Module 8

Quantifying how measurement precision affects one's ability to detect differences between measurands

Prof. Stephen B. Vardeman
Statistics and IMSE
Iowa State University

March 4, 2008

Detecting a Difference

The problem of determining whether "there is a difference" is fundamental to engineering and technology. For example

- in process monitoring, engineers need to know whether process parameters (e.g., the mean widget diameter being produced by a particular lathe) are at standard values or have changed.
- in evaluating whether two machines are producing similar output, one needs to assess whether product characteristics from the two machines are the same or are consistently different.
- when using hazardous materials in manufacturing, engineers need to compare chemical analyses for current environmental samples to analyses for "blank" samples, looking for evidence that important quantities of toxic materials have escaped a production process and thus increased their ambient level from some "background" level.

We consider the matter of the adequacy of a gauge or measurement system for the purpose of detecting a change or difference.

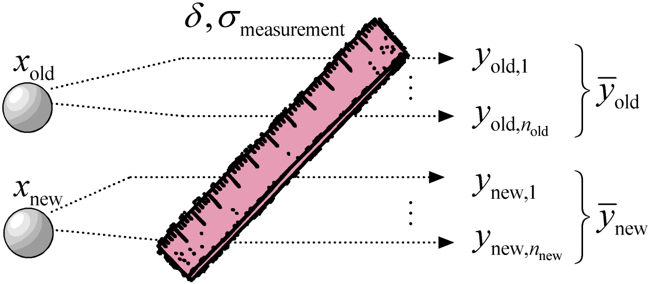
Comparing Two Objects Using a Linear Device

Let $\sigma_{\text{measurement}}$ stand for an appropriate standard deviation for describing the precision of some measurement system. (Depending upon the context this could be σ or $\sigma_{\text{R\&R}}$ from a gauge R&R study.) We will investigate the impact of $\sigma_{\text{measurement}}$ on one's ability to detect change or difference through consideration of the distribution of

$$\bar{y}_{\text{new}} - \bar{y}_{\text{old}} \ ,$$

where $\bar{y}_{\text{new}}$ is the sample mean of n_{new} measurements taken on a particular "new" object and $\bar{y}_{\text{old}}$ is the sample mean of n_{old} measurements taken on a particular "old" object. This is much like one of the scenarios discussed in Module 2B and the next panel illustrates the situation.

Comparing Two Objects Using a Linear Device (cont.)



$y_{old,i}$'s $\sim$ ind $\left(x_{old} + \delta, \sigma_{measurement}\right)$ independent of

$y_{new,i}$'s $\sim$ ind $\left(x_{new} + \delta, \sigma_{measurement}\right)$

The Difference Between Two Sample Mean Measurements (Made With a Linear Device)

Information on the "old" object can be strong enough that n_{old} can be thought of as essentially infinite, and thus $\bar{y}_{\text{old}}$ essentially equal to the mean of old observations, $x_{\text{old}} + \delta$. If $x_{\text{new}} + \delta$ is the mean of the new observations and new and old sample means are independent,

$$\text{E}(\bar{y}_{\text{new}} - \bar{y}_{\text{old}}) = (x_{\text{new}} + \delta) - (x_{\text{old}} + \delta) = x_{\text{new}} - x_{\text{old}} ,$$

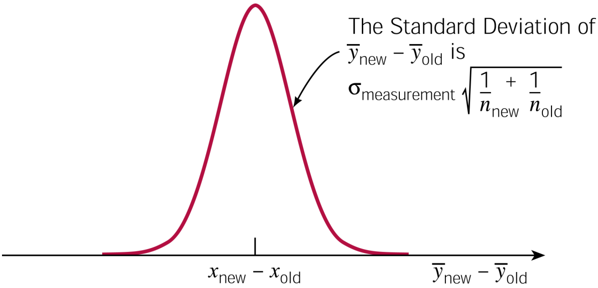
and

$$\text{Var}(\bar{y}_{\text{new}} - \bar{y}_{\text{old}}) = \sigma^2_{\text{measurement}} \left(\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}} \right) .$$

When the information on the old object is strong enough that n_{old} is essentially infinite, this reduces to

$$\text{Var}(\bar{y}_{\text{new}} - \bar{y}_{\text{old}}) = \sigma^2_{\text{measurement}} \left(\frac{1}{n_{\text{new}}} \right) .$$

The Distribution of the Difference in Two Sample Mean Measurements (Made With a Linear Device)



Note that if $x_{\text{new}} = x_{\text{old}}$ the distribution of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ is centered at 0.

Example 8-1

Chemical Analysis for Benzene. A standard deviation characterizing a laboratory's precision of measurement of benzene content of samples is $\sigma_{\text{measurement}} = .03\mu\,\text{g}/\,\text{l}$. To determine whether the amount of benzene in an environmental sample exceeds that in a "blank" sample (supposedly containing only background levels of benzene), the environmental sample will be analyzed $n_{\text{new}} = 1$ time and its measured content compared to the mean from $n_{\text{old}} = 5$ analyses of the blank sample. Then the random variable $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ has

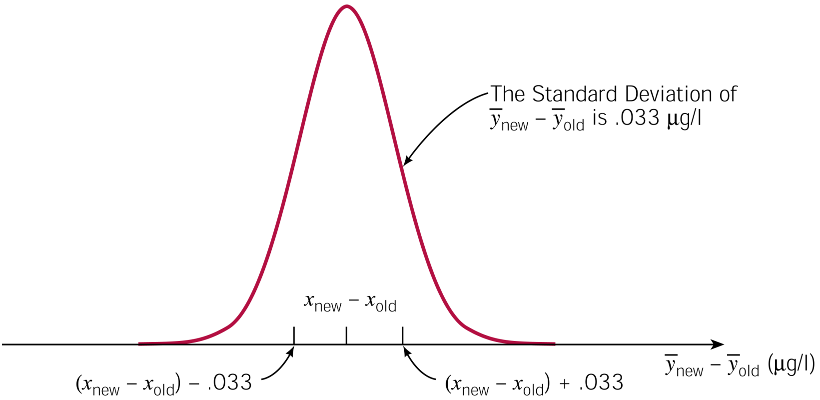
$$\text{E}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}}) = x_{\text{new}} - x_{\text{old}} \ ,$$

and

$$\sqrt{\text{Var}(\overline{y}_{\text{new}} - \overline{y}_{\text{old}})} = \sigma_{\text{measurement}}\sqrt{\frac{1}{1} + \frac{1}{5}} = .03\sqrt{\frac{6}{5}} = .033\mu\,\text{g}/\,\text{l}\,.$$

The next panel shows a distribution for $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$. (If information on the blank sample was essentially perfect, the standard deviation of $\overline{y}_{\text{new}} - \mu_{\text{old}}$ would be $.030\mu\,\text{g}/\text{l}$, and not much smaller than the value here.)

Example 8-1 (cont.)



Using the Distribution of the Difference in Means

The distribution of $\bar{y}_{\text{new}} - \bar{y}_{\text{old}}$ forms the basis for several common ways of evaluating the adequacy of a measurement technique to characterize a change or difference. One simple rule of thumb often employed by analytical chemists is that a difference $x_{\text{new}} - x_{\text{old}}$ must be on the order of 10 times the standard deviation of $\bar{y}_{\text{new}} - \bar{y}_{\text{old}}$ before it can be adequately characterized by a measurement process. For example, in the benzene analyses of Example 8-1 with sample sizes $n_{\text{new}} = 1$ and $n_{\text{old}} = 5$, this rule of thumb says that only increases in real benzene content of at least

$$10 \times .033 = .33\mu\,\mathrm{g}/\,\mathrm{l}$$

can be reliably characterized. This somewhat ad hoc guideline is a requirement that one's "signal-to-noise ratio" (ratio of mean to standard deviation) for determination of a difference be at least 10 before being comfortable with the resulting precision.

Using the Distribution of the Difference in Means to Set a Critical Limit

A second approach to using the distribution of $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ to describe one's ability to detect a difference between measurands involves some ideas from hypothesis testing. In interpreting an observed $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$, one might require that it be of a certain minimum magnitude before declaring that there is a clear difference between new and old objects. Suppose for the rest of this discussion that one is concerned about detecting an increase in response, i.e. the possibility that $x_{\text{new}} - x_{\text{old}} > 0$. It then makes sense to set some **critical limit, $\mathbf{L_c}$**, and to only declare that there is a difference if

$$\overline{y}_{\text{new}} - \overline{y}_{\text{old}} > L_{\text{c}} \ .$$

If one wishes to limit the probability of a "false positive" (i.e., a type I error) L_{c} should be large enough that the eventuality above occurs rarely when in fact $x_{\text{new}} = x_{\text{old}}$.

Using the Distribution of the Difference in Means to Set a Critical Limit

If $\bar{y}_{\text{new}} - \bar{y}_{\text{old}}$ is normally distributed, it is possible to use the fact that when $x_{\text{new}} = x_{\text{old}}$ the variable

$$\frac{\bar{y}_{\text{new}} - \bar{y}_{\text{old}}}{\sigma_{\text{measurement}}\sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}}$$

is standard normal to set a value for L_c. One may pick z_1 so that for standard normal Z, $P[Z > z_1] = \alpha$, for α a small number (of one's choosing). Then setting

$$L_c = z_1 \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}$$

the false alarm rate is no more than α.

Once one has established a critical value L_c (as above or otherwise) it is reasonable to ask what is the probability of detecting change of a given size. Again, assuming that $\overline{y}_{\text{new}} - \overline{y}_{\text{old}}$ is normally distributed it is possible to answer this question. That is, with

$$z_2 = \frac{L_c - (x_{\text{new}} - x_{\text{old}})}{\sigma_{\text{measurement}}\sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}} ,$$

the probability (depending upon $x_{\text{new}} - x_{\text{old}}$) of declaring that there has been a change (that there is a difference) is

$$\gamma = P[Z > z_2] ,$$

(for Z again standard normal).

Rather than computing for a given $x_{\text{new}} - x_{\text{old}}$ the probability of detecting a difference of that size, one can ask what difference in measurands would be required to produce a given (large) probability of detection. In analytical chemistry, such a value is given a special name. For a given standard deviation of measurement $\sigma_{\text{measurement}}$, sample sizes n_{new} and n_{old}, critical value L_{c} and desired (large) probability γ, the **lower limit of detection, $\mathbf{L}_d$,** of a measurement protocol is the smallest difference in measurands $x_{\text{new}} - x_{\text{old}}$ producing

$$P[\bar{y}_{\text{new}} - \bar{y}_{\text{old}} > L_{\text{c}}] \geq \gamma.$$

Where $\bar{y}_{\text{new}} - \bar{y}_{\text{old}}$ is normal (and z_2 chosen so that $\gamma = P[Z > z_2]$ is large),

$$L_{\text{d}} = L_{\text{c}} - z_2 \sigma_{\text{measurement}} \sqrt{\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}}}\ .$$

(z_2 above is typically negative so that L_{d} is typically larger than L_{c}.)

Example 8-1 (cont.)

Consider again the benzene analysis. Suppose that it is desirable to limit the probability of producing a "false positive" (a declaration that the environmental sample contains more benzene than the blank sample when in fact there is no real difference in the two) to no more than $\alpha = .10$. For Z standard normal, $P[Z > 1.282] = .10$. So an appropriate critical value is

$$L_c = 1.282(.030)\sqrt{\frac{1}{1} + \frac{1}{5}} = .042.$$

Suppose then one wants to evaluate the probability of detecting a difference in real benzene content of size $x_{\text{new}} - x_{\text{old}} = .02\mu\,\text{g}/\,\text{l}$ using this critical value. The above discussion shows that with

$$z_2 = \frac{.042 - .02}{.030\sqrt{\frac{1}{1} + \frac{1}{5}}} = .67,$$

the probability is only about

$$P[Z > z_2] = P[Z > .67] = .2514.$$

Example 8-1 (cont.)

There is a substantial (75%) chance of failing to identify a $.02\mu\,\mathrm{g}/\,\mathrm{l}$ difference in benzene content beyond that resident in the blank sample. This unpleasant fact motivates the question "How big does the increase in benzene concentration need to be in order to have a large (say 95%) chance of seeing it above the measurement noise?" Since

$$P[Z > -1.645] = .95 \quad ,$$

for $\gamma = .95$

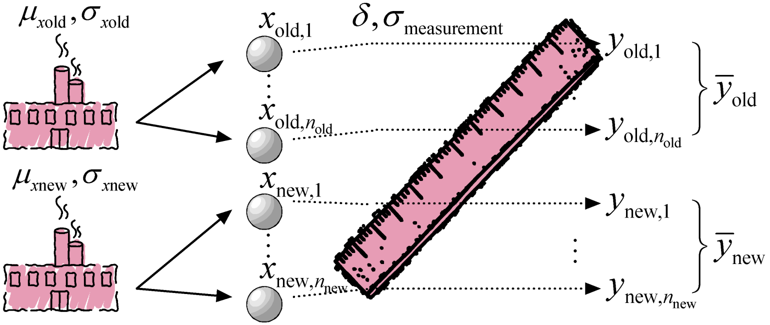
$$L_{\mathrm{d}} = .042 - (-1.645)(.030)\sqrt{\frac{1}{1} + \frac{1}{5}} = .096\mu\,\mathrm{g}/\,\mathrm{l}$$

is the lower limit of detection. A real difference in benzene content must be of at least this size for there to be a large (95%) chance of "seeing" it through the measurement noise.

Comparing Two Conditions

It is an extremely important distinction that the discussion here has been phrased in terms of detecting a difference between two particular objects and *not* between processes or populations standing behind those objects. Example 8-1 concerns comparison of a particular environmental sample and a particular blank sample. It does *not* directly address the issue of how the population of environmental samples from a site of interest compares to a population of blanks. In a manufacturing context, comparisons based on the foregoing material would concern two particular measured parts, not the process conditions operative when those parts were made. Only measurement variation has been taken into account, and not object-to-object variation for processes or populations the measured objects might represent.

The problem of comparing two conditions has today been pictured as:



$$y_{\text{old},i}\text{'s} \sim \text{ind}\left(\mu_{x\text{old}} + \delta, \sqrt{\sigma^2_{x\text{old}} + \sigma^2_{\text{measurement}}}\right) \text{ independent of}$$

$$y_{\text{new},i}\text{'s} \sim \text{ind}\left(\mu_{x\text{new}} + \delta, \sqrt{\sigma^2_{x\text{new}} + \sigma^2_{\text{measurement}}}\right)$$

Comparing Two Conditions (cont.)

In this second context,

$$\mathrm{E}(\bar{y}_{\text{new}} - \bar{y}_{\text{old}}) = (\mu_{x\text{new}} + \delta) - (\mu_{x\text{old}} + \delta) = \mu_{x\text{new}} - \mu_{x\text{old}} ,$$

and

$$\begin{aligned} \mathrm{Var}(\bar{y}_{\text{new}} - \bar{y}_{\text{old}}) &= \frac{\sigma^2_{x\text{new}} + \sigma^2_{\text{measurement}}}{n_{\text{new}}} + \frac{\sigma^2_{x\text{old}} + \sigma^2_{\text{measurement}}}{n_{\text{old}}} \\ &= \sigma^2_{\text{measurement}} \left(\frac{1}{n_{\text{new}}} + \frac{1}{n_{\text{old}}} \right) + \frac{\sigma^2_{x\text{new}}}{n_{\text{new}}} + \frac{\sigma^2_{x\text{old}}}{n_{\text{old}}} \end{aligned}$$

and (of course) item-to-item/measurand-to-measurand variability is present in $\bar{y}_{\text{new}} - \bar{y}_{\text{old}}$.

Comparing Two Conditions (cont.)

In the event that the "old" and "new" processes have comparable values of σ_x, the formulas and language used this module can be reinterpreted to allow application to the problem of detecting changes in a process or population mean, by replacing $\sigma_{\text{measurement}}$ with

$$\sqrt{\sigma_x^2 + \sigma_{\text{measurement}}^2}$$

Otherwise, a separate development of formulas is required. For example, if in the context of Example 8-1, blank samples are more homogeneous than are field samples from a particular site, an analysis parallel to that here but based on the two different values for σ_x will be needed for application to the problem of comparing a site mean benzene level to a blank mean level.