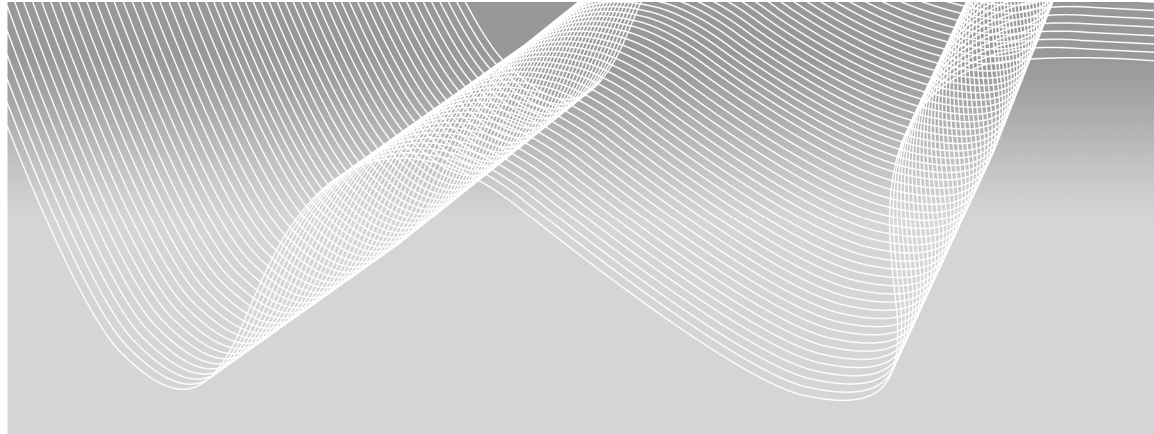
Part 1

Characterizing the Measurement Process

Chapter 1 **Basic Concepts of Measurement Capability** 3

Chapter 2 **Estimating Repeatability, Bias, and Linearity** 31

Chapter 3 **Estimating Reproducibility and Total Measurement Error** 109



Chapter 1

Basic Concepts of Measurement Capability

Introduction 4
Figure-of-Merit Statistics 5
- The P/T Ratio 5
- The SNR 6
- RR Percent 6
- Six Sigma Quality of Measurement 6
- Capability Potential versus P/T and SNR 7

Precision versus Accuracy in a Measurement Tool or Process 11
To Calibrate or Not 16
- Understanding Risk 16
- JMP Sample Size Calculations and Power Curves 18

Uncertainty in Estimating Means and Standard Deviations 21
- Confidence Interval for the Mean 22
- Confidence Interval for the Standard Deviation 24

Components of Measurement Error 28
Repeatability Error 28
Reproducibility Error 29
Linearity of Measurement Tools 29
Random versus Fixed Effects 30

Introduction

Standing between every manufacturing process and the observer is some form of measurement process. The observer never sees the actual manufacturing process clearly or separately—a measurement process or device always intervenes to provide the data necessary for interpreting process performance. Ideally the impact of the variation in the measurement device or process is negligible, but that might not be the case. A manufacturing process could be satisfying all Six Sigma requirements, but a measurement tool could be obscuring that truth. Equation 1.1 illustrates the relationship among perceived process variance, actual process variance, and the measurement tool contributions to the variance.

$$\hat{\sigma}^2_{perceived\ process} = \sigma^2_{actual\ process} + \sigma^2_{measurement\ tool} \qquad 1.1$$

Figure 1.1 provides a graphic illustration of this equation. Obviously, if the contribution due to variation in the measurement tool is extremely small, then the perceived process variation is approximately the same as the actual process variation.

Therefore, establishing a figure-of-merit statistic that identifies sources of variation for a measurement tool or process helps the observer decide whether the contribution of measurement error materially affects any interpretation of the manufacturing process.

Figure 1.1 Contribution of Measurement Error to Perceived Process Error

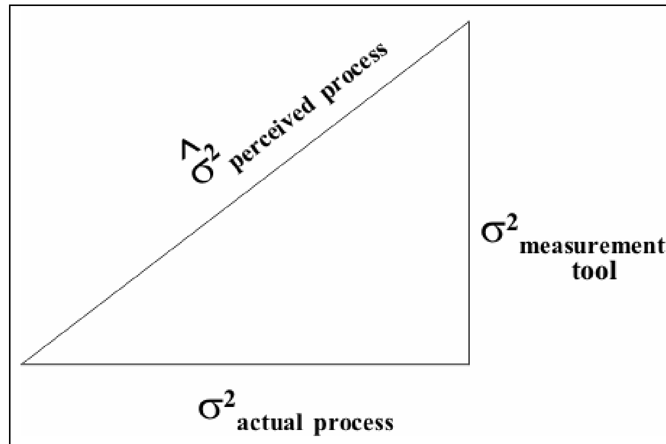


Figure-of-Merit Statistics

The P/T Ratio

The precision/tolerance ratio (P/T) is one convenient figure of merit used throughout this book for describing the capability of a measurement device. Equation 1.2 illustrates its calculation based on the inverse of the capability potential (Cp), where USL and LSL are the upper and lower specification limits, respectively, for the process in question—not the measurement process.

$$P/T = 100\left(\frac{6\sigma_{measurement}}{USL - LSL}\right) \qquad 1.2$$

If the P/T ratio for a measurement tool is ≤ 30 (see "Capability Potential versus P/T and SNR" in this chapter), then its contribution to the perceived process variation is negligible.

The SNR

In many cases, the specification limits for a process are arbitrary assignments, subject to change. An alternative figure of merit for a measurement tool is the signal-to-noise ratio (SNR). Equation 1.3 illustrates its calculation.

$$SNR = \frac{\hat{\sigma}_{process}}{\hat{\sigma}_{measurement}} \qquad 1.3$$

Because the perceived variation for the process probably contains a contribution from measurement error, a more elaborate calculation of SNR is applicable (Equation 1.4).

$$SNR = \frac{\sqrt{\hat{\sigma}^2_{process} - \hat{\sigma}^2_{measurement}}}{\hat{\sigma}_{measurement}} \qquad 1.4$$

RR Percent

When a measurement study involves only two factors, such as operator and part, JMP can prepare a report on the study that includes RR percent. This statistic compares the measurement variation to the total variation in the data and calculates a percent Gage R&R. Barrentine (1991) suggests guidelines for acceptable RR percent. This statistic is approximately 1/SNR (discussed in the previous section). See the *JMP Statistics and Graphics Guide* for more information. Such a simple metrology study occurs very seldom in the examples used here, so this book contains little additional discussion of this topic.

Six Sigma Quality of Measurement

Another approach to assessing the capability of a measurement process uses the capability analysis option associated with the display of distributions in JMP. A pop-up menu option on the distribution report generates a capability analysis of data collected from a measurement process (assuming the user has removed from that data all unusual values that are due to an assignable cause).

Part of that capability report includes the parts per million (ppm) of observations beyond the specification limits (in this case one uses the specification limits for the process being monitored, not the specification limits associated with the measurement tool). This approach is analogous to the calculation of the P/T ratio illustrated earlier. Also reported is Sigma Quality (provided the user has specified specification limits). Values of Sigma

Quality ≥ 6 indicate that the measurement tool is capable of handling the measurement task being examined. A section in Chapter 2 illustrates this approach.

Capability Potential versus P/T and SNR

To refresh the reader's memory, Equation 1.5 illustrates the calculation of capability potential (Cp).

$$Cp = \frac{USL - LSL}{6\sigma} \quad (1.5)$$

where *USL* and *LSL*, respectively, are the upper and lower specification limits for a process, and σ is the observed process standard deviation.

This statistic, discussed further in Chapter 6, estimates how much of the output of a process fits between defined specification limits. A Cp value of 1 predicts that some 99.73% of observations will fit within the specifications (see Chapter 6 for calculations supporting this statement). The JMP data tables referenced in the discussions in this and in following sections are available on the companion Web site for this book at http://support.sas.com/reece. Sample data tables are arranged by chapters in the discussion. To open a table using the JMP Starter window, select the **File** category, and then click **Open Data Table** (left panel in Figure 1.2). Alternatively, select **File→Open** in the menu bar at the top of the JMP window (right panel in Figure 1.2). Select the **Chapter 1** directory, then select the data table you want and open it (Figure 1.3).

Figure 1.2 Opening a Data Table in JMP

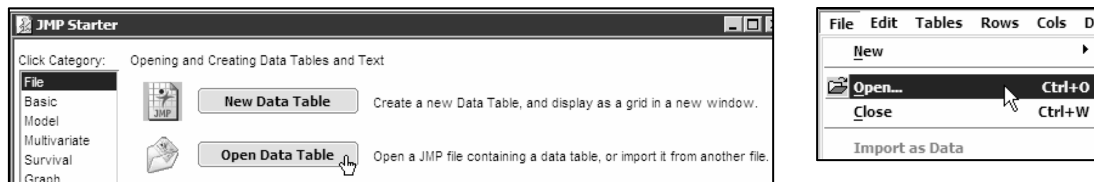


Figure 1.3 Selecting a Data Table

Open Data File

Look in: Chapter 1

Capability vs Measurment Error
Measurement Tools
MN, STDEV CI
Sample Size vs Risk and Delta

Figure 1.4 contains the table Capability vs Measurement Error.jmp. Entries in this table are simulations of the contributions of measurement error to an observed theoretical capability potential (Cp). In the left panel of the display, the symbol to the left of the name of a column indicates that the contents of that column have numeric or continuous modeling properties. The symbol to the right of the last four column names indicates that each column has an associated formula to compute its contents. To see that formula, right-click that symbol to reveal a menu (left panel of Figure 1.5). On that menu, select **Formula** to reveal the embedded calculation (right panel of Figure 1.5).

Figure 1.4 Capability vs Measurement Error.jmp

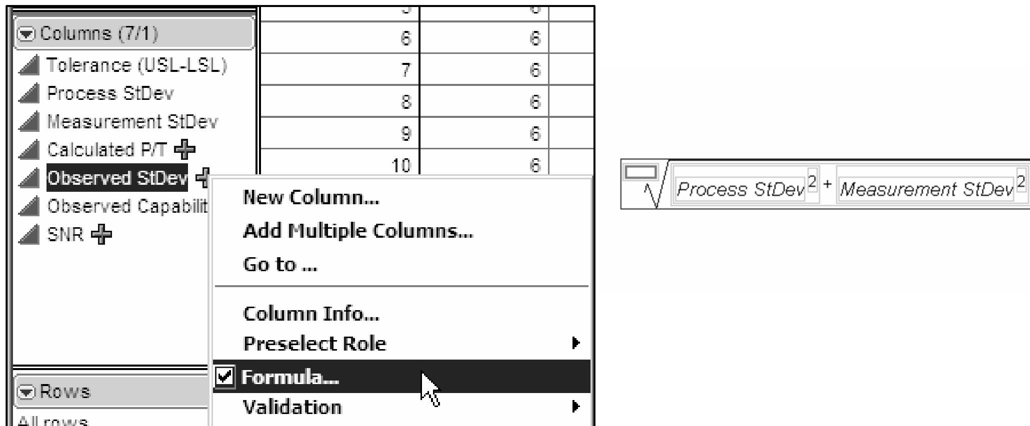
Capability vs Measurment Error

Columns (7/0): Tolerance (USL-LSL), Process StDev, Measurement StDev, Calculated P/T, Observed StDev, Observed Capability, SNR

Rows: All rows 10

	Tolerance (USL-LSL)	Process StDev	Measurement StDev	Calculated P/T	Observed StDev	Observed Capability	SNR
1	6	1	0.01	1	1.00005	1.000	100.00
2	6	1	0.05	5	1.00124922	0.999	20.00
3	6	1	0.1	10	1.00498756	0.995	10.00
4	6	1	0.15	15	1.01118742	0.989	6.67
5	6	1	0.2	20	1.0198039	0.981	5.00
6	6	1	0.25	25	1.03077641	0.970	4.00
7	6	1	0.3	30	1.04403065	0.958	3.33
8	6	1	0.5	50	1.11803399	0.894	2.00
9	6	1	0.75	75	1.25	0.800	1.33
10	6	1	1	100	1.41421356	0.707	1.00

Figure 1.5 Revealing a Column Formula



To explore how the observed capability potential (Cp) of a process varies with the measurement error encountered, generate overlay plots of observed capability (y) versus either the P/T ratio (x) or the SNR. Figure 1.6 illustrates setting up the graph for capability versus SNR using the menu bar at the top of the JMP window. Notice that the X axis has been converted to logarithmic scale during the setup.

Figure 1.6 Setting Up an Overlay Plot for Observed Capability versus SNR

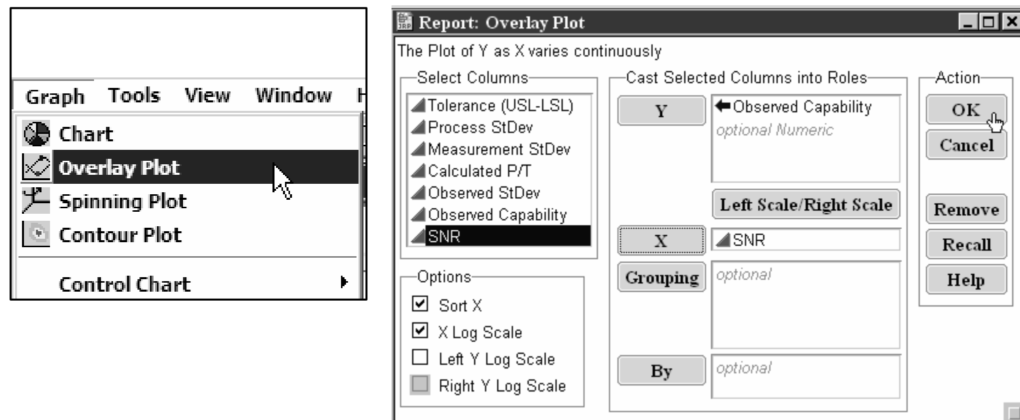


Figure 1.7 displays the overlay plot for observed capability versus P/T ratio; Figure 1.8 displays the overlay plot for observed capability versus SNR. Each graph received considerable modification to improve the displays.

Figure 1.7 Observed Capability versus P/T Ratio

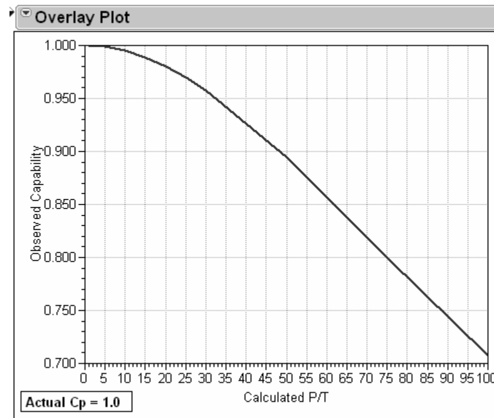
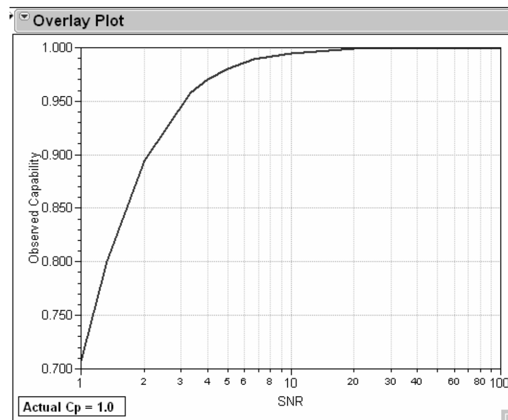


Figure 1.7 suggests that so long as P/T < 30, the measurement process does not materially affect the perception of process performance. Similarly, in Figure 1.8, as long as SNR is >~3, the same is true. Obviously, having a measurement of P/T << 30 or SNR > 5 is highly desirable.

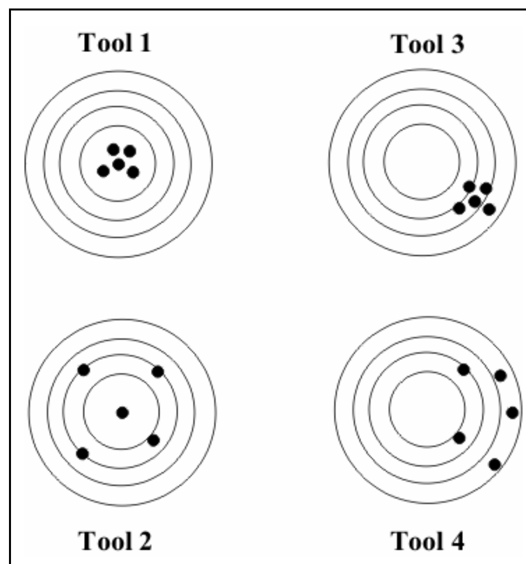
Figure 1.8 Observed Capability versus SNR



Precision versus Accuracy in a Measurement Tool or Process

To minimize the confusion that might exist between the terms “accuracy” and “precision,” consider the following definitions. *Precision,* as applied to a measurement tool or process, is a measure of the total amount of variation in that tool or process. Intuitively, the more precise a measurement tool or process, the more desirable it becomes. *Accuracy* is the difference between a standard or true value and the average of several repeated measurements using a particular tool or process. A measurement tool or process can be precise and accurate, or it can have any combination of those factors, including neither of them. Figure 1.9 illustrates these concepts using a marksmanship model.

Figure 1.9 Precision versus Accuracy (Marksmanship Model)



The target labeled “Tool 1” represents a measurement process that is both precise and accurate. That labeled “Tool 2” represents a measurement process that is not particularly precise, but does give accurate results on average—on average “we got ‘em.” The process or tool suffers from large variation, but no bias.

The image for “Tool 3” shows considerable precision, but the aim is off; the tool has considerable bias in its measurements and is not accurate. “Tool 4” is a disaster. Not only are the results not accurate, but they lack precision as well.

The JMP data table Measurement Tools.jmp, found in the directory **Chapter 1** on the companion Web site for this book at http://support.sas.com/reece and illustrated in Figure 1.10, contains simulated data from the four measurement tools just described. This table makes extensive use of the formula capabilities in JMP to simulate data.

Figure 1.10 Excerpt of Measurement Tools.jmp

Measurement Tools

Measurement Tools

Columns (5/0)
x
Tool 1
Tool 2
Tool 3
Tool 4

Rows
All rows 100
Selected 0
Excluded 0
Hidden 0
Labelled 0

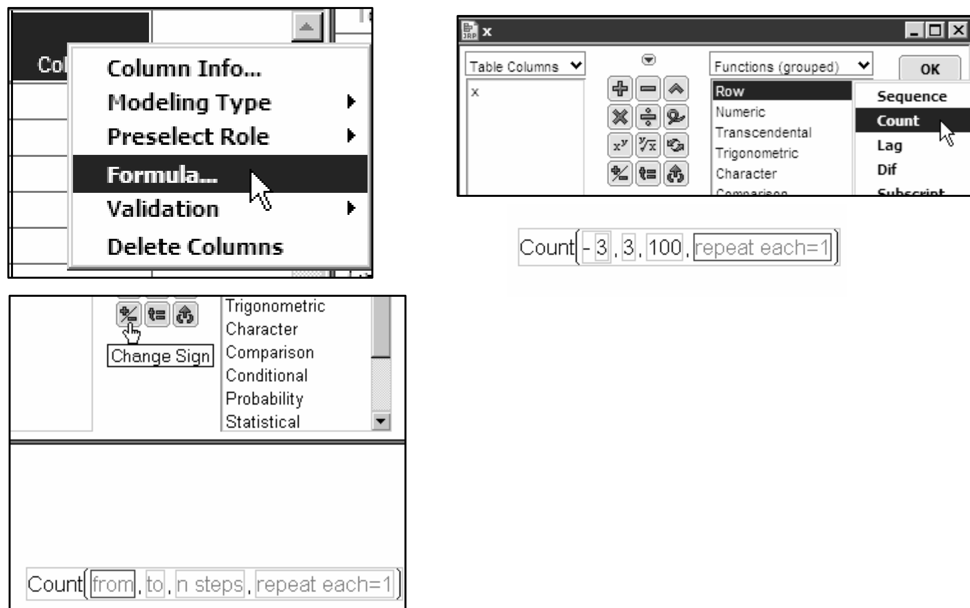
	x	Tool 1	Tool 2	Tool 3	Tool 4
1	-3	8.5855e-32	1.21518e-8	9.11e-171	2.1932e-43
2	-2.9393939	1.529e-30	2.49632e-8	7.847e-168	1.1882e-42
3	-2.8787879	2.5676e-29	5.05334e-8	6.373e-165	6.3429e-42
4	-2.8181818	4.0655e-28	1.00804e-7	4.881e-162	3.3367e-41
5	-2.7575758	6.07e-27	1.9815e-7	3.525e-159	1.7297e-40
6	-2.6969697	8.5455e-26	3.83824e-7	2.4e-156	8.8358e-40
7	-2.6363636	1.1344e-24	7.32635e-7	1.541e-153	4.4477e-39
8	-2.5757576	1.4199e-23	1.37804e-6	9.328e-151	2.2062e-38
9	-2.5151515	1.6759e-22	2.55421e-6	5.325e-148	1.0784e-37
10	-2.4545455	1.8651e-21	4.66521e-6	2.866e-145	5.1942e-37
11	-2.3939394	1.9572e-20	8.39661e-6	1.455e-142	2.4654e-36
12	-2.3333333	1.9366e-19	1.48921e-5	6.961e-140	1.1531e-35
13	-2.2727273	1.8068e-18	2.60272e-5	3.141e-137	5.3147e-35
14	-2.2121212	1.5896e-17	4.48247e-5	1.337e-134	2.4138e-34
15	-2.1515152	1.3186e-16	7.60723e-5	5.363e-132	1.0803e-33
16	-2.0909091	1.0314e-15	0.00012722	2.029e-129	4.7644e-33
17	-2.030303	7.6072e-15	0.00020965	7.237e-127	2.0706e-32
18	-1.969697	5.2904e-14	0.00034046	2.434e-124	8.8673e-32

To create a table like this:

1. Generate a new, blank table in JMP.
2. To this table add 100 rows using any of several approaches. The approach in this example was to use the **Add Rows** option under the **Rows** menu on the menu bar and change the default number of rows from 20 to 100.
3. Change the title of the first column to **x** by clicking in the column heading and typing the new heading.
4. Figure 1.11 illustrates adding a formula for that column. Because the table currently contains only one column, all actions affect that column. To access the options in the upper left panel, right-click the column and select **Formula**. In the window that

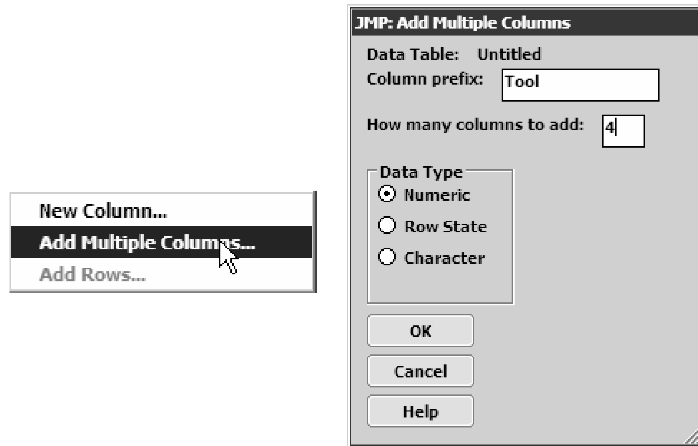
appears (upper right panel) select **Row→Count**. To start filling in the formula, select **Change Sign** in the formula editor window; enter **3** in the first highlighted box. Follow that entry with another **3** and **100** to complete the formula. Clicking **OK** in the editor window completes the process and writes the results to the table in the first column.

Figure 1.11 Creating a Count of Entries in the First Column of Measurement Tools.jmp



5. The next steps add additional columns to the table and label them as shown in Figure 1.10. The author chose to right-click in the empty space to the right of the first column to produce the options in the left panel of Figure 1.12. Selecting **Add Multiple Columns** brings up the window shown in the right panel of Figure 1.12. Completing this window as shown adds four new columns to the original table.

Figure 1.12 Adding Multiple Columns to Measurement Tools.jmp



6. Creating a formula for each of the four Tool columns using the **Normal Density** function (left panel of Figure 1.13) provides the entries in those columns. By default this function does not provide for entering the desired mean and standard deviation of a normal density function. To add that capability, press the comma key twice to bring up the right panel in Figure 1.13. Select the **x** column to fill in the first box, and then enter the appropriate mean and standard deviation for the column **Tool 1**. The means and standard deviations for each of the columns **Tool 1** through **Tool 4** are as follows:

Column	Mean	Standard Deviation
Tool 1	0	0.25
Tool 2	0	0.5
Tool 3	4	0.25
Tool 4	4	0.50

Figure 1.13 Generating Distribution Data for Measurement Tools.jmp

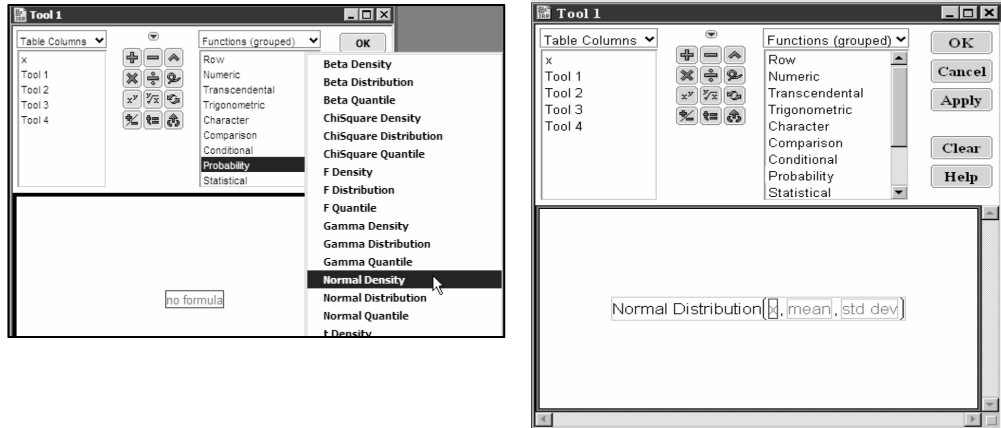
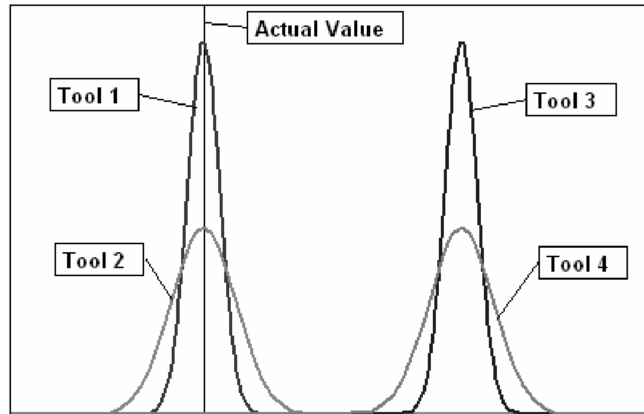


Figure 1.14 shows an overlay plot for the simulated distribution data in each column versus the values of **x**. This figure has undergone considerable modification to improve its appearance, including adding a reference line to the X axis as well as annotating the curves to identify the columns that produced them.

Figure 1.14 Precision versus Accuracy in Measurement Tools



To Calibrate or Not

When you have a measurement device, it is important to make sure it is actually measuring the object correctly. A calibration study is to determine how well a device carries out the measurement process.

In some situations, traceable standards might be available to calibrate a measurement process. The National Institute of Science and Technology (NIST), formerly the National Bureau of Standards (NBS), certifies a variety of standards. Or a measurement tool might have some calibration procedure embedded in its software. The problem is how to decide whether or not to calibrate an instrument at some point in its use. Obviously, uncalibrated instruments could indicate that a manufacturing process is producing material off target when it really is not.

Understanding Risk

Any decision made regarding calibrating an instrument contains risk. Figure 1.15 shows a truth table to help define the situations one might encounter. Although the average engineer or researcher probably does not realize it, making a decision about whether or not to calibrate a measurement tool involves generating two hypotheses—one the exact opposite of the other.

Figure 1.15 Truth Table

		TRUTH	
		Does not require calibration	Requires calibration
DECISION	Do not calibrate	*CORRECT DECISION*	β ERROR
	Calibrate	α ERROR	*CORRECT DECISION*

The *null hypothesis*, usually abbreviated H_0, states that "the measurement tool does not require calibration." Its direct opposite, designated the *alternate hypothesis* or *alternative hypothesis*, states that "the measurement tool requires calibration." An investigator gathers data and analyzes it to determine whether or not the facts support the alternate hypothesis. The result of that analysis dictates whether he or she rejects or fails to reject the null hypothesis.

Here are some trivial mnemonics (not necessarily statistically rigorous) that help keep the α and β risk straight:

- ART: **α**lpha – **R**eject the null hypothesis when it is **T**rue.
- BAF: **β**eta – **A**ccept the null hypothesis when it is **F**alse.

The *power* of a test is $1 - \beta$.

Another trivial method for helping to understand the types of risk is to liken α risk to seeing a ghost. Alternatively, β risk is akin to stepping off a curb in front of an oncoming truck. From Figure 1.15, if an engineer decides to calibrate a measurement tool when it does not require it (rejecting H_0 when it is actually true), he or she commits an α error. On the other hand, deciding not to calibrate a measurement tool when it actually requires it produces a β error.

Obviously, prudent investigators want to keep both risks small in their work, although some level of risk is always present. Calibrating a measurement tool unnecessarily might or might not be a serious problem, depending on the complexity and cost of the operation. If the combination of complexity and cost is large, then proceeding cautiously is good advice. If complexity and cost are trivial, then unnecessary calibrations will not necessarily produce a serious problem.

The nature of the problem is that the observer must compare a computed average to a standard value within the bounds of α and β risks and in the presence of some variation in observations. That is, the sample size required to detect a difference between an observed average and a standard value is directly proportional to the α and β risks allowed and to the ratio of the inherent variation in the observations to the size of the difference to detect. Diamond (1989) provides an expression for estimating the sample size for this problem shown in Equation 1.6.

$$N = \left(z_{\alpha/2} + z_{\beta} \right)^2 \left(\frac{\sigma^2}{\delta^2} \right) \qquad 1.6$$

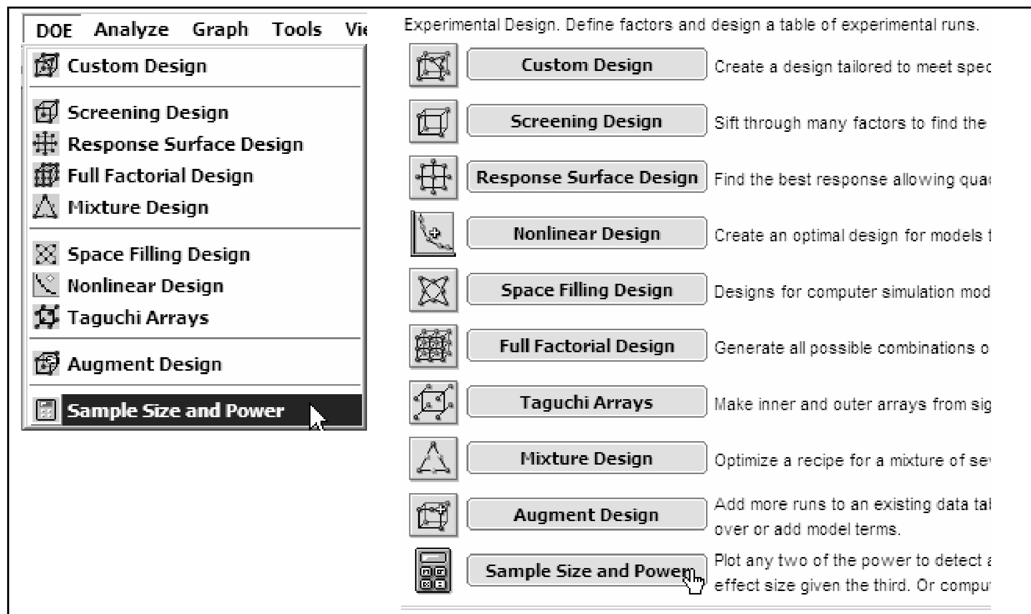
where $z_{\alpha/2}$ and z_{β} are values from the unit normal distribution corresponding to the risks accepted, σ is the variation in the data, and δ is the chosen difference to detect.

Rarely, if ever, will an investigator know precisely what the variation in the data will be before conducting any experiments; that value must be estimated from the experimental data after running the experiments. A very useful approach is to decide what difference relative to the inherent variation in the data is acceptable and solve the equation from that perspective.

JMP Sample Size Calculations and Power Curves

JMP contains a very useful utility to help an experimenter make reasonable decisions about sample size requirements for a variety of scenarios. The situation described here is actually the simplest among many. As shown in Figure 1.16, select **Sample Size and Power** on the **DOE** menu. Then select the first option, **One Sample Mean,** on the window that appears, because the problem under consideration is to compare an observed average of several observations to a target value.

Figure 1.16 Accessing Sample Size Calculations



Selecting **One Sample Mean** displays the window shown in Figure 1.17. The system sets a default α risk at 0.05, but the user can specify any value by editing the table. To use this system most effectively, set **Error Std Dev** to 1, and then specify the **Difference to detect** as some fraction or multiple of the error. As the window indicates, supplying two values calculates the third, whereas entering only one value (for example, a fraction or multiple of the unknown error) produces a plot of the other two.

Figure 1.17 Opening Window of Sample Size Calculation

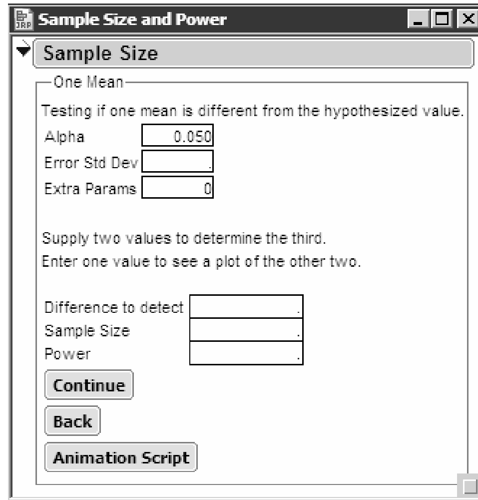


Figure 1.18 shows Sample Size vs Risk and Delta.jmp (also found in the **Chapter 1** directory on the companion Web site for this book at http://support.sas.com/reece) created by exercising the option of filling in **Difference to detect** and **Power** for a number of scenarios, given a value of 1 for **Error Std Dev**. This approach to determining sample size can be very useful in exploring various levels of risk and differences to detect what effect each has on the number of samples required to satisfy the conditions.

Figure 1.18 Sample Size vs Risk and Delta.jmp

Sample Size vs Risk and Delta

	Alpha	Error Std Dev	Difference	Power (1 - Beta)	Sample Size
1	0.05	1	1	0.95	15
2	0.05	1	0.5	0.95	54
3	0.05	1	0.1	0.95	1301
4	0.1	1	1	0.95	12
5	0.1	1	0.5	0.95	45
6	0.1	1	0.1	0.95	1084
7	0.05	1	1	0.9	13
8	0.05	1	0.5	0.9	44
9	0.05	1	0.1	0.9	1053
10	0.1	1	1	0.9	10
11	0.1	1	0.5	0.9	36
12	0.1	1	0.1	0.9	858

Sample Size vs Risk a

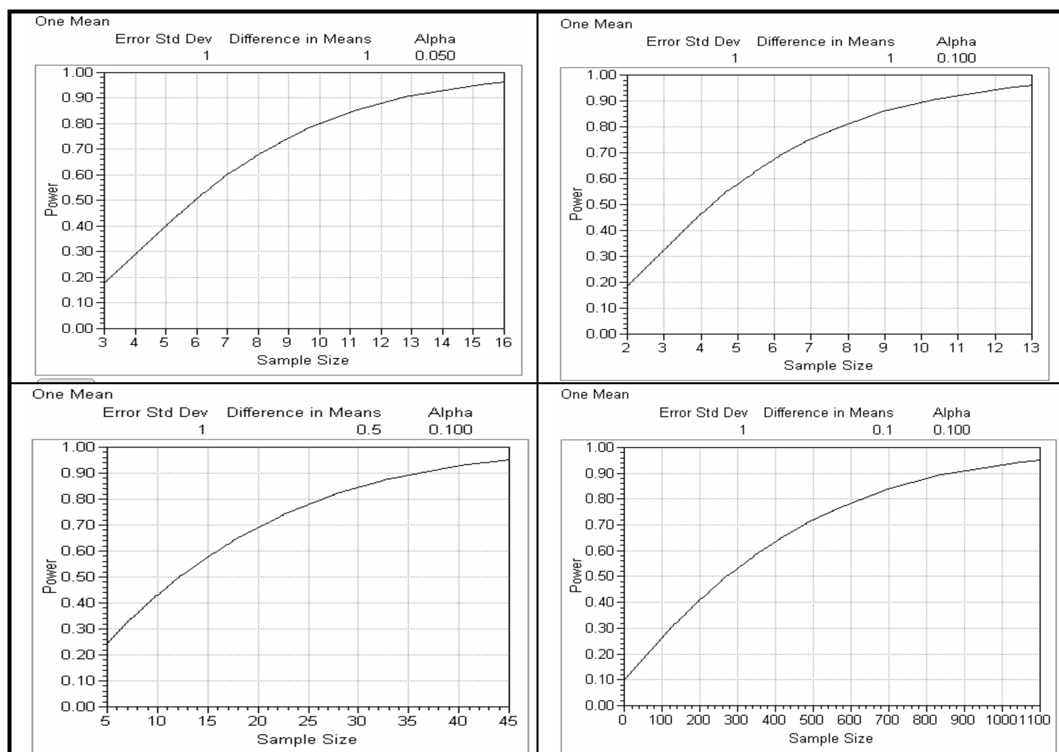
Columns (5/0)
- Alpha
- Error Std Dev
- Difference
- Power (1 - Beta)
- Sample Size

Rows

All rows	12
Selected	0
Excluded	0

An alternative approach for exploring sample size requirements specifies only the **Difference to detect** in Figure 1.19 in order to produce a plot showing how the other two parameters (**Sample Size** and **Power**) vary under set conditions of **Alpha** risk and **Difference to detect,** given an expected **Error Std Dev**.

Figure 1.19 Plots of Power versus Sample Size, Given Alpha and a Difference to Detect Relative to Error



Uncertainty in Estimating Means and Standard Deviations

Anyone who has had a basic course in statistical process control (SPC) or perhaps in some level of measurement capability instruction might have been struck by the large numbers of observations usually recommended. In SPC one must estimate the grand average of process output as well as an estimate of variation in that process. Similarly, in characterizing the capability of a measurement tool or process, one must estimate several possible sources of variation. Figure 1.20 shows MN, STDEV CI.jmp (also found in the **Chapter 1** directory on the companion Web site for this book at http://support.sas.com/reece). Throughout the table, the observed mean is 10 and the observed standard deviation is 1, but the sample sizes used to determine these observations vary from 2 to 100. The lower and upper 95% confidence interval (CI) boundaries are simulations based on established statistical concepts.

Figure 1.20 MN, STDEV CI.jmp

MN, STDEV CI
NOTE Data for CI graphs

Columns (7/0)
Sample Size
Observed Mean
Lower 95% CI Bound for mean
Upper 95% CI Bound for mean
Observed Std Dev
Lower 95% CI Bound for SD
Upper 95% CI Bound for SD

Rows
All rows 19
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Sample Size	Observed Mean	Lower 95% CI Bound for mean	Upper 95% CI Bound for mean	Observed Std Dev	Lower 95% CI Bound for SD	Upper 95% CI Bound for SD
1	2	10	1.0154	18.9846	1	0.4461	31.9102
2	3	10	7.5159	12.4841	1	0.5207	6.2847
3	4	10	8.4088	11.5912	1	0.5665	3.7285
4	5	10	8.7583	11.2417	1	0.5991	2.8736
5	6	10	8.9506	11.0494	1	0.6242	2.4526
6	7	10	9.0752	10.9248	1	0.6444	2.2021
7	8	10	9.1640	10.8360	1	0.6612	2.0353
8	9	10	9.2313	10.7687	1	0.6755	1.9158
9	10	10	9.2846	10.7154	1	0.6878	1.8256
10	15	10	9.4462	10.5538	1	0.7321	1.5771
11	20	10	9.5320	10.4680	1	0.7605	1.4606
12	25	10	9.5872	10.4128	1	0.7808	1.3912
13	30	10	9.6266	10.3734	1	0.7964	1.3443
14	35	10	9.6565	10.3435	1	0.8089	1.3102
15	40	10	9.6802	10.3198	1	0.8192	1.2840
16	45	10	9.6996	10.3004	1	0.8279	1.2632
17	50	10	9.7158	10.2842	1	0.8353	1.2461
18	60	10	9.7417	10.2583	1	0.8476	1.2197
19	100	10	9.8016	10.1984	1	0.8780	1.1617

Confidence Interval for the Mean

Equation 1.7 supplies the entries for the lower and upper 95% CI bounds for the mean (third and fourth columns of data in Figure 1.20).

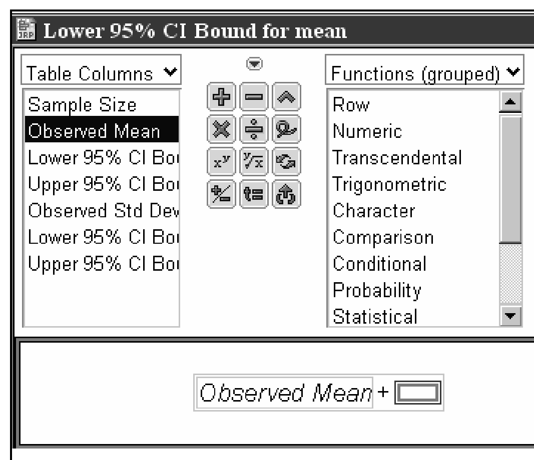
$$\bar{y} \pm t_c \left(\frac{s}{\sqrt{n}} \right) \qquad 1.7$$

where $\bar{y}$ is the observed average; t_c is a critical value of the Student's *t* based on n – 1 degrees of freedom (supported as *t* quantile in JMP); *s* is the observed standard deviation; and *n* is the sample size. The reader can display the formulas associated with data columns three and four to see the actual JMP implementation of this equation in each case.

Here are the major steps in creating the formula for the lower 95% CI bound for the mean.

1. After selecting the appropriate table column and choosing to generate a formula, select the **Observed Mean** column and add an element to it (Figure 1.21).

Figure 1.21 Adding an Element to the Observed Mean Column



2. Under the **Probability** option, select **t Quantile** (Figure 1.22). Figure 1.23 shows the resulting formula.

Figure 1.22 Selecting t Quantile

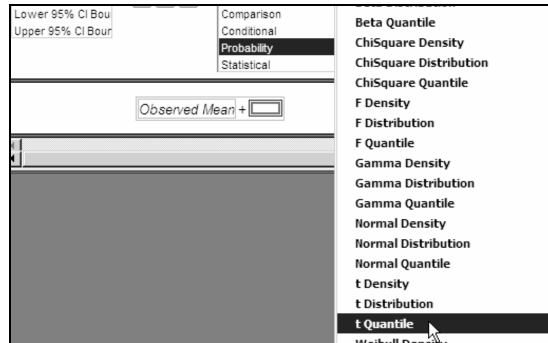
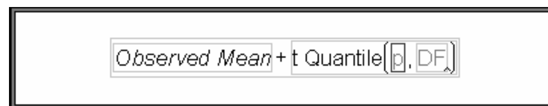


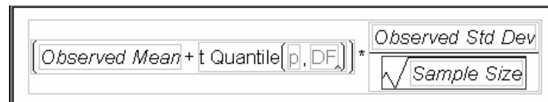
Figure 1.23 The Formula after Step 2



3. Select the multiplication operator from the choices provided in the formula window.

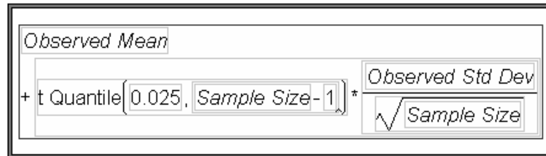
4. With the new box highlighted, select the division operator. The numerator in this expression is the **Observed Std Dev**, while the denominator is the square root of **Sample Size** (Figure 1.24).

Figure 1.24 The Formula after Step 4



5. The **DF** in the formula is **Sample Size – 1**; the **p** value is a number representing the fraction of a student's t distribution remaining in the left tail of that distribution as the lower bound of the 95% confidence interval—0.025. Figure 1.25 shows the completed formula.

Figure 1.25 The Completed Formula



The formula for the upper 95% CI boundary for the mean is identical to that just described except for the value entered for **p**. The value for this expression is 0.975, which reflects the area under the student's t distribution left of that upper boundary. Therefore, the area under the curve between the two limits is 0.975 – 0.025 = 0.95.

Confidence Interval for the Standard Deviation

In Figure 1.20, data columns six and seven, respectively, provide the upper and lower 95% CI bounds for the standard deviation. Equation 1.8 provides these values based on the varying sample sizes.

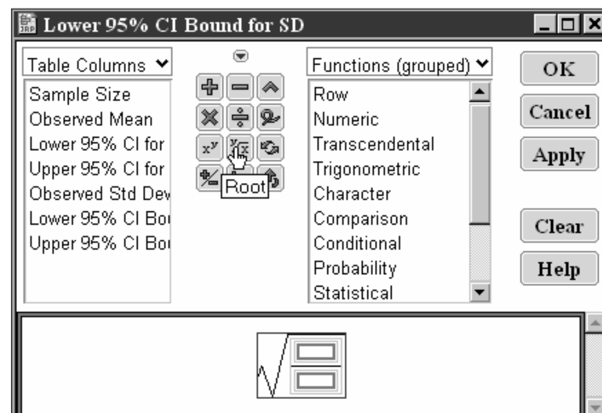
$$\sqrt{\left(\frac{\nu s^2}{\chi^2_{\nu,1-\alpha/2}}\right)} \le \sqrt{\sigma^2} \le \sqrt{\left(\frac{\nu s^2}{\chi^2_{\nu,\alpha/2}}\right)} \qquad 1.8$$

where ν represents the degrees of freedom in the sample (n – 1); s^2 is the square of the observed standard deviation; σ^2 is the true population variance; and χ^2 is a value from the chi-square distribution such that $1 - \alpha/2$ or $\alpha/2$ of the distribution remains to the left of that value ($\alpha = 0.05$ for a 95% confidence interval). The reader can display the formulas associated with data columns six and seven to see the actual JMP implementation.

The formulas embedded in these columns compute the lower and upper 95% confidence boundaries of the standard deviation (given as 1 in the table) based on the number of observations used to estimate it. Here are the steps used to create these formulas.

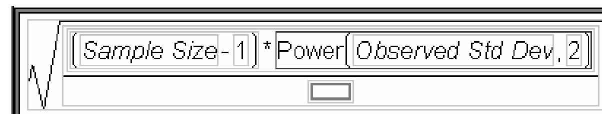
1. To create a formula with the creation of a division as the initial entry, select the ÷ symbol in the formula window. Convert that fraction to its square root using the √ symbol in the formula window (Figure 1.26).

Figure 1.26 Beginning the Formula



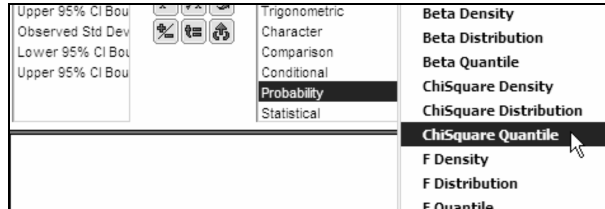
2. Add **Sample Size –1** to the numerator to compute the degrees of freedom in the estimate. Multiply this entry by the square of observed standard deviation. Use the $\mathbf{x^y}$ option in the formula window (Figure 1.27).

Figure 1.27 The Formula after Step 2



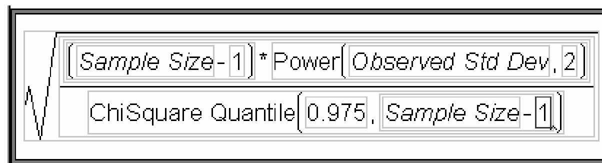
3. From the **Probability** menu, select **ChiSquare Quantile** (Figure 1.28).

Figure 1.28 Selecting ChiSquare Quantile



4. Fill in the appropriate value for **p** and **DF** in the expression. As shown in Equation 1.8 and as implemented in JMP, **p** refers to a value on the horizontal axis of the distribution such that a fraction of the area under the χ^2 curve lies to the left of that value. For a 95% confidence interval, $\alpha = 0.05$, so $1 - \alpha/2 = 0.975$. The **DF** for the estimate is one less than the sample size. For the upper boundary of the confidence interval, the **p** value is 0.025.

Figure 1.29 The Completed Formula



Figures 1.30 and 1.31 are overlay plots created in JMP that illustrate how the confidence intervals for the mean and standard deviation, respectively, respond to changes in the sample size used to estimate them.

Figure 1.30 Mean CI versus Sample Size

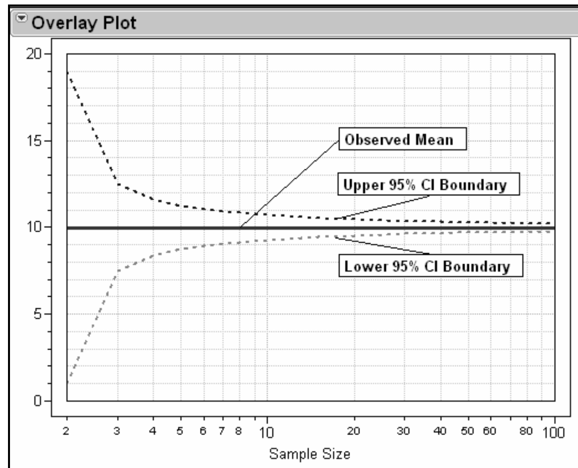
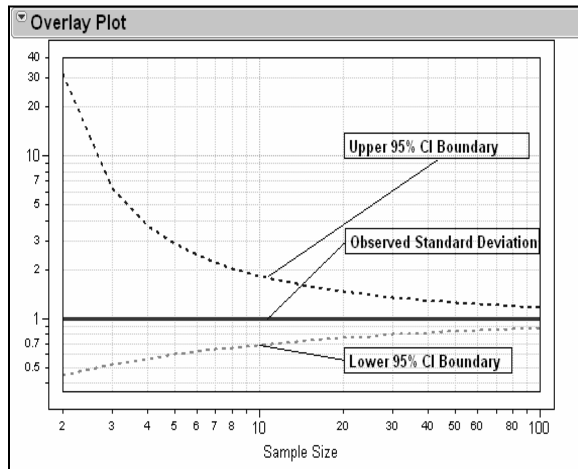


Figure 1.31 StDev CI versus Sample Size



Notice in both figures that the horizontal axes are logarithmic and that the vertical axis in Figure 1.31 is logarithmic. In Figure 1.30 the uncertainty in estimating a mean (confidence interval) begins to stabilize between sample sizes of 10 to 20. However, Figure 1.31 shows that considerable uncertainty exists in the estimate of a standard deviation until the sample size is 30 or more. This does not mean that an investigator cannot estimate these parameters with fewer samples. It means that the uncertainty in

estimates of either means or standard deviations based on small samples might be unacceptable, particularly in the case of standard deviations.

A measurement capability study primarily estimates variation in the measurement process; therefore, the investigator should try to accumulate at least 30 independent or replicate samples for each source of variation in that measurement process in order to have more sound estimates of the contributions of each.

Components of Measurement Error

The observed error in a process is the sum of the actual process variation and a measurement error, as shown in Equation 1.1 previously. Measurement error itself has two identifiable and measurable components as well: repeatability and reproducibility, as shown in Equation 1.9.

$$\sigma^2_{measurement} = \sigma^2_{repeatability} + \sigma^2_{reproducibility} \qquad 1.9$$

Repeatability Error

Repeatability error is the simplest measurement error to estimate, because it represents the ability of the measurement process to repeat values in a short period of time. In the semiconductor industry, for example, an investigator using an automated tool could generate data for this estimation in a matter of minutes. All one has to do is place the object to be measured in the tool and press the Measure button a number of times. Since this error is a variance, the experimenter should collect at least 30 readings in short order (Figure 1.31) in order to obtain a reliable estimate. In other cases involving operators measuring individual parts, this error represents the ability of operators to repeat their measurements on those parts, perhaps over a period of time.

The important point is that this error is an estimate of the variation of the measurement system under conditions of minimum perturbation. A guideline for the level of allowable repeatability error is for the P/T ratio to be ≤5. Many measurement experiments allow estimation of repeatability error along with estimation of reproducibility error and total measurement error. Conducting a simple preliminary experiment to estimate repeatability alone can pay dividends in that such a study might detect an inherent weakness or problem with the measurement system and do it relatively inexpensively. The sections “An Oxide Thickness Measuring Tool” and “Repeatability of an FTIR Measurement Tool” in Chapter 2 illustrate case studies where this was indeed the case.

Reproducibility Error

Reproducibility error includes all other variables chargeable to the measurement system, such as day-to-day variation or operator-to-operator variation. Again this type of study estimates standard deviations or variances, so an investigator planning this investigation must pay particular attention to generating enough data such that each measurement factor has enough degrees of freedom (replicates) associated with it to produce a reliable estimate. Therefore, a complete and robust measurement study to estimate total measurement error and to separate repeatability and reproducibility errors can require weeks to complete. This is not to say that the measurement study should dominate the work of individuals running a process. Rather, over a significant period of time, the study should include enough measurement episodes to enable reliable estimation of variances.

Generating a matrix of trials before starting the study and analyzing the result of filling in the measurement data with random numbers will provide information about the degrees of freedom associated with each factor of interest.

In the semiconductor industry, an investigator usually measures several points on a wafer. These points will yield different values because coating thicknesses, for example, will vary depending on location on the wafer surface. The variations among these points is not normally charged to the measurement process, because they are an artifact of the object being measured and not part of the measurement process itself. But repeated measurements of those same locations do contribute to measurement variability.

Linearity of Measurement Tools

As it applies to measurement processes, *linearity* is a measure of how stable measurement error and bias are over some range of values being measured. For example, a particular process might measure items of widely varying dimensions. The question to answer is whether or not the measurement error and bias is constant over that range. If it is not, then this type of study will make users aware of any additional limitations of their measurement process. Graphical representations can generate the required information. The section “Repeatability and Linearity of a Resistance Measurement Tool” in Chapter 2 provides an example that illustrates some of these principles.

Random versus Fixed Effects

As applied to measurement systems, *random effects* are sources of variation that might include repeated measurements, variation over time, variation due to operators, variation due to different measurement tools, or even variation due to supposedly identical objects. Classical statistical methods consider observations of such events as examples from a large population of possible events—such as an infinite number of repetitions of an event, an infinite number of days for a study, or an infinite number of operators doing a particular task. Realistically, a measurement study samples replications, time, and operators and assumes they constitute a sample from a larger population. Particularly in the case of operators, usually no more than a few are available or trained for a particular task. Therefore, even though the entire population of operators might be involved in a study, logic requires treating any contribution from them as a source of nuisance variation or random noise initially. If, for example, one or more operators demonstrably produce results different from the group, then further study of each operator becomes warranted to establish a cause and a possible correction. A later section illustrates this point.

In a measurement system, a factor is a *fixed effect* when that factor that is not normally considered a source of noise in the measurement system. In the semiconductor industry, examples of fixed effects would be the differences in individual measurement locations on a wafer, or differences between wafers with distinctly different properties such as film thicknesses. Ironically, factors designated fixed in one case (such as locations on a wafer surface) can become random in another, depending on the context of their analysis. For example, an experimental study of a manufacturing process might seek to minimize the variation of measurements found across a wafer surface. Although the investigation could continue to consider each wafer location a fixed effect, the objective of such a study generally is to minimize the contribution of those differences. Therefore, a logical approach considers the measurement sites sources of random error or nuisance variation in a process.

In the author's experience, whether to consider a particular effect random or fixed depends on the context of its effect on a process, whether a measurement process or a manufacturing process. JMP has excellent utilities for handling both types of effects, and later sections will make every effort to explain how and why to assign particular effects to a specific designation of random or fixed.