2010

Elementary Statistical Methods and Measurement Error

Stephen B. Vardeman
Iowa State University, vardeman@iastate.edu

Joanne Wendelberger
Los Alamos National Laboratory

Tom Burr
Los Alamos National Laboratory

Michael S. Hamada
Los Alamos National Laboratory

Leslie M. Moore
Los Alamos National Laboratory

See next page for additional authors

Authors
Stephen B. Vardeman, Joanne Wendelberger, Tom Burr, Michael S. Hamada, Leslie M. Moore, Marcus Jobe, Max Morris, and Huaiqing Wu

Elementary Statistical Methods and Measurement Error

Stephen B. Vardeman [*], Joanne R. Wendelberger [**], Tom Burr [**], Michael S. Hamada [**], Leslie M. Moore [**], J. Marcus Jobe [***], Max D. Morris [*], and Huaiqing Wu [*]

[*] Iowa State University, [**] Los Alamos National Laboratory, and [***] Miami University of Ohio

10.23.09

Abstract

How the sources of physical variation interact with a data collection plan determines what can be learned from the resulting data set, and in particular, how measurement error is reflected in the data set. The implications of this fact are rarely given much attention in most statistics courses. Even the most elementary statistical methods have their practical effectiveness limited by measurement variation; and understanding how measurement variation interacts with data collection and the methods is helpful in quantifying the nature of measurement error. We illustrate how simple one- and two-sample statistical methods can be effectively used in introducing important concepts of metrology and the implications of those concepts when drawing conclusions from data.

KEY WORDS: accuracy, bias, calibration, data collection, linearity, measurement error, precision, repeatability, reproducibility, statistical education.

1. Introduction

Good measurement is an essential part of collecting informative data, a vital ingredient of empirical learning. Measurement quality has many practical implications. In commerce, it is essential that a liter of fuel is a liter throughout the world, and that when a consumer's electric meter indicates that a kilowatt hour has been used, a kilowatt hour has really been delivered. In medicine, when a blood sample is sent to a laboratory for the

measurement of vitamin D concentration, it is important for setting of a proper patient treatment that the test result adequately reflects the underlying nature of the sample (Pollack (2009)).

Good measurement and good statistics go hand-in-hand. Meaningful inferences regarding real-world problems must be based on good measurements, while rigorous quantification of measurement quality depends upon statistical methods. In general, "measurement" is not an important emphasis in most statistics courses. In this article we argue that it can and should be.

The following methods are standard material in elementary statistics courses. There are the one-sample confidence limits for a mean,

$$\bar{y} \pm t \frac{s}{\sqrt{n}}, \tag{1}$$

and for a standard deviation,

$$s\sqrt{\frac{n-1}{\chi^2_{\text{upper}}}} \quad \text{and} \quad s\sqrt{\frac{n-1}{\chi^2_{\text{lower}}}}, \tag{2}$$

assuming a normal distribution. Only slightly less common are the two-sample confidence limits for the difference in the means of samples from two normal distributions,

$$\bar{y}_1 - \bar{y}_2 \pm \hat{t}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{3}$$

(where $\hat{t}$ denotes a t value with estimated degrees of freedom) and for the ratio of standard deviations from two normal distributions,

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

They are typically described as being applicable to "the" population (or process) mean(s) and "the" population (or process) standard deviation(s) and illustrated using some kind of (normal) "box(es) of tickets" model(s) (in the style of Freedman, Purves, and Pisani

(2007)).

Even in those rare cases where a statistics course raises the issue of measurement error, the actual impact of measurement error on what can be learned from the methods presented in (1) – (4) or what those methods can say about measurement error are not usually given careful treatment. This is unfortunate. These simple methods can be used to effectively communicate important points about measurement error and to quantify the impact of measurement error. We believe that a statistics education should include careful and meaningful discussions concerning what is estimated by the parametric functions $\mu, \sigma, \mu_1 - \mu_2,$ and σ_1 / σ_2 using (1) – (4) in the presence of measurement error, and how physical sources of variation affect data.

An outline of this article is as follows. In Section 2, we begin with some basic principles of measurement error and its quantification and modeling. In Section 3, we consider "one-sample" data collection schemes, how elementary inference methods are affected by measurement error, and how those methods can help quantify measurement error. Next, in Section 4, we consider more complicated situations, involving "two-sample" data collection plans and two-sample inference methods. In Section 5, we suggest several classroom activities to help bring alive the impact and quantification of measurement error. We conclude the article with a discussion in Section 6.

2. Some Basics

The “population" means and standard deviations we teach students to estimate are means and standard deviations *of the relevant data-generating mechanisms* (that include the relevant measurement processes). The assumption that the estimates correspond directly to quantities of subject matter interest is often not stated and sometimes not correct. For example, the population mean could refer to the mean weight of net contents in boxes of a type of cereal *plus* the mean measurement error in hypothetical repeated measurements of the net weight in one such box. Unfortunately, "the population" is "the *data*

population" whose characteristics can be substantially different from what is really of primary interest. In order to have a framework for discussing these concepts, we first set forth some basic terminology.

We introduce the term **measurand** to denote the numerical characteristic of an object being measured. When we apply a measurement process to a single object, we hope to learn about the value, x, of that measurand. We will use the word (measurement) **device** to include a fixed combination of physical measurement equipment, operator/data-gatherer identity, measurement procedure, and surrounding physical circumstances (such as time of day, temperature, etc.) used to produce a **measurement**, y, that is intended to represent the measurand. We shall demonstrate that the measurement potentially provides information not only about the measurand, but also about the measurement process itself. The difference between the measurement and the value of the measurand,

$$\varepsilon = y - x$$

is the **measurement error** produced in measuring x. If one adopts a probability model for y, the mean (which is obtained by taking the expected value) measurement error,

$$\mathrm{E}(\varepsilon) = \delta(x)$$

represents the **measurement bias** which quantifies **measurement accuracy. Note that the bias is potentially a function of** x, so different measurands might produce different biases. The standard deviation of the measurement error,

$$\sqrt{\mathrm{Var}(\varepsilon)} = \sigma_{\mathrm{meas}}(x),$$

(which again could depend upon x) quantifies **measurement precision**. Using this notation, the mean and standard deviation of the measurement y are

$$\mathrm{E}(y) = x + \delta(x) \quad \text{and} \quad \sqrt{\mathrm{Var}(y)} = \sigma_{\mathrm{meas}}(x) \ . \tag{5}$$

The possibility that measurement bias depends upon the measurand is problematic. In many measurement situations, substantial real world resources are spent trying to reduce the measurement bias through **calibration** studies and adjustments. When the bias is nonzero, it may be possible to assume that the bias is constant. In this case, some metrologists call the device making the measurement **linear** over a specified range of

measurands. The possibility that measurement precision depends upon the measurand is a more difficult matter and in general cannot be simply "adjusted away" by clever transformation of y.

If multiple measurands are under consideration, it is natural to think of x itself as random, with each measurand representing a sample from an underlying population. Where neither bias nor precision depends upon the measurand (i.e., $\delta(x) = \delta$ and $\sigma_{\text{meas}}(x) = \sigma_{\text{meas}}$), it may be appropriate to model a pair (x, ε) as independent with $\mathrm{E}(x) = \mu_x$ and $\mathrm{Var}(x) = \sigma_x^2$. One then has

$$\mathrm{E}(y) = \mu_x + \delta \quad \text{and} \quad \sqrt{\mathrm{Var}(y)} = \sqrt{\sigma_x^2 + \sigma_{\text{meas}}^2}\,. \tag{6}$$

The impact of measurement error displayed in (5) and (6) can be presented to elementary statistics audiences in terms of simple graphics like those in Figures 1 and 2.

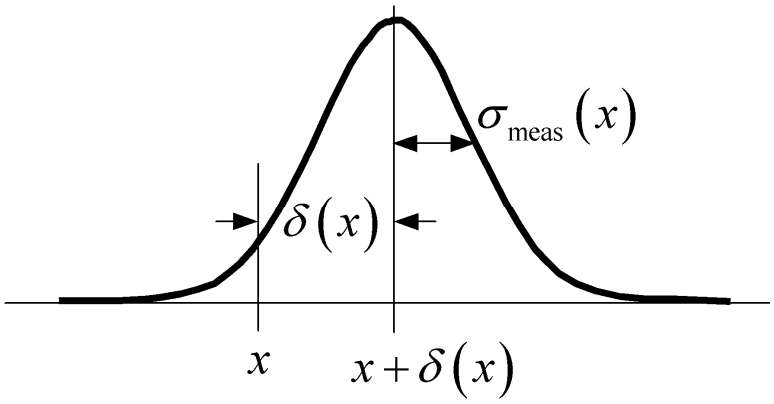


Figure 1: Distribution of a measurement y for a measurand x.

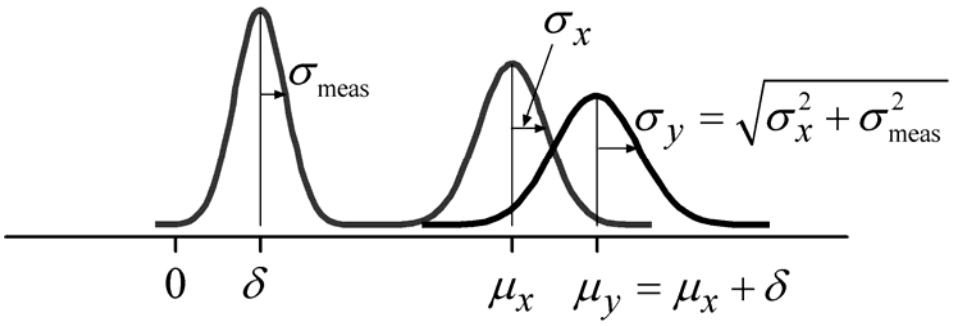


Figure 2: Random measurement error and measurand variation combine to produce observed variation.

All of this material is fairly standard, but it is often not presented in general statistics courses. What we propose here that goes beyond this well-known material is systematic attention to and elaboration of these ideas in the context of the methods given in (1) – (4).

3. One-Sample Inference and Measurement Error

A useful and common illustration associated with the one-sample methods in (1) and (2) involves a box of numbered tickets with μ and σ written on the side of the box, where μ and σ represent the mean and standard deviation of the numbers on the tickets in the box. A sample of n tickets is taken from the box which provides the numbers $y_1, y_2, \ldots, y_n$ that can be used to carry out inference for μ and σ. This picture raises the question, "Exactly what does the box represent?" and intentionally ignores details of data collection that are tied to what can actually be learned from the data.

Figures 3 and 4 are schematics of two quite different scenarios that are covered by the "single box of tickets model." In these and subsequent figures, the notation "$\sim \text{ind}(\cdot,\cdot)$" indicates independent draws from a distribution whose mean is the first argument and whose standard deviation is the second argument.

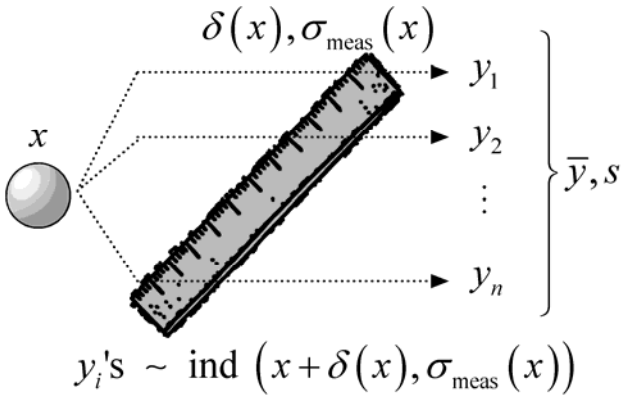


Figure 3: A single sample derived from n repeat measurements made on a single measurand with a fixed device.

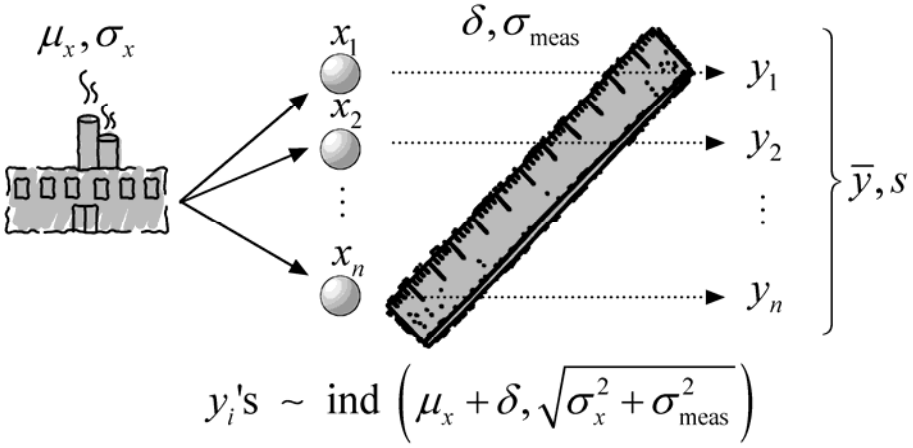


Figure 4: A single sample derived from single measurements made of n different measurands from a physically stable process with a fixed device (assuming device linearity $\delta(x) = \delta$ and constant measurement precision $\sigma_{\text{meas}}(x) = \sigma_{\text{meas}}$).

The first case, a single measurand measured repeatedly as depicted in Figure 3, is relatively simple, but not terribly common in practice. As a way to motivate this case, consider the situation of measuring the width of a single binder clip with a vernier micrometer that produced the measurements in Table 1.

Table 1: Repeated Width Measurements (in mm) of a Single Binder Clip

31.953
31.951
31.948
31.950
31.948
31.943
31.953

For this first case, application of formulas (1) and (2) produces inferences for, respectively,

a) $x+\delta(x)$, the measurand plus corresponding device bias, and

b) $\sigma_{\text{meas}}(x)$, a measure of intrinsic measurement variability for the given device at that measurand value *x*.

Notice that if what is being measured is a **standard** for which x is supposedly known, subtracting x from $\bar{y}$ in a) above produces a way to make an inference about the bias $\delta(x)$. This implies that the simple inference method in (1) already provides an elementary context for discussing the important topic of bias-adjustment in an elementary statistics course.

The second case, a sample of measurands measured by a single measurement system, is shown in Figure 4 and is much more complicated than the first. For the binder clip example, a sample of binder clips can be measured by a single micrometer to produce the measurements in Table 2.

Table 2: Single Width Measurements (in mm) for Ten Different Binder Clips

31.948
31.980
31.912
31.985
31.918
31.950
31.953
31.962
31.917
31.970

In general, if the measurand x is random,

$$\mathrm{E}(y) = \mathrm{E}(x+\varepsilon(x)) = \mathrm{E}(x) + \mathrm{E}(\mathrm{E}(\varepsilon(x) \mid x)) = \mathrm{E}(x) + \mathrm{E}(\delta(x))$$

and

$$\begin{aligned}\mathrm{Var}(y) &= \mathrm{Var}(x+\varepsilon(x)) = \mathrm{E}(\mathrm{Var}(x+\varepsilon(x) \mid x)) + \mathrm{Var}(\mathrm{E}(x+\varepsilon(x) \mid x)) \\ &= \mathrm{E}(\sigma^2_{\mathrm{meas}}(x)) + \mathrm{Var}(x+\delta(x)),\end{aligned}$$

where the "population" mean and standard deviation have the relatively simple forms given in (6) only when the measurement device is linear (i.e., $\delta(x) = \delta$) and has constant precision (i.e., $\sigma_{\mathrm{meas}}(x) = \sigma_{\mathrm{meas}}$). In this case, (1) and (2) respectively produce inferences for

a) $\mu_x + \delta$, the average measurand plus device bias, and

b) $\sqrt{\sigma_x^2 + \sigma^2_{\mathrm{meas}}}$, a measure of measurand variation inflated by measurement variation.

Notice that introductory treatments of statistical methods often use "random sample from physical population X" examples and implicitly assume that both $\delta = 0$ and $\sigma_{\mathrm{meas}} = 0$. This simplification misses the opportunity to introduce the concept of measurement error and teach what statistical methods can and cannot do. For example, the facts that "the mean" is really the mean of measurand plus bias and that no amount of sampling (or clever calculation) will eliminate the bias, are important practical points about both measurement and statistics. The fact that the interval in (2) applied directly to the

observations y will typically overstate the size of σ_x because of measurement variation is also important. If we are going to claim that our methods really help explain the world around us, it is important that we carefully discuss their limitations and implications.

Figure 5 illustrates a third way that a sample of n measurements y can be generated. Here, a single measurand is measured using n different devices. (In this and the remaining cases, we will not mention the binder clip example, but the reader can extend it to many of these cases.) Using the word "device" to represent all aspects of the process used to make a measurement, this kind of figure might illustrate a situation in which the same piece of measurement equipment is used by multiple operators to measure x.

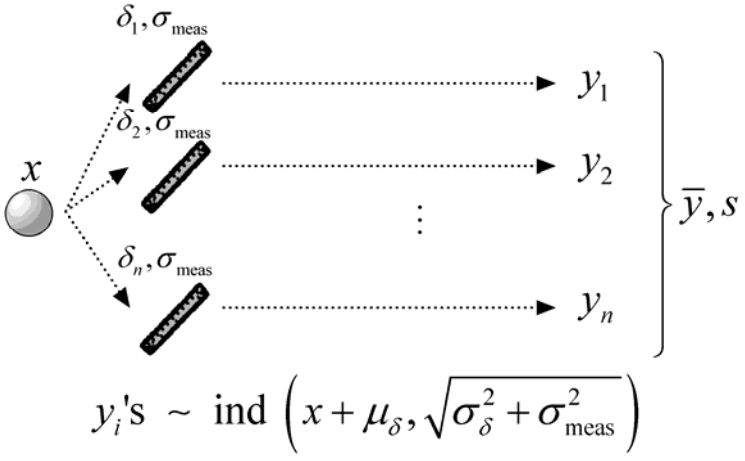


Figure 5: A single sample obtained from a single measurement of a fixed measurand made with each of n different devices from a large population of such devices with a common measurement precision σ_{meas}.

Notice that we may ignore any dependence of bias and precision on x here, but that random choice of measurement device makes the quantities δ and σ_{meas} random. Then

$$\mathrm{E}(y) = x + \mathrm{E}(\delta) = x + \mu_\delta \quad \text{and} \quad \mathrm{Var}(y) = \mathrm{Var}(\delta) + \mathrm{E}\left(\sigma_{\text{meas}}^2\right).$$

It is only when the devices have the same precision (for measuring x) that one gets the simpler expression $\sigma_y = \sqrt{\sigma_\delta^2 + \sigma_{\text{meas}}^2}$ shown in Figure 5.

The complicated nature of $\mathrm{E}(y)$ and $\mathrm{Var}(y)$ in the scenario covered by Figure 5 makes application of formulas (1) and (2) unappealing, but the discussion of this situation is still useful. Producing a "sample" of n observations by measuring the same measurand with n different measurement devices begins to illustrate the problems associated with switching measurement systems during a given empirical study. When the difference between devices is explained completely by the difference between the humans involved in measurement, Figure 5 motivates a discussion of so-called **repeatability** and **reproducibility** sources of variation, measured respectively by σ_{meas} and σ_{δ} in this example.

Together, the scenarios represented by Figures 3, 4, and 5 provide motivation for elementary qualitative discussions of the concept of **components of variance** and the need for data collection plans and calculation procedures that can be used to separate them. A combination of the elements of Figures 3 and 4 together with appropriate calculation that goes beyond the methods in (1) and (2) provide a way to separate the different components of variance, σ_x^2 and σ_{meas}^2. Similarly, some combination of the elements of Figures 3 and 5 together with appropriate calculation allows separation of σ_{δ}^2 and σ_{meas}^2. Further, if measurement systems are switched during a study, repeated measurements should be made for each measurement system in order to separately estimate their σ_{meas}^2's.

As a final example of the interaction of metrology concepts and simple one-sample statistical methods, the "paired data" scenario is presented in Figure 6.

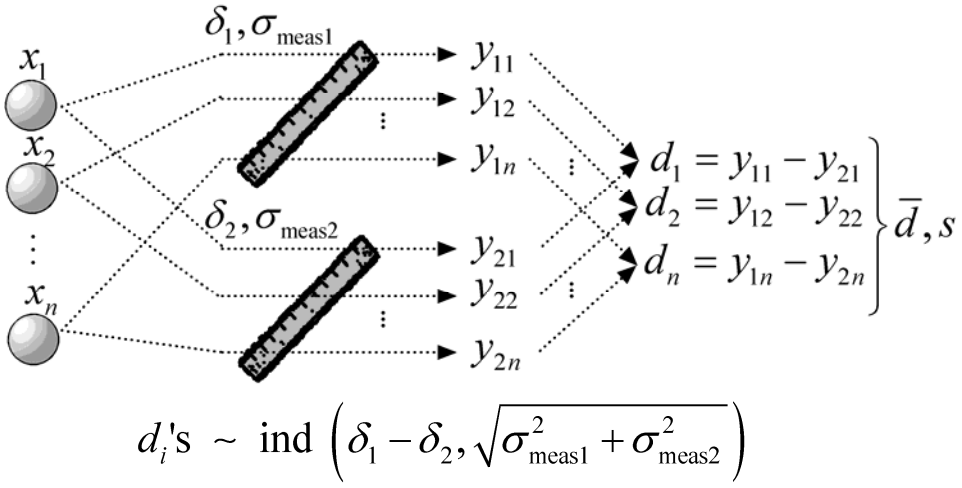
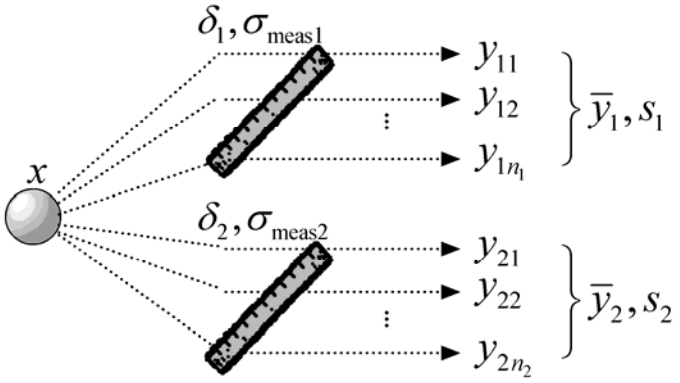


Figure 6: A single sample consisting of differences of measurements on n different measurands made using two linear devices ($\delta_1(x) = \delta_1$, $\delta_2(x) = \delta_2$) with constant measurement precisions ($\sigma_{\text{meas1}}(x) = \sigma_{\text{meas1}}$, $\sigma_{\text{meas2}}(x) = \sigma_{\text{meas2}}$).

Under the non-trivial assumptions of (independence and) device linearity and constant precision of measurement, device biases can be compared by applying (1) to the differences d regardless of the origin of the measurands. The relevant "population mean" is $\delta_1 - \delta_2$. If one of the devices involved provides "gold standard" measurements guaranteed to have no bias, then the bias of the other device can be estimated, and henceforth measurements can be replaced with bias-adjusted ones. This approach assumes that the device is well-adjusted, so the corresponding δ is guaranteed to be 0 relative to the gold standard.

4. Two-Sample Inference and Measurement Error

Figure 7 provides the two-sample version of the scenario in Figure 3.



y_{1i}'s ~ ind $\left(x+\delta_1, \sigma_{\text{meas1}}\right)$ independent of

y_{2i}'s ~ ind $\left(x+\delta_2, \sigma_{\text{meas2}}\right)$

Figure 7: Two samples of n_1 and n_2 measurements of a single measurand with two devices to compare two measurement devices.

The design shown in Figure 7 can be used to compare two devices, by focusing on a particular measurand. The difference in "population means" is $\mu_1 - \mu_2 = \left(x+\delta_1\right) - \left(x+\delta_2\right) = \delta_1 - \delta_2$ and the ratio of "population standard deviations" is $\sigma_{\text{meas1}} / \sigma_{\text{meas2}}$. Consequently, the methods in (3) and (4) produce direct inferences for the quantities $\delta_1 - \delta_2$ and $\sigma_{\text{meas1}} / \sigma_{\text{meas2}}$.

Assuming the devices are linear and have constant measurement precision, inferences made on the basis of a single measurand can be extended to all other measurands in the range of the devices. However, there is at least one important circumstance in which the data collection design represented by Figure 7 cannot be used to compare two measurement devices. That is the case where measurement is **destructive**, so multiple measurements cannot be made on a given item. In this case, Figure 8 shows the only practical kind of data collection design for comparison of two devices.

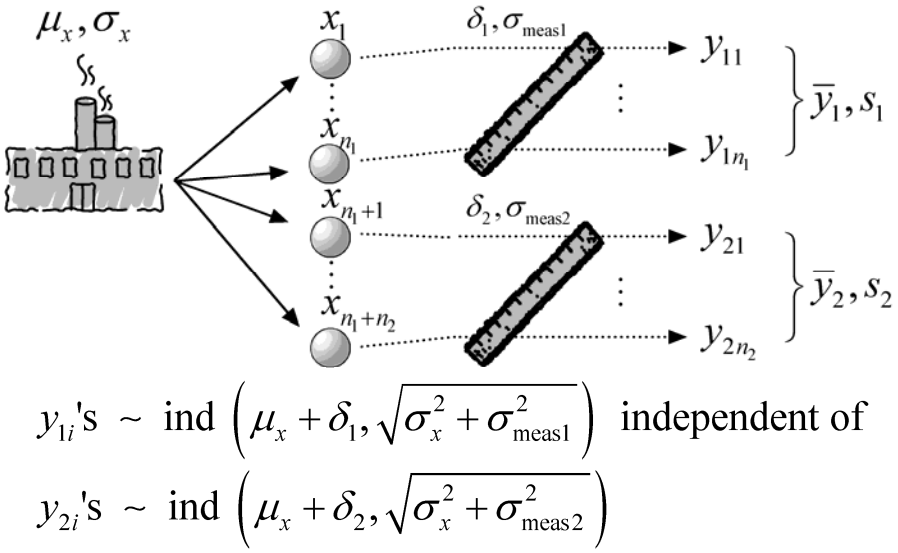
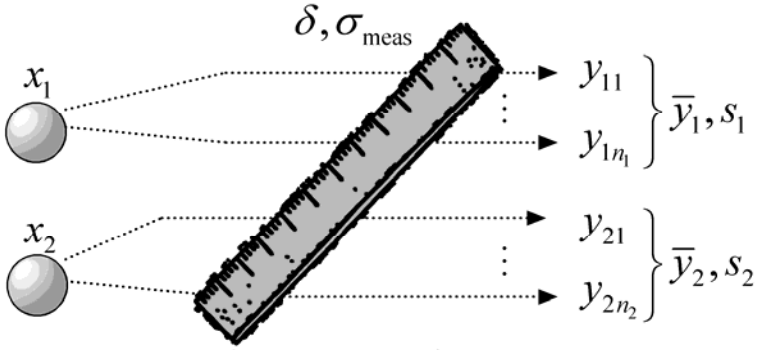


Figure 8: Two samples consisting of a single measurement made on $n_1 + n_2$ measurands from a stable process, n_1 from device 1 and n_2 from device 2 (assuming device linearities ($\delta_1(x) = \delta_1$, $\delta_2(x) = \delta_2$) with constant measurement precisions ($\sigma_{\text{meas1}}(x) = \sigma_{\text{meas1}}, \sigma_{\text{meas2}}(x) = \sigma_{\text{meas2}}$)).

As in the discussion of the scenario of Figure 4, the two "population" means are in general $\mathrm{E}(y_1) = \mathrm{E}(x) + \mathrm{E}(\delta_1(x))$ and $\mathrm{E}(y_2) = \mathrm{E}(x) + \mathrm{E}(\delta_2(x))$, and the corresponding variances, $\mathrm{Var}(y_1) = \mathrm{E}(\sigma^2_{\text{meas1}}(x)) + \mathrm{Var}(x + \delta_1(x))$ and $\mathrm{Var}(y_2) = \mathrm{E}(\sigma^2_{\text{meas2}}(x)) + \mathrm{Var}(x + \delta_2(x))$. These reduce to the simple forms given in Figure 8 for well-behaved measurement devices (i.e., $\delta_1(x) = \delta_1$, $\delta_2(x) = \delta_2$, $\sigma_{\text{meas1}}(x) = \sigma_{\text{meas1}}$, and $\sigma_{\text{meas2}}(x) = \sigma_{\text{meas2}}$). In any case, under such circumstances, the method in (3) provides inferences for comparing device biases. Comparison of the "population" standard deviations specified in Figures 7 and 8 leads to the important conclusion that

a) inferences for $\delta_1 - \delta_2$ based on (3) and a design portrayed in Figure 7 are likely to be more informative than ones based on a design portrayed by Figure 8, and
b) inferences based on (4) provide an indirect way to compare device precisions, but they will be clouded by measurand variation.

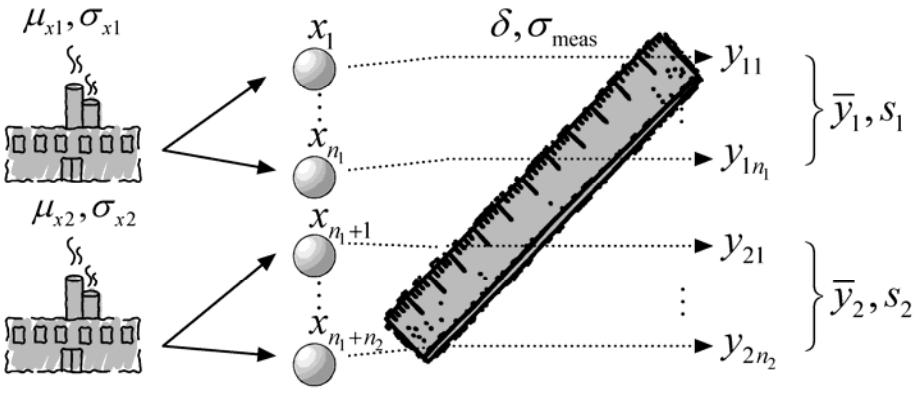
Finally, an additional pair of two-sample scenarios involving a single measurement device is portrayed in Figures 9 and 10.



$$y_{1i}\text{'s} \sim \text{ind}\left(x_1 + \delta, \sigma_{\text{meas}}\right) \text{ independent of}$$

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

Figure 9: Two samples consisting of repeat measurements of two different measurands made with one device (assuming device linearity and constant measurement precision).



$$y_{1i}\text{'s} \sim \text{ind}\left(\mu_{x1} + \delta, \sqrt{\sigma_{x1}^2 + \sigma_{\text{meas}}^2}\right) \text{ independent of}$$

$$y_{2i}\text{'s} \sim \text{ind}\left(\mu_{x2} + \delta, \sqrt{\sigma_{x2}^2 + \sigma_{\text{meas}}^2}\right)$$

Figure 10: Two samples consisting of single measurements made using a single device on multiple measurands produced by two stable processes (assuming device linearity and constant meausurement precision).

In the case portrayed by Figure 9, the difference of "population" means will be $\mu_1 - \mu_2 = \left(x_1 + \delta\left(x_1\right)\right) - \left(x_2 + \delta\left(x_2\right)\right)$. In the case that the device is linear (i.e., $\delta(x) = \delta$), this is simply $x_1 - x_2$ and the confidence limits given in (3) apply to the difference in measurands. (This actually holds even if the device does not have the constant precision indicated by the figure.) There is no practical interest in comparing "population" standard deviations via (4) in this case (unless one allows the possibility that $\sigma_{\text{meas}}\left(x_1\right) \neq \sigma_{\text{meas}}\left(x_2\right)$ and wishes to investigate whether measurement precision varies with measurand).

Figure 10 considers the comparison of process characteristics. If the device is linear (i.e., $\delta(x) = \delta$), (3) allows estimation of $\mu_1 - \mu_2 = \left(\mu_{x1} + \delta\right) - \left(\mu_{x2} + \delta\right) = \mu_{x1} - \mu_{x2}$, the difference in process means. However, even assuming constant measurement precision, a direct comparison of process standard deviations (σ_{x1} and σ_{x2}) is not available. This data collection scheme is probably the most commonly used example of "tickets from two boxes" presented in elementary statistics courses. Failure to make the point that the confidence limits in (4) are essentially for comparing σ_{x1} and σ_{x2} only when measurement noise is completely negligible invites misinterpretation of statistical analysis results.

5. Instruction

A version of the material presented here (including versions of the figures) has become a standard part of a statistical quality control course taught at Iowa State University. The course has a basic engineering statistics course as a prerequisite and goes on to consider more complicated data collection designs and statistical methods associated with both linear calibration and gauge "R&R" studies common in industrial contexts. Anecdotal experience in the quality control course indicates that the approach here is a sound one and offers promise for effective lectures as part of a statistics education. Also, as an

experiment, the one sample material in Section 3 was introduced in an elementary business statistics course at Miami of Ohio University, but the students generally had difficulties with the assigned homework. Consequently, while we would like to see this material be a part of a statistics education at any level, realistically, we expect that it should be a part of a second course in statistics, such as a methods course or a design of experiments course.

An instructor of this material needs to emphasize the importance of thinking about how items are sampled, the number of samples and how they are used, the number of measurement devices and the samples on which they are used, and the number and type of measurements on a given item. It is important to clearly state the goal of the data collection – are we comparing items, populations, measurements, or measurement devices? Because there are several scenarios that can result in the same number of observations, instructors also need to help their students see the differences, e.g., the difference between a single device making multiple measurements and multiple devices making a single measurement.

At Iowa State University, we use simple hands-on in-class measurement exercises involving simple/cheap plastic dial calipers and (hard to size) Styrofoam™ packing peanuts to give students a better understanding of some of the data collection plans illustrated here. Individual students are asked to measure one dimension of several different peanuts and to also remeasure that dimension of a single peanut a number of times. The data that is collected then serves as raw material for homework based on the ideas from this article. Students like this instructional device and find that it makes many of the concepts of metrology concrete. It also gives the instructor a specific case to refer to as lectures progress. Besides the packing peanuts example, the binder clip example introduced in Section 2 can be used with admittedly more expensive micrometers. Finally, another suggestion is to measure masses of items like binder clips with balances that would be readily available in a university's chemistry laboratory. Moreover, smaller standard masses can be purchased relatively inexpensively to assess bias of the balances.

6. Discussion

The particular methods represented by (1) – (4) do have technical limitations (particularly those for standard deviations). Miller (1986) provides some perspective on these limitations. However, our goal here is to convey the interplay between measurement and statistics, not the specific details. We believe the impact of measurement error and application of inference methods for quantifying measurement error should be an explicit part of a statistics education at all levels. While we have concentrated on the most elementary methods, there are more complicated issues that are both fascinating and important, but these issues can only be considered at higher levels.

Acknowledgements

We thank the editor and an associate editor for their helpful comments, which led to improvements in this article. This work was supported in part by NSF grant DMS #0502347 EMSW21-RTG awarded to the Department of Statistics, Iowa State University.

References

Freedman, D., Purves, R. and Pisani, R. (2007). *Statistics*, 4th Edition, New York: W.W. Norton & Co.

Miller, R. (1986). *Beyond ANOVA: Basics of Applied Statistics*, New York: John Wiley and Sons, Inc.

Pollack, A. (2009), "Quest Acknowledges Errors in Vitamin D Tests," *The New York Times*, January 8, 2009, New York Edition, B1.