Replication of Psychology of Personal Data Donation (2019, PLOS ONE)

William de Melo (wdemelo@ucsd.edu)

2024-10-12

Table of contents

Introduction

My choice for the replication project is Skatova and Goulding's Psychology of personal data donation.

This study stood out to me as an open source project with a topical and interesting subject matter. The purpose of the study is to find whether people were willing to donate their personal data to benefit research as well as their reasons for doing so. This kind of research is especially important because it focuses on the motives and ethics of a new kind of data collection; of

harvesting data collected from people simply going about their daily lives. For instance, in the study, the researchers ask participants if they would give their supermarket loyalty card data. This kind of data is very naturalistic and widely available, and could transform the way we conduct research, but we should understand why people would donate their data first to ensure these donations are ethical. Additionally, if they are ethical, we should then consider ways to promote the donation of data for research.

The instruments used by the researchers are:

- The Reasons for Data Donation scale (30 items), which assessed reasons for donating data,
- The Prosocial Tendencies Measure (23 items), which examines behaviors related to charitable activities,
- The Self-Report Altruism Scale (20 items), which has participants report the frequency at which they help others,
- The Interpersonal Reactivity Index (28 items), which records one's capacities in various social areas, and
- Goldberg's 35 bipolar markers, to measure the Big Five personality traits.

Analyses utilized in this study included hypothesis testing (chi-squared, F) to assess the differences between conditions and demographics, factor analyses on the results to establish themes behind why people would donate, and linear regression to determine the predictive power of the Reasons for Data Donation scale on willingness to donate. Of these analyses, the biggest hurdle of this replication will be learning how to conduct factor analyses. Additionally, it may be difficult to sustain participants' interest using such a lengthy battery of questions. However, I believe that in two months' time, this replication could feasibly be conducted.

Repository link: https://github.com/willdemelo/skatova2019

Original study: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0224240

Methods

Power Analysis

Original effect size, power analysis for samples to achieve 80%, 90%, 95% power to detect that effect size. Considerations of feasibility for selecting planned sample size.

Planned Sample

Planned sample size and/or termination rule, sampling frame, known demographics if any, preselection rules if any.

Materials

All materials - can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Procedure

Can quote directly from original article - just put the text in quotations and note that this was followed precisely. Or, quote directly and just point out exceptions to what was described in the original article.

Analysis Plan

Can also quote directly, though it is less often spelled out effectively for an analysis strategy section. The key is to report an analysis strategy that is as close to the original - data cleaning rules, data exclusion rules, covariates, etc. - as possible.

Clarify key analysis of interest here You can also pre-specify additional analyses you plan to do.

Differences from Original Study

Explicitly describe known differences in sample, setting, procedure, and analysis plan from original study. The goal, of course, is to minimize those differences, but differences will inevitably occur. Also, note whether such differences are anticipated to make a difference based on claims in the original article or subsequent published research on the conditions for obtaining the effect.

Methods Addendum (Post Data Collection)

You can comment this section out prior to final report with data collection.

Actual Sample

Sample size, demographics, data exclusions based on rules spelled out in analysis plan

Differences from pre-data collection methods plan

Any differences from what was described as the original plan, or "none".

Results

Data preparation

Data preparation following the analysis plan.

Confirmatory analysis

The analyses as specified in the analysis plan.

Side-by-side graph with original graph is ideal here

Exploratory analyses

Any follow-up analyses desired (not required).

Discussion

Summary of Replication Attempt

Open the discussion section with a paragraph summarizing the primary result from the confirmatory analysis and the assessment of whether it replicated, partially replicated, or failed to replicate the original result.

Commentary

Add open-ended commentary (if any) reflecting (a) insights from follow-up exploratory analysis, (b) assessment of the meaning of the replication (or not) - e.g., for a failure to replicate, are the differences between original and present study ones that definitely, plausibly, or are unlikely to have been moderators of the result, and (c) discussion of any objections or challenges raised by the current and original authors about the replication attempt. None of these need to be long.