Statistical Intervals: Confidence, Prediction, Enclosure

José G. Ramírez, PhD, W.L. Gore and Associates Inc.

Table of Contents

Abstract

Statistical intervals can be confusing, even in the minds of those who use them often. This paper uses an example to describe the differences between confidence intervals, prediction intervals and tolerance intervals.

1. Coaxial Cable Manufacturing Example

Let's say you are in charge of a coaxial cable manufacturing operation, and that one type of coaxial cable you make has a target resistance of 50 Ohms with a standard deviation of 2 Ohms. You recently took a random and representative sample of 40 cables from your production process over the course of a month to characterize the resistance measurements of the coaxial cable population. Using JMP® (**Analyze > Distribution**) you generate a histogram (Figure 1) along with basic statistics.

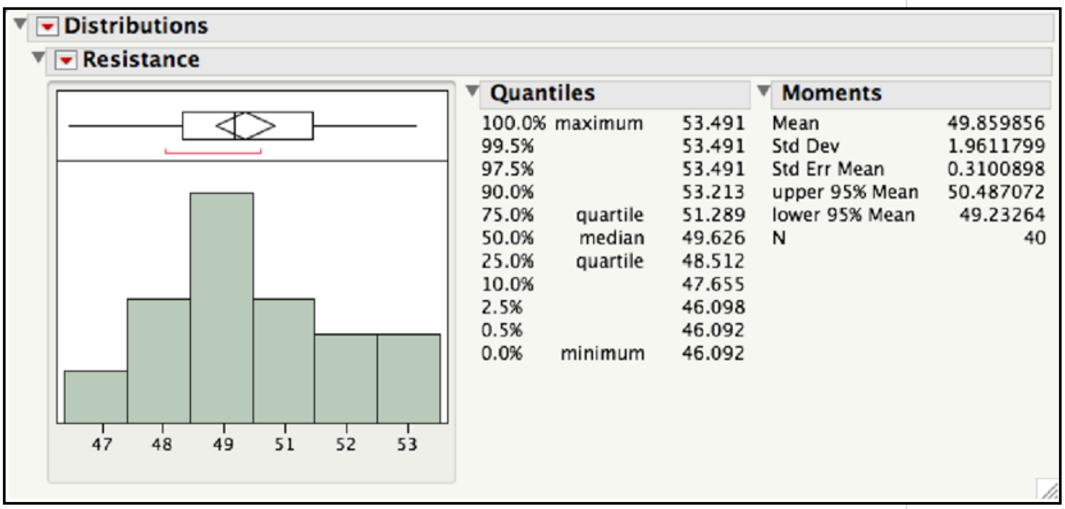


Figure 1. Histogram and Summary Statistics for Resistance Data.

The average resistance is 49.86 Ohms and the standard deviation is 1.96 Ohms. Based on these estimates of the mean and the standard deviation, it seems that you are meeting the target values of 50 Ohms and 2 Ohms, respectively. But what else can we say about the resistance data?

2. Confidence Intervals

A confidence interval is ideal for quantifying the degree of uncertainty around common parameters of interest such as the center of a sampled population, or its spread. For normally distributed data, a confidence interval for the mean looks like:

$$\bar{X} \pm t_{1-\frac{\alpha}{2},n-1} \, s\sqrt{\frac{1}{n}} \qquad (1)$$

As you can see, the confidence interval adds a margin of error to *X*, the estimate of the center of the population, which is a function of a given *t* distribution quantile with degrees of freedom equal to the number of observations minus one (*n-1*), the estimate of the standard deviation *s*, and the sample size *n*. Similarly, a confidence interval for the standard deviation gives lower and upper bounds for the variation of the standard deviation estimate. Yes, even the estimate of noise, *s*, has noise in it!

To generate a confidence interval in JMP, click on the red arrow next to the Resistance title above the histogram (see Figure 1) and select **Confidence Interval**. The default is 95 percent confidence intervals for both the mean and standard deviation as shown in Figure 2. The 1-Alpha = 0.95 indicates the 95 percent degree of confidence of the intervals.

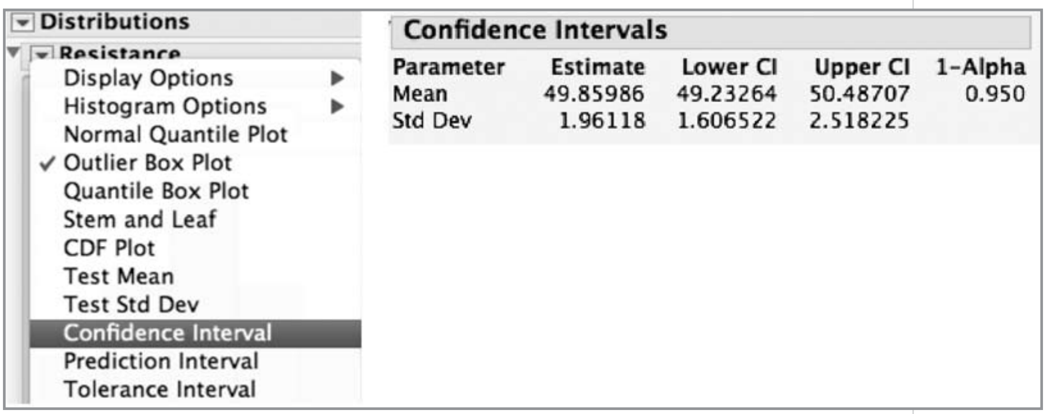


Figure 2. 95 Percent Confidence Intervals for the Mean and Standard Deviation.

The 95 percent lower confidence bound for the mean is 49.23 Ohms, while the upper 95 percent confidence bound is 50.49 Ohms. What this means is that with 95 percent confidence the true mean of the resistance measurements population lies in the interval [49.23 Ohms; 50.49 Ohms]. Because this interval contains the target value of 50 Ohms, we are 95 percent confident that we are meeting the target. For the standard deviation we can say with 95 percent confidence that the true standard deviation of the resistance measurements population lies in the interval [1.61 Ohms; 2.52 Ohms]. Again, as the target standard deviation of 2 Ohms falls within this interval, we can say that we are meeting the target with 95 percent confidence. But what does confidence mean?

2.1. The Meaning of "Degree of Confidence"

Our first reaction is to interpret "95 percent confidence" as "there is a 95 percent chance that the interval contains the true mean of my population." Unfortunately, this is not quite right. Think of these confidence intervals as "nets" that can capture the true population parameter (mean or standard deviation) so that for 95 percent confidence, on average, 95 percent of these "nets" are going to capture the true population mean, while 5 percent of them, on average, are going to fail, i.e., will not capture the true population mean. In other words, ***before*** you make the 95 percent confidence interval, there is a 95 percent chance that your "net" is going to work (capture the true population mean). ***After*** it was made, it either captured the true population mean or it did not. It's just like a game of poker: Before the hand is dealt you have a 42.3 percent chance of getting one pair, but after you get the hand, you either have the pair or you don't! Therefore, the degree of confidence is a statement about the quality (yield) of the procedure for generating the confidence interval. Equation (1) is the procedure for generating a confidence interval for the mean.

2.2. Examining Confidence Intervals

JMP provides a nice simulator for the confidence interval for the mean that helps us visualize and understand the interpretation of degree of confidence, as described above. The simulation, modified to mimic the resistance situation, generates 100 samples of size 40 from a normal distribution with mean 50 Ohms and standard deviation 2 Ohms. For each sample, a 95 percent confidence interval for the mean is calculated, using Equation (1), and the "yield" of this process (how many intervals out of a hundred contain the true mean of 50 Ohms) is calculated. Figure 3 shows one of these simulations for which 96 percent of the intervals include the true population mean of 50, in agreement with the 95 percent statement. Clicking on the **New Sample** button generates a new set of intervals. By doing this repeatedly, you will see that sometimes 94 percent of the intervals contain the true mean, others 97 percent, etc, but in the long run the average confidence turns out to be 95%.

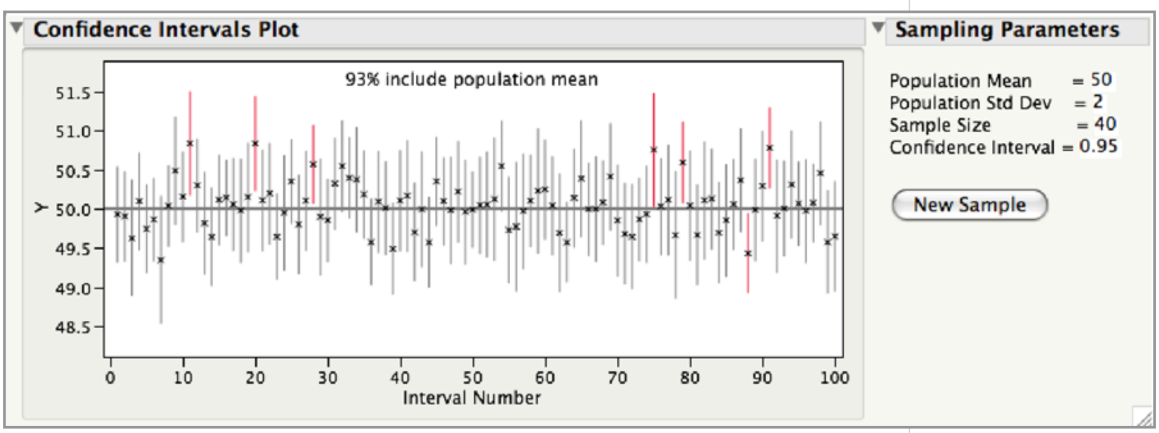


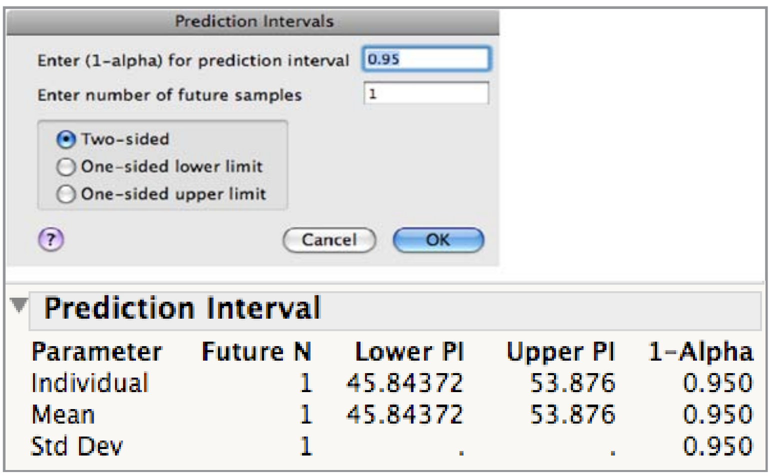
Figure 3. Confidence Interval for the Mean Simulation.

3. Prediction Intervals

What if you want to make a claim about the resistance of a future cable, or the average resistance of a group of cables that you are going to manufacture in the future? The confidence intervals for the mean and standard deviation (Figure 2) that you calculated refer to the population of cables manufactured during the month in which the 40 cables were sampled, not to an individual observation or group of observations in the future. A prediction interval for a *single* future observation resembles a confidence interval for the mean, but it is wider because it takes into account the prediction noise by adding a 1 to the expression inside the square root:

$$\bar{X} \pm t_{1-\frac{\alpha}{2},n-1}\, s\sqrt{1+\frac{1}{n}} \qquad (2)$$

Clicking again on the red arrow next to Resistance and selecting **Prediction Interval** allows you to generate a 95 percent prediction interval for one future observation (the default), as shown in Figure 4.



Parameter	Future N	Lower PI	Upper PI	1-Alpha
Individual	1	45.84372	53.876	0.950
Mean	1	45.84372	53.876	0.950
Std Dev	1	.	.	0.950

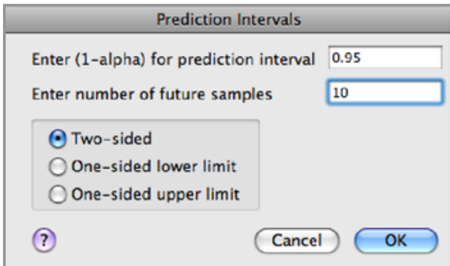
Figure 4. 95 Percent Prediction Interval for One Future Observation.

The 95 percent lower prediction bound is 45.84 Ohms, and the 95 percent upper prediction bound is 53.88 Ohms. The claim you can make is that, with 95 percent confidence, you expect a future coax cable to have a resistance between 45.84 Ohms and 53.88 Ohms. Note that since the prediction is for just 1 observation the “Individual” and “Mean” entries agree.

Let's say you get an order for a batch of 10 cables. What can we tell the customer to expect? In other words,

1. Where do you expect the resistance values of ALL 10 to be?
2. Where do you expect the average and standard deviation resistance of the 10 cables to fall?

This time we input 10 on **Enter number of future samples**.



Prediction Interval

Parameter	Future N	Lower PI	Upper PI	1-Alpha
Individual	10	43.95166	55.76806	0.950
Mean	10	48.45736	51.26235	0.950
Std Dev	10	1.046839	3.076734	0.950

Figure 5. 95 Percent Prediction Intervals for 10 Future Observations.

You expect the resistance values of ALL 10 future cables to be between 43.95 Ohms and 55.77 Ohms. This is a simultaneous interval in the sense that the bounds are derived so they contain the resistance values of ALL 10 future cables with the specified confidence, 95% in this case. Because of this, the interval is wider than the prediction interval for 1 future observation. If you are interested in a claim not about the 10 individual cables but the performance of the batch, then you expect the average resistance of the batch of 10 future cables to fall within 48.46 Ohms and 51.26 Ohms, and the standard deviation to be within 1.05 Ohms and 3.08 Ohms. Note that both the prediction intervals for the mean and the standard deviation of 10 future observations contain the targets of 50 Ohms and 2 Ohms.

What about a batch of 100 or 1000 cables? What about claims about the capability of your process? Although we can construct simultaneous prediction intervals for 100, 200,…, 1000 observations, as the number of future observations increases so does the width of the simultaneous prediction interval. Fortunately, there is another type of interval that helps us make claims about large batches or the capability of our process.

4. Tolerance Intervals

A tolerance interval is an *enclosure* interval for a specified proportion of the sampled population, *not* its mean or standard deviation. For a specified confidence level, you may want to determine lower and upper bounds such that 99 percent of the population is contained within them. Tolerance bounds allow you to set up specification limits by finding the lower and upper values, which correspond to stated yield or process capability goals.

It is common practice in engineering and science to add ±3sigma to our estimate of the mean to set a specification limit, for example. This is based on the fact that for normally distributed data a ±3sigma interval around the mean contains 99.73 percent of the population. However, this is true only (and this fact is seldom mentioned) if we know the true population parameters. This is very rarely the case, and we need to estimate the mean and variance from a (usually small) random and representative sample from our population. A tolerance interval is also constructed as the mean ± a multiple of the standard deviation, but it takes into account the estimation noise and the sample size via the function *g*(confidence, proportion, sample size).

$$\bar{X} \pm g_{(1-\alpha/2,p,n)}\, s \qquad (3)$$

A 95 percent ±3sigma-equivalent tolerance interval is then given by the equation X ± *g*(0.975, 0.9973, n) x *s*. Note that the 0.975 refers to the confidence, while the 0.9973 (3sigma equivalent) refers to the proportion of the population covered by the tolerance bounds. These two values make the tolerance interval a little confusing. Clicking on the red arrow next to the Resistance histogram and selecting **Tolerance Interval** brings up the Tolerance Intervals window to generate a 95 percent (**Specify Confidence (1-Alpha)**) tolerance interval to enclose 99.73 percent (**Specify Proportion to cover**) of the resistance population. Figure 6 shows a 95 percent tolerance interval that covers 99.73 percent of the resistance data population.

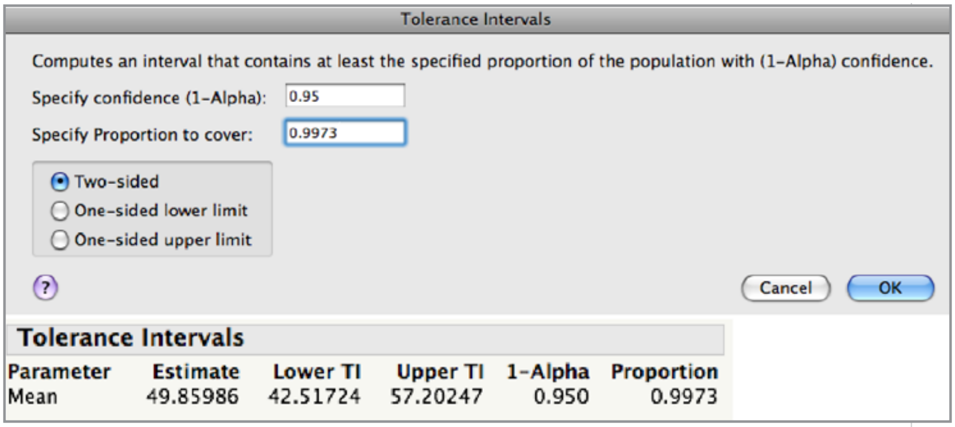


Figure 6. 95 Percent Tolerance Interval to Cover 99.73 Percent of the Resistance Population.

You expect 99.73 percent of your resistance measurements to be within 42.52 Ohms and 57.20 Ohms. The "traditional" mean ± 3sigma interval (= 49.86 ± 3x1.96) [43.98; 55.74] is narrower because it does not take into account the sample size (40) and the estimation noise. Based on these tolerance bounds you can develop specification limits for Resistance as [42 Ohms; 58 Ohms], which will contain > 99.73 percent of the resistance population measurements.

From the summary in Table 1, we can see that, with 95% confidence, the corresponding tolerance interval to enclose 99.73% of the population is narrower than the simultaneous prediction intervals for 100 and a 1000 observations. As the number of future observations increases so will the width of the simultaneous prediction interval. Therefore, for batch sizes greater than 50, it is better to use a tolerance interval than a simultaneous prediction interval.

5. How Much Can We Trust Our Claims?

Our calculated statistical intervals represent the state of our process at a given time; i.e., they are snapshots in time of what the process is doing. How do we guarantee the claims that we are going to make based on the statistical intervals? Statistical analyses usually depend on the homogeneity assumption that the data comes from a single universe rather than multiple ones. Another way of thinking about the homogeneity assumption is by asking the question: how stable is the “process” that generates our data?

A good way of checking for process stability is by means of a process behavior chart, or control chart (for more information on process behavior charts to check the homogeneity assumption please refer to Wheeler (2005)). Figure 7 shows a process behavior chart (Graph > Control Chart > IR) for the 40 individual resistance measurements that we collected for our study. Because the process behavior chart shows a reasonable degree of stability, we can say that the process is stable, consistent or predictable. We have some evidence then, that the data we used to develop our intervals comes from a predictable process; therefore, the claims we make based on the statistical intervals are valid.

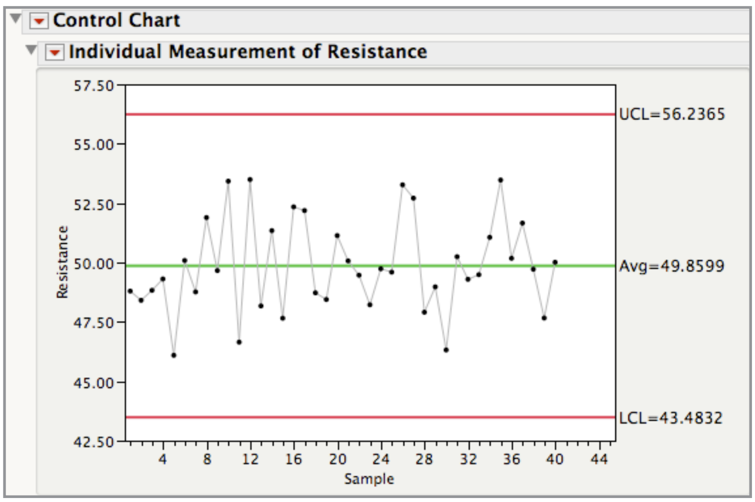


Figure 7. Process Behavior Chart for 40 Resistance Measurements

The Lower Control Limit (LCL), 43.48 Ohms, and Upper Control Limit (UCL), 56.24 Ohms define the natural process variation for the resistance data. So as long as the process remains stable we expect future resistance readings to be within 43.48 Ohms and 56.24 Ohms. Sounds like a tolerance interval, doesn’t it? In fact, the natural process limits can be thought of as a 95% tolerance interval but, since they are calculated using ±3sigma and not Equation 3, the coverage is usually less than 99.73%. For the LCL and UCL displayed in Figure 7 the tolerance interval coverage is about 98.38%.

6. Summary

Statistical intervals help us to quantify the uncertainty surrounding the estimates that we calculate from our data, such as the mean and standard deviation. The three types of intervals presented here – confidence, prediction and tolerance – are particularly relevant for applications found in science and engineering because they allow us to make very practical claims about our sampled data, as shown in Table 1. Because JMP makes it easy for us to obtain these intervals, all we have to do is apply them correctly.

Interval Type	Confidence	Lower (Ohms)	Upper (Ohms)	Interpretation
Confidence for the Population Mean	95%	49.23	50.49	The "true" mean of the resistance measurements is somewhere within these bounds.
Prediction for 1 Future Observation	95%	45.84	53.88	You expect one future resistance measurement to be within these bounds.
Prediction for 10 Future Observation	95%	43.95	55.77	You expect ALL ten future resistance measurement to be within these bounds.
Prediction for the Average of 10 Future Observations	95%	48.46	51.26	You expect the average of 10 future resistance measurements to be within these bounds.
Tolerance to Enclose 99.73% of the Population	95%	42.52	57.20	You expect 99.73% of the resistance measurements to be within these bounds.
Prediction for 100 Future Observation*	95%	42.32	57.40	You expect ALL hundred future resistance measure-ment to be within these bounds.
Prediction for 1000 Future Observation*	95%	40.81	58.91	You expect ALL thousand future resistance measure-ment to be within these bounds.

**Wider than the corresponding tolerance interval to enclose 99.73% of the population*

Table 1. Statistical Intervals and Interpretations for the Resistance Data.

References

Ramírez, José G., and Ramírez, Brenda S. To be published 2008. *Analyzing and Interpreting Continuous Data Using JMP: A Step-by-Step Guide*. Cary, NC: SAS Institute Inc.

Hahn, Gerald J., and William Q. Meeker. 1991. *Statistical Intervals: A Guide for Practitioners.* Hoboken, NJ: John Wiley and Sons, Inc.

Wheeler, Donald J. (2005). *The Six Sigma Practitioner's Guide to Data Analysis*. SPC Press, Knoxville, TN.