Five-minute Science Lesson:

A Brief Discussion of Some Causes of Testing Errors

Raymond Nelson [1]

Scientific tests are intended to quantify an unknown parameter of interest that cannot be easily subject to direct physical measurement (subject only to random measurement error) or deterministic observation (immune to random variation and immune to human influence). Scientific tests rely on proxy information that is correlated with, though not itself, the unknown parameter of interest. Scientific tests are inherently probabilistic, and are not expected to be infallible. Tests are expected to quantify – probabilistically – the strength of information to support a conclusion or the margin of uncertainty that surrounds a conclusion. Although not inclusive of errors that may result from test faking (countermeasures) or testing errors related to suitability or representation, following is a brief discussion of some causes of testing error.

Procedural error

One potentially obvious cause of testing errors can be the incorrect execution of the testing procedures or incorrect use of the test instrumentation. A traditional way of reducing procedural errors relied on professionalism and professional infrastructure to avoid or reduce test errors that may result from procedural errors. These methods can include the use of published standards, procedural rules, professional supervision, education and training, continuing education, quality control, and gaining extensive professional experience. Although important, these activities can also start to become economically burdensome. Some types of procedural errors can be reduced through automation. However, some testing procedures are not suitable for automation and there will most likely always remain some need for reliance on human professionals to accomplish subtle and complex human tasks in the testing context.

[1] Raymond Nelson is a research specialist with Lafayette Instrument Company. Mr. Nelson is a psychotherapist with and field polygraph examiner who has published numerous articles on many aspects of the polygraph test. Mr. Nelson is a past APA President, and is currently serving as an elected member of the APA Board of Directors. Mr. Nelson is one of the developers of the OSS-3 and ESS-M algorithms for test data analysis. The views and opinions expressed herein are those of the authors and not the APA or LIC.

Random error or un-controlled variation

Random error can be thought of as the normally expected variation in recorded data, numerical scores, and test/experimental outcomes that we would observe if we were to repeat a scientific test or experiment numerous times. An ideal test would always give the same result – regardless of human behavior and with no random variation. Use of frequentist confidence intervals and Bayesian credibility intervals is necessary because random variation seems to exist in every context in which measurements and data are recorded. One method of reducing random measurement error or uncontrolled variation as a source of testing error is to obtain more data. Using more data is form of reliance on the law-of-large-numbers (LLN) or the central-limit-theorem (CTL). The CTL tells us that although we often cannot measure an entire population we estimate an unknown population parameter as the means of the means from numerous random samples from the population. The LLN says more simply that larger random samples may sometimes more closely estimate reality. As the old saying goes – measure twice, cut once. Understanding random measurement error is an important objective of frequentist statistical theory, for which our tolerance for error due to random or uncontrolled variation is often expressed at the alpha = .05 level.

Systematic error

Another source of error can be thought of as systematic error. Understanding systematic error helps us to understand how strongly, even if imperfectly, a dataset or analytic result supports a particular hypothesis or categorical conclusion. How strongly does it constrain or allow the possibility that some other hypothesis or conclusions may actually be correct? Or, in more practical terms, how sure or confident can we be in the conclusion supported by the test data and analytic result? Systematic testing error is can is often estimated using Bayesian analysis. Systematic error can be thought of as an error in the underlying theory, procedures or testing apparatus.

Systematic error is reproducible error. Metaphors are sometimes useful to assist in developing our understanding of abstract concepts such as systematic error. For example: take a pistol to a target and make five holes. Aim for center-X. There will be a pattern of hole, and most likely – even with a skilled marksman – all the holes will not be at the exact same location (assume some reasonable distance). The different holes represent random error/variation. It is considered OK if the holes are close together – indicating a small amount of uncontrolled or random variation. But if the group of holes is clustered away from the center-X then that is the systematic error. In this weapon analogy it is easy to evaluate the systematic error and potential causes – and easy to make a few small adjustments to reduce the systematic error and put the next group of hole closer to or on top of the center-X. But the results may differ for a target at a different distance due to potential systematic differences in trajectory – and for which the influence of random variation may also become more obvious. It is useful to understand the difference between random error and systematic error in all testing and data analysis contexts.