第 3 回講演会概要

（概要作成：内閣府経済社会総合研究所社会指標ユニット）

講師：三輪哲　東北大学大学院教育学研究科　准教授 タイトル：「インターネット調査の偏りと補正」 日時：平成 26 年 9 月 18 日（木）　13 時～15 時 場所：中央合同庁舎　第 8 号館 4 階 426 会議室

1．はじめに

・最近の社会調査は回収率が低くなっている。回収率の低下の問題点は 2 つある。第 1 は、データが得られないことである。第 2 は、特定層（例えば、若年層など）に偏って回収率の低下が起きていることである。ランダムサンプリングをするのは、日本社会の縮図となるデータを得るためであるが、特定層のデータが回収できないということは、得られたデータが当てにならないということを意味している。

・インターネット調査の利点は、安価で素早いこと（2 千人規模の調査だと、通常の調査では、調査開始からデータの納品までに 2 か月はかかるが、インターネット調査なら約 3 日で納品される)、動画や音楽を入れたり質問が提示される順番をランダムに変えたりする（これにより質問の順番が与える影響を消すことができる）などの工夫ができることである。

・インターネット調査の問題点は、標本の代表性である。

・インターネット調査を、従来型の調査（訪問面接調査や訪問留置調査）に置き換えるべく、データに補正をかけるという発想があるが、私の見解は、補正はある程度可能であるが、限定的である、というものである。

・訪問調査とインターネット調査は、回答者がそもそも異なるため、今まで何十年も継続的に行ってきた訪問調査をインターネット調査に置き換えるのは、経年比較の観点からはお勧めしない。

2. 標本調査における誤差　＜スライド 3 ～ スライド 14＞

・誤差の発生メカニズムを述べる。

・「日本国民全員を捉えたい」という場合、日本国民全員が「目標母集団（理想上の母集団)」となる。しかし、住民基本台帳などに含まれていない人もいるため（「台帳の制約」があるため)、「調査母集団（実在の母集団)」は目標母集団より少し欠けたものになる。「調査母集団」から「標本抽出」をし、「計画標本（調査されるべき対象としての標本)」が出来上がる。ここで大幅に人数が減る。「実査」を行って得られるのが「有効標本（データセットに含まれる、分析可能な標本)」であるが、不在、転居などの理由で、さらに人数が減る。このように、本来知りたいのは「目標母集団」なのに、得られた「有効標本」は少数である。少ない「有効標本」だと「目標母集団」を十分に予測できない可能性がある。

・標本調査がもてはやされたのは、少ない標本を調べただけでも、母集団のことが「ある

程度」わかるからである。重要なのは「どの程度正確か」と「誤差をどう評価するか」、である。

・「誰が答えるか」に関わる誤差は以下の 3 つがある

第 1 は、「カバレッジ誤差」である。既述のように、台帳の制約から生じる誤差である。目標母集団に含まれているはずなのに、台帳等に記載されていないために、調査母集団に含まれていないケースである。台帳の掲載者と未掲載者で回答に質的な違いがある場合、重大な意味を持つことになる。また、無作為に固定電話に電話して対象者を選ぶ場合も、固定電話を持たない人が調査母集団に含まれていないため、カバレッジ誤差が生じていると言える。

第 2 は、「標本誤差」である。つまり、母集団のごく一部といえる標本を調べることに起因する誤差である。ただし、標本が無作為に抽出できていれば、標本誤差の出る確率は推計できるため、「飼いならされた誤差」とも言える。通常、統計的手法が評価する誤差とは標本誤差のみを指している。

第 3 は、「非回答誤差」である。回答者が調査に協力してくれないこと（例えば、居留守、調査拒否等）によって生じる誤差である。これは回答者と非回答者で、回答傾向が全く異なると大きな問題となる。

・なお、誤差は、以下の 2 つに分けられる。

第 1 は、「狭義の誤差」である。系統だった差ではなく、たまたま今回はプラスになった、たまたま今回はマイナスになったというような、ランダムなブレである（例えば、「生活満足度」の回答が、その時の気分によって変わってしまうことである）。この場合、推計値の精度は低下するが系統だった差は生じない。

第 2 は、「偏り」である。これはどちらか一方へと影響するものであり、系統だった差が生じてしまう（例えば、「生活満足度」を聞かれた時に、調査員に対して良い顔をしたいので、いつも「満足」と答えてしまうことである）。この場合、推計値に体系的な誤りが生じてしまう。

・以上を踏まえると、インターネット調査と、従来型調査の違いは、「偏り」としてとらえるべきである。そう考える理由は 3 つあり、第 1 は登録モニターに調査依頼をすることによってカバレッジ誤差が生じていること、第 2 は母集団リストがないので無作為抽出が出来ないこと、第 3 は先着順に回答を締め切っていること、である。特に第 3 は、非回答者をどう評価するかのむずかしさもある。これらを鑑みると、インターネット調査は「やる気のある人」が回答するものであり、インターネット調査と従来型調査の違いは、「狭義の誤差」というよりも「偏り」と呼ぶべきものである。

3. 比較調査設計と全体的な差異 ＜スライド 15 ～ スライド 44，以下では適宜スライド番号を記載する＞

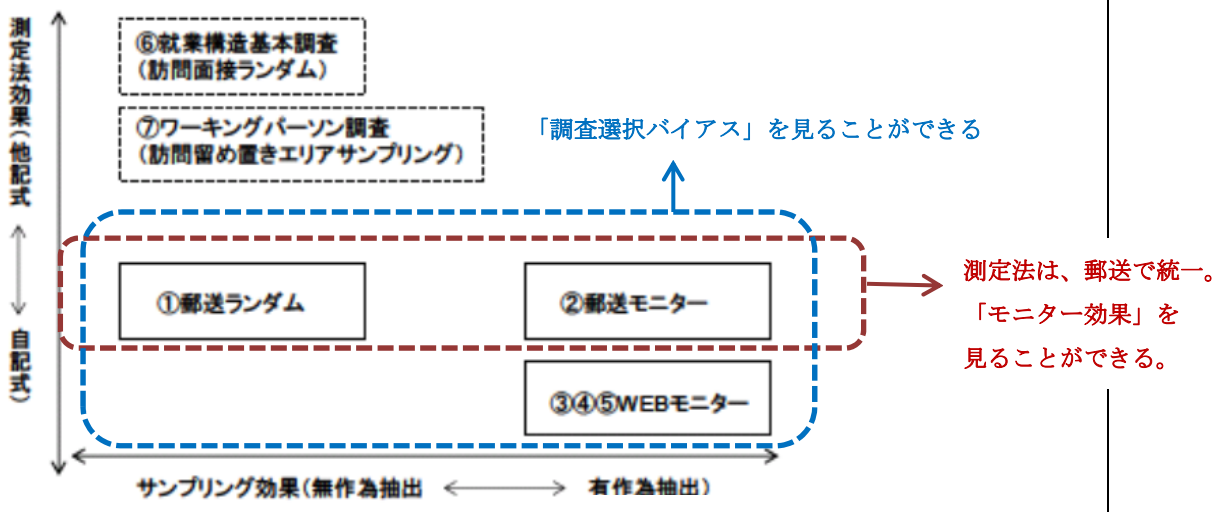
・以下では、東大社研とリクルートワークス研究所が行った共同調査の結果や、その他の既存の調査を用いる。

・東大社研とリクルートワークス研究所が行った共同調査では、下図の①から⑤の調査を実施した。

・本発表では、「インターネット調査の偏り」が、以下の 3 つを要因としているのかを検証する。
第 1 は「調査選択バイアス」である。どのような調査法にするか（WEB 調査か郵送調査か）によって、回答する人が違っているのではないか、ということである。（下図の①②③④⑤を参照）。
第 2 は「モニター効果」である。同じ調査法でも、モニター登録者と無作為抽出された人とでは、回答する人が違っているのではないか、ということである（下図の②と③④⑤を参照）。
第 3 は「モード効果」である。同じ有作為の回答者であっても紙に書くのとパソコンで入力するのでは回答が違うのではないか、ということである。

◆比較した調査について　（萩原, 2009）

＜スライド 16＞



注）黒の点線による枠線は既存の調査

図 1　調査の基本設計

・MDS（Multi Dimensional Scaling）分析を行った。結果を次ページに示すが、MDS 分析では、図の点が近いほど、相対的に似ていることを意味している。結果、3 つの WEB モニター調査から成る塊と、郵送調査の塊ができた。
・郵送の場合は、郵送モニター調査と郵送ランダム調査の距離が近しいことから、サンプリングの影響は乏しいことがわかる。
・比較対照群として、訪問留置法で行ったワーキングパーソン調査（WPS）から 600 人ずつ 30 個のデータを抜き出したが、ワーキングパーソン調査の塊に WEB モニター調査の塊や郵送調査の塊が入らない。つまり、WEB モニター調査と、訪問留置調査、郵送調査は、全く別ものであると言える。
・つまり、WEB モニター調査は、訪問調査の代表品にならないのである。
・では、なぜこれほどまでに違うのか？調査法によって、答えている人が違うのだろうか？

◆ＭＤＳ分析の結果（点が近いほど相対的に似ている）

＜スライド 21＞

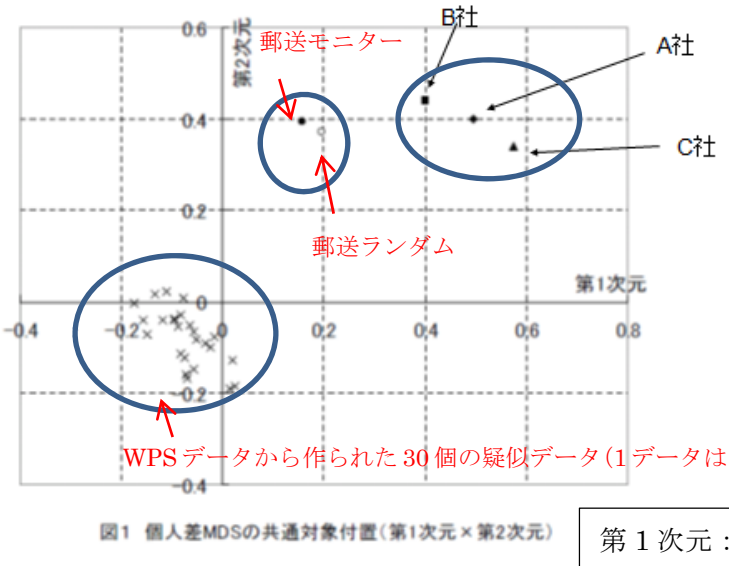


図1　個人差MDSの共通対象付置（第1次元×第2次元）

第 1 次元：仕事の意識
第 2 次元：基本属性

4. 答える人の違い

◆調査選択バイアス

どのような調査法にするかによって、回答者が違っているのではないか？＜スライド 26＞

・「今後、調査に協力しても良いと考える調査法」について、郵送ランダム調査の回答者、郵送モニター調査の回答者、3 つの WEB モニター調査の回答者に聞いた。その結果、郵送ランダム調査や郵送モニター調査の回答者は、「郵送調査に協力してもよい」の回答率が高く（郵送ランダムでは 39.7%、郵送モニターでは 60.1%）、WEB モニター調査の回答者は「WEB 調査に回答してもよい」の回答率が高い（3 つの WEB モニター調査すべてで 9 割を超えている）。

・郵送ランダム調査と郵送モニター調査を比較すると、郵送モニター調査は全体的に調査に協力的である。

・以上から、調査法によって回答者は異なる、と言える。

◆モニター効果
同じ調査法でもモニター登録者と無作為抽出された人とでは、回答する人が違っているのではないか？　＜スライド 27 ～ スライド 28＞

・スライド 27 とスライド 28 は、郵送ランダム調査と郵送モニター調査の結果を比較し、両者で統計的有意差が認められたもののみ数値を記載した。
・郵送ランダム調査と郵送モニター調査を比較した場合、属性に関しては「学歴」しか統計的有意差はない。しかし「社会意識・態度」に関しては統計的有意差がある。このようにモニター効果は、属性には働かず、意識に関する質問には働く。
・つまり、属性的には郵送ランダム調査の回答者と郵送モニター調査の回答者は似ているが、意識は両者では異なるのである。
・よって従来型調査を WEB モニター調査に変えるのは難しいと思われる。

5. 答え方の違い

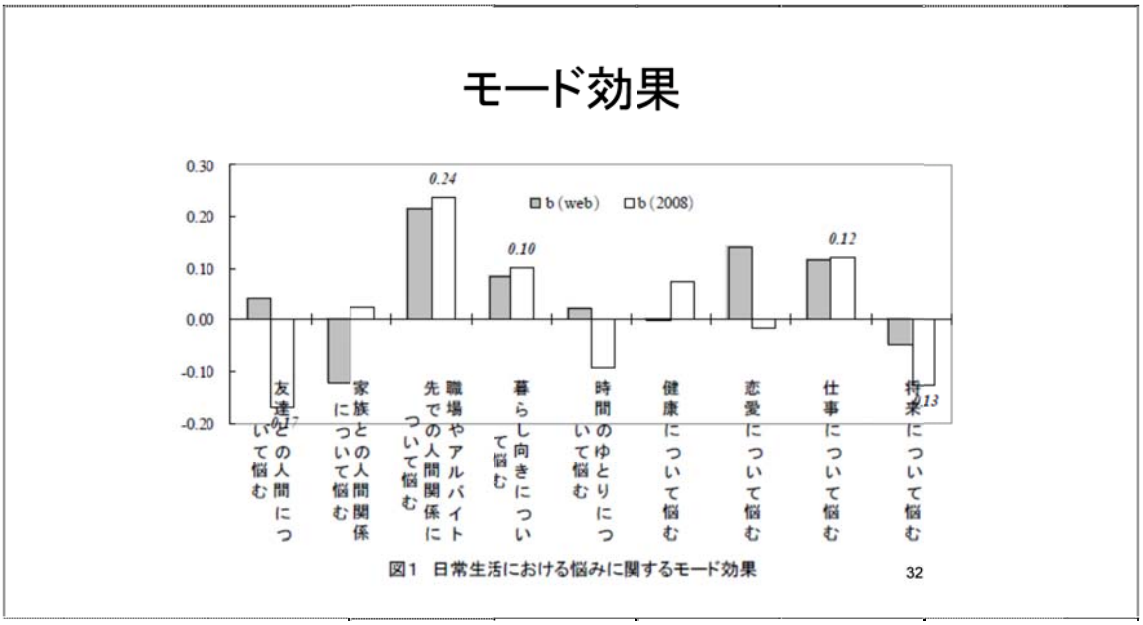
◆モード効果
同じ回答者であっても紙に書くのとパソコンで入力するのでは回答が違うのではないか

・東大社研高卒パネル調査を使って検証した。
・本パネル調査では 2006 年から、質問紙か WEB か、回答法を選択できるようにした。以下では、B さん、C さんのように、回答法が変わった人を抜き出して固定効果モデルを用いて分析する（ただし、B さん、C さんのように回答モードが変わった人は、少数である）。

	2006 年調査（24 歳時点）	2008 年調査（26 歳時点）
A さん	質問紙で回答	質問紙で回答
B さん	質問紙で回答	WEB で回答
C さん	WEB で回答	質問紙で回答
D さん	WEB で回答	WEB で回答

・次ページの黒いバー（WEB で回答したか紙に回答したかによる回答差）を見たところ、両者には統計的な有意差はない。つまり、質問紙調査と比べ、WEB 調査では賛成しやすいとか反対しやすいということはない。モード効果はなく、紙か WEB かは「偏り」ではなく「誤差」の範囲と言える。
・なお、図の白いバー（2006 年調査と 2008 年調査の回答傾向の違い）を見たところ、統計的有意差のある項目もあった。これは 24 歳時と比べ 26 歳時の生活環境が変化したことが影響を与えていると思われる。

＜スライド 32＞



注）
・黒いバーがプラスの場合：質問紙よりも、WEB の方が「そう思う」と回答する可能性が高い場合
・黒いバーがマイナスの場合：質問紙よりも、WEB の方が「そう思わない」と 〃
（質問紙と WEB を比べ、統計的有意差が出たものだけ数値を記載）

・白いバーがプラスの場合：2006 年に比べ 2008 年の方が「そう思う」と回答する傾向が強くなった場合
・白いバーがマイナスの場合：2006 年に比べ 2008 年の方が「そう思わない」と 〃
（2006 年と 2008 年を比べ、統計的有意差が出たものだけ数値を記載）

6. 補正の可能性

◆補正は時にうまくいく ＜スライド 36 ～スライド 37＞
・補正は、うまくいく時もあればうまくいかない時もある。
・『家族社会学研究』に掲載されたパネルデータ分析である。
・パネル第 1 派に比べパネル第 3 波では脱落者が出た。第 3 波データを、いかに第 1 波データに近づけるか IPW（Inverse Probability Weighting）で試みた。ウェイトを使用し補正をして、データが修正された好例である。

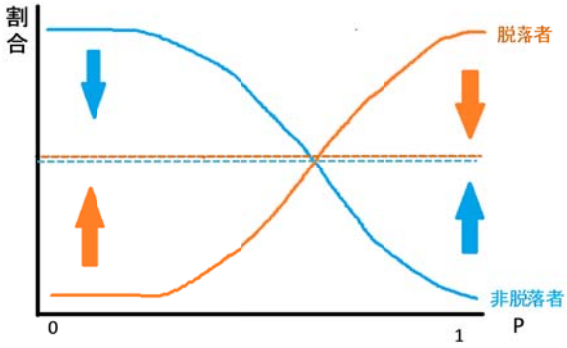
◆傾向スコア法

＝ロジスティック解析、プロビット分析等によって算出された予測確率を用いた補正法

（性別、学歴、子どもの有無等、たくさんの変数情報を１つの傾向スコアにまとめる）

＜板書＞

いかにも継続しそうな人の
データを低く見積もる。↓



いかにも脱落しそうな人の
データを高く見積もる。↑

注）

横軸 P：0～1 をとる。その調査から脱落しそうか否か。
（1 はいかにも脱落しそうな人、0 はいかにも脱落しなさそうな人）

縦軸：そうした人が占める割合。

赤のライン：脱落者。0 の人（脱落しなさそうな人は少なく、1 の人（脱落しそうな人）は多い。

青のライン：非脱落者（調査継続者）。　0 の人は多く、1 の人は少ない。

・P（どれだけ脱落しそうか）の値が青のライン、赤のラインで異なることが問題。

・傾向スコアを用いた補正法とは、

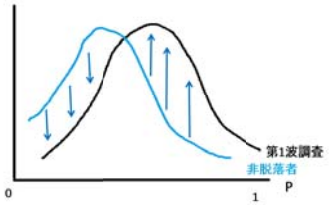
青（非脱落者）のラインに関して、

P が 1 に近い人（いかにも脱落しそうな人。例えば、若くて都会のオートロックマンション在住等）はウェイトをかけて高く見積もる。

P が 0 に近い人（いかにも脱落しなさそうな人。例えば、真面目で田舎在住で女性等）はウェイトをかけて低く見積もる。

赤（脱落者）のラインに関しても、同様のことをすると、青（非脱落者）のラインと P の分布が重なってくる。

・ただし、今回の補正に関しては、現実には、脱落者のデータはないので、脱落前（第 1 波調査）のフルサンプルを使って、非脱落者（調査継続者）のデータを補正した。

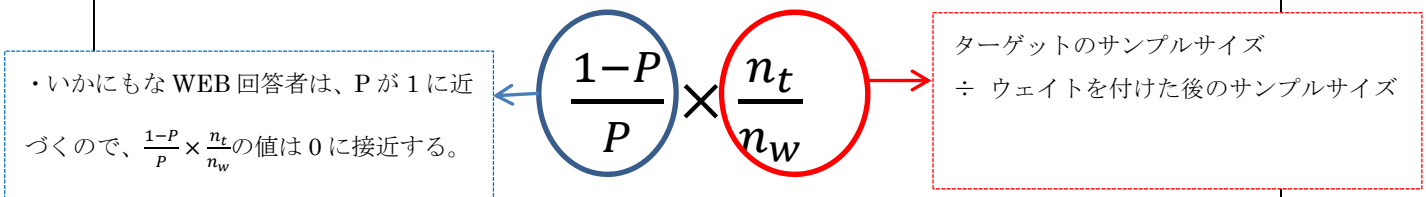


◆傾向スコアのセオリーの出し方　＜スライド 38＞

・星野崇宏氏（東京大学教育学研究科准教授）の提唱する傾向スコアを用いた補正は以下の通りである。
・まず第 1 段階として、「予備的実験調査（従来型調査と WEB 調査）」と「本調査（WEB 調査）」を行う。
・第 2 段階として、誰が WEB 調査に回答して、誰が従来型調査に回答するのかを、プロビット分析やロジスティック回帰分析で求め、予測を向上させるような変数セットを見つける。
・第 3 段階として、本調査として WEB 調査を行う。
・第 4 段階として、実験調査データと本調査データを使って傾向スコアを算出する。
・第 5 段階として、傾向スコアを使ってウェイトをかけることにより本調査データを補正する。

＜板書＞

傾向スコアによる補正式（ウエイトのかけ方）



◆クロスバリデーションを用いた傾向スコアの出し方　＜スライド 39＞

・セオリーの出し方（星野・森本）とは異なり、本調査を実施しなかった。
・第 1 段階として、予備的実験調査として、同じ質問から成る調査（ワーキングパーソン調査）を、「訪問留置調査」と「WEB 調査」で実施した。
・第 2 段階・第 3 段階として、訪問留置調査データと WEB 調査データを合成し、ランダムに 2 分した（＝クロスバリデーションを行った）。半分のデータセットは、傾向スコアを求めるための群であり、残り半分のデータセットは傾向スコアを用いて補正するための群（検証用データ）とした。
・第 4 段階として、上記の 2 郡のデータ（実験調査データと検証用データ）を用い、傾向スコアを算出した。
・第 5 段階として、傾向スコアをウェイトとして用い、検証用データを補正した。

◆クロスバリデーションを用いた傾向スコア法による補正の**結果①** ＜スライド 40 及び 45＞

・補正前と補正後の二乗誤差を見ると、どの質問項目も、補正後数値＜補正前数値であり、補正はおおむねうまくいっていると言える。

◆クロスバリデーションを用いた傾向スコア法による補正**の結果②** ＜スライド 41 及び 46＞

・WEB 調査の結果が従来型調査（訪問留置調査）の結果にどれくらい近づけたかを、各質問の平均値で見た。

・◆は補正前の数値で、■は補正後の数値である。対角線のライン（45 度線）は、留置調査と WEB 調査の平均値が一致することを意味するラインであり、■（補正後）が◆（補正前）よりも対角線に近づいていれば（平均値の一致に近づけば）、補正の成功を意味する。逆に、■（補正後）が◆（補正前）よりも対角線から遠のけば（平均値の一致から遠のけば）、補正の失敗を意味する。

・補正によって、狙っていた値（平均値）に近づけたか（45 度線に近づけたか）をみると、ほぼ成功している。

◆まとめ - 補正の可能性 ＜スライド 42 ～ スライド 44＞

・クロスバリデーションを使った補正によって、約 45%、誤差の 2 乗和を減少させることができた。ゆえに補正はある程度有効と言える。（ちなみに星野・森本は 68%誤差を減少させることができた）。しかし、クロスバリデーションで用いた傾向スコアを使って、全く別のデータを補正した場合、誤差の減少は 23%にとどまった。つまり、一度、傾向スコアを算出すればどんなデータにでも使える、というわけではない。

・楽観的なことを言えば、もっと良い傾向スコアの出し方（Imbens や Rubin のガイドライン）を参照すれば、もっとより良い補正ができる可能性はある。また「調査にどれくらいやる気があるか」の変数を使えれば、補正はもっとうまくいくと思われる。

・悲観的なことを言えば、調査票に含まれている変数しか補正に利用できないのだが、補正のためだけに用いる質問項目を調査票に入れる余裕は現実的にはない。また、補正の参照基準となる従来型調査が正しい結果かどうかはわからない。また、補正がうまくいくのは、補正のために使える情報は多く、補正をかけたい変数が少数の場合である。反対に、学歴・年齢・地域といった少ない変数で、多くの変数の補正をすることは（多くの対象を少ない情報で補正をするのは）難しい。傾向スコアには、1 個の変数に、性別、学歴、婚姻といったたくさんの変数の偏りを生み出す要因が集約されているのである。

・以上から、「補正はある程度可能だが、かなり限定的である」という楽観的ではない結論にたどり着いた。ただし、この見解は WEB 調査を否定するものではない。ただ、今まで何十年も継続的に行ってきた訪問調査をインターネット調査に置き換えるのは、継続性や経年比較の観点からはお勧めしない。

以上。