8.0 Nonresponse (From S. Lohr)

* 1. Effects of Nonresponse

Increasing sample size does nothing to reduce nonresponse bias – you may have more subjects (those that would respond to the survey), but you may also be increasing the potential bias due to nonresponse.

Analysis of complete records is representative of the population of persons that would respond to the survey – generally not the same population as the target population.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Stratum | Size | Total | Mean | Variance |
| Respondents | *NR* | *tR* |  |  |
| Nonrespondents | *NM* | *tM* |  |  |
| Entire Population | *N* | *t* |  |  |

The bias is approximately: , which will be small if the mean of nonrespondents is close to the mean of respondents or if the ratio of nonrespondents to population size is small.

8.2 Designing Surveys to Reduce Nonsampling Errors

Anticipating the reasons for nonresponse will help prevent high nonresponse rates – getting to know your target population.

Dillman (1995): factorial experiment in the 1992 Census Implementation Test

59.5

64.3

58.0

59.8

52.6

50.0

56.4

62.7

# Envelope

**Letter**

# Postcard

Yes

No

No

Yes

No

Yes

8.3 Callbacks and Two-Phase Sampling

1. Randomly sample *n* elements in the population; *nR* respond and *nM* do not respond.
2. Make a second call on a random subsample of 100*v*% of the *nM* nonrespondents in the sample, with the subsampling fraction of *v*.

We could estimate the mean and total value of nonrespondents’ responses

 

8.4 Mechanisms for Nonresponse

Responding can be thought of as a random variable that takes on two values

|  |  |  |
| --- | --- | --- |
| *Ri* = |  | 1 if unit *i* responds |
| 0 if unit *i* does not respond |

The probability that a unit selected for the sample will respond (propensity score for *i* )

 

In all model based approaches to dealing with nonresponse, *yi* is a response of interest, **x***i* is a vector of information know about element *i* in the sample. Information used in the survey design is also included in **x***i*.

a. Missing completely at random

 Where φ does not depend on **x***i*, *yi*, or the survey design.

 The respondents are representative of the selected sample.

 When nonresponse is ignored, the MCAR mechanism is implicitly adopted.

b. Missing at random given covariates (Ignorable nonresponse)

 If φ depends on **x***i* but not on *yi*, the data are missing at random.

 The nonresponse depends only on observed variables.

 Since we know the values of **x***i* we can successfully model the nonresponse.

c. Nonignorable Nonresponse

 If the probability of nonresponse depends on the value of a response variable and cannot be completely explained by the values of the **x**’s

Imputation

1. Deductive imputation

Logic defines the missing value

1. Cell mean imputation (not recommended)

Artificially reduces variability

Changes associations between variables

1. Hot Deck Imputation

Elements are divided into classes

Values from one responding element are substituted for each missing response

1. Sequential

Previous response in a given subclass is used

1. Random

Donor is randomly selected

1. Nearest neighbor

Closest in terms of preselected characteristics

1. Regression Imputation (recommended)

Predict the missing value from responses to questions that were answered by all (or nearly all) persons

Use the predicted value PLUS a randomly generated error term (to maintain original variability in responses) – based on the standard error of estimate.

1. Substitution

Replacing the case with missingness completely

1. Multiple imputation