

スノーボール・サンプルによる比率の推定

太郎丸 博

この文書は第 95 回数理社会学会大会 (2023/8/25-26 愛知大学) での萌芽セッションで使ったポスターの内容を増補し、少しだけ詳しく説明したものである。

1 問題

1.1 有意抽出からの一般化

社会学では有意抽出 (non probability sampling) が頻繁に用いられている。質的研究では数人から数十人程度の対象者が有意抽出される。量的研究でも、完全に等確率で対象者が抽出できることはなく、実際には回収率が 100% になることもない。これは社会学に限ったことではなく、経験データを扱う多くの学問分野で起きていることだと思う。それゆえ、有意抽出されたサンプルの特徴について知ることは、非常に重要だが、いくつかの例外 (Bethlehem 2015; Bethlehem, Cobben, and Schouten 2011) を除けば、そのような研究はマレである。

この小論では、スノーボールサンプリング (以下、SS と略称) した場合に、サンプルから計算した比率 (標本比率) が、母集団における比率 (真値) とどの程度乖離するのか、そして、サンプルから計算した 95% 信頼区間がどの程度正確なのか、モンテカルロシミュレーションを使って検討する。つまり、スノーボールサンプルからの一般化がどの程度正確なのかを調べる。

有意抽出を多用する質的研究の教科書では、質的研究の目的は一般化ではなく解釈だ、といったことが異口同音に書かれているので、このような教科書の著者にとっては、私の研究は邪道、誤った方向に進んでいる研究、ということになるのかもしれない。しかし、実際の質的調査を使った論文を読んでも、少数の対象者から得た知識をもとに、あいまいに一般化しているような結論を導く研究は少なくない。例えば、A というエスニック・グループに属する人々の語りを複数検討し、A は な意味世界の中で生きている、といった結論を導くような研究である。この研究で検討しているのは、A に属する数人の語りだけであり、その他の A の人々がどのような意味世界で生きているかはわからないはず

だが、「解釈」の名のもとに、Aの人々すべてが同じ意味世界に生きているかのような主張を、結論では導いてしまうのである。このような曖昧な一般化は、適切な場合もあれば、そうでない場合もあるかもしれない。いずれにせよ、私は、このような曖昧な一般化を批判したいのではなく、むしろ積極的に一般化の可能性を追求したいのである。私が批判したいのは、「一般化」と「解釈」を対置して、有意抽出にもとづく研究を一般化から遠ざけようとする人たちであり、量的研究でもそのような論調の人たちは珍しくない。今回の研究は主に質的研究を念頭に置いているが、とうぜんSSするならば、量的研究にもあてはまる。

1.2 スノーボールサンプリング

本論では、有意抽出の中でもスノーボールサンプリング(SS)に限定して検討する。SSとは、調査対象者に次の対象者を紹介してもらい、次の対象者に次の次の対象者を紹介してもらい、次の次の対象者に...といった連鎖を繰り返して対象者を抽出する方法である。実際には機縁法(convenience sampling)のほうが圧倒的に多く用いられている印象だが、機縁法よりもSSのほうが、対象者の選び方がはっきりしているので、シミュレーションするのが容易である。つまり、簡単にモデリングできそうだから、まずはSSについて考えてみた、ということである。

また、本稿では標本から比率の推定を行うと仮定する。抽象的ではわかりにくいので、母集団は元Jリーガーで、彼らのうち、現職が自営業である人の比率を推定したいとする。後で述べるように、誤差相関の問題を扱う上では、正規分布する変数を扱ったほうが簡単だが、社会学ではカテゴリカル変数を使うことが多く、量的変数でも正規分布していないことはしばしばある。現実的な仮定を置くことを選んだということである。質的研究者は比率の推定なんかしないが、「Aは.....な意味世界の中で生きている」という結論を導くためには、Aのほぼ100%の人が「.....な意味世界で生きている」というカテゴリに分類できることを論証する必要がある。これは推測統計学的には、比率の推定をしていることと大差ない。

1.3 母集団のネットワーク構造

SSの場合、母集団において、誰と誰が知り合いで、紹介可能なのか、ということが決定的に重要である。いわゆるネットワークの構造が、得られるサンプルの特徴に影響を及ぼすことは自明である。もしも、自営の人は自営の知人が相対的に多く、自営でない人(以下、その他と呼ぶ)は自営でない知人が多い場合(統計学的には誤差相関がある、という

ことである)、自営の人は自営を紹介しやすく、その他の人はその他を紹介しやすくなるので、それが元Jリーガーに占める自営率の推定をゆがませたり、信頼区間の幅を大きくしたりする可能性がある。

このように考えると、最初の対象者が誰かによって、サンプルの特徴が変わってくる可能性があることがわかるだろう。誤差相関が強い場合、最初の対象者が自営ならば、次の対象者も自営になりやすく、その次の対象者も ... といった事態が生じやすくなる。これがどのようなバイアスをもたらすのか検討するのがこの研究の目的である。

2 方法：シミュレーション

下記の条件で架空の母集団からスノーボールサンプリングし、比率の誤差の大きさを推定する。

- 母集団（Jリーグ引退者）の規模は1000人と仮定する。
- 母集団に占める自営業率 p は 0.03, 0.1, 0.3, 0.5 と仮定し、それぞれの場合について推定する。無作為抽出する場合、 $p = 0.5$ のとき自営ダミーの分散は最大になるので、標準誤差も最大になる。 p が 0.5 から遠ざかるほど標準誤差は減少する。有意抽出でも後で定義する「誤差」（標準誤差を一般化したもの）は母集団での標準偏差に比例して減少することは証明済みである（太郎丸 2021）。
- 母集団の構成員には $i = 1 \sim 1000$ の ID を振り、 $1 \sim p \times 1000$ を自営、それ以外はその他とする。
- 母集団はスモールワールド（Watts and Strogatz 1998）と仮定し（無向グラフ、平均知人数 = 6）、edge の架け替え率 q は 0.02, 0.1, 0.25, 0.5 の値をとると仮定し、それぞれの場合についてすべて計算する。
- $q = 0$ （レギュラーグラフ）の時には、 i は $i-3, i-2, i-1, i+1, i+2, i+3$ と知人（後述のように id の両端数は例外）であると仮定する。図 1 のいちばん左がレギュラーグラフである。白が自営、黒がその他であるとする。自営の一番はしの人（id = 1）は、3人としかつながっていない。これは R(R Core Team 2022)の sna パッケージ (Butts 2023) の rgws() 関数の仕様で、自営の上端とその他の上端のあいだに edge をひく場合もある（というか、ワッツたちはそうしている）。ただこちらのほうが、自営とその他の人のつながりが少なくなり、バイアスが強く出るのではないかと期待してこうしている。id = 1 の人は、自営業の中でも最も「その他」の人との距離（パスの長さ）が遠い人なので、レギュラーグラフでこの人からサンプリングをスタートすると、な

かなかその他の人を抽出できないだろう。ただし、エッジの架け替え率が大きくなるほど、「その他」とつながる自営は増えていき、ランダムネットワーク（すべての人のあいだに等確率でエッジを掛けたネットワーク）に近づく。ランダムネットワークで SS すると、一人目の対象者は有意抽出でも、二人目以降の対象者はランダムに選ばれる（エッジがランダムにかかっているから）。それゆえ、SS でも、だれからスタートしようと、 q が大きいほど無作為抽出した場合に近い推定量が得られると私は予測した。

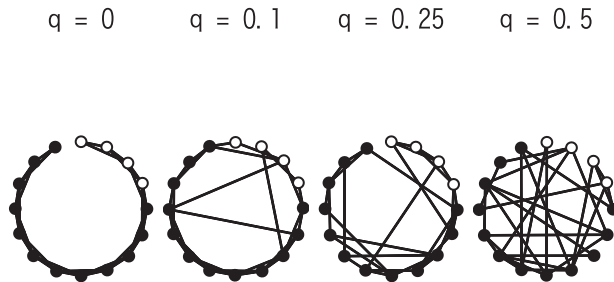


図1 スモールワールド・ネットワークの例

- ・ サンプルサイズ n は 5, 10, 20, 50, 100 人。
- ・ 最初の対象者は $i = 1$ の人（自営でその他から最遠）、または母集団で最も知人の多い人（複数いたら ID 番号の一番小さい人）とし、それぞれの場合の誤差を計算している。
- ・ 対象者は知人の中から無作為に一人選んで次の対象者として紹介すると仮定する。紹介する知人がいない場合は、その対象者よりも前に抽出された対象者の知人全体の中から無作為抽出する。
- ・ 推定方法：
 1. 標本比率 \hat{p} で母比率（真値） p を推定し、二項分布でその信頼区間を推定する。
 2. 切片のみの空間誤差相関モデルで、知人同士に誤差相関があると仮定（Ward and Gleditsch 2008）。重みづけ行列は、知人の場合 1、知人でない場合 0 をとる。R の spatialreg パッケージ（Bivand, Pebesma, and Gomez-rubio 2013）で計算した。

上の空間誤差相関モデルは、残差が正規分布することを仮定しているし、標準誤差などは N が十分に大きな値をとらないと不正確になる可能性がある。しかし、OLS でも二値

変数の予測をかなり上手にできることはよく知られているし (Horrace and Oaxaca 2006)、このモデルは比較的よく知られた推定法なので、まずはこの方法を使ってみた。

2.1 誤差・入区間率の計算

上で挙げたすべての条件組み合わせで 20 回抽出を繰り返す。真値が p 、架け替え率が q 、サンプルサイズが n 、推定法が k のときの誤差の大きさを e_{pqnk} と表記し、

$$e_{pqnk} = \sqrt{\frac{\sum(p - \hat{p})^2}{\text{繰り返し数}}}$$

と定義する。誤差の大きさは、標準誤差とほぼ同じ定義だが、有意抽出の場合、標本平均は一致推定量とはならない ($E(\hat{p}) \neq p$)。この「誤差」は無作為抽出した場合、標準誤差に近似する (付録参照)。

また、95% 信頼区間に真値が入った比率 (入区間率と呼ぶ) r_{pqnk} を計算した。点推定値と真値の誤差だけでなく、95% 信頼区間がどの程度正確に推定できるのかも重要なポイントだろう。正の誤差相関が強いほど、標準誤差は大きくなる (と思う、未確認)。それゆえ、誤差相関を仮定せずに単純に二項分布で信頼区間を推定すると、その幅を過小に推定してしまうのは自明である。ただ、どの程度過小になるのか知りたい。また、空間誤差相関モデルでも、サンプルサイズが小さかったり、残差が正規分布していなかったりした場合にどの程度正確な推定ができるのか、知りたい。

これらが、 p, q, n, k によってどのように変化するのか、検討する。その際に誤差と入区間率の平均値を検討する。例えば、架け替え率 q によって平均誤差がどう異なるのか調べたい場合、

$$e_{\cdot q \cdot} = \frac{\sum_p \sum_n \sum_k e_{pqnk}}{4 \times 5 \times 2}$$

を計算して、これが架け替え率によってどう違うか示す。ただし、 $e_{\cdot q \cdot}$ は架け替え率が q のときの平均誤差を示す。

3 結果

3.1 推定値と真値との誤差

図 2 は推定法別にサンプルサイズと平均誤差の関係を示したものである。どちらの方法で推定してもサンプルサイズが大きくなるほど平均誤差は縮小している。サンプルサイズが 5、10、20 の場合、単純に標本比率で推定したほうが誤差が小さいが 50 以上になると

誤差相関モデルでも同程度の誤差になる。

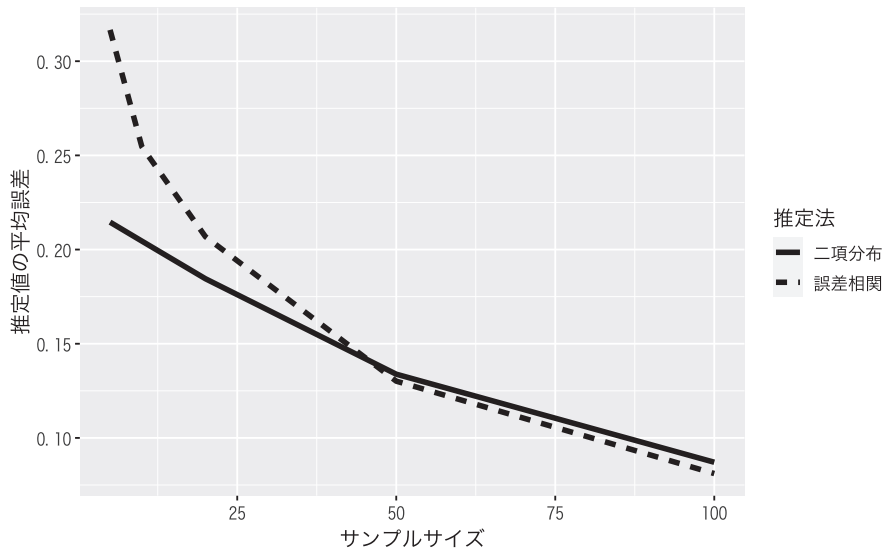


図2 推定法別、サンプルサイズと誤差

図3は、母集団のクラスタリングの程度（架け替え率 q が小さいほどクラスタリングは大きい）と、最初の対象者によって平均誤差がどのように異なるのか示したものである。架け替え率が大きくなるほど誤差は減少している。クラスタリングが大きい（つまり、自営は自営とつながる傾向が強い）と、最初の対象者と同じ現職の人ばかり、抽出してしまう可能性が高いということだろう。このような傾向は1番の人（つまり、その他の人とつながっている確率が最も低い人）から、抽出をスタートした場合に顕著だが、架け替え率が0.25や0.5だと、知人数が最大の人からスタートした場合とほとんど同じ誤差であることがわかる。

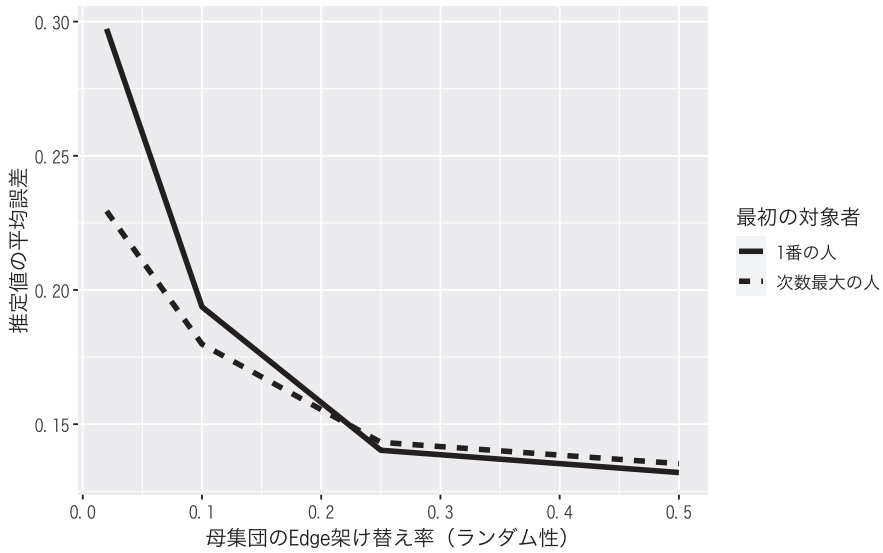


図3 最初対象者・N別平均誤差

図4は母集団での自営率と最初の対象者によって誤差がどう異なるのか示したものである。自営率が低い(つまり母分散が小さい)場合、次数の多い人からサンプリングをスタートしたほうが誤差が小さくなるが、自営率が0.5に近づく(つまり母分散が大きくなる)と、むしろ次数の多い人からスタートしたほうが誤差が大きくなった。表1にしめたように、この交互作用効果はOLSでは有意になる。直感的には常に次数の多い人からスタートしたほうが、さまざまな対象者にアクセスでき、誤差も小さくなりそうにおもすが、自営率が高いとなぜかそうはならない。理由は不明である。

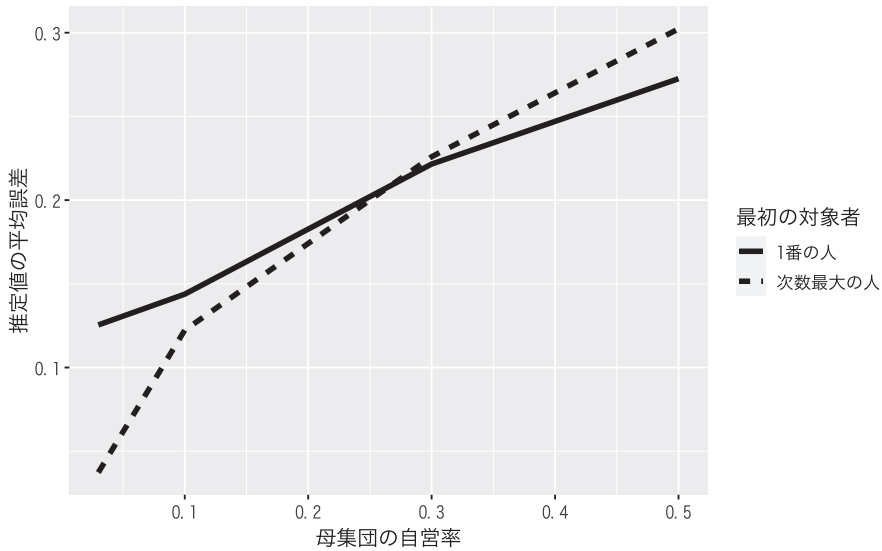


図4 最初対象者・母集団の自営率別

OLS で誤差の対数を予測した結果が表1である。OLS での推定が適切なのか自信がないのだが、いちおうやってみた。モデル1は主効果のみのモデル、モデル2は、すべての独立変数のあいだに一次の（二変数間）交互作用効果を仮定したモデルからスタートして、後退ステップワイズ法で、BIC が最大になるモデルを選んだ結果である。誤差を対数変換したのは、対数変換したほうが残差が正規分布に近似するし、決定係数も上がったからである。

サンプルサイズ n に関しては、予備的分析でそのまま n や $1/\sqrt{n}$ をかわりに投入してみたが、 \sqrt{n} を投入するのがもっとも決定係数が高かったため、これを採用している。比率の推定の際の標準誤差が $\sqrt{p(1-p)/n}$ であることを考えると、 $1/\sqrt{n}$ がもっともあてはまりが良いことが期待されたが、そうはならなかった（従属変数を対数変換しなくても同様の結果だった）。同様に、母自営率 p に関しては、予備的分析で、 p や \sqrt{p} を $\sqrt{p(1-p)}$ のかわりに投入してみたが、 $\sqrt{p(1-p)}$ の場合がもっとも決定係数が高かった。 $\sqrt{p(1-p)}$ も標準誤差からの類推で試してみたが、期待通りの結果であった。エッジの架け替え率 q も予備的分析で $\log q$ の代わりに投入してみたが、 $\log q$ のほうが決定係数が高かった。これについては特に理論的根拠はない。

Model 1 をみると、「次数最大の人からサンプリングをスタート」だけが有意ではなく、その他の変数の効果は有意であることがわかる。ただし、Model 2 の結果を見ると、「次数最大 ...」と三つの変数の交互作用効果が有意になっている。つまり、直感的には次数中

心性の高い人からスタートすると、誤差が小さくなりそうな気がしたが、ケースバイケースでかえって誤差が大きくなる場合もあり、平均的には ID = 1 番の人からスタートした場合と有意な差はないということである。

誤差相関モデルでの推定は、モデル 1 が示すように平均的には誤差が大きくなるが、モデル 2 の交互作用効果が示すように、サンプルサイズが大きいほど、母比率 p が 0.5 にちかづき、母分散が大きくなるほど、そして架け替え率が大きくなる（つまりランダムネットワークに近づく）ほど、誤差が小さくなる。おおむね誤差相関モデルの前提（大きなサンプルサイズと残差の正規分布）が満たされるほど、誤差が小さくなるようである。

表 1 OLS で誤差の対数を予測

	モデル 1	モデル 2
切片	0.74 *** (0.13)	-0.17 (0.19)
次数最大者からスタート	-0.08 (0.06)	-0.26 (0.15)
誤差相関モデル	0.28 *** (0.06)	1.76 *** (0.17)
\sqrt{n}	-0.12 *** (0.01)	-0.01 (0.03)
$\sqrt{p(1-p)}$	5.03 *** (0.24)	8.04 *** (0.42)
$\log q$	-0.24 *** (0.03)	-0.00 (0.05)
次数最大者 × 誤差相関		-0.38 *** (0.08)
次数最大者 × $\sqrt{p(1-p)}$		1.85 *** (0.32)
次数最大者 × $\log q$		0.14 *** (0.04)
誤差相関 × \sqrt{n}		-0.09 *** (0.02)
誤差相関 × $\sqrt{p(1-p)}$		-3.85 *** (0.32)
誤差相関 × $\log q$		-0.26 *** (0.04)
$\sqrt{n} \times \sqrt{p(1-p)}$		-0.37 ***

		(0.06)
$\sqrt{n} \times \log q$		-0.03 ***
		(0.01)
Adjusted R ²	0.68	0.85
N	320	320

3.2 入区間率

図5を見ると、入区間率はサンプルサイズが増えるほど改善するが、二項分布よりも誤差相関モデルのほうが正確であることがわかる。しかし、誤差相関モデルでも入区間率はせいぜい92%程度であり、95%よりもやや低い。つまり、信頼区間の幅が狭すぎることをわかる。

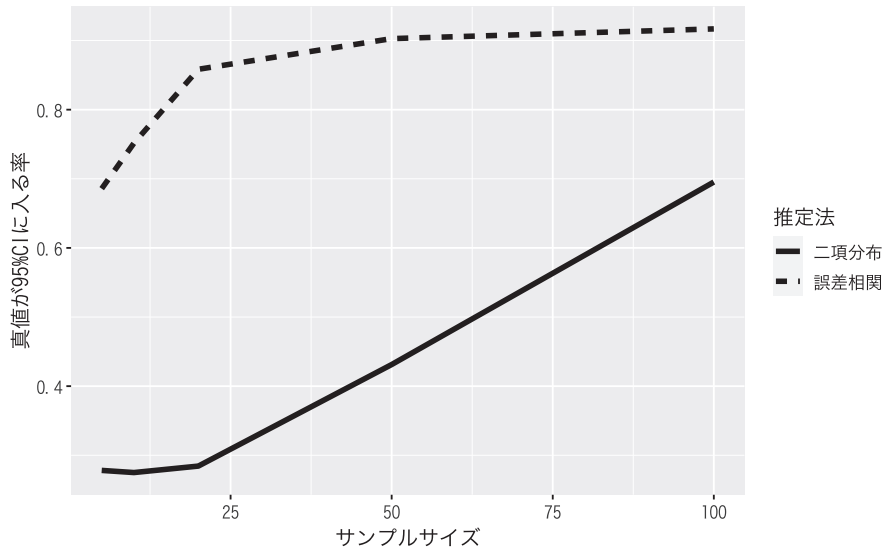


図5 推定法別サンプルサイズと真値が95%信頼区間に入る率

図6を見ると、エッジの架け替え率が大きくなるほど入区間率が上がる傾向があるのがわかる。また次数最大の人からスタートするより、ID = 1番の人からスタートした場合のほうが入区間率は上がるが、その差はせいぜい8ポイント程度であり、エッジの架け替え率や推定法の違いほどの効果はない。

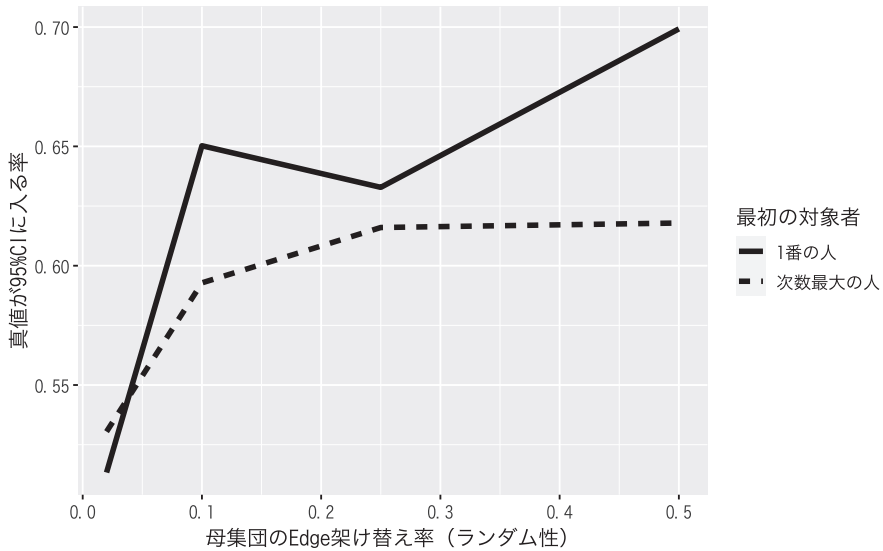


図6 推定法、最初の対象者、架け替え率別真値が95%信頼区間に入る率

図7は母自営率 p が大きくなるほど入区間率が小さくなる傾向を示しているが、これはサンプルサイズが50人以下のときであり、サンプルサイズが100人の場合、母自営率は入区間率にほとんど影響しないことがわかる。

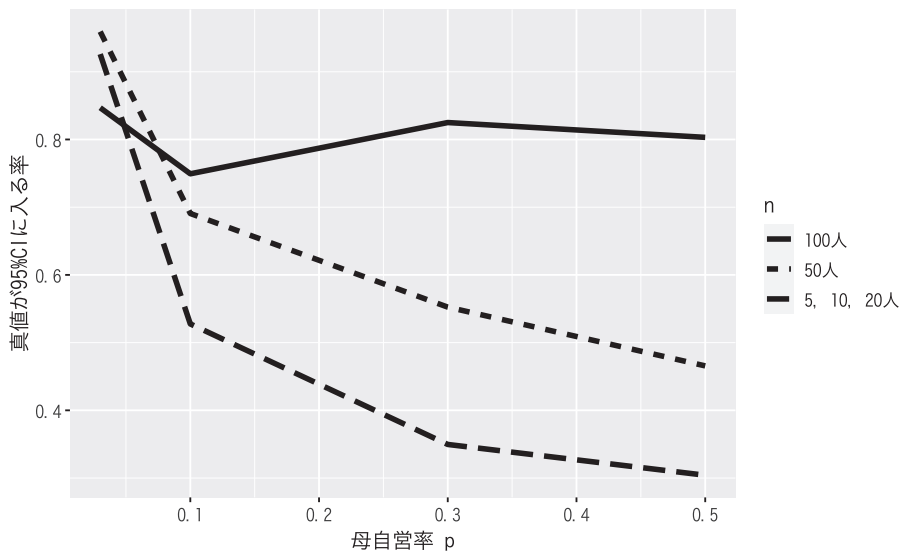


図7 母自営率と n 別真値が95%信頼区間に入る率

表1と同じ要領で、入区間率を推定した結果が表2である。「次数最大者からスタート」は、主効果も交互作用効果もBICの改善に寄与しないため、Model 2からは落ちている。

モデル2の交互作用効果を解釈しよう。誤差相関モデルのほうが入区間率が高いが、この優位はサンプルサイズが大きくなるほど縮まっていくことは、図4で確認したとおりである。

表2 OLSで入区間率予測

	Model 1	Model 2
切片	72.91 *** (5.04)	122.20 *** (5.97)
架け替え率 q	23.22 ** (7.65)	23.24 *** (6.11)
次数最大者からスタート	-2.49 (2.80)	
n	0.32 *** (0.04)	-0.24 * (0.10)
誤差相関モデル	43.73 *** (2.80)	-5.89 (7.02)
$\sqrt{p(1-p)}$	-138.19 *** (10.76)	-293.64 *** (14.99)
$n \times$ 誤差相関		-0.27 *** (0.06)
$n \times \sqrt{p(1-p)}$		1.94 *** (0.24)
誤差相関 $\times \sqrt{p(1-p)}$		167.01 *** (17.18)
Adjusted R ²	0.60	0.75
N	316	316

4 示唆・議論・課題

以上の結果は、いくつかの机上の仮定のもとに成り立っているもので、どこまで現実のサンプリングにあてはまるのかはわからない。しかし、ある程度、現実のサンプリングにも当てはまるのであれば、以下のような示唆が得られたことになる。

1. SSでも標本規模が大きくなれば誤差は減少する。50人ぐらいまでは顕著な減少が期待でき、これはランダムサンプリングと大差ないように思う。英米の質的研究では50人以上にインタビューすることが標準のようであるが、この慣行の合理性が本稿

のシミュレーションで裏付けられたことになる。

2. 比率を推定する場合、点推定値の誤差は、単純な標本比率のほうが、誤差相関モデルよりも小さい。しかし、95% 信頼区間は誤差相関モデルのほうが正確である。誤差相関モデルから得られる 95% 信頼区間を若干広くしたぐらいが、適切な信頼区間となるようである。
3. SS の出発点となる最初の対象者は、次数中心性の高い人であっても誤差は減少しない。こういった人に次の対象者を紹介してもらえれば、簡単に多数の対象者を見つけることができるかもしれないが、これは場合によってはサンプルのゆがみを大きくする場合もある。次数中心性の高い人から SS をスタートするメリットがあるのは、著しく自営率 p が 0 に近い場合や、母集団の誤差相関が強い（自営の知人はほとんど自営、その他の知人はほぼすべてその他）場合だけで、そうでなければむしろ誤差は大きくなる。平均的には、ほとんど自営の知人しかいない人 ($ID = 1$ の自営) からスタートした場合と差がなかった。

以上のような解釈に対しては、以下のような批判や課題が考えられる。

- ・ どの知人を紹介するかはランダム（等確率）であるがこのシミュレーションでは仮定したが、実際には協力してくれそうな人や社会的地位の高い人などが紹介されやすくなる場合もある。このような「協力してくれそう」や「社会的地位」が自営かどうかと相関していれば、これらは標本比率に影響を及ぼしうる。ただし、この問題はけっきょく、対象者が次の対象者を紹介するときに、どの程度の確率で自営を選ぶのか、という問題である。このシミュレーションでは、それは架け替え率 q によって決まるので、「協力してくれそう」や「社会的地位」が、自営が紹介される確率におよぼす影響は q で表現できる、という考え方もできそうである。ただ、いずれにせよ、どんな人が紹介されやすいのか、という問題は、さらに検討の余地があろう。
- ・ 今回はスモールワールドを仮定したが、違うネットワーク構造（例えばスケールフリー（Barabási and Albert 1999））でどうなるかは今後の課題である。
- ・ 次数中心性の高い人からスタートしても、平均的には誤差は減少しなかったが、違う中心性の高い人からスタートすれば誤差を減らせる可能性はある。例えば近接中心性⁽¹⁾の高い人からスタートすれば誤差は減るかもしれない。このような問題も今後の課題である。

⁽¹⁾ 中心性の概念の説明はネットワーク分析の教科書なら必ず扱っているが、R でのプログラミングも含めて鈴木（2017）を参考にした。

- 空間誤差相関モデルは残差が正規分布から離れるほど誤差が増える傾向が見られた。今回利用した推定法は、残差の正規分布を仮定しているので当然の結果だが、誤差相関プロビットモデル (Wang, Iglesias, and Wooldridge 2013) なら、このような問題を回避できるかもしれない。これも今後の課題である。

謝辞

本研究は JSPS 科研費 22K18552 の助成を受けた。

付録：「誤差」と標準誤差の関係

本論では比率について論じているが、ここでは平均について論じる。それは、標準誤差と「誤差」の関係を論じる場合、比率よりも平均について議論するほうが簡単だからである。比率とは 0 または 1 の値をとる二値変数の平均値なので、平均値に関して言えることは、ほぼ比率に関しても一般化できる。一般化できないのはサンプルサイズが小さい場合と、母比率が 0 または 1 に非常に近い場合だけである。

誤差の定義

Y という変数の母平均 \bar{Y} を推測したい。母集団から N 個の事例 i ($i = 1, \dots, N$) を抽出する。これを M 回繰り返す。 M は非常に大きい。 j 番目の標本の Y の平均を \bar{Y}_j ($j=1, \dots, M$)、 j 番目の標本の i 番目の事例の Y の値を Y_{ij} とすると、

$$\bar{Y}_j = \frac{\sum_{i=1}^N Y_{ij}}{N} \quad (1)$$

と定義する。また、「誤差」 $s(Y)$ は、

$$s(Y) = \sqrt{\frac{\sum_{j=1}^M (\bar{Y}_j - \bar{Y})^2}{M-1}} \quad (2)$$

と定義する。

定理

無作為抽出の場合、 M が十分に大きければ、本稿でいう「誤差」は標準誤差に近似する。

定理の証明

標準誤差とは標本平均(本稿の本論では標本比率)の標準偏差である。無作為抽出の場合、標本平均の期待値は母平均と一致する $E(\bar{Y}_j) = \bar{Y}$ 。これを式 (2) に代入すると、

$$s(Y) = \sqrt{\frac{\sum_{j=1}^M (\bar{Y}_j - E(\bar{Y}_j))^2}{M-1}} \quad (3)$$

だが、式 (3) は M が十分に大きければ \bar{Y}_j の標準偏差、すなわち標準誤差に近似する。以上で証明終わり。

文献

- Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439):509–12. doi: 10.1126/science.286.5439.509.
- Bethlehem, Jelke. 2015. "Essay: Sunday Shopping - the Case of Three Surveys." *Survey Research Methods* 9 (3):221–30. doi: 10.18148/srm/2015.v9i3.6202.
- Bethlehem, Jelke, Fannie Cobben, and Barry Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. John Wiley & Sons.
- Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gomez-rubio. 2013. *Applied Spatial Data Analysis with r*. 2nd ed. Springer.
- Butts, Carter T. 2023. *Sna: Tools for Social Network Analysis*.
- Horrace, William C., and Ronald L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model." *Economics Letters* 90 (3):321–27. doi: 10.1016/j.econlet.2005.08.024.
- R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 鈴木努. 2017. 『Rで学ぶデータサイエンス 8: ネットワーク分析 第2版』共立出版.
- 太郎丸博. 2021. 「有意抽出時の標本誤差とサンプルサイズ、母分散の関係」『第71回数理社会学会大会報告要旨集』.
- Wang, Honglin, Emma M. Iglesias, and Jeffrey M. Wooldridge. 2013. "Partial Maximum Likelihood Estimation of Spatial Probit Models." *Journal of Econometrics* 172 (1):77–89. doi: 10.1016/j.jeconom.2012.08.005.
- Ward, Michael D., and Professor Kristian Skrede Gleditsch. 2008. *Spatial Regression Models*. Thousand Oaks: Sage.
- Watts, Duncan J., and Steven H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (6684):440–42. doi: 10.1038/30918.