Approaches to Social Research

THIRD EDITION

Royce A. Singleton, Jr.
College of the Holy Cross

Bruce C. Straits
University of California, Santa Barbara

New York ■ Oxford
Oxford University Press
1999

Exhibit P-459

Copyrighted Material

Oxford University Press

Oxford New York
Athens Auckland Bangkok Bogotá Buenos Aires Calcutta
Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paulo Singapore Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Copyright © 1988, 1993, 1999 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.,
198 Madison Avenue, New York, New York 10016
http://www.oup-usa.org

Oxford is a registered trademark of Oxford University Press

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

Singleton, Royce.
Approaches to social research / Royce A. Singleton, Jr., Bruce C. Straits.—3rd ed.
p. cm.
Includes bibliographical references and indexes.
ISBN 0-19-510525-7
1. Social sciences—Research. 2. Social sciences—Methodology.
I. Straits, Bruce C. II. Title.
H62.S4776 1998
300'.72—dc21 97-36851
CIP

The best statistical measure of population heterogeneity for a quantitative variable is the **standard deviation**, conventionally symbolized by the Greek letter sigma (σ). There is a direct link between this measure and the concept of standard error, introduced earlier in our discussion of probability sampling theory. The standard error, recall, indicates the degree of error in a sample estimate (the "average" amount by which a sample estimate deviates from the population value it estimates). The formula used to calculate the standard error is $\sigma/\sqrt{N}$, that is, the standard deviation divided by the square root of the sample size N. From this formula one can see that the standard error is directly related to the heterogeneity of the population as measured by σ and is inversely related to sample size. Also, the formula itself is a statement of the first principle regarding sample size: the greater the heterogeneity of the population, the larger the sample necessary to achieve a given level of precision.

Desired Precision

Technically speaking, precision refers to the degree of variability or error in a sample estimate, hence to the standard error. Intuitively, however, the concept of precision is perhaps best conveyed by relating it to the size of the confidence interval used to estimate a population value. Thus it is more precise to say that the average number of beers consumed is likely to fall between 2.0 and 4.0 than to say that the average is likely to fall between 1.0 and 5.0. For a given confidence level (say, 95 percent), the size of the confidence interval is directly related to the standard error; that is, the smaller the standard error, the smaller the confidence interval and the *more* precise the sample estimate. Of course, the larger the sample, the smaller the standard error. Therefore, it follows that the larger the sample, the greater the precision of the sample estimate.

Two facts about this relationship are especially noteworthy because they defy intuition to a certain extent. First, ordinarily it is the absolute size of the sample rather than the proportion of the population sampled that determines precision. As long as the population is relatively large, the proportion of the population sampled has a negligible effect on precision. For example, in 1990 the population of Vermont was a little over 560,000 and the population of Massachusetts was about 6 million. Now, if one were to take a simple random sample of 2000 in each of these states, the sample proportion of the total population would be 1 of every 280 persons in Vermont and 1 of every 3000 persons in Massachusetts. Yet, based on these samples, an estimate (say, of average income) would be just as precise for Massachusetts as for Vermont.

We can get a mathematical understanding of this relationship by reexamining the formula for the standard error. Although the formula given above shows that the standard error is determined only by the standard deviation and sample size, this formula actually applies to populations of theoretically infinite size. For finite populations, the formula should be multiplied by a correction factor equal to $\sqrt{1-f}$, where f is the **sampling fraction**, or proportion of the population included in the sample (Kish, 1965:43–44). Notice, however, that if the sampling fraction is very near zero, then the correction factor becomes $\sqrt{1}$, or 1, which has no effect on the

standard error. In most practical examples, the population is so much larger than the sample that f is extremely small—near zero—and the correction factor can be ignored. Only with small populations would we expect f to be much larger than zero; however, for a small population a small f implies a small sample size, and it is the latter that has the greater impact on the standard error.

Although precision is governed primarily by the absolute numerical size of the sample rather than by the proportion of the population sampled, the sample need not be enormous in size to yield very precise results. This is the second crucial fact about precision and sample size. The sampling error tends to be quite small for a sample of size 2000 to 3000, and increasing the sample size beyond this number decreases the error by so little that it usually is not worth the additional cost.

The mathematical explanation for this once again can be found in the standard error formula. Notice that the standard error goes down as the *square root* of the sample size goes up. Because of the square root function, each time we wish to decrease the standard error by one-half we must increase the sample size fourfold. At this rate, the precision gained with increased sample size reaches a point of minute, diminishing returns after a few thousand cases. Consider, for example, how sample size affects the standard error of a percentage, such as the estimated percentage of the vote a candidate will receive in an election. With two candidates and an evenly split vote, the standard error will be 5.0 percent for a sample size of 100. This error decreases to 2.5 percent when the sample reaches 400, and to 1.0 percent when the sample reaches 2500, as Table 6.4 shows. To get the error down to 0.5 percent would require a sample size of 10,000.

The reader should now begin to understand how election forecasters can make accurate predictions using samples of a few thousand out of millions of voters. The huge size of the population has no effect on the precision of sample estimates. And a sample size of 2000 to 3000 is large enough to predict accurately all but the closest elections.

We hasten to add that 2000 to 3000 should not be regarded as *the* standard size for reliable sample results. Not only does sample size depend on factors other than precision, but necessary levels of precision also vary widely from one study to the next. At one extreme, the CPS requires a sample of some 60,000 housing units, a sample "large enough so that the sampling errors of the total estimates of unemployment are only about 0.1%" (Sudman, 1976:3). At the other extreme, thirty cases

TABLE 6.4. Standard Error of a Percentage of 50 Percent, Broken Down by Sample Size*

Sample size	*Standard error (percent)*
100	5.0
400	2.5
2,500	1.0
10,000	0.5

*Standard errors are smaller for percentages greater than or less than 50 percent.

generally is regarded as minimally adequate for statistical data analysis, although most social researchers would probably recommend at least one hundred.

Sampling Design

The type of sampling design also affects decisions about sample size. Recall that one way of increasing precision other than selecting a larger sample is to use stratified rather than simple random sampling. In other words, for the same level of precision, a stratified random sample requires fewer cases than a simple random sample. A cluster sample, on the other hand, requires a somewhat larger number of cases for precision equal to that of a simple random sample. A nonmathematical explanation for these comparisons rests on the concept of heterogeneity.

Remember that stratified random sampling provides greater sampling efficiency (or precision) when the stratifying variable is related to the variable one is estimating. If it is related, much of the variation or heterogeneity in the estimated variable resides in differences between the strata, and each stratum tends to be relatively homogeneous. With respect to beer consumption, for example, there is much less variability among men or among women than between men and women. By sampling every stratum, the effect of stratified random sampling is to eliminate this *between* source of heterogeneity, leaving only the variation within the strata. And because of their homogeneity, each stratum requires a relatively small sample.

Although cluster sampling has the advantage of low cost, it lacks efficiency in precision. In two-stage cluster sampling, for example, there is variability both between and within the clusters. By taking a sample of clusters, the "between source" of heterogeneity that was eliminated in stratified sampling is still present in cluster sampling. If the variability between clusters tends to be large compared with the variability within the clusters, then the sampling error could be considerable, depending on how many clusters are selected. (For this reason, we want clusters to be heterogeneous compared with differences between the clusters—just the opposite of the strategy in stratified sampling.) Furthermore, each stage in a multistage cluster design contributes a source of variability or error to the total sample; consequently, the more stages, the larger the total sampling error tends to be, and the larger the sample required for a given level of precision.

Available Resources

Each individual case requires an expenditure of available resources. Consequently, at some point cost must enter into the equation for determining sample size. When a fixed amount of money and/or time has been allocated for a project, Sudman (1976:88–89) offers the following rule of thumb for survey research: allocate one-half of both money and time to data collection (and the other half to data analysis). Once the data collection procedure is specified, the sample size can be determined on a time and cost per case basis. That is, the number of cases will be equal to the total time for data collection divided by the time per case, or the total funds for data collection divided by the cost per case.

Number of Breakdowns Planned

The number of variables and variable categories into which the data are to be grouped and analyzed also must be taken into account in determining sample size. In general, the more breakdowns planned in the analysis—the more complex the relationships under investigation or the more distinct subcategories of separate interest—the larger the sample must be. Consider, for example, what happens when a sample of 1000 is divided into males/females, then into blacks/whites, then into people over/under 18 years of age, then into urban/rural residence. If we were interested in describing rural black young men, we might find that a breakdown of the 1000 cases looks something like this:

Sample	1000
Males	489
Black males	60
Young black males	20
Rural young black males	4

With just four cases available, the sample clearly would be too small for a reliable analysis of this particular subgroup. To avoid this kind of problem, it is important to estimate the number of breakdowns to be made during data analysis and to make sure that the total sample size will provide enough cases in each subcategory. If the requisite sample size turns out to be rather large, it might be more efficient to sample the relevant subcategories separately, as in stratified and quota sampling.

Final Notes on Sampling Errors and Generalizability

In sections on probability sampling and sample size considerations, we spoke repeatedly of sampling error—the deviation of a sample estimate from the true population value. The sampling error referred to in our discussions is random error, produced by the random selection of elements. Though unavoidable in sampling, this error can be estimated and reduced by increasing the size or efficiency of one's sample. There is, however, another type of sampling error, called **sample bias**,[11] which is nonrandom, difficult to detect, and often much more damaging to sample accuracy.

In probability sampling designs, the two most common sources of sample bias are **coverage error** due to incomplete sampling frames and **nonresponse bias** due to incomplete data collection. As an example of coverage error, we mentioned earlier that telephone directories provide inadequate sampling frames to the extent that they exclude the poor who cannot afford telephones and the more wealthy who tend to have unlisted numbers. The problem of nonresponse bias arises when, through refusals to cooperate, unreturned questionnaires, missing records, or some other means, the sample turns out to be a fraction of the number of cases originally se-