Non-Response Bias

Nathan Berg
University of Texas, Dallas, Richardson, Texas, USA

Glossary

bias The expected difference between an estimated characteristic of a population and that population's true characteristic.

item non-response A non-response to a particular survey item accompanied by at least one valid measurement for the same respondent, for example, leaving just one item on a questionnaire blank or responding to some questions by saying, "I don't know," but providing a valid response to other questions.

non-response A survey response that falls outside the range of responses that survey designers consider to be valid.

unit One observation or a single vector of measurements, usually corresponding to a particular individual at a given point in time; many units make up a sample.

unit non-response Refusal or failure to provide any valid responses by someone who survey designers intended to include in the survey.

Non-response bias refers to the mistake researchers expect to make in estimating a population characteristic based on a sample of survey data in which, due to non-response, certain types of survey respondents are underrepresented.

Motivation for Analyzing Non-Response Bias

To illustrate and underscore the importance of analyzing non-response bias, consider the following scenario. A researcher working for a marketing firm wishes to estimate the average age of New Yorkers who own telephones. In order to do this, the researcher attempts to conduct a phone survey of 1000 individuals drawn from the population of phone-owning New Yorkers by dialing randomly chosen residential phone numbers. However, after 1000 attempts, the researcher is in possession of only 746 valid responses because 254 individuals never answered the phone and therefore could not be reached. At this point, the researcher averages the ages of the 746 respondents with valid responses and considers whether this average is likely to be too high or too low. Does one expect the 254 non-responders to be roughly the same age as respondents who answered their phones?

After thinking it over, the researcher concludes that the average age of the 746 responders is a biased estimate because the surveys were conducted during business hours when workers (as compared to older retirees) were less likely to be at home. If working age respondents are underrepresented, then the average among the 746 valid age responses is biased upward. In this case, the difference between the biased average and the true but unobserved average age among all telephone owners is precisely non-response bias.

Social scientists often attempt to make inferences about a population by drawing a random sample and studying relationships among the measurements contained in the sample. When individuals from a special subset of the population are systematically omitted from a particular sample, however, the sample cannot be said to be random in the sense that every member of the population is equally likely to be included in the sample. It is important to acknowledge that any patterns uncovered in analyzing a nonrandom sample do not provide valid grounds for generalizing about a population in the same way that patterns present in a random sample do. The mismatch between the average characteristics of respondents in a nonrandom sample and the average characteristics of the population can lead to serious problems in understanding the causes of social phenomena and may lead to misdirected policy action. Therefore, considerable attention has been given to the problem of non-response bias, both at the stages of data collection and data analysis.

Classifying Types of Error and Bias

Sampling Error

Anytime we generalize about a population based on a sample, as opposed to conducting a complete census of the population, there is an unavoidable possibility of mistaken inference. As such, sampling error arises even under the best of circumstances simply because, due to chance, averages of variables in a random sample are not identical to the corresponding averages in the population. Fortunately, sampling error typically disappears as the sample size increases. More important, sampling error does not lead to bias, because population characteristics can be estimated in such a way that the probability-weighted average of possible overestimates and underestimates is precisely zero.

Nonrepresentative Samples

It is important to distinguish from sampling error an entire family of nonsampling errors that arise when a sample is selected from a population in such a way that some members of the population are less likely to be included than others. In such cases, the sample is said to be nonrandom, or nonrepresentative, with respect to the population we intend to study. In contrast to sampling error, a nonrepresentative sample generally leads to biased estimation.

A number of factors may cause a sample to be nonrepresentative. One possibility is that, because of a flawed survey design, the survey simply fails to reach certain segments of the population. In the previous example, a daytime phone survey tended to underrepresent people who work, just as a survey of rural-area dwellers, or of car owners, would underrepresent users of public transportation.

Systematic mistakes by surveyors in coding survey responses can also lead to non-representative samples. The key issue is whether such mistakes are correlated with the characteristics of the individual being surveyed. For instance, a surveyor who, in the course of interviewing survey respondents, sometimes gets carried away discussing sports and forgets to record the respondent's last few responses will end up with a sample in which sports fans are underrepresented among the complete survey responses.

Perhaps the most common reason for nonrepresentative samples, however, is the behavior of survey respondents themselves. Oftentimes, the very fact of being a non-responder correlates with other characteristics of interest. When it does, non-response inevitably leads to nonrandom sampling and creates the potential for biased estimation of the characteristics under study. Researchers working with survey data must always consider the possibility that certain types of individuals are more likely to refuse to respond. This problem is acute when one of the key variables of interest determines, in part, who is more likely to select themselves out of a sample by not answering a survey question.

It is often suspected, for example, that individuals with high incomes are less likely to voluntarily disclose their income, biasing survey-based estimates of income downward. Similarly, those engaged in illicit drug activity, fearing the consequences of divulging potentially incriminating information, are probably less likely to participate in a survey about drug use, leading, again, to the potential for systematic underestimation. A slightly more subtle example is the case of estimating the percentage of a population that supports one of two political candidates. Apathetic voters are often thought to be the least likely to cooperate with political pollsters, even though many of them will in fact vote. Basing election forecasts on a sample of only those who agree to answer the poll can be misleading because the opinions of apathetic voters are underrepresented in pollsters' samples.

Dealing with Nonrepresentativeness before or after Data are Collected: Sample Design and Data Analysis Stages

To deal with nonrepresentative samples, it is helpful to distinguish two broad stages in a social science research project: data collection and data analysis. Some researchers conduct surveys themselves and therefore have direct control over the details of data collection. Others work with data sets originally collected by someone else, in which case the researcher exerts no direct control over the data collection stage.

For those who have a say about how the data are to be collected, it is crucial to try to foresee potential flaws in order to reduce the likelihood of bias. A vast literature exists on the topic of survey design, covering everything from the wording of survey questions to the issue of how many times those who do not answer the phone on a phone survey ought to be called back. Sometimes surveys can be designed in such a way as to provide a means of estimating the non-response bias associated with a particular data collection technique, for example, by comparing the results of face-to-face and phone interviews.

Many researchers in the social sciences, rather than collecting new data themselves, study data that have been collected by others, such as the U.S. Census, the Current Population Survey, and the General Social Survey. As a secondary data analyst, the researcher must decide what to do about survey respondents who failed to answer particular questions, referred to as the missing data problem. An additional issue is what to do about the target respondents who did not participate in the survey at all.

Similarity among Biases with Different Labels

One finds many different labels for biases that are, in fact, instances of one common problem—trying to learn about a population based on a nonrepresentative sample. It is helpful to see the underlying similarity among biases that arise from nonrepresentative samples because a successful approach to dealing with bias in one context often can be applied to new settings. In particular, survey data with missing responses can frequently be analyzed using techniques from the statistical and econometric literature under the heading measurement error. Terms such as "noncompletion bias" or "volunteer bias," referring to the nonrepresentative sample problem that arises when only special kinds of respondents actually complete a survey questionnaire or to situations in which the subpopulation of volunteers is substantively different from the rest of the population, should be viewed as essentially the same problem.

The connection between non-response bias and selection bias warrants special mention. Non-response is special kind of the selection problem of the type analyzed in the work of James Heckman. Thus, selection bias, when referring to the mechanism by which some survey respondents choose not to answer survey questions (thereby selecting themselves out of the sample), overlaps with what is defined here as non-response bias. Heckman interprets the selection problem more generally as a kind of econometric misspecification. As illustration, it is useful to consider a regression model used frequently by labor economists in which the expected wage depends on demographic variables as well as other factors thought to influence workplace productivity. If no account is taken of the mechanism by which only special kinds of individuals choose to become workers and therefore wind up being included in the sample (implying that regressors are correlated with the error term in the regression model), then the econometric model is, in Heckman's words, "misspecified," leading to misspecification bias.

Misreporting versus Non-Response

When those collecting data ask respondents to report on their own behavior in connection with activities such as cheating, personal finance, sex, or alcohol and drug use, some respondents, instead of refusing to answer, will misreport their behavior. When interpreted at face value, a sample in which certain kinds of individuals tend to misreport will not accurately represent the population under study. As with nonrepresentative samples caused by non-response, misreporting usually leads to bias, which can be referred to as misclassification bias, misreporting bias, contaminated data bias, or simply response bias. The task of the researcher is to consider how such misreporting will influence estimates of key population characteristics.

Analysis of Survey Data with Missing Responses

Item versus Unit Non-Response

An important distinction to make regarding non-response is item versus unit non-response, a distinction that turns on whether there is at least one survey item for which a valid response was obtained or whether the entire unit is missing. When entire units are missing from a sample, no test or correction for bias is available without obtaining additional data about the targeted respondents who did not respond to the initial survey. In contrast, samples with item non-response may allow for unbiased estimation because partially completed responses from item non-responders may be used to control for differences across responders and item non-responders. This section discusses techniques for computing unbiased estimates using samples which feature item non-responses.

Little and Rubin's Missing Data Framework

Roderick Little and Donald Rubin, individually and in joint work, have written a number of frequently cited articles on the subject of analyzing data with missing values. Their approach is quite general and applies directly to most situations that applied researchers working with survey data are likely to face.

Imputation

One approach to dealing with missing survey responses is to somehow fill in the missing values, imputing good guesses in place of missing survey entries. Some researchers, for instance, may replace missing measurements with the average value across the complete cases. A more sophisticated approach involves replacing missing values with estimates based on prediction equations that are fitted with the complete cases and subsequently used to predict missing values using the partial responses of item non-responders. After imputing values to fill in the missing data, data analysis proceeds using traditional estimation techniques.

A serious drawback to this technique is that the precision of the estimates computed using the data set with imputed values will be overstated for two reasons. First, imputed values generally are computed by averaging over other observations and, therefore, will be more tightly clustered about the mean than a fresh collection of bona fide observations would be. And second, the use of traditional statistical techniques after imputing values for missing entries in the data matrix will be based on an overstated sample size because a sample of N observations, some of which have been imputed, will contain fewer than N independent pieces of information. This means that standard errors will be too small and that the nominal size of significance tests will be inflated.

Weighting

Another approach to working with incomplete data involves discarding partial observations and assigning a weight to each complete observation so that the weighted sample better represents the average characteristics of the population. For instance, with a sample of 68 men and 32 women in which women appear to be underrepresented, one might consider placing additional weight on female units in the sample, perhaps based on the gender ratio from the U.S. Census, in order to reduce bias. In principle, weighting should correct for bias that arises from estimation based on nonrepresentative samples. A severe complication, however, is knowing how to compute standard errors that accurately account for the imprecision in the weights themselves. Doing so is notoriously difficult. Therefore, many authors, including Little and Rubin, recommend against weighting techniques. Those authors point out that the most common approach to non-response is simply to discard incomplete responses, effectively giving each of the complete sample units the same weight. Except for the unusually lucky case in which the complete-only subsample is a truly random sample of the population, this technique, although simple to use and widely practiced, leads to biased estimates.

The Maximum-Likelihood Approach

The maximum-likelihood approach is, far and away, the preferred approach to correcting for non-response bias, and it is the one advocated by Little and Rubin. The maximum-likelihood approach begins by writing down a probability distribution that defines the likelihood of the observed sample as a function of population and distribution parameters θ. If x_1 and x_2 represent responses to two different survey questions by a single individual, the likelihood associated with a complete response may be expressed as $f(x_1, x_2; \theta)$, where f is the joint probability density function of x_1 and x_2. For individuals who only report x_1, the likelihood associated with x_1 is $\int_{-\infty}^{\infty} f(x_1, x_2; \theta)dx_2$, which can, under the assumption of joint normality, be simplified to a more convenient form. In this way, a likelihood function is specified that includes terms corresponding to each observation, whether completely or only partially observed. The likelihood objective is then maximized with respect to θ, which produces estimates of the desired characteristics, enjoying all the well-known properties of maximum-likelihood estimation.

Most important among those properties, maximum-likelihood estimates converge to the true value of θ under the assumption that the probability distribution is correctly specified. Maximum-likelihood estimates are also asymptotically normal and asymptotically efficient, meaning that, for large samples, the maximum-likelihood estimate of θ is approximately normal and is the best use of the information contained in the sample. In addition to these advantages, the maximum-likelihood approach makes it possible to estimate fairly elaborate multi-equation models in which the probability that an individual fails to respond depends on other observable variables. Within such a framework, it is often possible to construct a quantitative test of the missing-at-random hypothesis, implemented as a straightforward test of an appropriate parameter restriction. The main drawback to maximum-likelihood estimation is that strong assumptions are required about the distribution of the process generating the survey responses. Still, the advantages are usually thought to outweigh the drawbacks, making it the approach of choice for many quantitative researchers.

Missing-at-Random, Missing-Completely-at-Random, Mixture Modeling, and Multiple Imputation

A frequently mentioned distinction in the missing-data literature involves the two terms, missing-at-random and missing-completely-at-random. If the probability of non-response for a variable Y is the same for every unit of observation in the population, then Y is said to be missing-completely-at-random. If, on the other hand, the probability of non-response systematically relates to other variables in the model, but not to the value of Y itself, then Y is said to be missing-at-random. Defining the random variable $R = 0$ if Y is missing, and $R = 1$ otherwise, another important distinction can be expressed as follows. Selection models require the user to observe the conditional distribution $Y \mid R = 1$ and model the conditional probability $R = 1 \mid Y = y$, whereas mixture models require observing $Y \mid R = 1$ and modeling $Y \mid R = 0$.

Other Perspectives on Correcting for Non-Response Bias

Lawrence Marsh and his co-authors have proposed a number of non-response models and developed associated maximum-likelihood estimators that appear to work well in practice. Marsh's work, in addition to providing straightforward maximum-likelihood estimators of non-response bias, compares the performance of maximum-likelihood-based corrections for non-response bias against those associated with alternative techniques of estimation, such as maximum entropy, finding consistent support for the maximum-likelihood approach. These results rest on the existence of auxiliary relations that determine the missing response mechanism. In the absence of auxiliary relations, Lien and Rearden's 1988 article shows that, when the missing observation is the dependent variable in a limited dependent-variable model, nothing is gained by applying maximum-likelihood-based corrections. Thus, special caution is warranted when estimating a model in which the dependent variable is frequently missing.

Measuring Non-Response Bias

Validation

Validation is a general approach to testing for non-response bias that almost always involves comparing two different samples drawn from the same population. The technique of validation permits us to measure non-response bias, to test the hypothesis of no bias, and to identify which variables, if any, are correlated with non-response. This approach is only feasible, however, if we are lucky enough to have two samples drawn from the same population.

Given a pair of samples, it is usually clear, either from the number of missing entries or from descriptive notes attached to the data, which data set has a lower non-response rate. The general philosophy of validation assumes that the sample with the lower non-response rate is, for all practical purposes, the "reliable" one. Accepting this view, significant departures among the observations in the "unreliable" sample relative to the average characteristics in the "reliable" sample can then be attributed to non-response bias, providing a qualitative measure (too high vs. too low) along with a quantitative measure of the severity of the problem.

For instance, it is well accepted that face-to-face interviews typically draw a higher response rate than phone surveys do. Now suppose we draw two samples of measurements on ethnicity, one face-to-face and the other by phone, and discover that the fraction of Asian Americans in the phone data is one-half that of the face-to-face interview data. Taking the estimated racial composition of those who respond to the face-to-face interview as the reliable benchmark, we might plausibly infer that Asian Americans are twice as likely to non-respond in a phone survey compared to other types of Americans. The qualitative finding that phone survey data may underrepresent Asian Americans is valuable in qualifying further estimates of characteristics on which Asian Americans are known to be different from other Americans. Beyond this, the magnitude of the difference, in this case a factor of one-half, can be used to place additional weight on the phone responses of Asian Americans in order to correct for the fact that they tend to be underrepresented in phone surveys.

Sex researchers, who must routinely deal with survey data suffering from very high non-response rates, have applied validation to gain a feel for the ways in which the respondents in their data are different from the U.S. population at large. A straightforward approach is to compare, say, the age distribution among sex survey respondents with the age distribution of the population of Americans as measured by the U.S. Census. Sex survey respondents, in fact, appear to be younger than average Americans are.

Validation is virtually the only way to learn about the characteristics of unit non-responders because, by definition, there is no information on unit non-responders in the rounds of data collection in which non-response occurs. One 1999 study by Heather Turner used validation techniques to uncover some surprising distinctions that need be made among those who are typically categorized together as non-responders. She identified two types of non-responders, differentiating those who refused to participate twice from those who could not be contacted after 17 attempts. Using data from other sources and from follow-up interviews, she discovered that those non-responders who directly refused to participate in the survey tended to be older, attended church more often, and were more skeptical about the confidentiality of interviews.

In an important finding, rich with policy implications, she produced evidence suggesting that, in contrast to the low-risk lifestyles of those who directly refuse to participate, the difficult-to-reach non-responders tended to have significantly more sexual partners and higher frequencies of risk factors for AIDS. This demonstrates how difficult it can be to generalize about non-responders and make reliable guesses as to whether non-response bias skews estimates up or down.

Measuring non-response bias in telephone surveys is a frequent concern for polling organizations and those conducting market research by telephone. A fundamental issue confronting anyone attempting to learn about the entire population of Americans based on a phone survey is the fact that not all American households have telephones. Previous attempts to measure the characteristics of nontelephone households indicate considerable differences with respect to phone-owning households across a number of important characteristics such as the propensity to have health insurance.

In a novel approach to measuring non-response bias published in 1995, Scott Keeter sought to estimate telephone noncoverage bias by conducting a series of phone surveys on the same randomly drawn sample of phone numbers at several points in time. Among those reached at any given time were, of course, some households who had only recently gained access to a telephone. And among those reached in earlier rounds of phone surveying were some households whose number later became disconnected. Labeling those who gained or lost telephone service at least once as "transient" and comparing the number of transients in his sample with government and industry estimates of how many American households are nontelephone households, Keeter determined that transients make up roughly one-half of all nontelephone households. Moreover, the demographic characteristics of nontelephone households recorded in other surveys appeared to match those of the transient group in Keeter's study, bolstering confidence in the ability of existing non-response-corrected phone survey methodologies to produce meaningful insights into the characteristics of American households in general.

Another area of policy research in which non-response bias can play an especially important role is that of valuing natural resources. Developers and government officials often attempt to study the benefits and costs of a proposed building project and must, at some point, put a dollar value on natural resources, including wetlands, endangered animals, and undeveloped green space. Similarly, officials at the Environmental Protection Agency and environmental economists confront the challenge of assessing the value of parks, wildlife, and air quality. Such endeavors must deal with the question of how to reliably elicit valuations that somehow reflect the aggregate preferences of residents. The basic idea is to use samples of citizens to estimate the worth of natural resources in the eyes of an average citizen.

It is fairly obvious that the problem of nonrepresentativeness will have a direct effect on such valuations. Suspecting that those who agree to participate in environmental surveys have higher than average subjective assessments of the value of natural resources, researchers in this area worry that non-response bias may lead to overstated valuations. In a 1993 article, John Whitehead and his colleagues employed a combination mail and phone survey design in an attempt to produce a bias-corrected valuation of a wetlands preservation project. Using the validation principle, these authors attempted to measure differences between non-responders and responders, both in terms of average demographic characteristics and in terms of willingness to pay for environmental amenities. Validation did, in fact, uncover a disparity between those who initially refused to participate and those who participated without hesitation. Although a non-responder with identical observable characteristics was found to be no less willing to pay than a similar responder, the group of eager respondents included more highly educated individuals and more males. After adjusting for non-response bias, the estimated aggregate willingness to pay fell by 33%.

In addition to its application in studying unit non-response, the logic of validation can also be applied to learn about item non-responders. Emil Kupek's 1998 article used a large national sex survey in Britain to study the covariates of item non-response. Kupek partitioned his sample into subsamples based on how reluctant individuals were in answering specific questions about their sexual behavior. Specifying the dependent variable to be a measure of each individual's reluctance to respond, Kupek estimated a model relating other demographic variables to the probability of item non-response. Non-responders in Kupek's sample turned out to be less educated and to include relatively more nonwhites. Perhaps surprisingly, factors such as gender, declared religious affiliation, age, and marital status seemed to have little effect on the probability of non-response. As in this study, simply establishing which variables correlate with non-response can amount to a key step in thinking through the broader consequences of non-response and, in particular, whether our nonrandom sample will actually lead to bias in estimating the population characteristics of interest.

Designing Surveys So That Non-Response Bias Can Be Estimated

An extensive body of research exists analyzing survey methods, seeking to refine their capacity to overcome potential sources of bias. The results, so far, however, are not reassuring. Survey responses are, without question, very sensitive to the way in which they are elicited. This phenomenon underlies disparaging remarks we frequently hear directed at survey findings in general, such as: "By changing the wording, anything can be shown with surveys." Although this statement is undoubtedly an exaggeration, the sensitivity of survey results to the fine detail of survey design has been demonstrated in numerous academic studies.

Hurd *et al*.'s 1998 study uses experimental evidence to analyze survey non-response and presents a thorough discussion of survey-response sensitivity in the context of estimating aspects of consumption and savings behavior. The order of survey questions, the gender of the surveyor, rewordings such as "10% survived" instead of "90% died," and a number of other seemingly innocuous differences in the implementation of surveys can sharply affect the average response. Compared to mail surveys, face-to-face interviews are known to produce higher reported rates of activities with a high degree of social approval such as volunteering, going to church, and engaging in safe rather than unprotected sex. Non-response rates can also vary dramatically depending on whether data are collected using phone, mail, or face-to-face interviews.

Complicating the picture is that these sensitivities to survey design are not always uniform across all segments of the population. For instance, it has been demonstrated that response rates for whites in face-to-face versus mail surveys are about the same, yet they differ significantly for African American respondents. Such findings underscore the delicate nature of survey design while raising important issues of interpretation that demand consideration even at subsequent stages of data analysis.

Randomized Response

The method of randomized response explicitly aims at reducing non-response and misreporting on survey items that concern sensitive topics. The idea behind randomized response is to introduce random questions or random coding procedures into the construction of response data so that it is impossible for the surveyor to infer the respondent's original response by looking at the data recorded for that individual. A survey question on illegal drug use might employ the following survey design. With probability $1-q$, respondent i is asked, "Have you ever taken an illegal drug," from which the response datum, $y_i = 1$, is recorded if the answer is "Yes," and $y_i = 0$ otherwise. But with probability q, the response datum is coded $y_i = 1$ no matter i's answer (or without ever asking i the sensitive question). The advantage of the randomized design is its capacity to convince respondents that it is safe to truthfully disclose private information. If $y_i = 1$, it may be that i answered "Yes," or it may be that i happened to fall in the $(q \times 100)\%$ of the sample for whom y_i is automatically coded 1.

From randomized response data, an unbiased estimator of the true frequency of drug use is easy to compute, assuming that randomization induces perfect compliance (i.e., full response and no misreporting). Denote the true rate of drug use as λ. Because (equation (1)),

$$Ey_i = (1-q)\lambda + q, \tag{1}$$

the estimator

$$\hat{\lambda} = \frac{\frac{1}{N}\sum_i^N y_i - q}{1-q} \tag{2}$$

is unbiased. The price to be paid for introducing randomization, however, is a reduction in the precision of estimation, as can be seen by examining the variance formula for $\hat{\lambda}$.

Multivariate versions of randomization are also possible. Fox and Tracy's 1986 monograph, *Randomized Response: A Method for Sensitive Surveys*, provides further details. The goal of randomization, in all its forms, is to reduce respondents' skepticism about the confidentiality of their responses. Whether randomization accomplishes its goal is open to debate, however, because it is not clear whether respondents understand randomization sufficiently well or trust the survey designers to follow through with an honest implementation.

A Budget Constraint Means a Trade-off between Sampling Error and Bias Reduction

Different survey designs have different price tags and, although more data are always desirable, it is not always obvious how to efficiently allocate spending on data collection given a fixed project budget. In designing surveys with the intention of reducing non-response bias in mind, there is often a nontrivial trade-off to consider when selecting a mix of survey techniques. For a given sum of money, an inexpensive mail survey will probably draw a sample with a higher number of units, thereby reducing sampling error. However, a smaller sample collected using face-to-face interviews will probably enjoy the advantage of a lower unit non-response rate. Thus, we are faced with trading off greater precision (increasing the sample size) against a greater chance that non-response bias will contaminate estimation. In this situation, a sound approach generally involves selecting a mix of sampling techniques that will lead to fairly precise estimates while providing reasonably good controls for non-response bias.

Parsing the Meaning of the "Don't Know" Response

A problem faced by most applied researchers working with survey data is interpreting the meaning of the response "Don't know" to a survey question. Those involved at the survey design stage often contemplate whether one should prompt those who respond "Don't know" to relent and provide a valid answer. Interestingly, there is debate about whether such prompting is a good idea or not. Insofar as prompting induces random guessing, it is not helpful. But when additional prompting succeeds at extracting additional information rather than noise, our estimation should, in principle, improve.

For example, public opinion researchers have demonstrated that opinions about political candidates elicited from respondents who say they know nothing about those candidates are, in fact, meaningful indicators of future voting behavior rather than random noise. But in other settings, the evidence points in the opposite direction. As a general rule, the responses of reluctant responders that we collect by means of a special technique of elicitation should be interpreted cautiously, with full acknowledgement that they probably contain more noise than the responses of other respondents.

In some contexts, it may be useful to try identifying multiple subgroups among item non-responders. The issue at stake is the extent to which we can generalize about non-responders. Qualitative information about non-response bias is particularly helpful in instances in which it can be presumed that non-response bias mitigates against finding a significant difference, referring here to an estimated characteristic such as average income across two groups. In such a case, without doing anything special to correct for bias, discovering a significant difference is especially persuasive, in spite of and, in part, because of the bias. But in other settings, rather than helping to converge to a simple conclusion, gathering additional information about non-responders may complicate the analysis, raising additional questions and revealing the folly of generalizing about non-responders as if they were a homogeneous subset of the population. Oftentimes, they are not.

Panel Data and Attrition

A panel data set contains multiple observations on a fixed group of individuals from whom measurements are collected at several points in time. That is, a random list of individuals is initially chosen, and then those same individuals are surveyed multiple times over the course of months or years. Rather than the snapshot view offered by a cross section in which each observation corresponds to a unique individual, a panel contains a time series for a collection of individuals, which allows researchers to study population characteristics through time.

A frequent problem with panel data is attrition, meaning that some respondents surveyed in the initial period later drop out. Respondents who drop out can be thought of as those who begin as fully cooperative responders but later become non-responders either by choice or circumstance. In this context, non-response bias is sometimes referred to as attrition bias.

Survey panel respondents may be classified as either full-time (those who remain in the sample at all point in time), monotonic attritors, or nonmonotonic attritors. "Nonmonotonic" refers to a respondent who becomes a non-responder at some point in time and then rejoins the survey. When all three types are present in a panel, a three-category logit or probit analysis can demonstrate relationships between the probability of attrition and variables that do not change with time. Simpler still, researchers sometimes run a sequence of regressions and examine the effect on regression coefficients of including or excluding attritors. By creating dummies for full-time, monotonic attritors, and nonmonotonic attritors and interacting those dummies with the regressors of interest, standard t tests on interaction terms can produce evidence that attrition is causing bias. As an example, Burkam and Lee's 1998 article applied these techniques to a panel of U.S. high school students, discovering that gender significantly affects the probability of attrition and also that attrition bias leads to an overstatement of black-white disparity on academic achievement tests.

In another useful example of how to deal with attrition, Fitzgerald *et al.*'s 1998 article estimated a structural model of attrition and studied the severity of attrition bias as it related to a number of standard demographic variables using the Michigan Panel Study on Income Dynamics (PSID). An annual survey panel used frequently by labor economists, the PSID loses roughly 12% of the participants each year. More than 20 years after its inception, fewer than 50% of the original participants remain. Although the observed characteristics of attritors are noticeably different from full-time respondents, coefficient estimates in a variety of models using the PSID, according to Fitzgerald *et al.*, appear to change little when attempts are made to correct for attrition bias. This is good news for researchers attempting to generalize about labor markets in the United States based on the PSID.

Summary

If non-responders are different from responders in ways critical to the main research questions under investigation, the possibility of non-response bias needs to be taken seriously. Whether designing a survey or analyzing previously collected data that have already been collected, a number of useful techniques may be applied to test for and possibly correct for non-response bias. In the data analysis stage, it is usually best, when feasible, to specify a separate equation for the non-response process and estimate all the parameters simultaneously by maximum likelihood. In particular applications, it can be useful to exploit other authors' approaches to dealing with the problem of a nonrepresentative sample, even when the problem is not explicitly referred to as non-response bias. Rather than attempting to solve the problems created by non-response, it is often acceptable simply to be sensitive to the potential problems and state the likely effect of non-response on reported estimates. Careful attention to the problem of non-response is a critical step in conducting high-quality research using survey data.

See Also the Following Articles

Attrition, Mortality, and Exposure Time • Longitudinal Studies, Panel • Maximum Likelihood Estimation • Randomization • Surveys • Weighting

Further Reading

Burkam, D. T., and Lee, V. E. (1998). Effects of monotone and nonmonotone attrition on parameter estimates in regression models with educational data: Demographic effects on achievement, aspirations, and attitudes. *J. Hum. Resources* **33,** 555–575.

Fitzgerald, J., Gottschalk, P., and Moffitt, R. (1998). An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics. *J. Hum. Resources* **33,** 25–74.

Fox, J. A., and Tracy, P. E. (1986). *Randomized Response: A Method for Sensitive Surveys.* Sage, Beverly Hills, CA.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47,** 153–161.

Hurd, M. D., McFadden, D., Chand, H., Gan, L., Merrill, A., and Roberts, M. (1998). Consumption and savings balances of the elderly: Experimental evidence on survey response bias. In *Frontiers in the Economics of Aging* (J. P. Smith, ed.), pp. 387–391. University of Chicago Press, Chicago, IL.

Keeter, S. (1995). Estimating telephone noncoverage bias with a telephone survey. *Public Opinion Q.* **59,** 196–217.

Kupek, E. (1998). Determinants of item nonresponse in a large national sex survey. *Arch. Sex. Behav.* **27,** 581–589.

Lee, B. J., and Marsh, L. C. (2000). Sample selection bias correction for missing response observations. *Oxford Bull. Econ. Statist.* **62,** 305–322.

Lien, D., and Rearden, D. (1988). Missing measurements in limited dependent variable models. *Econ. Lett.* **26,** 33–36.

Little, R. J. A., and Rubin, D. B. (1990). The analysis of social science data with missing values. In *Modern Methods of Data Analysis* (J. Fox and J. S. Long, eds.), pp. 374–409. Sage, Newbury Park, CA.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* John Wiley and Sons, New York.

Turner, H. A. (1999). Participation bias in AIDS-related telephone surveys: Results from the National AIDS Behavioral Survey (NABS) Non-Response Study. *J. Sex Res.* **36,** 52–66.

Whitehead, J. C., Groothuis, P. A., and Blomquist, G. C. (1993). Testing for non-response and sample selection bias in contingent valuation: Analysis of a combination phone/mail survey. *Econ. Lett.* **41,** 215–220.