

調査観察データの統計科学ゼミ-第 18-a 回-

Masamichi Ito

Osaka University Graduate School of Human Sciences
Adachi Lab M1

February 13, 2020

① 6.1 インターネット調査について

② 6.2 選択バイアスとしての理解

③ References

① 6.1 インターネット調査について

② 6.2 選択バイアスとしての理解

③ References

インターネット調査の普及

近年、インターネット調査はメジャーに。しかし...

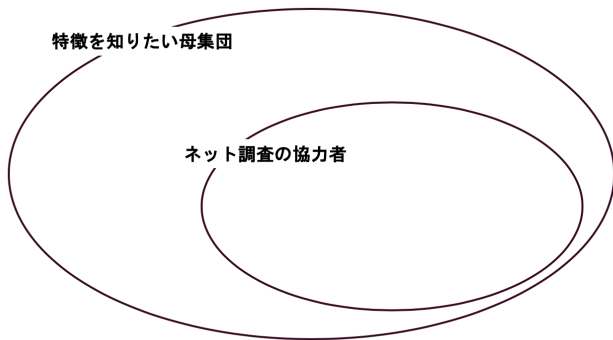


Figure: ネット調査協力者集団の偏り

ネット調査では、バイアスがかかる
→ バイアス補正が必要

インターネット調査の普及

インターネット調査の利点

- ① 訪問調査などの従来型調査に比べてコスト低い
← 費用が回答者数に比例しない
- ② 回収や集計結果が得られるまでのスピードが非常に早い
- ③ 回答に応じたインタラクティブな調査が可能
→ モニター調査や会場テストの有力な代替案となる

→ これまで市場調査を行っていなかった中小企業が実施を検討したり、大手企業でも定例的な調査に加えて実施されるようになった。

さらに、

- 社会状況やニーズの動向、流行が非常に短期間に変化していく中で、市場の変化を素早く理解し行動に移すことの重要性が高まってきた。
- 国民のプライバシー意識の向上により、訪問調査への拒否率上昇、住民基本台帳の閲覧実質無理

クローズ型パネル調査について

インターネット調査は大別して、**オープン型**と**クローズ型**に分類できる。

オープン型

調査を行っていることをホームページで広告し、偶然その HP を閲覧した回答者が自主的に回答する形式。**特定の期間中に目標とするサンプルサイズを得ることが保証されず、どんな人からの回答が得られるか事前に予想することができない**ため、市場調査ではあまり利用されていない。

クローズ型

後述する 4 つのステップを経て行う調査方法であり、**リソースタイプの調査**(大隅, 2002a) とも呼ばれる。

クローズ型パネル調査について

クローズ型パネル調査は、以下の4つの段階を経て実行される。

クローズ型パネル調査の実行

- ① 調査会社が事前に「様々な調査に協力する可能性があるモニター」をリクルートする。応募してきたモニターのリストを調査協力者のパネル（以降、調査協力パネルと呼ぶ）として整備する。
- ② 調査案件ごとに、調査の目的に応じて調査パネルから調査回答者の候補を抽出する。
- ③ 調査回答者の候補に対してメール等で調査依頼をする。
- ④ 依頼に応じた調査回答者の候補に対してメール等で調査依頼をする。

テキストでは、こちらのインターネット調査に限定した議論が行われている。

1 6.1 インターネット調査について

2 6.2 選択バイアスとしての理解

3 References

確率抽出と有意抽出

インターネット調査は流行ってはいるが、**標本抽出の代表性という観点では問題アリ**。バイアスすごいのである。その説明の前に...

確率抽出 (probability sampling)

各調査者毎に抽出される確率 (**包含確率**(inclusion probability)) を事前に設定し、その確率で抽出する。包含確率を利用した重み付け集計を行うことで母平均等の不偏推定量を計算可能

有意抽出 (purposive sampling)

確率抽出を行わない抽出法で、大きくは、

- 広告に応募した人を標本とする**募集法**
- 紹介を頼って回答者をふやす**縁故法**, **スノーボール法**
- 標本として選ぶべき対象の属性とサンプルサイズを事前に指定し、その条件が満たされるまで選出する**割り当て法**

と分類される。

クローズ型インターネット調査での3つのバイアス

(続き) 従来型の調査は**確率抽出**，ネット調査は**有意抽出**(特に募集法)

ここで，クローズ型インターネット調査には3つの選択バイアスが生じると考えられる。(図 6.1 を見ながら)

- ① 調査協力パネルに参加する協力者の**候補をサンプリングする際**の選択バイアス
(無作為抽出やエリアサンプリングしてない，募集法，professional respondents)
- ② 協力者の候補が**調査協力パネルに参加する時点**での選択バイアス
(謝礼が十分か，回答者は忙しくないか)
- ③ 調査協力パネルへの登録者が各調査への**回答に応諾するかどうか**によって生じる選択バイアス
(対象者の忙しさや，調査への関心で，実際に回答してくれるかどうか変わる)

3つのバイアス (続き)

(続き)

- この3つのうち、「1. 候補をサンプリングする際のバイアス」は調査会社の努力で低減可能.
- 「インターネットの普及率が向上すれば、ネット調査のバイアスの問題はなくなる？」
→ そんなことはない. 特に 2,3 は個人の回答への協力という自由意志の問題

→ バイアスを取り除く方法を考えるべきである.

欠測データの枠組みによるバイアスの理解

有意抽出によるバイアスも欠測データの枠組みを用いて明確化可能

	ネット調査の標本($z = 1$)	無作為抽出の標本($z = 0$)
ネット調査での回答: y_{web}	ネット調査でのネット調査の結果(1)	無作為抽出標本でのネット調査の結果(2)
従来型調査での回答: y_{real}	ネット調査標本での従来型調査の結果(3)	無作為抽出標本での従来型調査の結果
共変量項目	両群で共通	両群で共通

Figure: 図 6.2 ネット調査と従来型調査

濃い灰色の部分はネット調査だけ実施された場合は欠測となる。
ネット調査では、(1)の部分だけが観測される、でも知りたいのは、(4)

欠測データの枠組みによるバイアスの理解

欠測データの枠組みで考えると、ネット調査と従来型調査の違いは、

- ① 標本の違い (両者の抽出方法の違い)
インターネット調査の標本から得られたデータは先ほど述べた 3 つの段階からなる選択バイアスを有する
- ② 回答方法の違い (一般には, $y_{web} \neq y_{real}$)
調査モードの違いと呼ばれるもの. 具体的にはネット調査でのラジオボタンのクリックによる回答と、訪問調査での多岐式の回答では異なる可能性がある.

欠測データの枠組みによるバイアスの理解

所有や経験など、事実に基づく項目であれば、あまりこの調査モードの効果は存在しないと考えて良い。

→ 調査モードの効果は存在しない ($y_{web} = y_{real}$) とすれば、図 6.2 の (4) と (2)(同様に (1) と (3) も) は等しいと考えられる。

この時 (1) $E(y_{web}|z=1)$ から、(4) $E(y_{real}|z=0)$ をなるべくよく予測する問題は、(1) から、(2)「もし無作為抽出標本がインターネット調査に回答した場合の回答」 $E(y_{web}|z=0)$ をできるだけ精度良く予測する問題に置き換えられる。

欠測データの枠組みによるバイアスの理解

この問題は、3.6 節の表と同じ問題構造である。→TET の考え方を応用すれば良い。よって、(2) は、

$$\hat{E}(y_{web}|z=0) = \frac{\sum_{i=1}^N \frac{z_i(1-e_i)}{e_i} y_{webi}}{\sum_{i=1}^N \frac{z_i(1-e_i)}{e_i}}$$

で推定可能である。二重にロバストな推定量は、

$$\hat{E}^{DR}(y_{web}|z=0) = \frac{1}{N} \sum_{i=1}^N \left[\frac{z_i(1-e_i)y_{webi}}{e_i} + \left(1 - \frac{z_i(1-e_i)}{e_i} g(x_i, \hat{\beta}_1) \right) \right]$$

となる。ここで、 $g(\mathbf{x}, \hat{\beta}_1)$ は、 $z=1$ での y_{web} の \mathbf{x} への回帰関数であり、 $E[y_{web}|z=0, \mathbf{x}] = E[y_{web}|z=1, \mathbf{x}] (= g(\mathbf{x}, \hat{\beta}_1))$ となること (SI 条件の成立) が必要である。

① 6.1 インターネット調査について

② 6.2 選択バイアスとしての理解

③ References

- 岩崎 (2015) 統計解析スタンダード 統計的因果推論 (→ おすすめ. 読みやすい文章構成)
- 高井, 星野, 野間 (2016) 「調査観察データ解析の実際 欠測データの統計科学-医学と社会科学への応用-」
- 星野 (2009) 「調査観察データの統計科学-因果推論・選択バイアス・データ融合-」

Thank you for your attention!!

Any Questions?

メモに使ってね！！