University of Waterloo W. H. Cherry

ERROR: Categorization of Error in the Social Sciences

The *idea* of error and its categorization in data-based investigating is widely recognized but terminology differs among disciplines; Table HL8.1 at the right compares the terminology in these Statistical Highlights with that in the social sciences. Advantages of the former are:

- ○ several words with the same (or similar) meanings are avoided;
- ○ sources of different categories of error are distinct and well defined;
- ○ four populations and the sample, the individual case and repetition, and the real world and the model, are clearly distinguished.

SOURCE: (for social sciences terminology): Singleton, R.A., Straits, B.C. and M.M. Straits: *Approaches to Social Research*. Second Edition, Oxford University Press, New York, 1993, pages 114-118, 171-172, 185-187, 216, 393, 402, 433 and 452-456.

Table HL8.1: ERROR TERMINOLOGY

Statistical Highlights	Social Sciences
Study error	External validity, sample bias, generalizability
Non-response error	Sample bias
Sample error	Sampling error, generalization
Measurement error	(Construct validity)
Comparison error	Internal validity
Model error	-----
Measuring inaccuracy	Measurement (in)validity, systematic measurement error
Measuring imprecision	Measurement (un)reliability, random measurement error

NOTE: 1. Because they unnecessarily duplicate terminology in these Statistical Highlights, words (or phrases) to **avoid** include:

- ∗ **Applicability** (of an Answer) refers to study error and/or sample error.
- ∗ **Generality** (of an Answer) usually refers to sample error; in DOE, **generality** (or a **wider inductive basis**) may refer to whether the Plan involves a factorial treatment structure so that interaction effect(s) can be estimated.
  - – **Generalizability** refers to *study* error and **generalization** to *sample* error (pp. 185-187 in the Source above).
- ∗ **Reliability** [usually] refers to adequate precision (attained by managing *im*precision) [sometimes to adequate accuracy].
- ∗ **Sensitivity** (ability to detect an effect) refers to adequate precision (attained by managing *im*precision).
- ∗ **Strength** (of an Answer) means precision so **weakness** means *im*precision.
- ∗ **Trustworthiness** (of an Answer) means accuracy so **untrustworthiness** means *in*accuracy.
- ∗ **Validity** (of an Answer) means accuracy so **invalidity** means *in*accuracy.

These matters and, more generally, the terminology of these Highlights are the subject of the Glossary in Statistical Highlight #91; background information about statistical distinctions, and notation which helps to maintain them, are given in Section 3 on pages HL91.2 and HL91.3 in Statistical Highlight #91.

- ● *Construct validity* – the meaning of a response to a measuring instrument and a source of measurement error – is barely considered in these Highlights (*e.g.*, see Note 8 at the top of page HL2.12 in Statistical Highlight #2 – see also pages 124 to 129 in the Source given above).

Appendix [optional reading]

The idea of **error** [introduced near the middle of the first side of Statistical Highlight #87 (and elsewhere, like page HL18.1 in Statistical Highlight #18)] is the difference between a point estimate of a numerical Answer and the true value.

More specifically, **overall error** is the difference between the (numerical) Answer provided by data-based investigating and the (unknown) answer that reflects the *actual* (or '*true*') state of affairs in the population or process.

Definitions of *our* six categories of error (listed in the first six lines of the left-hand column of Table HL8.1 above) are:

- ∗ **Study error:** the difference between [the (true) values of] the study population/process attribute and the target population/-process attribute. [The population-process distinction is discussed in Statistical Highlight #94.]
- ∗ **Non-reponse error:** the difference between [the (true) values of] the respondent population and study population attributes.
- ∗ **Sample error:** the difference between [the (true) values of] the sample attribute and the study population/process attribute.
- ∗ **Measurement error:** the difference between a measured value and the true (or long-term average) value of a *variate*.
  - – **Attribute measurement error:** the difference between a measured value and the true (or long-term average) value of a [population/process or sample] attribute.
- ∗ **Comparison error:** for an Answer about an $\mathbf{X}$-$\mathbf{Y}$ relationship that is based on comparing attributes of groups of elements or units with different values of the focal variate(s), comparison error is the difference from the *intended* (or *true*) state of affairs arising from:
  - – differing distributions of lurking variate values between (or among) the groups of elements or units OR – confounding.

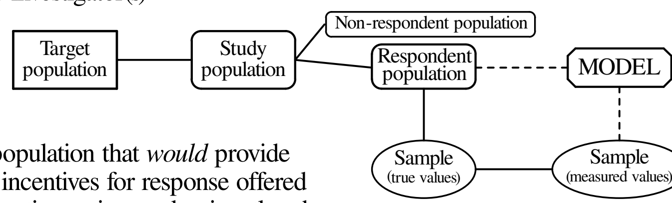
  The alternate wording of the last phrase accommodates the equivalent terminologies of lurking variates and confounding; in a particular context, we use the version of the definition appropriate to that context:
  - ● 'lurking variates' can more readily accommodate phenomena like Simpson's Paradox – see Statistical Highlight #51;
  - ● 'confounding' is more common in the context of comparative Plans – see Statistical Highlights #63 and #3.
- ∗ **Model error:** the difference between the model and its modelling assumptions and the actual state of affairs in the real world; modelling assumptions in introductory courses typically include:
  - ○ equiprobable selecting of units for the sample;
  - ○ the form of the structural component of the response model;
  - ○ the normality of each residual;
  - ○ probabilistic independence of the residuals;
  - ○ equal standard deviations of (response) variate values among different groups of units.

(*continued overleaf*)

NOTES: 2. Study, non-response and sample error are defined in terms of *attributes* of groups of elements whereas measurement error involves *individual* measurements – this is why the additional (sub)category of *attribute* measurement error is needed.

3. We need *both* true values and long-term average values in the two measurement error definitions because:
- 'true' values for quantities like length, mass and time (and the many quantities derived from them) can be invoked because there are **standards** (*i.e.*, certified *known* values) for such quantities – see Statistical Highlight #42;
- long-term average values may be all we have available when, for instance, investigating the distribution of responses to a questionnaire with particular question wording and/or question order.

The schema at the right below shows five groups of elements which *we* distinguish for data-based investigating and which are involved in the definitions of our six error categories given overleaf on the lower half of page HL8.1. Also, the model is included in the schema, as a *link* between the sample and the respondent population.



∗ **Target population**: the group of elements to which the investigator(s) want Answer(s) to the Question(s) to apply.

∗ **Study population**: a group of elements *available* to an investigation.

∗ **Respondent population:** those elements of the study population that *would* provide the data requested under the incentives for response offered in the investigation [such incentives arise predominantly when the elements are people, but missing data may also arise when elements are *in*animate].

∗ **Non-respondent population:** those elements of the study population that would *not* provide the data requested under the incentives for response offered in the investigation – see also Note 4 below.

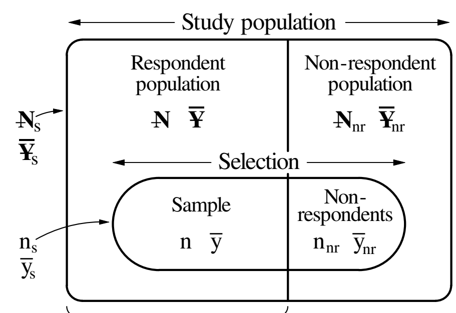
∗ **Sample**: the group of units selected from the respondent population *actually used* in an investigation – the sample is a *sub*-set of the respondent population (as the vertical line in the schema above reminds us).

NOTE: 4. As indicated in the diagram at the right, *we* consider the *study* population to be made up of the *respondent* and *non-respondent* populations. The set of units selected from the study population is the *selection*, and comprises the *sample* (from the respondent population) and the *non-respondents* (from the *non*-respondent population). The diagram has *two* categories of symbols:
- the Ns and ns refer to *numbers of elements or units*;
- the $\overline{Y}$s and $\bar{y}$s are *averages* of a response variate Y of the elements or units.

The relationships among the numbers of elements or units are:

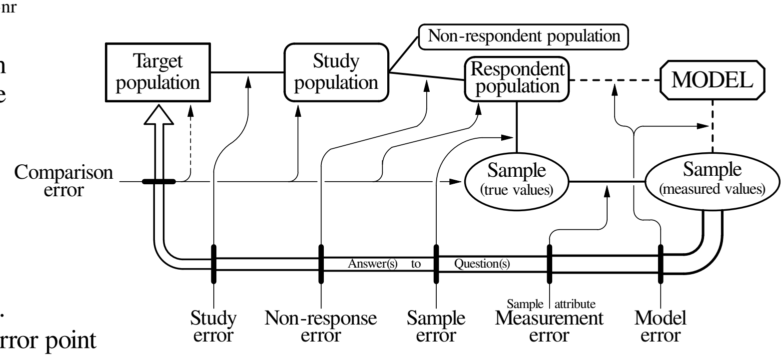
$$\begin{array}{ccccc} \text{Study population} & = & \text{Respondent population} & + & \text{Non-respondent population} \\ N_s & = & N & + & N_{nr} \end{array}$$

$$\begin{array}{ccccc} \text{Selection} & = & \text{Sample} & + & \text{Non-respondents} \\ n_s & = & n & + & n_{nr} \end{array}$$



Statistical theory, particularly of survey sampling, is developed mainly in the context of the *respondent* population, often with*out* recognizing it explicitly.

The impact of our six categories of error can conveniently be displayed as a development of the upper schema at the right.



- Five error category names are given across the bottom of the schema at the right, although 'measurement error' is really 'sample attribute measurement error.'
- The arrow rising from each error category name shows its place of impact in the schema.
  - The four arrows arising from comparison error point to *boxes* representing *groups* of elements or units (a population or a sample) rather than, as for the other five error categories, to *lines joining boxes*; the comparison error arrow at the right is to be taken as pointing to *both* sample ellipses.
    - *Multiple* comparison error arrows are a consequence of its different manifestations in different Question contexts, as summarized in Table HL60.2 on the lower half of page HL60.4 in Statistical Highlight #60.
- The broad arrow from the sample ellipse of measured values back to the target population represents Answers(s) to the Question(s) about the target population that are *inferred* from measured sample data on response (and, usually, explanatory) variates.
  - The thick lines crossing this broad arrow at the arrows rising from each error category represent *limitations* imposed by error on Answer(s); the progressive *de*crease in width of the broad arrow after each error category reinforces this idea.