

IE 361 Module 2

Basic Concerns of Metrology and Some Simple Applications of
Elementary Probability and Statistics to Metrology

Reading: Section 2.2 *Statistical Quality Assurance for Engineers*
(Sections 2.1-2.2.1 of Revised *SQAME*)

Prof. Steve Vardeman and Prof. Max Morris

Iowa State University

August 2008

Formulating Appropriate Measures of Process Performance

In a quality assurance/process improvement context, measures of process performance should be *customer-oriented*.

A nice example of this point is an early attempt by one of the "Big 3" US auto makers to track initial customer satisfaction. This company first tried using corporate warranty costs (charged to the company by dealers doing in-warranty service) as its measure of satisfaction. That was clearly not anything a customer cares about. A more appropriate measure that was later adopted was "per vehicle visits to a dealer for warranty service."

Huge modern marketing efforts are aimed at finding out what is important to customers. In successful organizations, these help define what measures of process performance are employed and monitored by the organizations.

Formulating Appropriate Measures of Process Performance

In any data collection context (including a process improvement effort) *direct* measures of a quantity are always preferable to relatively *indirect* ones. If one wants to know about y , it is better to try to measure y than to measure w that is presumed to be related to y . If, for example, one is designing a controller for a clothes dryer, somehow measuring weight loss as water is removed from a load is preferable to measuring air temperature of the dryer's exhaust, even if it is easier to do the latter than to figure out how to do the former. In the above auto initial customer satisfaction context, some direct inquiries to customers about satisfaction is probably preferable to reliance on only the per vehicle service calls measure.

The matter of measuring process performance brings us naturally to the general topic of measurement or the science of "metrology."

Basic Terminology and Concerns of Metrology

A quantity for which one hopes to establish a true or correct value is called a *measurand*. (A measurand can be anything from the surface roughness of a machined metal part, to the torque delivered by a pneumatic torque gun, to the color of a printed carton.) "The problem" is that in the real world, a measurand can rarely (if ever) be known perfectly. That is, there is almost always some *measurement error* associated with attempts to evaluate a measurand. A measurement y never perfectly reflects a measurand x . This is illustrated in cartoon form in the next figure. (We are here going to use the generic term "device" to describe a particular combination of equipment, people, procedures, etc. used to produce a measurement.)

Basic Terminology and Concerns of Metrology

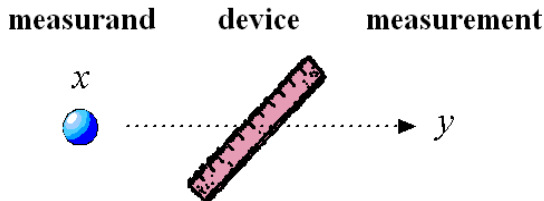


Figure: A Cartoon Illustrating How a Measurand x is Measured With Error to Produce a Measurement y

If one is going to do real world quality assurance, one must then have a working understanding of the nature of measurement error, how to quantify it, and how to account for its presence in assessing, monitoring, and improving quality. Probability and statistics are essential tools in understanding, quantifying, modeling, and assessing the impact of measurement error.

Basic Terminology and Concerns of Metrology

Basic issues in metrology are:

- Validity (is a measurement device or system really tracking what one wants it to track?)
- Precision (consistency of measurement)
- Accuracy (getting the "right" answer on average)

In developing an effective measurement system, these issues are addressed in the order given above. While these terms may appear to have similar colloquial meanings, they are technically quite different and it is essential that you use the correct terminology in IE 361.

Basic Terminology and Concerns of Metrology

A common analogy that should help you keep straight the difference between precision and accuracy is the analogy between measurement and target shooting.

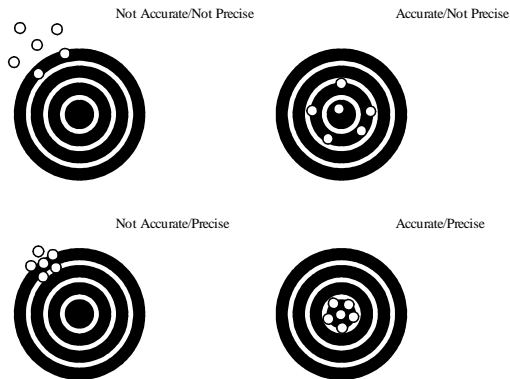


Figure: Measurement/Target Shooting Analogy

Basic Terminology and Concerns of Metrology

Good organizations devote significant resources to ensuring that their measurement systems are properly maintained and adequate to support process monitoring and improvement efforts. The comparison of the output of a measurement device to "known"/"standard" values (and subsequent adjustment of how that device reads to the standards) is known as *calibration*. Calibration has to do with assuring measurement *accuracy*.

A Simple Measurement Model—Using Probability to Describe Measurement Error

A basic statistical/probabilistic model for measurement is that what is measured, y , is the *measurand*, x , plus a normal random measurement error, ϵ , with mean δ and standard deviation σ_{device} .

$$y = x + \epsilon$$

Pictorially this is as in the following figure.

A Simple Measurement Model

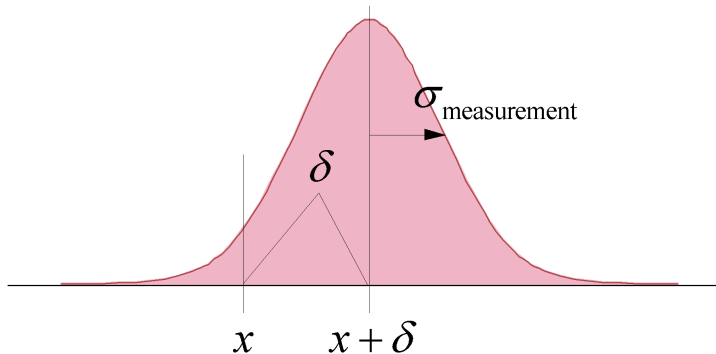


Figure: A Basic Probability Model for Measurement

A Simple Measurement Model

The difference between the measurand and the mean measurement is the *bias* in measurement, δ . Ideally, this is 0 (and it is the business of calibration to attempt to make it 0). At a minimum, measurement devices are designed to have a "*linearity*" property. This means that that over the range of measurands a device will normally be used to evaluate, if its bias is not 0, it is at least constant (i.e. δ does not depend upon x). This is illustrated in cartoon form in the next figure (where we assume that the vertical and horizontal scales are the same).

A Simple Measurement Model

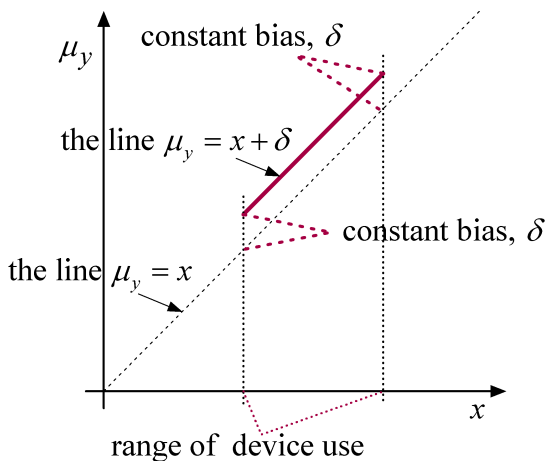


Figure: A Cartoon Illustrating Measurement Device "Linearity"

A Simple Measurement Model

When one makes multiple measurements on the same measurand using a fixed device, one has not a single, but multiple data values, y_1, y_2, \dots, y_n . It is standard to model these repeat measurements as *independent random draws from the same distribution* (with mean $x + \delta$ and standard deviation σ_{device}) illustrated on panel 10. We will abbreviate this model assumption as

$$y_i \sim \text{ind} (x + \delta, \sigma_{\text{device}})$$

and illustrate it in the following figure.

A Simple Measurement Model

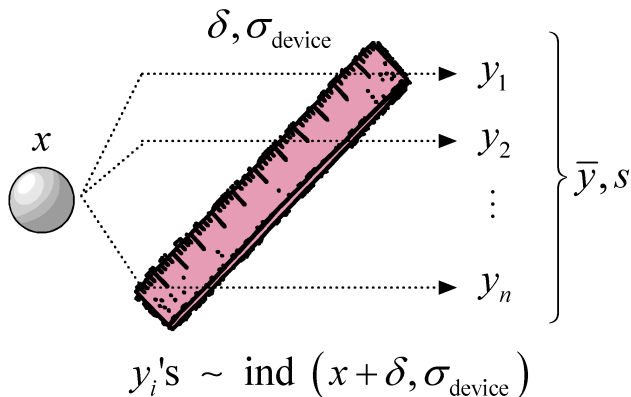


Figure: Cartoon Illustrating Multiple Measurements on a Single Measurand

A Simple Measurement Model

Note in passing (we'll have more to say about this directly) that this figure makes clear that the sample mean from repeat measurements of this type, \bar{y} , can be expected to approximate $x + \delta$ (the measurand *plus bias*). Further, the sample standard deviation, s , can be expected to approximate the device standard deviation, σ_{device} .

Most often in quality assurance applications, there is not a single measurand, but rather multiple measurands representing some important feature of multiple items or batches produced by a production process. In this context, we extend the basic measurement model by assuming that x varies/is random. (Variation in x is "real" process variation, not just measurement variation.) In fact, if the device is linear and the measurand is itself normal with mean μ_x and standard deviation σ_x and independent of the measurement error, we then have

$$y = x + \epsilon$$

A Simple Measurement Model

with mean

$$\mu_y = \mu_x + \delta$$

and standard deviation

$$\sigma_y = \sqrt{\sigma_x^2 + \sigma_{\text{device}}^2} > \sigma_x$$

(so observed variation in y is larger than the process variation because of measurement noise). Notice that these relationships are consequences of the basic (Stat 231) facts that if U and V are independent random variables $\mu_{U+V} = \mu_U + \mu_V$ and $\sigma_{U+V}^2 = \sigma_U^2 + \sigma_V^2$. (These are the "laws of expectation (or mean) and variance." If they don't look completely familiar, you should find and review them in your basic statistics text.)

A Simple Measurement Model

It is quite common in quality assurance applications to select n items from the output of a process and make a single measurement on each, with the intention of drawing inferences about the process. If the process can be thought of as physically stable and the device is linear, it is standard to model these measurements as *independent random draws from the same distribution* (with mean $\mu_x + \delta$ and standard deviation $\sqrt{\sigma_x^2 + \sigma_{\text{device}}^2}$). We can abbreviate this model assumption as

$$y_i \sim \text{ind} \left(\mu_x + \delta, \sqrt{\sigma_x^2 + \sigma_{\text{device}}^2} \right)$$

and illustrate it in the following figure.

A Simple Measurement Model

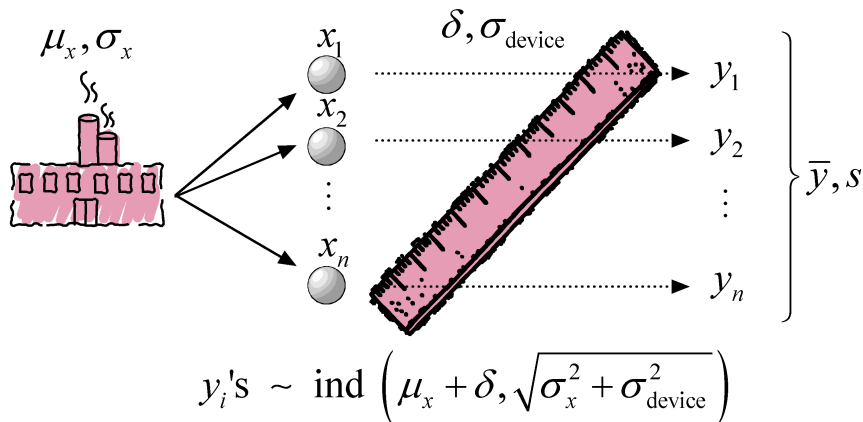


Figure: Cartoon Illustrating Single Measurements on Multiple Items From a Stable Process (Assuming Device Linearity)

A Simple Measurement Model

Panels 14 and 18 are *not* the same, even though they both lead to a "sample" of n measurements y . Here, \bar{y} , can be expected to approximate $\mu_x + \delta$ (the process mean plus bias) while the sample standard deviation, s , can be expected to approximate a combination of the device standard deviation and the process standard deviation.

A comparison of panels 14 and 18 illustrates the very important and basic insight that:

How sources of physical variation interact with a data collection plan governs what of practical importance can be learned from a data set, and in particular, how measurement error is reflected in the data set.

This principle governs the practical use of even the very simplest of statistical methods. We begin to illustrate this matter in the balance of this section (and expand on this theme even more fully in the next module).

Measurement Error and One Sample Inference

"Ordinary" confidence interval formulas (taught in Stat 231 and any other basic statistics course) based on a model for y_1, y_2, \dots, y_n of sampling from a normal distribution with mean μ and standard deviation σ , are

$$\bar{y} \pm t \frac{s}{\sqrt{n}} \text{ for estimating } \mu$$

and

$$\left(s \sqrt{\frac{n-1}{\chi_{\text{upper}}^2}}, s \sqrt{\frac{n-1}{\chi_{\text{lower}}^2}} \right) \text{ for estimating } \sigma$$

(It is your responsibility to review these on your own if they do not look absolutely familiar.) The "news" here is that sources of physical variation (and in particular, sources of measurement error and item-to-item variation) and data collection plans interact to give practical meaning to " μ " and " σ " and to govern what of practical importance can be learned from the application of these formulas.

Measurement Error and One Sample Inference

Let us consider what these elementary inference formulas and the data collection plans illustrated on panels 14 and 18 yield, beginning with the situation of repeat measurements on a fixed measurand illustrated on panel 14. Notice that in this context, if n **repeat measurements of a single measurand**, y_1, y_2, \dots, y_n , have sample mean \bar{y} and sample standard deviation s

- applying the t confidence interval for a mean, one gets an inference for

$$x + \delta = \text{measurand plus bias}$$

So

Measurement Error and One Sample Inference

- - in the event that the measurement device is known to be well-calibrated (one is sure that $\delta = 0$, there is no systematic error), the limits $\bar{y} \pm ts/\sqrt{n}$ based on $\nu = n - 1$ df are limits for x
 - in the event that what is being measured is a *standard for which x is known*, one may use the limits

$$(\bar{y} - x) \pm t \frac{s}{\sqrt{n}}$$

(once again based on $\nu = n - 1$ df) to estimate the device bias, δ

- applying the χ^2 confidence interval for a standard deviation, one has an inference for the size of the device "noise," σ_{device}

Measurement Error and One Sample Inference

Next consider what can be inferred from **single measurements made on n different measurands** y_1, y_2, \dots, y_n from a stable process with sample mean \bar{y} and sample standard deviation s as illustrated on panel 18.

- The limits $\bar{y} \pm ts/\sqrt{n}$ (for t based on $n - 1$ degrees of freedom) are limits for

$\mu_x + \delta =$ the mean of the distribution of true values plus bias

- Note also that the quantity s estimates $\sigma_y = \sqrt{\sigma_x^2 + \sigma_{\text{device}}^2}$, **that really isn't of fundamental interest**. So there is little point in direct application of the χ^2 confidence interval for a standard deviation in this context.

Measurement Error and One Sample Inference

Note in passing (we'll say more about this later) that since

$$\sigma_x = \sqrt{(\sigma_x^2 + \sigma_{\text{device}}^2) - \sigma_{\text{device}}^2} = \sqrt{\sigma_y^2 - \sigma_{\text{device}}^2}$$

estimates of specimen-to-specimen variation (free of measurement noise) might be based on estimates of σ_{device} (based on data like those on panel 14) and σ_y (based on data like those on panel 18). See display (2.3), page 20 of *SQAME* in this regard.

Example Below are $m = 5$ "measurements" made by a single analyst on a single physical sample of material using a particular assay machine.

1.0025, .9820, 1.0105, 1.0110, .9960

Measurement Error and One Sample Inference

These have mean $\bar{y} = 1.0004$ and $s = .0120$. Consulting a χ^2 table using $\nu = 5 - 1 = 4$ df, we can find a 95% confidence interval for σ_{device}

$$\left(.0120\sqrt{\frac{4}{11.143}}, .0120\sqrt{\frac{4}{.484}} \right) \text{ i.e. } (.0072, .0345)$$

(One moral here is that ordinary small sample sizes give very wide confidence limits for a standard deviation.) Consulting a t table also using 4 df, we can find 95% confidence limits for the true value for the specimen plus instrument bias ($x + \delta$)

$$1.0004 \pm 2.776 \frac{.0120}{\sqrt{5}} \text{ i.e. } 1.0004 \pm .0149$$

Measurement Error and One Sample Inference

Suppose that subsequently, samples from $n = 20$ different batches are analyzed and $\bar{y} = .9954$ and $s_y = .0300$. The t confidence interval

$$.9954 \pm 2.093 \frac{.0300}{\sqrt{20}} \quad \text{i.e.} \quad .9954 \pm .0140$$

is for $\mu_x + \delta$, the process mean *plus* any assay machine bias/systematic error.