



# Australian Bureau of Statistics

## Statistical Language - Types of Error



### Statistical Language



### Types of Error

#### What is error?

**Error (statistical error)** describes the difference between a value obtained from a data collection process and the 'true' value for the population. The greater the error, the less representative the data are of the population.

Data can be affected by two types of error: sampling error and non-sampling error.

#### What is sampling error?

**Sampling error** occurs solely as a result of using a sample from a population, rather than conducting a census (complete enumeration) of the population. It refers to the difference between an estimate for a population based on data from a sample and the 'true' value for that population which would result if a census were taken. Sampling errors do not occur in a census, as the census values are based on the entire population.

Sampling error can occur when:

- the proportions of different characteristics within the sample are not similar to the proportions of the characteristics for the whole population (i.e. if we are taking a sample of men and women and we know that 51% of the total population are women and 49% are men, then we should aim to have similar proportions in our sample);
- the sample is too small to accurately represent the population; and
- the sampling method is not random.

Sampling error can be measured and controlled in random samples where each unit has a chance of selection, and that chance can be calculated. In general, increasing the sample size will reduce the sample error.

#### What is non-sampling error?

**Non-sampling error** is caused by factors other than those related to sample selection. It refers to the presence of any factor, whether systemic or random, that results in the data values not accurately reflecting the 'true' value for the population.

Non-sampling error can occur at any stage of a census or sample study, and are not easily identified or quantified.

Non-sampling error can include (but is not limited to):

- **Coverage error:** this occurs when a unit in the sample is incorrectly excluded or included, or is duplicated in the sample (e.g. a field interviewer fails to interview a selected household or some people in a household).
- **Non-response error:** this refers to the failure to obtain a response from some unit because of absence, non-contact, refusal, or some other reason. Non-response can be complete non-response (i.e. no data has been obtained at all from a selected unit) or partial non-response (i.e. the answers to some questions have not been provided by a selected unit).
- **Response error:** this refers to a type of error caused by respondents intentionally or accidentally providing inaccurate responses. This occurs when concepts, questions or instructions are not clearly understood by the respondent; when there are high levels of respondent burden and memory recall required; and because some questions can result in a tendency to answer in a socially desirable way (giving a response which they feel is more acceptable rather than being an accurate response).
- **Interviewer error:** this occurs when interviewers incorrectly record information; are not neutral or objective; influence the respondent to answer in a particular way; or assume responses based on appearance or other characteristics.
- **Processing error:** this refers to errors that occur in the process of data collection, data entry, coding, editing and output.

## Why does error matter?

---

The greater the error the less reliable are the results of the study. A credible data source will have measures in place throughout the data collection process to minimise the amount of error, and will also be transparent about the size of the expected error so that users can decide whether the data are 'fit for purpose'.

### Examples of question wording which may contribute to non-sampling error.

Memory recall:

"How many kilometres did you travel in July last year?"

Socially desirable questions:

"Do you regularly recycle your waste paper and plastics?"

Under reporting:

"How many glasses of alcohol do you drink per week?"

Over reporting:

"How much did you win from gambling last week?"

Leading question:

"Do you think the government is doing enough to stop the increase in violent crime on our streets?"

Double-barrelled question:

"Are you happy with the price of, and services offered by, your gym membership?"

**Recommended: Read Measures of Error next**

## Further information:

---

### External links:

- [Errors in Statistical data](#)
- 

**[Return to Statistical Language Homepage](#)**

This page last updated 3 July 2013



Unless otherwise noted, content on this website is licensed under a Creative Commons Attribution 2.5 Australia Licence together with any terms, conditions and exclusions as set out in the website Copyright notice. For permission to do anything beyond the scope of this licence and copyright terms contact us.