

比率の差の検定における不合理

鈴木, 譲
九州大学大学院人間環境学研究院人間科学部門共生社会学講座

<https://doi.org/10.15017/4771885>

出版情報 : 人間科学共生社会学. 9, pp.1-8, 2019-03-20. Faculty of Human-Environment Studies,
Kyushu University
バージョン :
権利関係 :



比率の差の検定における不合理

鈴木 譲

要 旨

本稿では、比率の差の検定の具体的事例を用いて、統計的仮説検定が持つ不合理性を明らかにする。周知のように、比率の差の検定は、2つの母集団における比率が等しいかどうかを調べるための手法である。実際の社会調査において扱う母集団は、言うまでもなく有限母集団であり、比率が完全に等しくなること自体がそもそもまれな現象である。極端な場合には、比率が決して等しくならないような場合すら多々ある。このような場合に、比率の差の検定を行うこと、さらに一般的に言えば、母集団の比率が等しいかどうかを確率的に問うこと自体が無意味である。問題にすべきなのは、比率が等しいかどうかではなく、等しくないとしても、どの程度の差異があるのかである。

キーワード：統計的仮説検定、帰無仮説、比率の差

1 はじめに

統計的仮説検定は社会学を初めとする社会科学において広く用いられている手法であるが、この手法に対してはこれまでも多くの批判が行われてきた (Kline 2004; Morrison and Ramon 1970)。また、社会統計学の教科書においても、この手法の安易な適用については注意が喚起されている (Bohrnstedt and Knoke 1988: 198)。

ただ、これまでの批判は主に有意水準の設定、母集団の選定、標本抽出の手法などに焦点が当てられており、抽象的な議論に終始している感が否めない。そこで、本稿では具体的な問題設定を行い、このような場合に比率の差の検定を行うと一体何が問題なのか、さらに言うならば、このような場合に比率の差の検定はいかに不合理な手法であるかを端的に指摘し、その上で統計的仮説検定の持つ一般的問題点を指摘する。

2 比率の差の検定：確率論的議論

統計的仮説検定の持つ不合理性を示すために、比率の差の検定を用いる典型的な事例を考える。今、6,248 人の男性からなる母集団 A と、6,247 人の女性からなる母集団 B を想定し、この

2つの母集団でフランス大統領の名前を知っている人の比率が等しいかどうかを調べるとする。標本調査を行い、500人の標本をそれぞれの母集団から無作為抽出し調べたところ、すべてが有効回答で、標本における比率はいずれも0.2であったとする。

この条件のもとで、比率の差の検定の標準的な手順を用いて、2つの母集団の比率が等しいかどうかを調べ、どのような結論が得られるか、そして、その結論がいかに不合理であることを示す。

この問題設定は、比率の差の検定の典型的な事例である。帰無仮説は「2つの母集団AとBにおいて、フランス大統領の名前を知っている人の比率が等しい」という命題である。標本の大きさはいずれも500、つまり $N_A = N_B = 500$ であり、標本における比率はいずれも0.2、つまり $P_A = P_B = 0.2$ である。標本全体における比率も0.2、つまり $P = 0.2$ である。

この時、以下の統計量 Z を計算し、標本数が十分大きいとして、標準正規分布による Z 検定を行うわけであるが、明らかに $Z=0$ である。以下の式では、連続性の補正を行っていないが、 $P_A = P_B$ であるから、連続性の補正を行ったとしても、 $Z=0$ である。

$$Z = \frac{|P_A - P_B|}{\sqrt{P \cdot (1-P) \cdot \left(\frac{1}{N_A} + \frac{1}{N_B}\right)}} = 0$$

次に、設定した有意水準に応じた臨界値を求め、計算した統計量がこの臨界値よりも大きければ、棄却域に入るとして、その有意水準で帰無仮説を棄却するわけである。

しかし、有意水準にかかわらず臨界値は常に正であるから、計算した統計量が0である以上、この問題設定ではどのように有意水準を設定しても帰無仮説を棄却することはできない。つまり、比率の差の検定を用いる限り、帰無仮説に関して何ら有益な判断を下すことはできないことになる。なお、信頼性区間を用いたとしても、得られる結論は同じである。この問題設定では有意水準にかかわらず、信頼性区間は0を含んでいる。

3 帰無仮説の論理的真偽：決定論的議論

前節では、比率の差の検定の手順に従い、この設定では有意水準にかかわらず帰無仮説を棄却できず、何ら有益な結論は導けないことを示した。ところがこの問題設定では、そもそも帰無仮説が論理的に偽の命題であることが証明できる。つまり、この2つの母集団においては、比率は決して等しくならないことが論理的に分かるのである。ここで示した設定は、典型的な比率の差の検定の事例に見えるが、実はこのような状況で、比率の差の検定を行うことは全く無意味である。

この点を説明するために、2つの母集団AとBの大きさを S_A と S_B とし、それぞれの母集団において比率を計算する条件に対応する要素の個数、つまり分子の値を a と b とする。この時、母集団における比率 p が等しいとすれば、明らかに次式が成り立つ。

$$\rho = \frac{a}{S_A} = \frac{b}{S_B}$$

今、非常に小さい母集団 A と B を考え、 $S_A = 10$ 、 $S_B = 15$ とすると、上の式は次のように表される。

$$\rho = \frac{a}{10} = \frac{b}{15}$$

この式が成り立つような a, b, ρ の組合せは、表 1 の 6 通りである。

表 1 比率の等しくなる組合せ

a	b	ρ
0	0	0
2	3	1/5
4	6	2/5
6	9	3/5
8	12	4/5
10	15	1

比率が等しくなる ρ の値は、0/5 から 5/5 までの 6 通りであるが、ここで分母の 5 は母集団の大きさ 10 と 15 の最大公約数に対応している。一般に S_A と S_B の最大公約数を C とすれば、比率が等しくなる ρ の値は、0/ C から C/C までの $C+1$ 通りである。

従って、 S_A と S_B が互いに素である場合、つまり、両者の最大公約数が 1 である場合を考えると、比率が等しくなるのは $\rho=0$ もしくは $\rho=1$ のいずれかに限られることが分かる。これは、比率を計算する条件に対応する人が誰もいないか、それとも、全員が対応しているかの 2 通りを意味している。

ここで、初めの問題設定に戻ると、 S_A と S_B の値は以下の通りである。

$$S_A = 6,248 = 2^3 \times 11 \times 71 \qquad S_B = 6,247$$

S_B の値 6,247 は、実は素数であり、1 と自分自身以外には約数を持たない。従って、 S_A と S_B は互いに素であり、母集団 A と B で比率 ρ が等しくなるのは、誰もフランス大統領の名前を知らないか、全員がフランス大統領の名前を知っているか、このいずれかしかない。しかし、問題設定にある通り、各母集団から 500 人の標本を抽出して調べたところ、比率はともに 0.2、つまりどちらの標本においても、100 人は知っていて、400 人は知らなかったわけであるから、このようなことはあり得ない。つまり、このような状況では、2 つの母集団での比率 ρ は決して等しくなることが、論理的に証明されたわけである。言い換えれば、ここでの帰無仮説は論理的に偽の命題である。

4 問題点の指摘

この問題設定では帰無仮説は論理的に偽の命題である。従って、推測統計・比率の差の検定を用いて帰無仮説を確率的に分析する意味は全くない。この議論が意味することは、以下の三点にまとめられる。

第一に、母集団が明確に定義されていない、あるいは、母集団の大きさが明確ではないにもかかわらず、推測統計を用いる場合が現実には多い。しかしながら、これでは帰無仮説が真になる可能性を保証できない。つまり、帰無仮説が真になる可能性があるかどうかすら調べられない。なぜならば、調べるのに必要な情報が欠如しているからである。このような状態で、推測統計を用いることは適切とは言いがたい。

もちろん帰無仮説が真か偽かが論理的には分からないから、確率論に依拠する推測統計を用いるわけであるが、少なくとも「帰無仮説が真の可能性はある」ことは論理的に保証されていなくてはならない。「帰無仮説が真の可能性はある」と「帰無仮説が真である」とは全く別の命題である。言い換えれば「帰無仮説が真であるか偽であるか」は分からないわけであるが「帰無仮説が真である可能性も偽である可能性もある」ことは分かっているとはいえない。母集団の大きさが分からなければ、そもそもこの点を確認することができない。

第二に、帰無仮説が論理的に偽である場合、問題なのは2つの母集団で比率が等しいかどうかではない。等しくないことはすでに証明されている。問題なのは、等しくないとしても、どの程度の相異があるかである。しかし、この点に関して、推測統計は何ら有益な情報を与えてくれない。

第三に、仮に帰無仮説が論理的に真の可能性があったとしても、我々が本当に知りたいのは2つの母集団で比率が等しいかどうかではない。比率が等しくなるのがごくまれな現象であることは、初めから分かっている。

たとえば、2つの母集団 A と B の大きさがそれぞれ 10 と 15 の場合、2つの母集団における比率の組合せは全部で $11 \times 16 = 176$ 通りあるが、比率が等しくなるのはこのうち 6 通りしかない。従って、単純に考えても比率が等しくなる場合は $6/176 \approx 0.034$ 、つまり 3.4% となる。

一般に、2つの母集団 A, B の大きさを S_A と S_B とし、その最大公約数を C とすれば、2つの母集団での比率が等しくなる組合せが、すべての比率の組合せに占める割合は、次式で表される。

$$\frac{C+1}{(S_A+1)(S_B+1)}$$

S_A と S_B が与えられた場合、その最大公約数 C の最小値と最大値を考えてみる。まず最小値は 1 であり、これは S_A と S_B が互いに素の場合である。また最大値は、 S_A と S_B が等しい場合で、 $C = S_A = S_B$ となる。この時、2つの母集団での比率が等しくなる組合せが、すべての比率の組合せに占める割合は $1/(C+1)$ となる。

たとえば、ともに大きさが 100 の母集団であれば、この値は約 0.99% となり、標本調査をす

るまでもなく、帰無仮説が成り立つ場合がまれであることが想像できる。問題なのは、ほとんどの場合に等しくないとしても、どの程度の相異があるかである。しかし、この点に関しては推測統計は何ら有益な情報を与えてくれない。有意水準を用いた議論の焦点は「比率が等しくない」ということをどの程度の信頼性を持って言えるかであり、「等しくない」ということの程度、つまり相異の程度を問題にしているのではない。

5 まとめ

言うまでもなく、安易に推測統計を用いることは、計量分析の質の低下を招くことになる。本稿で問題にしているのは、帰無仮説が成り立つという仮定のもとで、計算した統計量が標本分布の中でどのように評価されるかという「帰無仮説棄却の論理」の不合理性ではなく、帰無仮説の設定自体に潜む不合理性である。そもそも偽である命題を確率的に分析することに意味はない。問題にすべきは、偽である場合の詳細、つまり、どの程度の差異があるのかである。しかし、この点に関しては、比率の差の検定は何ら有益な情報を与えてくれない。比率の差の検定が問題にしているのは、あくまで比率が等しいかどうかであり、等しくない場合の差異の程度に関しては全く無力である。

さらに言えば、実際に推測統計を用いる場合、母集団が明確に定義されていないか、あるいは、その大きさが不明確な場合が多々あるが、これは帰無仮説が真になる可能性の判定すら行えない状況にあることを意味している。

「推測統計は確率論を用いるので、記述統計よりも高度な分析手法であり、推測統計を用いれば記述統計よりも詳細・正確・有益な分析結果が得られる」というのは幻想に過ぎない。推測統計を安易に用いることにより、社会調査における計量分析の質が低下してしまうことこそ問題であり、我々はこの危険性を認識すべきである。

今後の改善の方向性としては、以下の3つの方法が考えられる

- (1) 確率論に依拠して現在の推測統計を引き続き用いる。ただし、母集団や無作為抽出などに関する条件を厳密に吟味する。
- (2) 確率論に依拠するが、現在の推測統計は用いない。無作為抽出された標本の記述統計から母集団の状態を推測する論理を、新たに構築する。
- (3) 確率論には依拠せず、標本は無作為ではなく代表性を重視して抽出する。標本の代表性と大きさ、および、標本における記述統計から、母集団の状態を質的に推測する。

最初の方法では、母集団の定義や大きさを厳密に吟味すれば、今回指摘したような誤り、つまり、帰無仮説が論理的に偽の命題であるにもかかわらず推測統計を用いるような誤りは回避できる。しかしながら、統計的仮説検定の論理自体に内在する問題点に関しては解決されない。統計的仮説検定では、標本の記述統計にもとづいて統計量を計算し、帰無仮説が真と仮定した条件の下で、この統計量が標本分布の棄却域に属する確率を問題にする。しかしながら、この

確率と、帰無仮説の真偽の可能性との間に、そもそも有意義な関係があるのかについては、異論も存在する。

二番目の方法は、まだ新たな論理が構築されていないので、すぐには利用できない。

三番目の方法は、確率論に依拠せず、標本の記述統計そのものを重視するやり方である。この方法では、標本は無作為抽出によらず、代表性を重視して、母集団の縮図となっているような形で抽出する。標本の記述統計から母集団の状態を推測する場合、その信頼性の拠り所となるのは、標本の代表性と標本の大きさである。

なお、標本の無作為性と代表性とは全く別の性質である。前者は、標本の抽出過程に関する性質であるのに対して、後者は標本の抽出結果に関する性質である。たとえば回収率が低いことは、無作為性に関しては問題となるが、代表性に関しては問題とはならない。

三番目の方法は確率論に依拠しないので、何パーセントで有意、といった形では簡単には判断ができない。しかしながら有意性による判断では、手順が定式化している反面、本来は重要な指標を、有意性という必ずしも適切とは言えない基準によって、切り捨てている可能性がある。このリスクの深刻さを考慮すると、推測統計は「行わないよりは行った方が良い」手順と言うよりも、むしろ「行わない方が良い」、更に言えば「行うべきではない」手順と考えられる。

これまで社会学において数多くの計量分析が行われてきたにもかかわらず、得られた知見の蓄積が余り見られないのは、一つには推測統計への過度の依存に原因があるとも考えられる。

補足1 一般的な比率の検定における帰無仮説

比率の検定としては、クロス表を用いたカイ自乗検定（独立性の検定）もその一種である。比率の差の検定では、帰無仮説は2つの比率が等しいことであったが、カイ自乗検定では、帰無仮説は複数の比率がすべて等しくなることである。従って条件はより厳しく、帰無仮説が成り立たない可能性はより大きくなる。

今、クロス表の行数を n 、列数を m とする。比率のクロス表として、各行ごとに行の合計値に対する各セルの比率を計算したものを考える。このクロス表で帰無仮説が成り立つ、つまり比率が等しくなるとは、 m 個の列の内、どの列をとっても、その列に関しては同じ値が n 個並んでいるという意味である。もちろん、各列ごとにその値は一般には異なる。クロス表の各行の合計値は n 個あるが、これら n 個の値の最大公約数を C とする。ここで、帰無仮説が成り立つような比率の値は、 C を分母とする場合に限られる。また、帰無仮説が成り立つ比率の組合せの総数 x は列数 m に応じて、次のようになる。

$$m=2 \quad x=C+1$$

$$m=3 \quad x=\sum_{i=1}^{C+1} i$$

$$m \geq 4 \quad x = \sum_{j_1=1}^{C+1} \sum_{j_2=1}^{j_1} \cdots \sum_{j_{m-2}=1}^{j_{m-3}} j_{m-2}$$

行の合計値である n 個の値が互いに素であれば $C=1$ であり、帰無仮説が成り立つのは比率が 0 か 1 の場合しかない。この場合、帰無仮説が成り立つ比率の組合せは、全部で m 通りである。

比率の差の検定の場合は、2 つの母集団において、たとえばフランス大統領の名前を知っているか知らないかであるから、クロス表としては 2 行 2 列である。 $m=2$ であるから、帰無仮説が成り立つ、つまり比率が等しくなる組み合わせは $C+1$ 通りとなる。

次にすべての比率の組合せを考える。 $1 \leq i \leq n$ とし、各行の合計値を S_i とする。各行における比率の組合せの総数を y_i とすると、 y_i は列数 m に応じて次のようになる。

$$m=2 \quad y_i = S_i + 1$$

$$m=3 \quad y_i = \sum_{j=1}^{S_i+1} j$$

$$m \geq 4 \quad y_i = \sum_{j_1=1}^{S_i+1} \sum_{j_2=1}^{j_1} \cdots \sum_{j_{m-2}=1}^{j_{m-3}} j_{m-2}$$

従って、クロス表全体でのすべての比率の組合せ y は、次のようになる。

$$y = \prod_{i=1}^n y_i$$

クロス表全体のすべての比率の組合せの中で、帰無仮説が成り立つような比率の組合せの占める割合は x/y となる。

この割合は、厳密には確率ではない。たとえば、大きさが 10 と 15 の 2 つの母集団の場合、比率が等しくなる組合せは 6 通りで、すべての比率の組合せ 176 通りの内の 3.4% である。母集団の各人がフランス大統領の名前を知っている確率は分からないから、10 人中 k 人 ($0 \leq k \leq 10$) の人がフランス大統領の名前を知っている確率も分からず、それぞれの比率の組合せを同等に見なして、3.4% を帰無仮説が成り立つ確率と見なすのは無理である。この場合、帰無仮説は確率事象としてはとらえられない。より正確に言えば、帰無仮説を確率事象として定義するための適切な確率空間が設定されていない。

帰無仮説の発生確率を正確にとらえるのは無理であっても、帰無仮説の成立が極めてまれな現象であることは容易に想像でき、棄却できるかどうかを問題にしてもさほど意味があるとは思えない。

もし、すべての比率の組合せを同等と見なして、帰無仮説の成り立つ確率を 3.4% と考えるならば、条件付き確率として第一種の過誤 $\alpha = \text{Prob}(\text{sample} | H_0)$ を考えることができる。しかし、そもそも帰無仮説の成り立つ確率が分かっているのならば、帰無仮説が棄却できるかどうかを調べるために標本調査をする必要はなくなってしまう。

補足2 有限母集団と無限母集団

有限母集団で比率を考える限り、今回示したように、決して比率が等しくない事例が存在する。ここで、推測統計を用いる場合には、想定しているのはあくまで無限母集団であるから、実際の有限母集団での比率は問題ではない、という主張について考えて見る。確かに、理論的には母集団は無限母集団である。しかし、我々が問題にしている社会現象では母集団は必ず有限である。もちろん有限母集団であっても、十分に大きい有限母集団から、十分に大きい標本を抽出するならば、中心極限定理などを用いた近似を行うことは可能である。しかし、これは確率分布の近似に関する議論である。統計的仮説検定においては、母集団における比率は確率変数ではなく、未知の定数である。分子は条件に合致する母集団の要素数、分母は母集団の全要素数であり、言うまでもなくこの比率の値は決定論的に一意に定まる。統計的仮説検定を用いて社会現象を分析するにあたって、我々が問題にしているのはこの比率であり、理論的な無限母集団の比率ではない。比率の検定が問題にしているのが無限母集団の比率だというのは、そのような無限母集団に関する分析が、我々が実際に分析対象としている有限母集団にどのような意味を持つのかが分からなくなってしまう。

文 献

- Bohrnstedt, George W. and David Knoke, 1988, *Statistics for Social Data Analysis*, second edition, Itasca, IL: F.E. Peacock.
- Kline, Rex B., 2004, *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*, Washington, D.C.: American Psychological Association.
- Morrison, Denton E. and Ramon E. Henkel eds., 1970, *The Significance Test Controversy – A Reader*, Chicago, IL: Aldine.