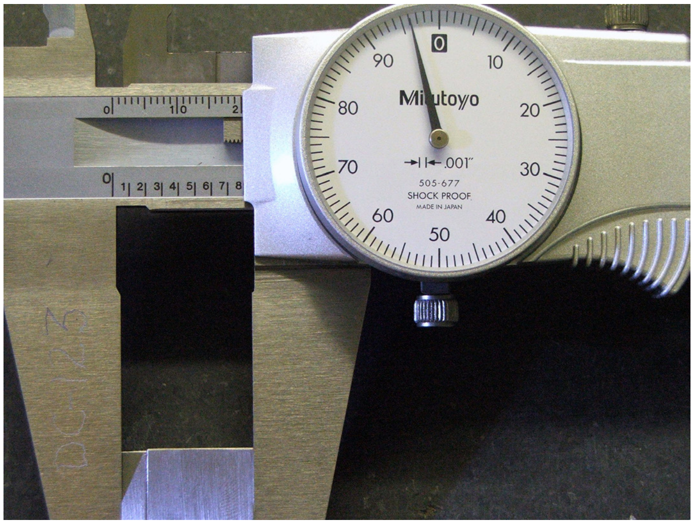
Minitab®

Gage R&R: Are 10 Parts, 3 Operators, and 2 Replicates Enough?

Most quality improvement projects have a clear goal, like reducing defects, improving a response, or making a change that benefits customers. Teams often want to jump right in and start gathering and analyzing data so they can solve the problems. Checking measurement systems first, with methods like Gage R&R, may seem like a waste of time.

But a Gage R&R study is a critical step in any statistical analysis involving continuous data, because it allows you to determine if your measurement system for that data is adequate or not.



If your measurement system can't produce reliable measurements, any analysis you conduct with those measurements is likely meaningless. After all, if you can't trust your measurement system, then you can't trust the data it produces.

Gage R&R can help you answer questions such as:

- Can my measurement system discriminate between parts?
- Is the measurement system variability small compared to the manufacturing process variability?
- How much variability in my measurement system is caused by differences between operators?

You can also use Gage R&R to determine where any weaknesses are. For example, you can use Gage R&R to figure out why different operators reported different readings.

The Standard Approach to Gage R&R

"You take 10 parts and have 3 operators measure each 2 times."

This approach to a Gage R&R experiment is so common, so accepted, that few people ever question whether it is *effective*. We're going to explore that in this paper.

Assessing a Measurement System with 10 Parts

First, let's look at how accurately you can assess your measurement system with just 10 parts.

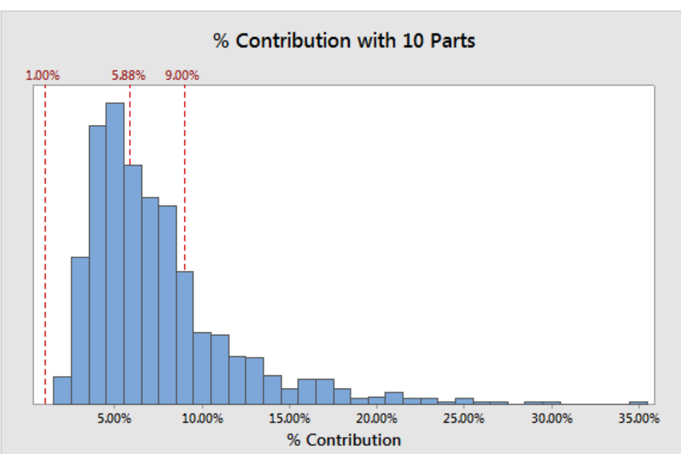
We'll focus on the %Contribution for GageR&R, which tells us how much of the variation in your process can be attributed to your measurement system. Anything under <1% is typically considered excellent, and >9% is poor.

We simulated the results of 1,000 Gage R&R studies with the following underlying characteristics:

- There are no operator-to-operator differences, and no operator*part interaction.
- The measurement system variance and part-to-part variance used would result in a %Contribution of 5.88%, between the popular <1% and >9% guidelines.

Based on these 1,000 simulated Gage studies, what do you think the distribution of %Contribution looks like? Do you think it is centered near the true value (5.88%), or do you think the distribution is skewed? And if so, how much do you think the estimates vary?

The distribution, with the guidelines and true value indicated, is shown in the graph below.



The good news is that it is roughly averaging around the true value.

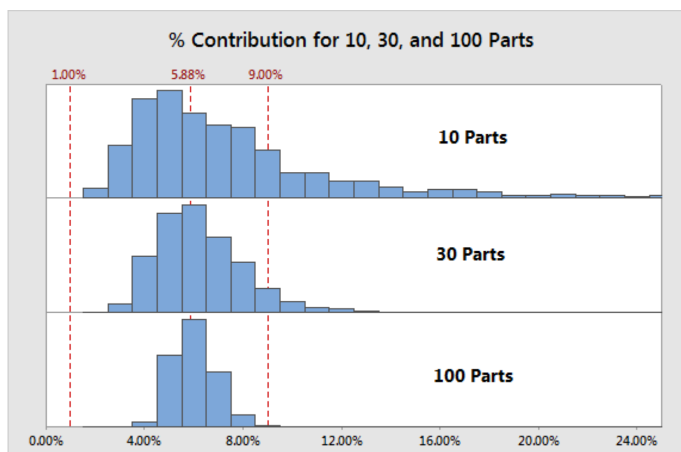
However, the distribution is highly skewed—a decent number of observations estimated %Contribution to be at least double the true value, with one estimating it at about *six times* the true value! In addition, the variation is huge. In fact, about 1 in 4 studies would have resulted in failing this gage.

Now, a standard gage study is no small undertaking. A total of 60 data points must be collected, and once randomization and "masking" of the parts is done, conducting the study itself can be quite tedious (and possibly annoying to the operators).

So just how many parts would we need to obtain a more accurate assessment of %Contribution?

Assessing a Measurement System with 30 Parts

We simulated another 1,000 gage studies, this time using 30 parts (that's a total of 180 data points). Then for good measure, we went ahead and simulated 1,000 gage studies using 100 parts (600 data points). So now consider the same factors from before for these counts—mean, skewness, and variation. The following graph shows the distributions of all three sets of simulations.



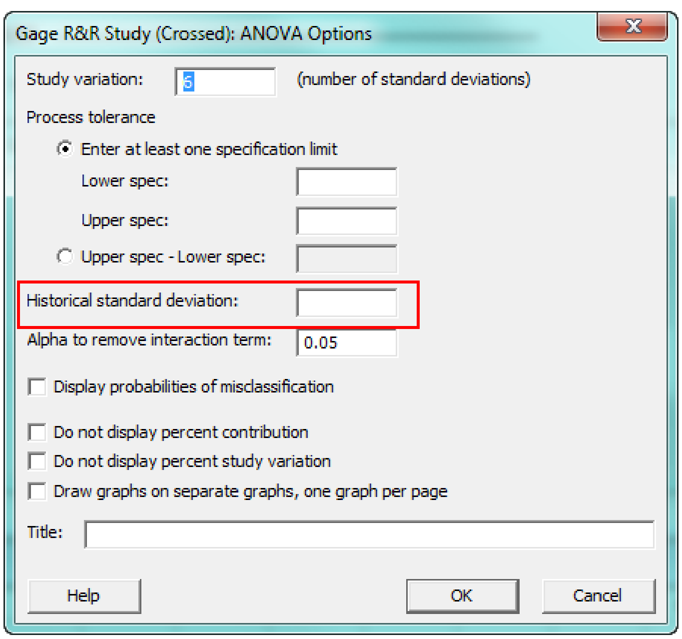
Mean is easy: if it was centered before, it's probably centered still.

Now let's look at skewness and how much we were able to reduce variation. Skewness and variation have clearly decreased, but perhaps you supposed variation would decrease more than it did. Keep in mind that %Contribution is affected by your estimates of repeatability and reproducibility as well, so increasing the number of parts will only tighten this distribution by so much. But still, even gage studies that use 30 parts—which would be an enormous experiment to undertake—still result in this gage failing 7% of the time!

So what is a quality practitioner to do?

Here are two recommendations. First, consider %Process. Often we are evaluating a measurement system that has been in place for some time, and we are simply verifying its effectiveness. In this situation, you can use the historical standard deviation as your estimate of overall variation, instead of relying on your small sampling of parts to come up with an estimate. This can eliminate much of the variation caused by the same sample size of parts.

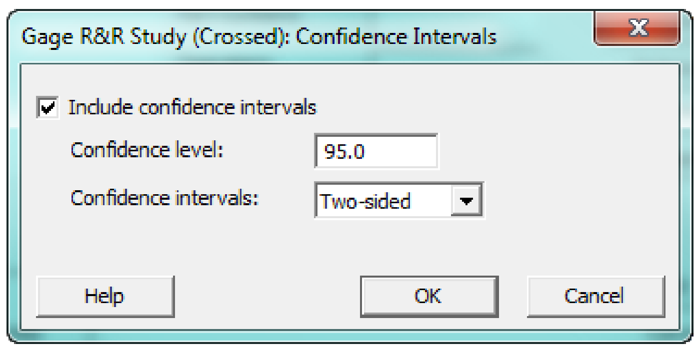
In Minitab's Gage R&R tools, it's a simple matter to enter your historical standard deviation in the Options subdialog:



Now your output will include an additional column of information called %Process. This column is the equivalent of the %StudyVar column, but uses the historical standard deviation (which comes from a much larger sample) instead of the overall standard deviation estimated from the data collected in your experiment:

Source	StdDev (SD)	Study Var (6 × SD)	%Study Var (%SV)	%Process (SV/Proc)
Total Gage R&R	0.30237	1.81423	27.86	30.24
Repeatability	0.19993	1.19960	18.42	19.99
Reproducibility	0.22684	1.36103	20.90	22.68
Operator	0.22684	1.36103	20.90	22.68
Part-To-Part	1.04233	6.25396	96.04	104.23
Total Variation	1.08530	6.51180	100.00	108.53

A second recommendation is to include confidence intervals in your output. When using Minitab for Gage R&R, this can be done in the Conf Int subdialog:



Including confidence intervals in your output doesn't inherently improve the wide variation of estimates the standard gage study provides, but it does force you to recognize just how much uncertainty there is in your estimate. For example, consider this output based on the gageaiag.mtw sample dataset included with Minitab, with confidence intervals turned on:

Gage R&R

Source	VarComp	95% CI	%Contribution (of VarComp)	95% CI
Total Gage R&R	0.09143	(0.052, 2.124)	7.76	(2.14, 66.18)
Repeatability	0.03997	(0.030, 0.056)	3.39	(0.89, 7.34)
Reproducibility	0.05146	(0.013, 2.084)	4.37	(0.77, 64.72)
Operator	0.05146	(0.013, 2.084)	4.37	(0.77, 64.72)
Part-To-Part	1.08645	(0.512, 3.631)	92.24	(33.82, 97.86)
Total Variation	1.17788	(0.602, 4.435)	100.00	

For some processes you might accept this gage based on the %Contribution being less than 9%. But for most processes you really need to trust your data, and the 95% CI of (2.14, 66.18) is a red flag which indicates you really shouldn't be very confident that you have an acceptable measurement system.

Now let's consider the other two factors in the standard Gage experiment: 3 operators and 2 replicates. What if, instead of increasing the number of parts in the experiment, you increased the number of operators or number of replicates?

Again, we are interested in the effect on %Contribution for Gage R&R, the estimate of overall Gage variation. Increasing operators will give you a better estimate of of the operator term and reproducibility, and increasing replicates would get you a better estimate of repeatability—but we want to look at the overall impact on the assessment of the measurement system.

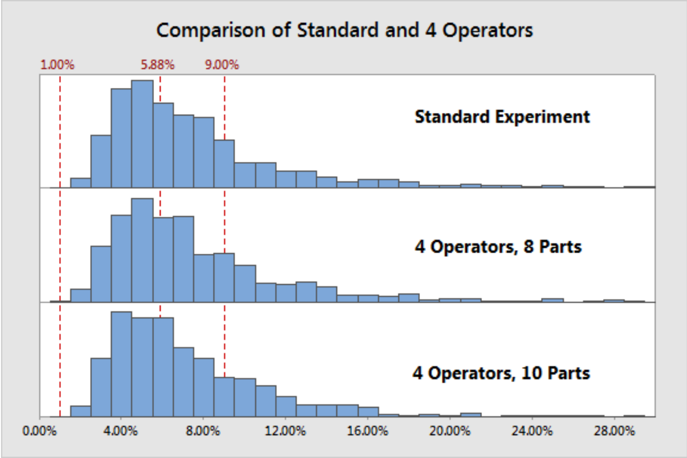
Operators

First we will look at operators. We performed two additional sets of simulations. In one, we increased the number of operators to 4 and continued using 10 parts and 2 replicates, for a total of 80 runs. In the other, we increased to 4 operators and still used 2 replicates, but decreased the number of parts to 8 to get back close to the original experiment size (64 runs compared to the typical 60).

The graph and output shown at the top of the next column is a comparison of the standard experiment and each scenario laid out here.

It may not be obvious in the graph, but increasing to 4 operators while decreasing to 8 parts actually *increased* the variation seen in %Contribution... so despite requiring 4 more runs, this is a poorer choice.

Further, the experiment that involved 4 operators but maintained 10 parts (for a total of 80 runs) showed no significant improvement over the standard study.

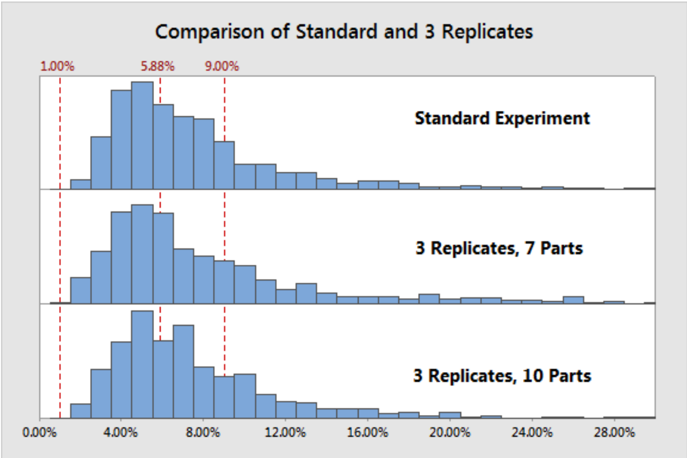


Variable	N	Mean	StDev	Minimum	Median	Maximum
10 PctContribution	1000	0.07473	0.04272	0.01696	0.06352	0.35230
4 8 2 PctContribution	1000	0.07894	0.05432	0.01338	0.06524	0.50001
4 10 2 PctContribution	1000	0.07353	0.04204	0.01582	0.06244	0.35894

Replicates

Now let's look at replicates in the same manner. In one run of simulations, we will increase the number of replicates to 3 while continuing to use 10 parts and 3 operators (for a total of 90 runs). In another, we will increase number of replicates to 3 and operators to 3, but reduce parts to 7 to compensate (which results in a total of 63 runs).

Now we can compare the standard experiment to each of these scenarios:



Variable	N	Mean	StDev	Minimum	Median	Maximum
10 PctContribution	1000	0.07473	0.04272	0.01696	0.06352	0.35230
3 7 3 PctContribution	1000	0.08394	0.06359	0.01219	0.06355	0.58203
3 10 3 PctContribution	1000	0.07629	0.04466	0.01839	0.06630	0.45971

We see the same pattern we observed when we increased the number of operators. Increasing to 3 replicates while compensating by reducing to 7 parts (for a total of 63 runs) significantly increases the variation seen in %Contribution. Increasing to 3 replicates while maintaining 10 parts shows no improvement.

Conclusions about Operators and Replicates in Gage Studies

As stated above, we're only looking at the effect of these changes on the *overall* estimate of measurement system error. So while increasing to 4 operators or 3 replicates either showed no improvement in our ability to estimate %Contribution—or actually made it worse—you could encounter a situation where you are willing to sacrifice that in order to get more accurate estimates of the individual components of measurement error. In that case, one of these designs might be a better choice.

For most situations, if you're able to collect more data, increasing the number of parts used remains your best choice for obtaining better estimates of %Contribution.

However, this raises a separate question: How do we select those parts...?

This document compiles material first published on the Minitab Blog. The authors of the original posts are Michelle Paret, technical sales manager at Minitab, and Joel Smith, director of rapid continuous improvement at Dr Pepper Snapple Group and former technical sales manager at Minitab.