

# IE 361 Module 3

More on Elementary Statistics and Metrology

Reading: Section 2.2 *Statistical Quality Assurance for Engineers*  
(Section 2.2 of Revised *SQAME*)

Prof. Steve Vardeman and Prof. Max Morris

Iowa State University

August 2008

# Basic One and Two Sample Inference Formulas

In Module 2 we reminded you that based on a model for  $y_1, y_2, \dots, y_n$  of sampling from a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ , elementary confidence limits are

$$\bar{y} \pm t \frac{s}{\sqrt{n}} \quad \text{for estimating } \mu \quad (1)$$

and

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

(where degrees of freedom for  $t$ ,  $\chi_{\text{upper}}^2$ , and  $\chi_{\text{lower}}^2$  are  $n-1$ ) and then made some simple applications of these limits in contexts that recognize measurement error.

# Basic One and Two Sample Inference Formulas

Parallel to the one sample formulas are the two sample formulas of elementary statistics. These are based on a model that says that

$$y_{11}, y_{12}, \dots, y_{1n_1} \quad \text{and} \quad y_{21}, y_{22}, \dots, y_{2n_2}$$

are independent samples from normal distributions with respective means  $\mu_1$  and  $\mu_2$  and respective standard deviations  $\sigma_1$  and  $\sigma_2$ . In this context, the so-called "Satterthwaite" approximation gives limits

$$\bar{y}_1 - \bar{y}_2 \pm \hat{t} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{for estimating } \mu_1 - \mu_2 \quad (3)$$

where appropriate "approximate degrees of freedom" for  $\hat{t}$  are

$$\hat{v} = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{s_1^4}{(n_1 - 1)n_1^2} + \frac{s_2^4}{(n_2 - 1)n_2^2}}$$

# Basic One and Two Sample Inference Formulas

$$\hat{v} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^4}{(n_1 - 1)n_1^2} + \frac{s_2^4}{(n_2 - 1)n_2^2}}$$

(This is a formula that you may not have seen in an elementary statistics course, where only methods valid when one assumes that  $\sigma_1 = \sigma_2$  are typically presented.) It turns out that above  $\min((n_1 - 1), (n_2 - 1)) \leq \hat{v}$ , so that a simple conservative version of this method uses degrees of freedom

$$\hat{v}^* = \min((n_1 - 1), (n_2 - 1))$$

Further, in the two-sample context, standard elementary confidence limits for comparing standard deviations are

$$\frac{s_1}{s_2} \cdot \frac{1}{\sqrt{F_{(n_1-1), (n_2-1), \text{upper}}}}} \quad \text{and} \quad \frac{s_1}{s_2} \cdot \frac{1}{\sqrt{F_{(n_1-1), (n_2-1), \text{lower}}}}} \quad \text{for} \quad \frac{\sigma_1}{\sigma_2} \quad (4)$$

# Basic One and Two Sample Inference Formulas

If these formulas do not look familiar, you should immediately stop and review them. Their use can, for example, be seen in Chapter 6 of Vardeman and Jobe's *Basic Engineering Data Collection and Analysis* or in the Stat 231 text. Here we will consider a variety of applications of them to problems that arise in metrological studies for quality assurance. Our basic objective is to amply illustrate (and help you develop the thought process necessary to successfully employ) the 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.*

# Some More Applications of Simple One-Sample Methods of Inference

In Module 2, we considered inference based on a "single sample" made up of either

- 1 repeat measurements on a single measurand made using the same device
- 2 single measurements made on multiple measurands coming from a stable process made using the same linear device

Here we present two more applications of one sample formulas.

# Application to a Single Sample Consisting of Single Measurements of a Single Measurand Made Using Multiple Devices (From a Large Population of Such Devices)

There are quality assurance contexts in which an organization has many "similar" measurement "devices" that could potentially be used to do measuring. In particular, a given piece of equipment might well be used by any of a large number of operators. (Recall that we are using the word "device" to describe a particular combination of equipment, people, procedures, etc. used to produce a measurement. So, in this language, different operators with a fixed piece of equipment are different "devices.") A way to try to compare these devices would be to use some (say  $n$  of them) to measure a single measurand. This is illustrated in the next figure.

# Single Measurements of a Measurand From Multiple Devices

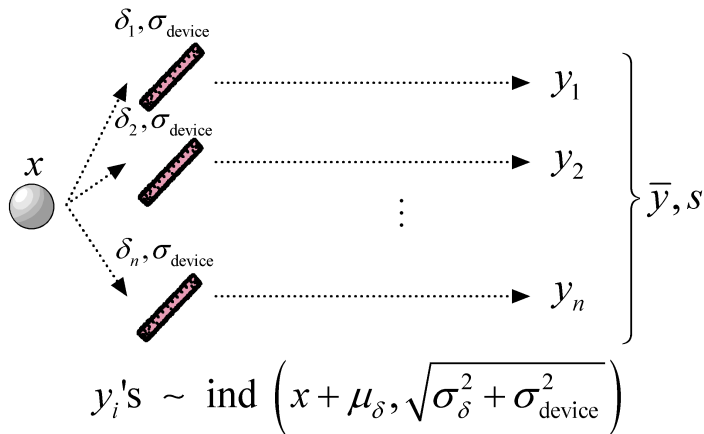


Figure: Cartoon Illustrating the Measurement of One Item Using Multiple Devices (From a Large Population of Such) Assuming a Common  $\sigma_{\text{device}}$

# Single Measurements of a Measurand From Multiple Devices

In this context, a measurement is of the form

$$y = x + \epsilon$$

where  $\epsilon = \delta + \epsilon^*$ , for  $\delta$  the (randomly selected) bias of the device used and  $\epsilon^*$  a measurement error with mean 0 and standard deviation  $\sigma_{\text{device}}$  (representing a repeat measurement variability for the particular device).

So one might write

$$y = x + \delta + \epsilon^*$$

Thinking of  $x$  as fixed and  $\delta$  and  $\epsilon^*$  as independent random variables ( $\delta$  with mean  $\mu_\delta$ , the average device bias, and standard deviation  $\sigma_\delta$  measuring variability in device biases) the laws of mean and variance from elementary probability then imply that

$$\mu_y = x + \mu_\delta + 0 = x + \mu_\delta \quad \text{and}$$

$$\sigma_y = \sqrt{0 + \sigma_\delta^2 + \sigma_{\text{device}}^2} = \sqrt{\sigma_\delta^2 + \sigma_{\text{device}}^2}$$

# Single Measurements of a Measurand From Multiple Devices

as indicated on panel 8. The average measurement is the measurand plus the average bias and the variability in measurements comes from both variation in device biases and the intrinsic imprecision of any particular device.

In a context where a schematic like panel 8 represents a study where several operators each make a measurement on the same item using a fixed piece of equipment, the quantity

$$\sqrt{\sigma_{\delta}^2 + \sigma_{\text{device}}^2}$$

is a kind of overall measurement variation that is sometimes called " $\sigma_{R\&R}$ ", the first "R" standing for **repeatability** and referring to  $\sigma_{\text{device}}$  (a variability for fixed operator on the single item) and the second "R" standing for **reproducibility** and referring to  $\sigma_{\delta}$  (a between-operator variability).

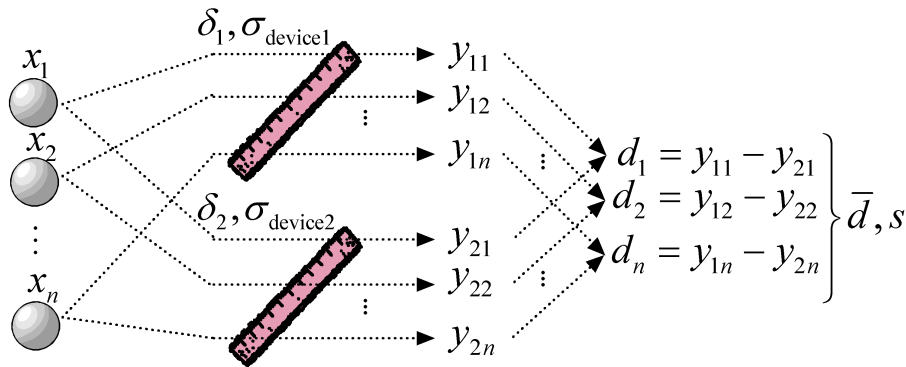
# Single Measurements of a Measurand From Multiple Devices

With  $\mu_y$  and  $\sigma_y$  identified, it is clear what the one sample confidence limits (1) and (2) estimate in this context. Of the two, interval (2) for " $\sigma$ " is probably most important, since  $\sigma_y$  is interpretable in the context of an "R&R" study, while  $\mu_y$  typically has little practical meaning. It is another question (that we will address in future modules with somewhat more complicated methods) how one might go about separating the two components of  $\sigma_y$  to assess the relative sizes of "repeatability" and "reproducibility" variation.

# Application to a Single Sample Consisting of Differences in Measurements on Multiple Measurands Made Using Two Devices (Assuming Device Linearity)

Another way to create a single sample of numbers is this. With two devices available and  $n$  different measurands, one might measure each once with both devices and create  $n$  differences between device 1 and device 2 measurements, as a way of potentially comparing the two devices. This possibility is illustrated in the next figure.

# Differences in Measurements on Multiple Measurands



$$d_i\text{'s} \sim \text{ind} \left( \delta_1 - \delta_2, \sqrt{\sigma_{\text{device1}}^2 + \sigma_{\text{device2}}^2} \right)$$

**Figure:** Cartoon Illustrating the Creation of Paired Differences from  $n$  Measurands Measured With Two Devices (Assuming Linearity and Constant  $\sigma_{\text{device}}$ 's)

# Differences in Measurements on Multiple Measurands

In this context, a difference is of the form

$$d = y_1 - y_2 = (x + \epsilon_1) - (x + \epsilon_2) = \epsilon_1 - \epsilon_2$$

and (again applying the laws of mean and variance from elementary probability) it follows that

$$\mu_d = \delta_1 - \delta_2 \quad \text{and} \quad \sigma_d = \sqrt{\sigma_{\text{device1}}^2 + \sigma_{\text{device2}}^2}$$

as indicated on panel 13. So applying the  $t$  interval for a mean (1), the limits

$$\bar{d} \pm t \frac{s}{\sqrt{n}}$$

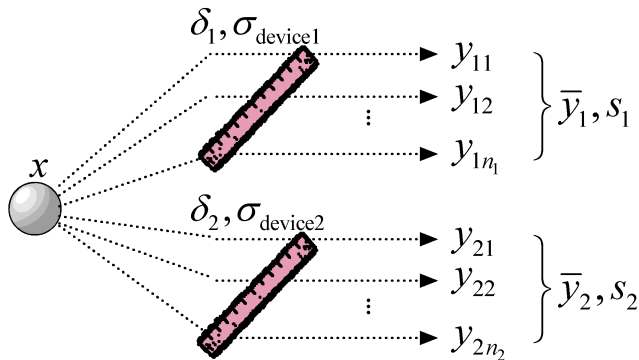
provide a way to estimate  $\delta_1 - \delta_2$ , the difference in device biases.

# Applications of Simple Two-Sample Methods of Inference

Application to Two Samples Consisting of Repeat Measurements of a Single Measurand Made Using Two Different Devices

One way to create "two samples" of measurements is to measure the same item repeatedly with two different devices. This possibility is illustrated in the next figure.

# Measurements of a Single Measurand From Two Devices



$y_{1i}$ 's  $\sim \text{ind} (x + \delta_1, \sigma_{\text{device1}})$  independent of

$y_{2i}$ 's  $\sim \text{ind} (x + \delta_2, \sigma_{\text{device2}})$

**Figure:** Cartoon Illustrating Repeat Measurement of the Same Measurand With Two Devices

# Measurements of a Single Measurand From Two Devices

Direct application of the two-sample confidence interval formulas show that the two-sample Satterthwaite approximate  $t$  interval (3) provides limits for

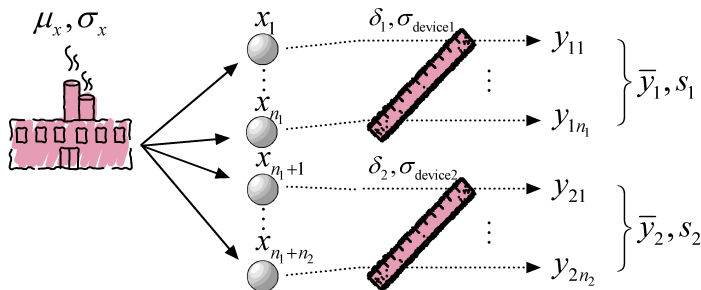
$$\mu_1 - \mu_2 = (x + \delta_1) - (x + \delta_2) = \delta_1 - \delta_2$$

(the difference in device biases), while the  $F$  interval (4) provides a way of comparing device standard deviations  $\sigma_{\text{device1}}$  and  $\sigma_{\text{device2}}$ . This data collection plan provides for straightforward comparison of the characteristics of the two devices.

# Application to Two Samples Consisting of Single Measurements Made With Two (Linear) Devices On Multiple Measurands From a Stable Process (Only One Device Used for a Given Measurand)

There are quality assurance contexts in which measurement is *destructive* and can not be repeated for a single measurand, and nevertheless one needs to somehow compare two different devices. In such situations, the only thing that can be done is to take items from some large pool of items or from some stable process and (probably after randomly assigning them one at a time to one or the other of the devices) measure them and try to make comparisons based on the resulting samples. This possibility is illustrated in the next figure.

# Single Measurements From Two Linear Devices On Multiple Measurands From a Stable Process



$y_{1i}$ 's  $\sim$  ind  $\left( \mu_x + \delta_1, \sqrt{\sigma_x^2 + \sigma_{\text{device1}}^2} \right)$  independent of

$y_{2i}$ 's  $\sim$  ind  $\left( \mu_x + \delta_2, \sqrt{\sigma_x^2 + \sigma_{\text{device2}}^2} \right)$

Figure: Cartoon Illustrating Two Samples Composed of Single Measurements on Different Measurands From a Stable Process

# Single Measurements From Two Linear Devices On Multiple Measurands From a Stable Process

Direct application of the two-sample Satterthwaite approximate  $t$  interval (3) provides limits for

$$\mu_1 - \mu_2 = (\mu_x + \delta_1) - (\mu_x + \delta_2) = \delta_1 - \delta_2$$

(the difference in device biases). So, in even in contexts where measurement is destructive, it is possible to compare device biases. It's worth contemplating, however, the difference between the present scenario and the immediately preceding one (represented by panel 16).

The measurements  $y$  in panel 16 are less variable than are the measurements  $y$  here in panel 19. That is evident in the standard deviations shown on the figures and follows from the fact that in the present case (unlike the previous one) measurements are affected by unit-to-unit/measurand-to-measurand variation. So all else being equal, one should expect limits (3) applied in the present context to be wider/less informative than when applied to data collected as in the last application.

# Single Measurements From Two Linear Devices On Multiple Measurands From a Stable Process

That is completely in accord with intuition. One should expect to be able to learn more useful to comparing devices when the same item(s) can be remeasured than when it (they) can not be remeasured.

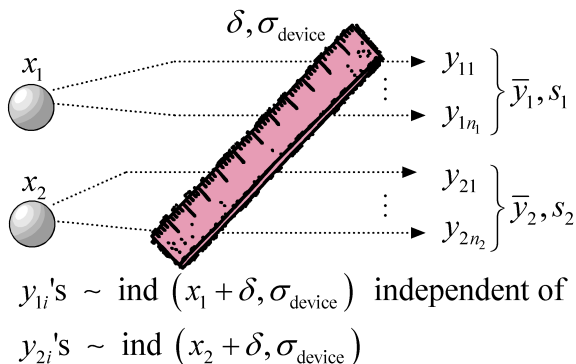
Notice that if the  $F$  limits (4) are applied here, one winds up with only an indirect comparison of  $\sigma_{\text{device1}}$  and  $\sigma_{\text{device2}}$ , since what can be easily estimated is the ratio

$$\frac{\sqrt{\sigma_x^2 + \sigma_{\text{device1}}^2}}{\sqrt{\sigma_x^2 + \sigma_{\text{device2}}^2}}$$

# Application to Two Samples Consisting of Repeat Measurements Made With One Linear Device On Two Measurands

A basic activity of quality assurance is the comparison of nominally identical items. Accordingly, another way to create two samples is to make repeated measurements on two measurands with a single device. This is illustrated in the next figure.

# Repeat Measurements of Two Measurands With One Linear Device



**Figure:** Cartoon Illustrating Repeated Measurement of Two Measurands With a Single Linear Device

# Repeat Measurements of Two Measurands With One Linear Device

In this context, it is clear that

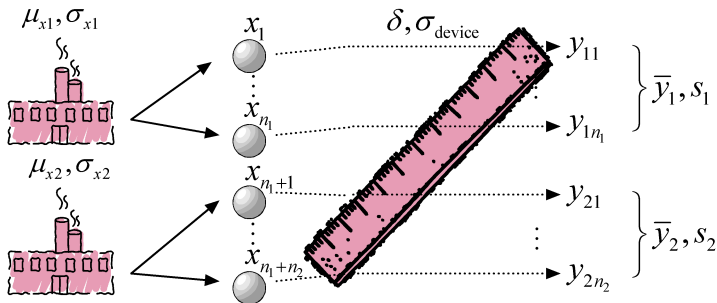
$$\mu_1 - \mu_2 = (x_1 + \delta) - (x_2 + \delta) = x_1 - x_2$$

so that application of the two-sample Satterthwaite approximate  $t$  interval (3) provides limits for the *difference in the measurands*, and hence a direct way of comparing the measurands. The device bias affects both samples in the same way and "washes out" when one takes a difference. (This, of course, assumes that the device is linear, i.e. that the bias is constant.)

# Application to Two Samples Consisting of Single Measurements Made Using a Single Linear Device on Multiple Measurands Produced by Two Stable Processes

Another basic activity of quality assurance is the comparison of nominally identical *processes*. Accordingly, another way to create two samples is to make single measurements on samples of measurands produced by two processes. This possibility is illustrated in the next figure.

# Single Measurements Made Using a Linear Device on Measurands Produced by Two Stable Processes



$y_{1i}$ 's  $\sim$  ind  $\left( \mu_{x_1} + \delta, \sqrt{\sigma_{x_1}^2 + \sigma_{\text{device}}^2} \right)$  independent of

$y_{2i}$ 's  $\sim$  ind  $\left( \mu_{x_2} + \delta, \sqrt{\sigma_{x_2}^2 + \sigma_{\text{device}}^2} \right)$

Figure: Cartoon Illustrating Single Measurements on Multiple Measurands From Two Processes Made Using a Single Linear Device

# Single Measurements Made Using a Linear Device on Measurands Produced by Two Stable Processes

In this context, it is clear that

$$\mu_1 - \mu_2 = (\mu_{x1} + \delta) - (\mu_{x2} + \delta) = \mu_{x1} - \mu_{x2}$$

so that application of the two-sample Satterthwaite approximate  $t$  interval (3) provides limits for the *difference in the process mean measurands* and hence a direct way of comparing the processes. Again, the device bias affects both samples in the same way and "washes out" when one takes a difference (of course, still assuming that the device is linear, i.e. that the bias is constant).

If the  $F$  limits (4) are applied here, one winds up with only an indirect comparison of  $\sigma_{x1}$  and  $\sigma_{x2}$ , since what can be easily estimated is the ratio

$$\frac{\sqrt{\sigma_{x1}^2 + \sigma_{\text{device}}^2}}{\sqrt{\sigma_{x2}^2 + \sigma_{\text{device}}^2}}$$

# Summary

Again, the major purpose of this module has been to illustrate (and help you develop the thought process necessary to successfully employ) the 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.*

The way we've approached this is through considering what the familiar elementary one and two sample methods of statistical inference provide under different data collection plans.