

クラウドソーシングを活用したオンライン実験における
実験参加タイミングによる被験者の偏りの把握
Understanding subject bias due to experiment participation timing
in online experiment using crowdsourcing

日室 聡仁[†] 後藤 晶[‡]
Akihito Himuro Akira Goto

1. はじめに

情報技術の進展に伴い、社会科学の分野でクラウドソーシングを活用して被験者を集めて ICT を活用してオンライン上で実験するオンライン実験が注目を集めている。この実験手法は一般的な実験室実験に比べ、多様な被験者を集めることができ、その結果、サンプリングバイアスを軽減し、実験によって得られる知見の頑健性を高めることができる。本研究では、クラウドソーシングで募集直後に集まる被験者だけで実験をするとサンプリングバイアスが発生する可能性を懸念して、参加タイミングによるサンプリングバイアスの可能性を検証した。本校では検証結果および検討した対応策を報告する。

2. 研究背景

筆者らは同一条件のオンライン実験を過去に複数回実施している。第 1 回目のオンライン実験は 2022 年 5 月 25 日に Yahoo クラウドソーシングを活用し 450 人の被験者を集めて実施し、第 2 回目のオンライン実験は 2022 年 7 月 11 日に Yahoo クラウドソーシングを活用して 450 人の被験者を集めて実施した。第 2 回目のオンライン実験では第 1 回目に参加した 450 人の被験者には募集をかけずに実験を実施した。この 2 つのオンライン実験は同条件の実験であるが、募集開始から規定人数が集まるまでの平均時間に 14 分 58 秒の差があり、実験途中での脱落率も 12% の差が確認され、オンライン実験での被験者の振る舞いも 2 つの実験で違いが確認された。これらの差は統計的にも有意な差があることが確認された。

これらの差は第 2 回目のオンライン実験で第 1 回目に参加した被験者には募集をかけずに実施したことが影響したのではないかと筆者らは考えている。つまり、第 1 回目のすぐに集まる被験者と第 2 回目のなかなか集まらない被験者との間には集団の違い(サンプリングバイアス)があるのではないかと考えている。本研究では Yahoo クラウドソーシングを活用したオンライン実験における参加タイミングによるサンプリングバイアスが本当に存在するのかを検証し、適切なサンプリング方法を検討する。

3. 実験設計

Yahoo クラウドソーシングを活用したオンライン実験における参加タイミングによるサンプリングバイアスが本当に存在するのかを検証するために、Yahoo クラウドソーシングで長時間被験者を募集し、集まった被験者を早着順に

450 人ずつグループ分けし、各グループを比較することですぐに集まる被験者群とそうではない被験者群に違いがあるのかを把握する。グループ分けの閾値である 450 人は 2 章の研究背景で説明したオンライン実験に合わせて設定する。グループは group1(募集開始から 1~450 番目に参加した被験者)、group2(募集開始から 451~900 番目に参加した被験者)、group3(募集開始から 901~1350 番目に参加した被験者)、other(group1~3 以外の被験者)の 4 群に分けて比較する。

実験では CRT 認知反応テスト^{[1][2]}と Big Five 性格診断^[3]を計測する。計測項目は 2 章の研究背景で上げたオンライン実験の結果より、早く集まる被験者のほうが実験内容を理解して回答している傾向、協調性が低い傾向、2 つの傾向があると仮定し計測する尺度を選択した。

実験は 2022 年 7 月 15 日(金)17:00 から Yahoo クラウドソーシングを活用して被験者を募り実施した。被験者には成果報酬として 25paypay ポイント(25 円相当)を実験後に支払うようにした。実験システムは oTree^[4]を用いて実装した。

4. 実験結果

4.1 被験者統計情報

実験の結果 4301 名の被験者のデータを収集した。被験者の平均年齢は 46.36 歳で標準偏差は 11.43 であった(年齢未回答の被験者を除く)。性別は男性 2355 名、女性 1360 名、回答しない 58 名であった。各グループの被験者募集から実験参加までの平均秒数は、group1 が 130 秒(2 分 10 秒)、group2 が 607 秒(10 分 7 秒)、group3 が 1482 秒(24 分 42 秒)、other が 9285 秒(2 時 34 分 45 秒)であった。

4.2 脱落率の比較

group1 は 450 人中 18 人(4.0%)が脱落、group2 は 450 人中 43 人(9.6%)が脱落、group3 は 450 人中 48 人(10.7%)が脱落、other は 2949 人中 43 人(14.1%)が脱落していた。各グループの脱落数・終了数の分布をカイ二乗検定で検定すると $p < 0.001$ と分布に有意な差があることが確認された。

4.3 CRT 認知反応テストの比較

各グループの CRT 認知反応テストの正解率を表 1 に示す。Frederick^[1]が提案した問題 1~3 については、group1 の正解率が一番高く、徐々に正解率が下がる傾向があることが確認された。また、問題 1~3 については各問題の各グループの正解・不正解の分布をカイ二乗検定で検定すると有意な差が確認された。次に、Toplak^[2]らが提案した問題 4~7 については問題 6,7 のみ group1 の正解率が一番高いことが確認された。問題 4~7 について、各問題の各グループの

[†] NEC ソリューションイノベータ株式会社

NEC Solution Innovators, Ltd.

[‡] 明治大学 Meiji University

正解・不正解の分布をカイ二乗検定で検定した結果、有意な差はみられなかった。

最後に、募集開始から実験までの秒数と問 1~7 の正解数との相関を確認すると相関係数は-0.07 と相関は確認されなかった。

問	group1	group2	group3	Other	カイ二乗検定
1	40%	37%	29%	28%	p=0.000
2	72%	65%	63%	59%	p=0.000
3	59%	52%	49%	47%	p=0.000
4	57%	65%	59%	59%	p=0.115
5	55%	55%	57%	55%	p=0.832
6	60%	58%	55%	54%	p=0.106
7	65%	62%	63%	59%	p=0.061

表 1: 各グループの CRT 認知反応テスト正解率

4.4 Big Five 性格診断の比較

各グループの尺度平均値を表 2 に示す。外向性は group1 が一番低く、テューキーの多重比較検定の結果、group1 と other に有意な差があることが確認された。協調性も group1 が一番低く、テューキーの多重比較検定の結果、group1 と group2, group3, other に有意な差があることが確認された。誠実性、神経症的傾向、経験への開放性についてはテューキーの多重比較検定の結果、有意な差があるグループは確認されなかった。

最後に、募集開始から実験までの秒数と各尺度との相関係数を確認すると、相関係数は最大 0.12、最小-0.04 と相関は確認されなかった。

項目	group1	group2	group3	Other
外向性	6.69	7.00	7.09	7.05
協調性	9.06	9.51	9.46	9.45
誠実性	7.79	7.78	7.73	7.77
神経症的傾向	8.44	8.12	8.32	8.44
経験への開放性	7.51	7.59	7.51	7.60

表 2: 4.4 Big Five 性格診断の結果(平均値)

5. 考察

脱落率と CRT 認知反応テストと Big Five 性格診断のすべてで各グループに有意な差が確認されたことより、クラウドソーシングで募集直後に集まる被験者だけで実験をするとサンプリングバイアスが発生する可能性があるとの示唆を得た。

脱落率や CTR 正解率が実験参加タイミングの早い群のほうがよくなる原因としては、実験参加タイミングの早い人のほうがクラウドソーシングで報酬を得ることに貪欲で、実験に対して真摯に対応しないと今後の募集がなくなるリスクを考えて行動した結果ではないかと考えられる。

Big Five 性格診断で外向性と協調性が実験参加タイミングの早い群のほうが低い傾向がある点についてはクラウドソーシングで報酬を得ることに貪欲な人の傾向が反映されたものと考えられる。

CRT 正解率や Big Five 性格診断と実験参加タイミングとで相関関係が確認されなかった原因としてはクラウドソーシングで報酬を得ることに貪欲な人が募集直後に参加しや

すい傾向はあるもののプレも大きいため、相関分析では関係性を確認できなかったと考えられる。

実験結果より、クラウドソーシングを活用したオンライン実験において実験参加タイミングを考慮せずに到着順に数百人程度のデータを集めて分析するとサンプリングバイアスが発生し、オンライン実験の特徴である多様な被験者を集めて実験結果の頑健性を高められるというメリットを享受できない可能性があることが示唆された。この結果を踏まえて、どのように対応すればよいかを検討する。

6. サンプリングバイアスへの対応策

募集から参加までのタイミングが早い被験者と遅い被験者が混ざったサンプルを収集する必要があるだろう。実現方法としては、事前にクラウドソーシング上にいる被験者一覧を取得するプレ実験を 1 週間程度かけて実施し、プレ実験で得られた被験者一覧からランダムに本実験参加者を選択し、選択された被験者のみに本実験への参加を依頼するのも一つの解決策であろう。この方法をとれば、応募にすぐ反応する被験者もなかなか反応しない被験者も一定の数が集まり、多様な被験者がサンプリングされたデータでの分析ができ、結果として実験によって得られる知見の頑健性を高めることができるだろう。

7. おわりに

本研究では、クラウドソーシングで募集直後に集まる被験者だけで実験をするとサンプリングバイアスが発生する可能性を懸念して、参加タイミングによるサンプリングバイアスの可能性を検証した。結果として参加タイミングが早い群の実験脱落率が低い、CRT 認知反応テストの正解率が高い、big5 性格診断の外向性と協調性が低いことが確認された。この差は参加タイミングが早い群のほうクラウドソーシングで報酬を得ることに貪欲な被験者が多いために発生したのではないかと推測する。参加タイミングによって集まる人の傾向が違うため、実験によって得られる知見の頑健性を高めるには参加タイミングの偏りをなくす必要があると考える。具体的な対応方法としては事前にクラウドソーシング上にいる被験者一覧を取得するプレ実験を 1 週間程度かけて実施し、プレ実験で得られた被験者一覧からランダムに本実験参加者を選択し、選択された被験者のみに本実験への参加を依頼するべきであると考えられる。

謝辞

本研究は JSPS 科研費 JP22K18153 によるデータを活用しました。

参考文献

- [1] Shane Frederick, "Cognitive Reflection and Decision Making", Journal of Economic Perspectives, vol. 19, no. 4, pp. 25-42(2005)
- [2] Maggie E. Toplak, Richard F. West, Keith E. Stanovich, "Assessing miserly information processing: An expansion of the Cognitive Reflection Test", Thinking & Reasoning, Volume 20, 2014
- [3] 小塩 真司, 阿部 晋吾, Pino Cutrone, "日本語版 Ten Item Personality Inventory (TIPI-J) 作成の試み", パーソナリティ研究 2012 年 21 巻 1 号 p. 40-52
- [4] Chen, D. L., Schonger, M., & Wickens, C., "oTree—An open-source platform for laboratory, online, and field experiments", Journal of Behavioral and Experimental Finance, Volume 9, March 2016, Pages 88-97