DOCUMENT RESUME

ED 388 679 TM 023 704

AUTHOR Jones, James A.

TITLE An Illustration of the Danger of Nonresponse for Survey Research.

PUB DATE Apr 95

NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).

PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Data Collection; *Effect Size; Monte Carlo Methods; Research Methodology; *Research Problems; *Response Rates (Questionnaires); Responses; *Statistical Bias; *Surveys

IDENTIFIERS Research Replication

ABSTRACT

In the educational literature, responses to surveys commonly serve as the source of data for many empirical articles. Whenever a survey is used as a source of data, the response rate can greatly affect the potential generalizability of the findings. Using Monte Carlo methods, this study examined the effects on sample estimates of the population mean and standard deviation for 3 levels of effect size differences between the responders and nonresponders (0.0, 0.25, and 0.50). Two data sets were used: 400 normally distributed random values and 200 responses to an item on a Likert-type scale. The number of replications for each condition was 5,000. The proportion of population values contained within a 95% confidence interval of the sample estimates was then calculated with respect to the mean and standard deviation. For the 0.0 effect size conditions, all response rate levels produced expected proportions of samples containing the population values. Increased effect size differences combined with reduced response rate levels resulted in biased estimates, particularly for the mean. Although return rates of 70% have been recommended as adequate, response rates of at least 90% may be needed if moderate effect size differences are suspected between responders and nonresponders. (Contains 4 tables and 14 references.) (Author/SLD)

AN ILLUSTRATION OF THE DANGER OF NONRESPONSE FOR SURVEY RESEARCH

by

James A. Jones, Ph.D.
Ball State University

This paper was prepared for the Annual Meeting
of the
American Educational Research Association
April 18-22, 1995
San Francisco, CA

Abstract

In the educational literature, responses to surveys commonly serve as the source of data for many empirical articles. Whenever using a survey as a source of data, the response rate can greatly affect the potential generalizability of the findings. Using monte carlo methods, this study examined the effects on sample estimates of the population mean and standard deviation for three levels of response rate (50%, 70%, and 90%) crossed by three levels of effect size differences between the responders and nonresponders (0.0, .25, .50). Two data sets were used: 400 normally distributed random values and 200 responses to an item on a Likert-type scale. The number of replications for each condition was 5,000. The proportion of population values contained within a 95% confidence interval of the sample estimates were then calculated with respect to the mean and standard deviation. For the 0.0 effect size conditions, all response rate levels produced expected proportions of samples containing the population values. Increased effect size differences combined with reduced response rate levels resulted in biased estimates, particularly for the mean. Although returns rates of 70% have been recommended as adequate, response rates of at least 90% may be needed if moderate effect size differences are suspected between responders and nonresponders.

Introduction

In the educational literature, responses to surveys served as the source of data for well over 50% of the empirical articles (Aiken, 1987; Hartman, Fuqua, and Jenkins, 1986). Whenever using a survey as a source of data, one issue of great concern for the researcher should be the response rate (West, 1991) because the sample of survey respondents is almost always used by the researcher as the basis for generalizations to the population from which the sample was drawn. The validity of these generalizations depends upon how representative the sample is of its population. Since nonresponders often differ in meaningful ways from those who do respond, samples that have large proportions of nonresponses are unlikely to be representative of the population (Hartman et al., 1986).

To ensure the population estimates produced by a sample are unbiased, Gay (1987) recommended a 70% return rate while Kerlinger (1986) stated that a return rate of at least 80 to 90% is required. Using a mathematical approach, Aiken (1981) calculated that 82.6% was the minimum response rate level needed to ensure that the sample of responses does not vary significantly from a sample that included both the responders and nonresponders. Based on his examination of real data sets, Berdie (1990) has suggested that response rates of 65 to 75% should be adequate. For example, Berdie (1990) found that the percent of respondents choosing a given category within an item appeared to be within 4.8% of actual values even when the response rate was as low as 52%. He also found that the mean for an item showed little change with increases in response rate beyond a return rate of 75%. Blixt (1990) has demonstrated, however, that a 70% return rate is adequate only when the degree of bias for the respondents is slight as compared to the whole sample. In other words, when the responses of those who returned the survey differ only slightly from the nonrespondents, a 70% return rate is adequate. When the bias of the respondents is severe, a 90% return rate is needed to produce unbiased parameter estimates (Blixt, 1990). While the range of recommended response rates varies from 65 to 90%, even the minimal goal of 65% may be difficult to achieve in practice since survey return rates of less than 50% are common (Kerlinger, 1986).

The purpose of this study is to demonstrate with monte carlo techniques the potential impact nonresponse can have on the generalizability of survey results. The population mean and standard deviation will be estimated from samples of varying response rates and effect size differences between survey responders and nonresponders. These estimates will then be compared to the known population values by counting the number of samples containing the population values within a 95% confidence interval (1.96 standard errors) of the estimates.

Method

Using a monte carlo method, 5000 samples of 400 normally distributed random values were split into three levels of response rates (50%, 70%, and 90%) and crossed with three levels of effect size differences between responders and nonresponders (0.0, .25, and .50). The response rate levels fall within the range of values likely to be seen in educational research (Gay, 1987) and the effect sizes of 0.0, .25, and .50 would represent no group difference, a small difference, and a medium group difference, respectively, for a t-test (Howell, 1987). Although the response levels were exact (i.e., in the 50% response rate condition, all samples had an exact response rate of 50%), the levels in the effect size conditions were approximate. For example, in the .25 effect size condition the mean effect size difference was approximately .25. Essentially, the sample effect sizes were distributed in a roughly symmetric about the target for that condition. This was done to heighten the realism of the simulation.

In the 0.0 effect size condition, values were randomly blanked out until the samples of 400 values were reduced to the desired level of response (i.e., 50%, 70%, or 90%). For the .25 and .50 effect size conditions, the situation was more complex since the respondents and nonrespondents (the values that were blanked out) needed to show systematic differences over the long run. To accomplish this, the probability of being a nonrespondent was altered depending on where the case fell within the distribution of scores. Using a method developed by Jones (1993), a cut-point above the sample mean was calculated and a selection method employed so that cases with values falling below this cut-point had a lower probability of being blanked out than those cases that had values above it.

In addition to the normally distributed random values, a second data set was put through the same algorithm. This data set consisted of responses to one of the items from the Hillside Assessment of Perceived Intelligences (HAPI) questionnaire from a previous study by Shearer and Jones (1994). The subjects consisted of 200 undergraduate students who responded to the question "Do you often have favorite tunes on your mind?" and their responses were limited to only five nonmissing values (i.e., 1=Every once in awhile, 2=Sometimes, 3=Often, 4=Almost all the time, and 5=All the time). The item was chosen because it had a roughly symmetric distribution about the middle value. This real data set with its limited response options is in contrast to the randomly generated data, which were continuous and unbounded.

Results

Randomly Generated Data

Prior to manipulation, the mean of the 5000 sample means was -.0003 with a standard error of .0499 and the mean of the standard deviations was 1.0000 with a standard error of .0355 for the randomly generated data. These values would appear to have approximated the population parameters of 0, .0500, 1, and .0354 for the mean, standard error of the mean, standard deviation, and standard error of the standard deviation, respectively. The response rate and effect size manipulations were then introduced. The resulting mean effect size differences and standard deviations for the 0.0, .25, and .50 conditions by response rate are shown in Table 1.

Table 1

Resulting Effect Size Differences Between Respondents and Nonrespondents Using Randomly Generated Data

	Mean Effect Size Standard Deviation N of Samples	Response Rate 50%	 70%	 90%	
Effect Size	0.0	-.0001 .1015 5000	-.0009 .1084 5000	.0005 .1656 5000	-.0002 .1284 15,000
	.25	.2735 .0989 5000	.2753 .1100 5000	.2449 .1726 5000	.2646 .1319 15,000
	.50	.4954 .0944 5000	.4972 .1075 5000	.5075 .1772 5000	.5000 .1316 15,000
		.2563 .2253 15.000	.2572 .2309 15.000	.2510 .2690 15,000	

The theoretical standard errors associated with each sample were then computed for the mean and standard deviations using the following formulas from Ferguson (1981):

$$SE_{mean} = \frac{\hat{\sigma}}{\sqrt{N}} \qquad SE_{Std\ Dev} = \frac{\hat{\sigma}}{\sqrt{2N}}$$

The proportions of sample estimates that contained the population value within a 95% confidence interval were calculated with respect to the mean and standard deviation.

As shown by Table 2, the combination of low response rate and differences between respondents and nonrespondents can severely bias the parameter estimates for the mean and standard deviation. In the worse scenario shown, a 50% response rate with a moderate difference between respondents and nonrespondents, only 4% of the samples contained the population mean within a 95% confidence band. The standard deviation would appear to be slightly more robust, with 25% of the samples containing the population standard deviation within a 95% confidence band in the 50% response rate and .50 effect size condition.

Table 2

Proportion of Sample Estimates Containing the Population Mean and Standard Deviation within a 95% Confidence Interval Using Randomly Generated Data

	Proportion of Means Proportion of Standard Deviations	Response Rate 50%	 70%	 90%	
Effect Size	0.0	.95 .95	.96 .95	.95 .95	.95 .95
	.25	.48 .83	.69 .89	.93 .94	.70 .89
	.50	.04 .25	.25 .62	.82 .90	.37 .59
		.49 .68	.63 .82	.90 .93	

Note: Within the cells, the proportion of samples containing the population mean within the confidence interval is given by the top number while the proportion for the standard deviation is given by the bottom number.

Real Data Set

The mean of the 200 respondents for the selected item was 3.28, the standard deviation was 1.0758, and the percent of responses for values 1 through 5 were as follows: 4%, 19%, 38.5%, 22%, and 16.5%, respectively. As with the randomly generated data, the response rate and effect size conditions were simulated by blanking out responses. For the 0.0 effect size condition, responses were randomly blanked out to create the nonrespondents. For the .25 effect size condition, the probability of the values 1, 2, 3 being blanked out was reduced, and in the .50 effect size condition, the probability of values less than 5 being blanked out was reduced. For example, in the 0.0 effect size condition with a 50% response rate, the probability of a value less than 4 being blanked out was .31; for the .25 effect condition and 50% response rate, this dropped to .20; and for the .50 effect condition and 50% response rate, the probability was now only .07. The resulting means and standard deviations for the respondents, and the mean effect size differences and standard deviations for the effect size and response rate conditions are shown in Table 3.

Table 3

Resulting Means and Standard Deviations for Respondents and Effect Size Differences Between Respondents and Nonrespondents Using Real Data Set

		Response Rate			
	Mean for Respondents Mean Standard Dev. Mean Effect Size Standard Dev.	50%	70%	90%	
	0.0	3.2803 1.0753 -.0006 .1431	3.2791 1.0758 .0029 .1563	3.2801 1.0758 -.0006 .2385	3.2798 1.0756 .0006 .1842
Effect Size	.25	3.1480 1.0439 .2455 .1413	3.2006 1.0590 .2459 .1574	3.2528 1.0708 .2529 .2410	3.2005 1.0579 .2481 .1851
	.50	3.0103 .8923 .5015 .1164	3.1185 .9800 .5003 .1503	3.2264 1.0479 .4980 .2602	3.1184 .9734 .4999 .1861
		3.1462 1.0038 .2488 .2450	3.1994 1.0383 .2497 .2553	3.2531 1.0648 .2501 .3199	

Note: Within the cells, the top number is the mean of the sample means for respondents, the second is the mean of the standard deviations, the third is the mean effect size difference between respondents and nonrespondents, and the bottom number is the standard deviation of the effect size differences. The mean and standard deviation for the complete sample was 3.28 and 1.0758, respectively. Each cell was replicated 5,000 times.

The theoretical standard errors associated with each sample based on the response rate condition were then computed for the mean and standard deviations. The proportions of sample estimates that contained the population value within a 95% confidence interval were calculated with respect to the mean and standard deviation. The results can be seen in Table 4.

Table 4

Proportion of Sample Estimates Containing the Population Mean and Standard Deviation within a 95% Confidence Interval Using Real Data Set

	Proportion of Means Proportion of Standard Deviations	Response Rate 50%	 70%	 90%	
Effect Size	0.0	.9926 .9970	.9998 .9998	1.0000 1.0000	.9975 .9989
	.25	.8392 .9794	.9710 .9992	1.0000 1.0000	.9334 .9929
	.50	.0770 .1240	.5148 .4559	.9998 1.0000	.5305 .6098
		.6329 .7001	.8285 .9015	.9999 1.0000	

Note: Within the cells, the proportion of samples containing the population mean within the confidence interval is given by the top number while the proportion for the standard deviation is given by the bottom number.

The real data set also showed that the combination of low response rate and differences between respondents and nonrespondents resulted in biased estimates of the population mean and standard deviation. In comparison to the randomly generated data, however, the level of bias would not appear as severe when the effect size difference is low and the response rate is above 50%. For example, in the .25 effect size and 70% response rate condition the proportion of samples containing the population mean within a 95% confidence interval was .69 for the randomly generated data and .97 for the real data set. Tables 2 and 4 both show, however, that with an effect size of .50, severe bias in the estimates of the population mean and standard deviation can occur when the response rate is below 90%.

Discussion

For the 0.0 effect size condition, a 50%, 70%, or 90% response rate produced estimates of the population mean and standard deviation within their respective standard errors for both the randomly generated and real data sets. Also, only a 90% return rate would still seem adequate for estimates of the population mean or standard deviation when there is a .50 effect size difference between the respondents and nonrespondents. The randomly generated data in the .50 effect size condition with a 90% return rate resulted in 82% of the sample estimates containing the population mean within a 95% confidence interval, while a 50% return rate resulted in only 4% of the samples containing the population mean within the confidence interval around the estimate. Similarly for the real data set, the proportion of samples containing the population mean within the confidence interval dropped from near 1.00 with a 90% response rate to just .08 with a 50% response rate. The standard deviation estimates did not appear to be as affected by response rate and effect size differences as the population mean estimates. The proportions of sample estimates containing the population value for the standard deviation did not appear to drop dramatically for the 50% or 70% response rate conditions until the effect size reached .50.

In general, the results of this study support the Blixt (1990) recommendation that a 90% response rate is needed when survey respondents and nonrespondents vary by more than a slight amount. Even with a 90% response rate, it would appear prudent to use caution when making generalizations to the population if large differences between respondents and nonrespondents are suspected. On the other hand, if the assumption that nonrespondents were MCAR, or missing completely at random (Little and Rubin, 1987), is a tenable one, the results of this study would suggest that the disadvantage of a 50% response rate as compared to a 90% rate is that the confidence bands are numerically wider as a function of the lower n. The proportions of samples with the population values falling within the confidence interval for their estimated parameters only appeared to create differences among the response rate conditions when a systematic difference between respondents and nonrespondents was introduced. Certainly, Berdie's (1990) recommendation of a 65 to 75% response rate is reasonable as long as effect size differences between respondents and nonrespondents are small.

As a final note, it should be kept in mind that this study was only a simulation of what may occur with real data sets. It was limited to the extent that the randomly generated and actual data can be considered representative of data gathered by researchers. In particular, this study assumed that the samples were randomly drawn from the population, and in the case of the randomly generated data, the variable of interest to the researcher was normally distributed. Also, the differences between responders and nonresponders were formed by a relatively simplistic mathematical function. Essentially, the probability of responding was related to being below a given value in the distribution. A possible example of this type of simple function may be one of the findings for a study concerning parents' reports of child support that survey nonrespondents tended to have small or no awards or payments of support as determined by court records (Schaeffer, Seltzer, and Klawitter, 1991). In other surveys, however, nonresponders may differ from responders on several characteristics in a myriad of complex functions such that the potential impact on any single variable could be considerably more difficult to gauge than was the case for the data used in this study.

REFERENCES

Aiken, L. R. (1981). Proportion of returns in survey research. *Educational and Psychological Measurement, 41,* 1033-1038.

Aiken, L. R. (1987). The problem of nonresponse in survey research. *Journal of Experimental Education, 56,* 116-119.

Berdie, Doug R. (1990). High interview response rates: Much ado about nothing? Paper presented at the annual conference of the American Educational Research Association, Boston, MA.

Blixt, S. L. (1990). Response bias as a function of response rate. Paper presented at the annual conference of the Mid-Western Educational Research Association, Chicago, IL.

Shearer, C. B., & Jones, J. A. (1994). The validation of the Hillside Assessment of Perceived Intelligences (HAPI): A measure of Howard Gardner's theory of multiple intelligences. Paper presented at the annual conference of the American Educational Research Association, New Orleans, LA.

Ferguson, G. A. (1981). *Statistical analysis in psychology and education.* New York: McGraw Hill.

Gay, L. R. (1987). *Educational research.* Columbus: Charles E. Merrill.

Hartman, B. W., Fuqua, D. R., & Jenkins, S. J. (1986). The problems of and remedies for nonresponse bias in educational surveys. *Journal of Experimental Education, 54,* 85-90.

Howell, D. C. (1987). *Statistical methods for psychology.* Boston: Duxbury Press.

Jones, J. A. (1993). *Comparison of selection, mixture, and simplified selection modeling for varying patterns of nonresponse in survey instruments.* Unpublished Doctoral Dissertation, Kent State University, Kent, OH.

Kerlinger, F. N. (1986). *Foundations of behavioral research.* New York: Holt, Rinehart and Winston.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Schaeffer, N. C., Seltzer, J. A., & Klawitter, M. (1991). Estimating nonresponse and response bias. *Sociological Methods & Research, 20,* 30-59.

West, L. J. (1991). A frequently violated requirement in survey research. *Journal of Education for Business, 66,* 134-135.