

Bias in Internet-Based Surveys

- A Comparative Study Using the Census and an Interview-Based Survey-

Masaki Taniguchi

President, Nippon Institute for Research
Advancement/ Professor, Graduate Schools for
Law and Politics, The University of Tokyo

Shoko Omori

Research Coordinator & Research Fellow,
Nippon Institute for Research Advancement

Given the fact that they are generally less expensive to conduct, Internet-based surveys have today become more widely used than interview-based and mail-based surveys, which use random sampling methods to select survey participants. However, questions have been raised concerning the representativeness of samples in Internet-based surveys.

In order to clarify the characteristics of Internet-based survey respondents, the project discussed here conducted an interview-based survey using random sampling and an Internet-based survey with target collection numbers for gender and age group, using the same survey items and the same survey timing. The study then compared the results of both surveys with census results.

The analysis revealed that, compared to the census results, the Internet-based survey showed bias in items such as area of residence and educational background, while the interview-based survey showed bias in home ownership rate. Correction based on propensity scores using census questionnaire data indicated that while there were limitations in correction for basic attributes, there was a certain correction effect for variables such as Internet usage time.

Key Points

In order to clarify the characteristics of Internet-based survey respondents, NIRA conducted an interview-based survey and an Internet-based survey simultaneously. These surveys employed the same questions. The results of both surveys were then compared with census results.

- Even when allocated by gender and age, Internet-based survey respondents tended to be more likely to be residents of large cities and to be better educated than the census population. In addition, 9-16% of the respondents were satisficers (i.e., respondents who offer quick answers to save effort). These respondents tend to offer haphazard or random responses.
- The interview-based survey using random sampling also had a bias that tended to over-represent people with more stable living conditions than the Japanese population as a whole. That is, the percentage of homeowners was higher than it actually is, and the percentage of those who continue to live in the same residence without moving was also higher.
- When the Internet-based survey was weighted using a propensity score method (a correction method based on census data), no significant changes were observed in either the interview-based or Internet-based surveys in the distribution of such variables as the rate of long-term political partisanship (support for a specific political party), the rate of support for the Cabinet, the degree of life satisfaction, and business confidence. On the other hand, a correction effect of more than 15 minutes in the interview-based survey and several minutes in the Internet-based survey occurred in the average daily Internet usage time.
- The results of this project and their implications for future social surveys include the following points.
 - ① When conducting surveys employing random sampling, in particular interview-based surveys, bias can be corrected to some extent by including in the questionnaire the same items that form part of the census, such as type of residence and information regarding change of residence, and by applying ex-post-facto correction such as a raking method.
 - ② When conducting an Internet-based survey, it is advisable not only to allocate the target number of respondents in advance, but also to set a larger target

number of respondents to ensure that after the satisficers' responses are removed, the component ratio of at least the respondents' sex, age, and region of residence will align with census data.

- ③ If the matters to be measured by an Internet-based survey are affected by the fact that survey respondents are Internet users (for example registered survey monitors), then appropriate ex-post-facto corrections should be made in line with actual statistics for Internet usage time, etc. (e.g., telework usage trends, metaverse penetration rates, etc.).
- ④ We should not have excessive expectations that a one-time Internet-based survey can measure “what we want to know.” Repeating a survey using the same methodology allows us to observe public opinion through changes from the previous survey (e.g., increase or decrease in the rate of support for the Cabinet) and make it possible to infer public opinion by means of understanding trends in bias (e.g., the difference between forecasts of the number of votes for specific parties based on survey data and actual election results).