012）の趣旨におおいに賛同しつつ，効果量の信頼区間の解説については質・量ともに補強が必要と思ったのが，南風原（2014b）を執筆した動機の1つであった。

一方，今回の特集で焦点が当てられているベイズ推論，ベイズ統計モデリング，オープンサイエンスについては，私は特に専門ではなく，各論文を読んで各領域の最近の発展ぶりを知ったところである。そのようなことから本論文は，冒頭で述べた「統計改革」あたりを足場にし，上述の村井・橋本（2018），ベイズ的仮説検定を扱った岡田（2018），そしてベイズ統計モデリングについて包括的に解説した清水（2018）を主に取り上げて感想等を述べ，最後にこれからの心理統計教育

について一言述べることで，今回のコメント依頼への返答としたい。

2.「統計改革」における定量的評価のハードル

村井・橋本（2018）は，検定力分析に基づくサンプルサイズ設計を推奨している。検定力は母集団効果量がゼロであるという帰無仮説が偽であるときに，その帰無仮説を正しく棄却する確率，すなわち検定結果が有意になる確率であり，それは，他の要因に加えて，母集団効果量の大きさによって変化する関数である。母集団分布やランダムサンプリングの仮定が満たされているとすると，サンプルサイズと有意水準を決めれば，実験や調査を行う時点で検定力は未知だが確定はしている。その検定力を推定するというのもありうる話だが，「検定力分析」という場合には，通常，そのような実際の検定力を推定することが目的ではなく，母集団効果量の値を「想定」して，それに対する関数としての検定力を計算するのが目的である。

彼らは，杉澤（2017）を引用して，母集団効果量の値の想定の仕方として以下の2つを挙げている（ここで表現は少し変えている）。

①実際の母集団効果量として予想される値
②理想の母集団効果量として望まれる値

このうち①は，上述の「実際の検定力」を推定することに近い。そして，効果量の予想が正確であればあるほど，それに対して計算される検定力は実際の検定力に近いものになる。その計算される検定力が十分に高い値になるようにサンプルサイズを設計する，というのが①の方法である。

この方法に従えば，予想される効果量が小さければサンプルサイズを大きくし，とにもかくにも実際に高い確率で有意になるように研究を設計することになる。しかし，その結果として有意な結果が得られても，「それはそうでしょう」ということで，エキサイティングなこと，情報的に新たに得られることはほとんどない。（ただ，実際には多くの場合，研究者が確保したいのは，この意味での検定力であろう。）

②は，そのままでは，研究上の意味付けがやや難しい。私が意味があると思うのは，②と似ている面もある次の方法である。

③母集団効果量として検出するに値する意味のある値

これは，①とは異なり，意味のないほどの微小な効果量であれば，むしろ「有意でない」という結果が望ましく，意味のある程度の効果量であれば「有意である」という結果が望ましいという考え方である。また，これは実際の効果量の値とは直接関係なく，検出に値するものとして研究者が設定する値である点も①とは異なる。この方法は，「サンプルサイズを大きくすれば何でも有意になるから，検定は意味がない」という批判にも応えるものである。

この③の進め方は，ロジックとしては良いのだが，現実的な問題として，「意味のある値とは？」という設定が難しい。村井・橋本も「効果量の設定はとりわけ難しい」と言っている通りである。彼らは，「その点，信頼区間に基づくサンプルサイズ設計の方が直感に沿う面があり，広がりやすいのではないか」と述べているが，結局のところ，たとえば相関係数のように比較的解釈の容易な効果量であれば，「検出するに値する意味のある値」も「望ましい信頼区間の幅」も，設定はさほど難しくない。これに対し，ある独立変数の偏決定係数（その独立変数を投入する前の残差分散に対して，その独立変数の投入によって減じられる残差分散の割合）などになると，値そのものの解釈が必ずしも容易でないため，「検出するに値する意味のある値」も「望ましい信頼区間の幅」も設定がしにくいことになる。なお，サンプルサイズの設計を検定力に基づいて行うか，信頼区間に基づいて行うかは，研究として検定と区間推定のどちらを目指しているかによって決まる。

関連して，村井・橋本は，「効果量の推奨が始まって以降，統計的仮説検定の結果に，効果量を添えている論文が増えたが，実のところ，アクセサリー的であることも多い」と指摘している。これも，相関係数などであれば，有意であることに加えてその値が報告されることは，ほとんどの場合，アクセサリー以上の意味をもつだろう。一方，解釈が容易でない効果量は，アクセサリーに成り下がる危険もあるが，岡田（2013, p. 237）や

南風原（2014b, p. 51）が述べているように，当該の研究の中ですぐに有効な活用ができない場合でも，先行研究など他の研究の効果量との比較を通して有用な知見が得られる可能性もある。したがって，効果量を結果の基本的な情報として報告することが望ましいことは間違いない。

なお，村井・橋本は，「頻度論の場合とベイズ的アプローチの場合双方の結果が儀礼的に併記されるのではという未来予測」を語っているが，私はこれについては懐疑的だ。今回の彼らの論文でも検定結果とともにベイズ確信区間を算出しているところがあるが，理論的に検定と整合するのは信頼区間であり，ベイズを併記するのは違和感がある。また，ここも信頼区間を併記するのは「儀礼」ではなく，結果の安定性等，実質的に有用な意味を伝えるためである。

以上，大久保・岡田（2012）で言う「統計改革」の範囲でコメントをしたが，村井・橋本も指摘しているように，検定は「2分法的思考を好む人間」にとっては使いやすいものである。逆に，効果量は，検定力分析で用いるにしろ，点推定・区間推定の対象にするにしろ，検定結果ほど扱いやすくはない（この点は，ベイズ推定も同様である）。検定の2値判断に比べ，「効果量の定量的評価」が難しいことが「統計改革」のハードルであるが，それは挑み続ける価値のあるハードルである。

3. ベイズ的2値判断への期待と課題

前項で取り上げた村井・橋本（2018）は，ベイズ的アプローチに関して，「極めて特殊な仮説たる帰無仮説から自由になれることは大きい」と述べている。しかし，岡田（2018）は，頻度論的統計で重要な位置づけにある仮説検定に背を向けるのではなく，そこで従来から指摘されてきた問題に対し，ベイズ的アプローチによって解決する試みを紹介している。その点では，冒頭で述べた「統計改革」の延長上の議論としてとらえることができる。また，2値判断を求める研究者の要望にも応えるものであり，その点でも堅実で有用な議論である。

また前項で，「サンプルサイズを大きくすれば何でも有意になるから，検定は意味がない」という批判に言及したが，言うまでもなく，これは帰無仮説が偽である場合の話である。「帰無仮説は厳密には常に偽であるから……」という説明がなされることがあるが，それは正しくない。たとえば，心理学の例ではないが，私にとっては関係の近い人の研究という意味で身近に感じている例として，2015年にノーベル物理学賞を受賞した東京大学の梶田隆章氏が反証した「ニュートリノには質量がない」という仮説がある。結果的にその仮説が偽であることを証明したわけだが，それまではずっと厳密に真であると信じられていた仮説である。ちなみに梶田氏の研究では，「ニュートリノには質量がない」という仮説のもとでは，3×10^{-12} の確率でしか生じない結果が得られたことが，その仮説を否定する根拠となっている（梶田，2015, p. 149）。もちろん，その確率というのは，おなじみの p 値である。本コメント論文の本筋からやや離れるが，「従来の帰無仮説検定でもノーベル賞は取れる」というエピソードである。

岡田（2018）は，このように真でありうる帰無仮説と対立仮説のうちのどれを採択すべきかという検定問題を，「ベイズファクターによる評価」という統一的な方法で解決する方法を示している。この方法は，頻度論的な検定が，「帰無仮説は棄却することはできるが，積極的に採択することはできない」という，棄却と採択について非対称であった問題を克服し，帰無仮説と対立仮説を同じ土俵上で比較することを可能にする。これは「革命的」と言ってよいほどの大きな特長である。しかも計算上の難しさも，近年の研究で解決されつつあるということで，非常に有望なアプローチと言えるだろう。

ただ，私のように古くからベイズ統計に接してきた者としては，やはり事前分布のことが気になる。岡田（2018）はその問題についても，客観ベイズ的な考え方に基づく既定事前分布などについて丁寧な解説を加えているが，研究の結論が事前分布の設定に依存することは変わらない。もちろんこのことはベイズの利点でもあるのだが，一方で，頻度論からベイズへの移行を妨げる大きな要因であったことも事実である。統計学的に望ましい性質をもつ事前分布が，すべての意味において最適であるとは限らない。また，デフォルトの事前分布を採用することは研究者にとっては楽

で，ベイズ統計の普及にも役立つだろうが，それはベイズ統計の本来あるべき姿と言えるだろうか。前項の最後に，検定の2値判断に比べ，「効果量の定量的評価」が難しいことに触れたが，ベイズによる仮説の2値判断では，事前分布の設定問題という形で，同様に難しい定量的評価を求められているとも言える。

関連して，これは岡田論文のことではないが，「ベイズ統計では仮説が真である確率がわかる」という趣旨の表現を目にすることがある。確かに「仮説が真である確率」を扱うのではあるが，それは，「事前確率をこのように設定したら，事後確率がこのように計算される」という以上のものではない。また，たとえば，「仮説が真である確率は .90 である」という言明は，今回たまたま得られたデータからそのように計算されたということであり，サンプリングによって変動する値であることは，頻度論的な統計指標の場合と同じである。新しい手法が勢いをもって普及していくとき，ややもするとそうした基本的なことが置き去りにされることがあるので，留意点として記しておく次第である。

4. モデリングによる研究法の変革

心理学の研究のうち仮説検証的な量的研究は，一般に，リサーチ・クエスチョンからスタートして仮説を導出し，次に具体的な研究デザインを立案して，そのデザインのもとで仮説に基づく具体的な予測を立てる。そしてデータを収集して分析し，予測との整合性を調べることを通して仮説の検証を行う，という流れで行われる。この中で統計の出番は最後のデータの収集と分析の段階である。もちろん，多様な統計法の考え方をリサーチ・クエスチョンの設定や仮説の導出，そして研究デザインの立案に役立てることはあるが（南風原，2011），そこでは統計が補助的に利用されるにとどまる。

これに対しベイズ統計モデリングは，清水（2018）が述べているように，行動の生起メカニズムそのものを確率的に表現するものであり，これは上述の研究の流れの中では，仮説の導出と予測の段階に該当するものである。つまり，研究の対象となる心理メカニズム，行動生起メカニズムの表現に統計を用いるものであり，これまでの研究における統計の活用目的とは大きく異なる。通常のデータ分析における検定・推定も，母集団を「データ発生装置」とみなし，そこから確率的に発生したデータによってサンプルが構成されると考える確率モデルに基づいている（南風原，2002）。しかし，ここでいう統計モデリングはもっと包括的なものであり，最後のデータ分析の際に設定されるものも含みながら，むしろ仮説そのものの表現に主眼がおかれ，そのモデルに基づく具体的予測を行うことを目的としている。その意味で，ベイズ統計モデリングの導入は研究法を変革するもの，より正確には，研究法の中での統計の位置づけを変革するものと言える。

ベイズ統計モデリングは，本特集号で竹澤（2018）が心理学におけるその必要性を論じているモデリングの中でも最も注目されているものであり，竹澤の言う，より厳密なモデル，より強い理論に向けての活用を期待したい。

そうした研究法の議論以外で，統計計算的に興味深いのは，従来，ベイズ統計は頻度論的統計に比べて計算が難しいことが欠点とされていたのに対し，むしろ事前分布を入れてベイズ的にモデリングすることで，MCMC 法に基づく汎用的な計算プログラムが利用できて，計算がより簡単にできるということである。この逆転現象もまた「革命的」と言ってよいのかもしれない。

一方，この簡便さは若干の危うさも感じさせる。「確率モデルを書いてデータを渡すだけで，パラメータがほぼ自動的に推定できてしまう」（清水，2018）という説明を読むと，パス図を書いてクリックすれば因果モデルが検証できる，という風潮のあった構造方程式モデリング（SEM）を想起させる。SEM を用いた研究の中には，SEM ならではの素晴らしい研究も多く発表された半面，検証されるモデルそのものが何を根拠として構築されたものか判然としないような研究もあった。同様に，ベイズ統計モデリングを用いた研究発表の中にも，思いつき程度の粗っぽいモデルも散見される。前項で述べたこととも関係するが，事前分布も十分な検討なく，機械的に設定しているようなものもある。ただ，SEM については清水も言及しており，「SEM を反面教師にしながら」と書いているので，心配は不要かもしれな

い。また，単なるモデル適合度でなく，具体的な予測についてテータで検証できるところも，SEMとは異なる重要な点だろう。

5. これからの心理統計教育

本特集号は，従来の検定の限界をベイズ的に克服する方法や，心理メカニズムのベイズ統計モデリングなど，心理統計の明るい未来が描かれていて希望をもたせてくれる。こうしたベイズ的要素を含む新しい展開を，これからの心理統計教育にどのように反映させていけばよいだろうか。

心理学で使用される初等統計のテキストで，1章をベイズ統計の解説に割いた先駆的な例としては芝・渡部（1984）などがある。ただ，この点は南風原（2014b）の中のベイズの章も同様だが，テキストの最後におかれていて，どちらかと言うと付録的な扱いであり，これでは十分とは言えない。心理統計を履修した者がベイズ統計とは何であるか，どのような利点があり，どのように活用されているかを知り，必要に応じて自らの研究に活かすための少なくとも基盤が築けるよう，より積極的な取り組みが望まれる。豊田（2016, 2017）のように，もっぱらベイズで，というのも1つの考え方だが，頻度論的統計をその限界やベイズ統計との関係を含め，学習していくことはこれからも必要である。南風原・平井・杉澤（2009，トピック5-2）は，最尤法について初めて学ぶ章で，最尤法が場合によっては適切でない推定値を与える例を示して「最尤法の限界とベイズ推定」について解説している。このような形で異なるアプローチを対比させながら，それらの理解を深めていくのも有効な方法であろう。トピックとしては，たとえば「帰無仮説の検定とベイズ的仮説検定」，「マルチレベル分析とベイズ階層モデル」といったものが考えられる。

大学や大学院の限られた履修時間の中で，どのように心理統計教育を展開していくか，心理統計がますます充実してきたいま，心理学ワールドの共通の課題として考えていかなければならない。

文　献

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.

南風原朝和（1991）有意性検定からの脱却は可能か 日本教育心理学会第33回総会発表論文集，L17–L18. Retrieved from https://www.jstage.jst.go.jp/article/pamjaep/33/0/33_L17/_pdf/-char/ja

南風原朝和（2002）心理統計学の基礎―統合的理解のために　有斐閣.

南風原朝和（2011）臨床心理学をまなぶ7　量的研究法 東京大学出版会.

南風原朝和（2014a）分散分析を基礎から見直す―有意性検定による「推測革命」と近年の「統計改革」 基礎心理学研究，*32*, 217–222. doi: 10.14947/psychono.KJ00009351487

南風原朝和（2014b）続・心理統計学の基礎―統合的理解を広げ深める　有斐閣.

南風原朝和・平井洋子・杉澤武俊（2009）心理統計学ワークブック―理解の確認と深化のために　有斐閣.

梶田隆章（2015）ニュートリノで探る宇宙と素粒子 平凡社.

三浦麻子・岡田謙介・清水裕士（2018）巻頭言　統計革命―特集号の刊行にあたって―　心理学評論，*61*, 1–2.

村井潤一郎・橋本貴充（2018）統計的仮説検定を用いる心理学研究におけるサンプルサイズ設計　心理学評論，*61*, 116–136.

岡田謙介（2013）心理学研究における効果量の活用と報告―APAの指針をふまえて　教育心理学年報，*52*, 234–237. doi: 10.5926/arepj.52.234

岡田謙介（2018）ベイズファクターによる心理学的仮説・モデルの評価　心理学評論，*61*, 101–115.

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.

芝 祐順・渡部 洋（1984）統計的方法Ⅱ　推測　増訂版　新曜社.

清水裕士（2018）心理学におけるベイズ統計モデリング　心理学評論，*61*, 22–41.

杉澤武俊（2017）検定力分析に基づくサンプルサイズ設計　村井潤一郎・橋本貴充（編著）心理学のためのサンプルサイズ設計入門（pp. 21–41）講談社.

竹澤正哲（2018）心理学におけるモデリングの必要性 心理学評論，*61*, 42–54.

豊田秀樹（2016）はじめての統計データ分析―ベイズ的〈ポストp値時代〉の統計学　朝倉書店.

豊田秀樹（2017）新訂　心理統計法―有意性検定からの脱却　放送大学教育振興会.

―2018. 5. 7　受理―