

Precision and Bias

Measurement - Method and Process

What are precision and bias? Walter Shewhart said

“It is important to realize ... that there are two aspects of an operation of measurement; one is quantitative and the other qualitative. One consists of numbers or pointer readings such as the observed lengths in n measurements of the length of a line, and the other consists of the physical manipulations of the physical things by someone in accord with instructions that we shall assume to be describable in words constituting a text [Shewhart 1939, p.130]

This leads to two concepts of measurement; a method of measurement and a measurement process. “A measurement method consists of a prescription or written procedure by which one can go about the business of making measurements on the properties of some physical material. ... A measurement process includes: (a) measurement method, (b) system of causes, (c) repetition, and (d) capability of control. A measurement process we could call a realization of a method in terms of particular individuals, particular equipment, and particular material to be tested.” [Murphy, 1961]

Properties of the Measurement Process

Statistical Control

If we accept the concept of a measurement process we then move to thinking of controlling that process. We can borrow terms from the language of statistical quality control. To qualify as a specification of a method of measurement, a set of instructions must be sufficiently definite to insure the statistical stability of repeated measurements of a single quantity, that is, derived measurement processes must be capable of meeting the criteria of *statistical control*. [Eisenhart, 1963 and Murphy, 1961]

Statistical control connotes a sense of consistency or predictability, but there is much more to it. A series of repeat measurements of a quantity provide a logical basis for predicting the behavior for future measurements of the same quantity by the same measurement process if and only if these measurements may be regarded as a random sample from a “population” or “universe” of all conceivable measurements of the quantity by the specified measurement process. In practice, the measurements can be assumed to be a sample until evidence suggests otherwise. Control chart techniques are used to monitor the process and provide the evidence of lack of control.

Limiting Mean

Why this concern about consistency? Aren't repeated measurements the same? Yes, if the measurement is coarse enough. Measurement of a person's height to the nearest meter results in the same measurement almost every time, measurement of that height to the nearest mm. does not. The inability to get exactly the same measurement of a particular object at every attempt is explained by saying the measurements are affected by *errors*, which are interpreted as manifestations of variations in the executions of the

process of measurement resulting from “imperfections of instruments, imperfect technique, less than ideal environs”. So, if the measurements aren’t the same, how tall is that person? Obviously, the numbers aren’t completely random. There is consistency on, at least a coarse scale, i.e. “taller” people get bigger numbers. Are we searching for a “true value”? Although an easy concept to express, the “true value” is, in most circumstances, unknowable. So where does this “in control process” lead? It leads to a “limiting mean”.

Consider an “in-control” measurement process that is producing a series of measurements. If one were to calculate the arithmetic mean after each new measurement there is a mathematical theorem that guarantees the sequence of cumulative means converges to a limit. In fact, virtually all such sequences from this process will converge to the same limit. This limiting mean μ , the value of which each individual measurement is trying to express, can be regarded not only as the “center of gravity” of the infinite conceptual population of all measurements that might conceivably be generated by the measurement process under the specified circumstances, but also as the value of the quantity as determined by this measurement process.

What good is this rigor? These two ideas, random samples from a population and a limiting mean, establish the framework that enables one to make quantitative inferential statements about the value of the limiting mean from a finite set of numbers produced by the measurement process. However, the step from the inferential statements to statements about the “true value” of the quantity measured must be based on subject matter knowledge and skill, but not on statistical methodology.

Definition of Error

The error in any measurement of a particular quantity is the difference between the measurement and the “true value” of this quantity. However, since the true value is ordinarily unknown and unknowable the error will also be unknown and unknowable. Limits to the error of a single measurement may be inferred from (a) the *precision*, and (b) bounds on the *systematic error*, of the measurement process from which it was produced. There is some risk of being incorrect because the inference is based not on information from a specific measurement but only on the “typical case” of the errors that are characteristic of measurements from the measurement process.

Systematic Errors

When the limiting mean is not equal to the true value, the concerned measurement process is said to have *systematic error* or *bias*. The systematic error will usually have both constant and variable components. Generally, there are obvious and not so obvious sources of bias. Known constant bias can be eliminated or adjusted for. Unknown and/or variable sources need to be identified. Bias adversely affects the statistical control of a measurement process. Control charts for the process mean and process range are excellent tools for the detection of systematic errors. If the truth is not known then bias cannot be determined.

Precision

The *precision* of a measurement process is the degree of agreement among measurements obtained from the measurement process being evaluated under prescribed conditions. The process must be in a state of statistical control, otherwise the precision of the process has no meaning. There is no implication of closeness to the true value. Repeated observation is fundamental to determining precision. No information can be gained from a single measurement. A measurement method cannot be described as precise because the method can only be realized in the context of a measurement process and as process parameters change precision may change.

Accuracy

The accuracy of a measurement is the closeness of agreement between the test result and the true value. This is straight forward. The accuracy of a measurement process is the degree of agreement of a set of measurements with the true value of the quantity being measured. This is not so straight forward. There are two schools of thought regarding accuracy. One school argues that a process is accurate if the average of all measurements is close to the truth, regardless of the closeness of each individual measurement. The second school insists that accuracy should imply that any given measurement is very likely to be close to the true value. The first school disregards any process precision and asks for low bias. The second requires both low bias and high precision. To avoid confusion the ASTM asks for description of processes in terms of precision and bias only.

Precision in a Measurement Process

Precision changes with changes in the process. Change operators, materials, temps, equipments and the precision can change. Although efforts should be made to remove or minimize the effects of such changes, their complete removal will be virtually impossible. This is especially true when the measurement method is used in more than one laboratory. This leads to two other concepts.

Repeatability

Repeatability concerns the variability between independent measurement results obtained within a single laboratory in the shortest practical period of time by a single operator with a specific set of test apparatus using test specimens taken from a single quantity of homogenous material.

Reproducibility

Reproducibility deals with the variability between single test results obtained in different measurement processes, each of which has applied the test method to test specimens taken at random from a single quantity of homogeneous material.

Reproducibility may be obtained for different conditions (such as operators) within a single laboratory, or for different laboratories

Quantifying Precision (or Imprecision)

Interlaboratory Study (Practice 691)

To obtain reliable estimates of repeatability and reproducibility, it is necessary to conduct an interlaboratory study in which several “in-control” laboratories use the prescribed test method to measure random samples of a homogeneous material. Once the test results are collected precision can be quantified. A specific plan for an ILS will depend upon the materials and measurement processes being evaluated. Practice 691 provides a general outline. Consulting groups that have ILS experience and/or a statistician can be very helpful.

After the data are collected the “repeatability”, the “reproducibility”, and the bias are calculated.

Repeatability

The basic measure of precision is the standard deviation. (Actually a measure of imprecision, in that high values mean low precision) The standard deviation, s_r , is calculated: (for a collection of measurements x_1, x_2, \dots, x_n)

$$s_r = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

If several sets of measurements are available from different processes using the same test method. The results can be pooled to obtain an overall measure of repeatability. Let $j = 1, \dots, p$ index the process and let $i = 1, \dots, n_j$ index the results from the j th process, Then the pooled repeatability is:

$$s_r = \sqrt{\frac{\sum_{j=1}^c \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{\sum_{j=1}^c (n_j - 1)}}$$

Reproducibility

For the sake of discussion, let's assume we are looking for the reproducibility between p laboratories. Further, we assume each lab produces n test results. From the labs' data we compute the pooled estimate of repeatability. The laboratory means are assumed to vary according to a normal distribution whose average is estimate by the average of all test results and whose standard deviation is given by s_L . Although s_L will be unknown it can be estimated because any laboratory sample mean will differ from the overall mean from two causes, s_L and s_r . Thus

$$(s_{\bar{x}})^2 = (s_L)^2 + (s_r)^2 / n$$

The n in the denominator is due to the averaging of the individual test results.

The reproducibility is a function of both the laboratory standard deviation, s_L , and the repeatability, s_r . They are related this way:

$$(s_R)^2 = (s_L)^2 + (s_r)^2$$

Substitution for s_L gives a direct calculation for s_R :

$$(s_R)^2 = (s_{\bar{x}})^2 + (s_r)^2(n-1)/n$$

Since interlaboratory variability cannot be less than intralaboratory variability, if s_R is less than s_r , s_R is set equal to s_r .

Indexes of Precision

Once the standard deviations are calculated one may report s , $2s$, or $2.8s$. $2s$ is the two standard deviation limit (generally one ought to expect about 95% of all observations to fall within $2s$ of the mean). Similarly, $2.8s$ is the two “difference standard deviation” limit, i.e. approximately 95% of all results will differ by less than $2\sqrt{2}s$.

Quantifying Bias.

If the true value or an accepted reference value is available the bias is the difference between the average of all test results and the reference value.

Variations in precision and bias.

Changes in the process due to material, operators, equipment, or environment change both precision and bias. Once the method and material are established, variations in process means due to qualitative factors can be viewed as systematic errors.

Writing Precision and Bias Statements

A Precision and Bias statement should include the following:

A statement of the test program:

- test materials,
- number of laboratories,
- number of test results per lab per material,
- the interlaboratory study with the analysis of the data.

A statement of precision:

- repeatability and reproducibility limits, with sd's reported in the text.

A description of and data from additional studies

A statement about bias:

- effects of changing properties on results,
- theoretical considerations for contributions to bias,
- estimates of maximum amount of bias.

References

Ku, Harry H. Editor. *Precision measurement and Calibration: Selected NBS Papers on Statistical Concepts and Procedures*. NBS Special Publication 300, vol. 1, 1969

Eisenhart, Churchill, Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems. *Journal of Research of the National Bureau of Standards-C.*, Vol. 67C, No. 2, 1963 (Reprinted in Ku, 1969)

Murphy, R.B., On the Meaning of Precision and Accuracy. *Materials Research and Standards*, ASTM, April. 1961. (Reprinted in Ku, 1969).

Youden, W.J., How to Evaluate Accuracy. *Materials Research and Standards*, ASTM, Apr. 1961. (Reprinted in Ku, 1969)

Youden, W.J., Experimental Design and ASTM Committees. *Materials Research and Standards*, ASTM, Nov. 1961 pp. 862-867 (Reprinted in Ku, 1969)

Practice E177, *Annual Book of ASTM Standards*, ASTM, Vol 14.02, pp 83-94, 1994.

Practice E691, *Annual Book of ASTM Standards*, ASTM, Vol 14.02, pp. 426-445, 1994.