

不良回答検出のためのアンケート設計方法について

伊藤健太郎[†] 須子統太[†] 小林学[‡]早稲田大学社会科学部[†] 早稲田大学データ科学センター[‡]

1. はじめに

近年, Web アンケート等によるアンケート調査が様々な場面で活用されている. しかし, Web アンケートには不良回答が混入することが多く, アンケートの信頼性に問題がある. 不良回答に対する一般的な対処方法として, IMC(Instructional manipulation checks) や DQS(Directed Questions Scale)といったアンケート設計時に設問を追加することで不良回答を検出する手法が提案されており, 実際のアンケートに対して適用することで実験的に検出確率について議論されている[1][2][3].

本研究では, 不良回答混入モデルを仮定したもとの, 不良回答検出のためのアンケート設計方法の統一的な枠組みを示し, いくつかの具体的な設計方法について回答検出率の理論的な評価を行う.

2. 設問と不良回答モデルについて

2.1. 情報設問

本来入手したい情報のために行うアンケート質問のことを情報設問と呼ぶ. 本研究では簡単のため「はい」, 「いいえ」の2択の設問が K 個ある場合を想定する. k 番目の設問に「はい」と回答する場合を $x_k = 1$ と表現し, 「いいえ」と回答する場合を $x_k = 0$ と表現する. この時ある回答者の全ての情報設問に対する回答列を $\mathbf{x} = (x_1, x_2, \dots, x_K)$ と表記する.

2.2. 検査設問

情報設問への回答が不良回答かどうかを検出するために付与した設問のことを検査設問と呼ぶ. 本研究では検査設問も簡単のため「はい」, 「いいえ」2択とし, 検査設問が N 個ある場合を考える. n 番目の設問に「はい」と回答する場合を $c_n = 1$ と表現し, 「いいえ」と回答する場合を $c_n = 0$ と表現する. この時, ある回答者の全ての検査設問に対する回答列を $\mathbf{c} = (c_1, c_2, \dots, c_N)$ とする.

2.3. 不良回答モデル (ランダム回答モデル)

設問に対して「はい」「いいえ」の選択肢についてランダムに回答するものを本研究の不良回答モデルとする. このとき不良回答の各設問に対する回答確率について次式が成り立つと仮定する.

$$\Pr\{x_k = 1\} = p, \quad k = 1, \dots, K, \quad (1)$$

$$\Pr\{c_n = 1\} = p, \quad n = 1, \dots, N, \quad (2)$$

3. 検査設問の作成方法

n 番目の検査設問について情報設問の全回答列 \mathbf{x} から関数 $c_n = f_n(\mathbf{x})$ で一意に正しい回答 c_n が定まる検査設問を作成することとする.

具体例として本研究では基本的な検査パターンとして以下の6通りを考える.

- ① n 番目の検査設問を必ず「はい」と回答するように指定(IMC, DQS に対応)

$$f_n(\mathbf{x}) = 1 \quad (3)$$

- ② n 番目の検査設問を必ず「いいえ」と回答するように指定(IMC, DQS に対応)

$$f_n(\mathbf{x}) = 0 \quad (4)$$

- ③ n 番目の検査設問の回答を k 番目の情報設問の回答と異なる回答を選択するように指定

$$f_n(\mathbf{x}) = 1 - x_k \quad (5)$$

- ④ n 番目の検査設問の回答を k 番目の情報設問の回答と同じ回答を選択するように指定

$$f_n(\mathbf{x}) = x_k \quad (6)$$

- ⑤ i 番目と j 番目の情報設問の回答の両方が「はい」の時, n 番目の検査設問を「いいえ」と回答するように指定. または i 番目と j 番目の情報設問の回答が異なる時, n 番目の検査設問を「はい」と回答するように指定. または i 番目と j 番目の情報設問の回答の両方が「いいえ」の時, n 番目の検査設問を「いいえ」と回答するように指定.

$$f_n(\mathbf{x}) = x_i + x_j \text{ mod } 2 \quad (7)$$

ただし mod 2 は 2 で割った余りを求める演算を表す.

- ⑥ i 番目と j 番目の情報設問の回答の両方が「はい」の時, n 番目の検査設問を「はい」と回答するように指定. または i 番目と j 番目の情報

Method of detecting defective responses from a questionnaire data on the basis of questionnaire design

[†] Kentaro Ito, Tota Suko

School of Social Sciences, Waseda University

[‡] Manabu Kobayashi

Center for Data Science, Waseda University

設問の回答が異なる時、 n 番目の検査設問を「いいえ」と回答するように指定。または i 番目と j 番目の情報設問の回答の両方が「いいえ」の時、 n 番目の検査設問を「はい」と回答するように指定。

$$f_n(x) = x_i + x_j + 1 \pmod{2} \quad (8)$$

以上のような規則で作成した検査設問に対して、情報設問の回答列 \mathbf{x} に対応する正しい検査設問の回答列 $\mathbf{f}(\mathbf{x})$ を次式で定義する。

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_N(\mathbf{x})) \quad (9)$$

従ったある回答者の情報設問の回答列が \mathbf{x} で検査設問の回答列が \mathbf{c} だったとき、 $\mathbf{f}(\mathbf{x}) \neq \mathbf{c}$ の場合、その回答列を不良回答とみなす。

4. 不良回答検出率

不良回答モデルに従ったある不良回答 \mathbf{x}, \mathbf{c} に対し、次式で不良回答検出率 L を定義する。

$$L = Pr\{\mathbf{x}, \mathbf{c} | \mathbf{f}(\mathbf{x}) \neq \mathbf{c}\} \quad (10)$$

ただし \mathbf{x} と \mathbf{c} はそれぞれ情報設問と検査設問の回答列に対する確率変数ベクトルを表す。

5. 数値計算

$K = 2, N = 1$ で確率 p を変動させた際の各パターンの不良回答検出率を図1で示す。

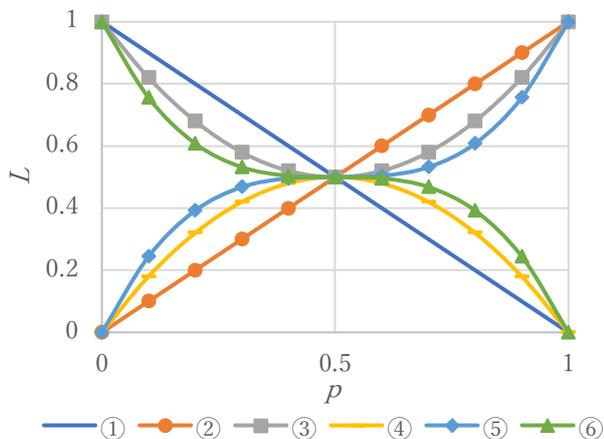


図1: 検査設問を1つ付与した際の不良回答検出率の変化 (数字は検査パターンに対応)

図1より検査設問を1つ付与する際は、 $p < 0.5$ の場合はパターン①を、 $p > 0.5$ の場合はパターン②を付与するのが最も回答検出率が高くなるが、種々の p を考慮すると③を付与する場合は平均的に良いことが分かる。

次に $K = 4, N = 2$ で確率 p を変動させた際、検査パターンの代表的な組み合わせに対する不良回答検出率を図2に示す。図2より検査設問を2つ付与した際に対象とする情報設問に重複がない場合、 $p < 0.5$ の場合はパターン①を2つ、 $p >$

0.5 の場合はパターン②を2つ付与するのが最も回答検出率が高くなるが、種々の p を考慮すると③を2つ付与する場合は平均的に良い。

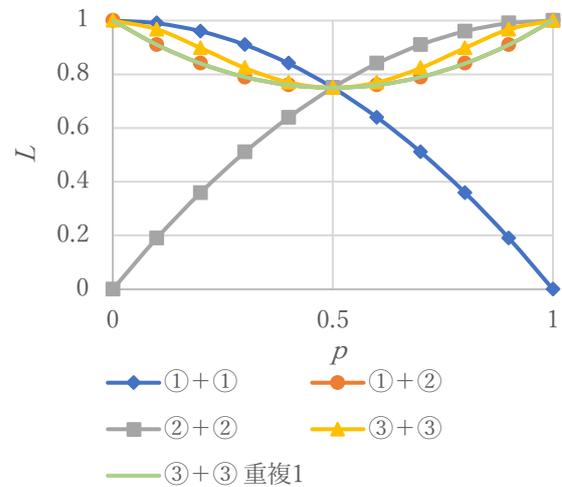


図2: 検査設問を2つ付与した際の不良回答検出率の変化

6. まとめ

本研究では不良回答モデルを仮定したもとの、不良回答検出のためのアンケート設計方法の統一的な枠組みを示し、いくつかの具体的な設計方法について不良回答検出率の評価を行った。

今後の課題は選択肢が「はい」、「いいえ」の2択以外の多択の設問に対しても対応可能な、より普遍的な枠組みの提案が挙げられる。

謝辞

本研究の一部は、日本学術振興会科学研究費基盤研究(C)一般(No. 21K11796), (No. 19K04914)により行われた。

参考文献

- [1] Oppenheimer, D. M., Meyvis, T., & Davidenko, N., "Instructional manipulation checks: Detecting satisficing to increase statistical power," *Journal of Experimental Social Psychology*, 45, pp. 867–872, 2009.
- [2] Maniaci, M. R., & Rogge, R. D., "Caring about carelessness: Participant inattention and its effects on research," *Journal of Research in Personality*, 48, pp. 61–83, 2014.
- [3] 三浦麻子, 小林哲郎, "オンライン調査における努力の最小限化が回答行動に及ぼす影響," 特集 調査法研究の新展開 2, 45巻, 1号, pp. 1–11, 2018.