

Measurement System Analysis And Destructive Testing

USE A NESTED MODEL WHEN TESTING A SMALL HOMOGENOUS BATCH.

By Douglas Gorman, Minitab, and Keith M. Bower, Minitab

A measurement system analysis is a vital component for many quality improvement initiatives. It is important to assess the ability of a measurement system to detect meaningful differences in process variables. In many instances, it is possible to design an experiment in which the level of variation attributed to the operator, parts and the operator by part interaction effect can be assessed. This is possible when all operators have the ability to measure each part and is typically the experimental design assessed during the measure phases of the define-measure-analyze-improve-control cycle.

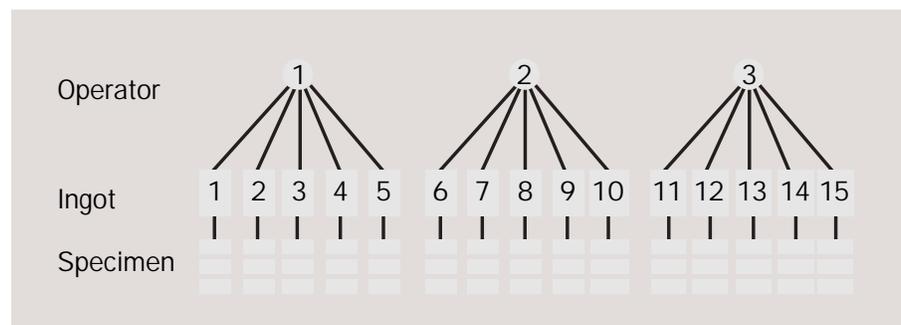
For many measurement systems, however, the part being measured is affected in some manner. As such, you cannot assume all operators in the study can assess the same part and reasonably obtain similar results.

Look at destructive testing, for example. The part characteristic such as tensile strength, impact strength or burst pressure of a vessel is measured as the part is destroyed. Once the measurement is obtained for a particular part, that part is no longer available for additional measurements with the same or different operators.

There are statistical methods available to estimate the repeatability and reproducibility (R&R) components in destructive scenarios if a key, and perhaps controversial, assumption is made. The assumption is that it is possible to identify a batch of parts enough alike that it is reasonable to consider them the same part. This means the measurement characteristic of interest is identical for each part in the group—the batch is homogenous. This assumption is important because the observed within-batch variability is used to estimate the repeatability of the measurement system. If this assumption is reasonable you can consider conducting an R&R study.

The homogenous batch size is an important consideration in designing and analyzing a destructive R&R study. A more traditional or crossed design and analysis may be appropriate when the batch is large enough to assign at least two parts from each batch to each operator. This is because

Figure 1. Nested Model Design



each operator can test each batch multiple times (the batch is crossed with the operator). When the batch can be crossed with the operator, this experimental design allows estimation of the operator by batch interaction.

When the homogenous batch is small and multiple parts from the batch cannot be given to each operator, an appropriate way to deal with the situation is to use a nested or hierarchical model. The model, the assumptions behind the model and the interpretation of results from a practical example are discussed using the following scenario.

Scenario

Consider a measurement system in which operators are required to measure the impact strength of steel. The impact test measures the energy required to break a notched metal bar of specified geometry. The test specimens are prepared from ingots randomly selected from a wider population of ingots. The experimenters believe specimens created from the same ingot are more homogenous than specimens created from different ingots. However, only three specimens can be prepared from each because of the ingot's size.

Due to the destructive nature of the test and the small batch size, it is not possible for each operator to measure each ingot multiple times. Therefore, you cannot use a crossed gage R&R study. Instead, you must use the nested gage R&R study, which is available in Minitab, release 13.

Figure 2. Nested Gage R&R Study Dialog box

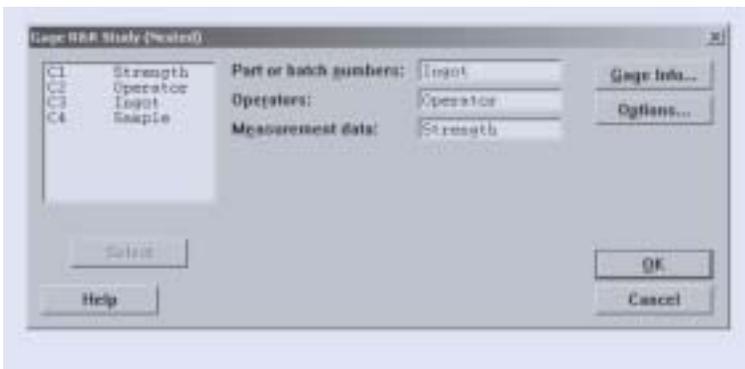


Table 1. Standard Deviation Results

Source	Standard deviation (SD)	Study variation (5.15*SD)	Percentage study variation (%SV)
Total gage R&R	0.73754	3.79831	64.31
Repeatability	0.24241	1.24840	21.14
Reproducibility	0.69656	3.58729	60.74
Part to part	0.87823	4.52288	76.58
Total variation	1.14684	5.90624	100.00

In this example there are three randomly selected operators in the study. Each operator is given five randomly selected ingots. Since it is possible to obtain three test samples from each ingot, each operator will record 15 measurements, leading to a total of 45 observations in this experiment.

Model

The statistical model for this nested design is: $strength_{ijk} = \mu + operator_i + ingot_{j(i)} + \epsilon_{(ij)k}$ where $i = 1,2,3$; $j = 1,2,3,4,5$; $k = 1,2,3$. For more information on the mathematical theory and practical application of nested designs, refer to *Design and Analysis of Experiments*¹ and *Statistics for Experimenters*.²

In this setup, the ingots are coded one through five. Of course, with the nested model in the above formula, the same ingot would not be measured across each operator. We may think, however, in terms of “ingot one for operator one” and “ingot one for operator two.” The arbitrary way in which we can code the 15 ingots under each operator (see Figure 1) indicates this model is nested, not crossed.

Results

Executing the commands: `stat > quality tools > gage R&R study (nested)`, you can complete the dialog box shown in Figure 2. Following the usual guidelines for the total R&R percentage study, the results suggest this measurement system needs improvement (see Table 1). A measurement system is generally considered unacceptable when this metric exceeds 30%, and in this case it is 64.31%. The largest relative contributor to the measurement system variability is the reproducibility

error of 60.74%. The default Minitab graphical output is investigated in the next section to guide improvement efforts with regard to the operator effect.

Note the standard deviations and study percentages in Table 1 (p. 17) do not add up to the totals in the final row. This is because the variances are additive, but the standard deviations are not. For example, $3^2 + 4^2 = 5^2$, but $3 + 4 \neq 5$.

Looking at the corresponding analysis of variance results in Table 2, you will see the operator effect is statistically significant at the $\alpha = 0.05$ level. This is because the p-value = 0.04481 < 0.05. You can, there-

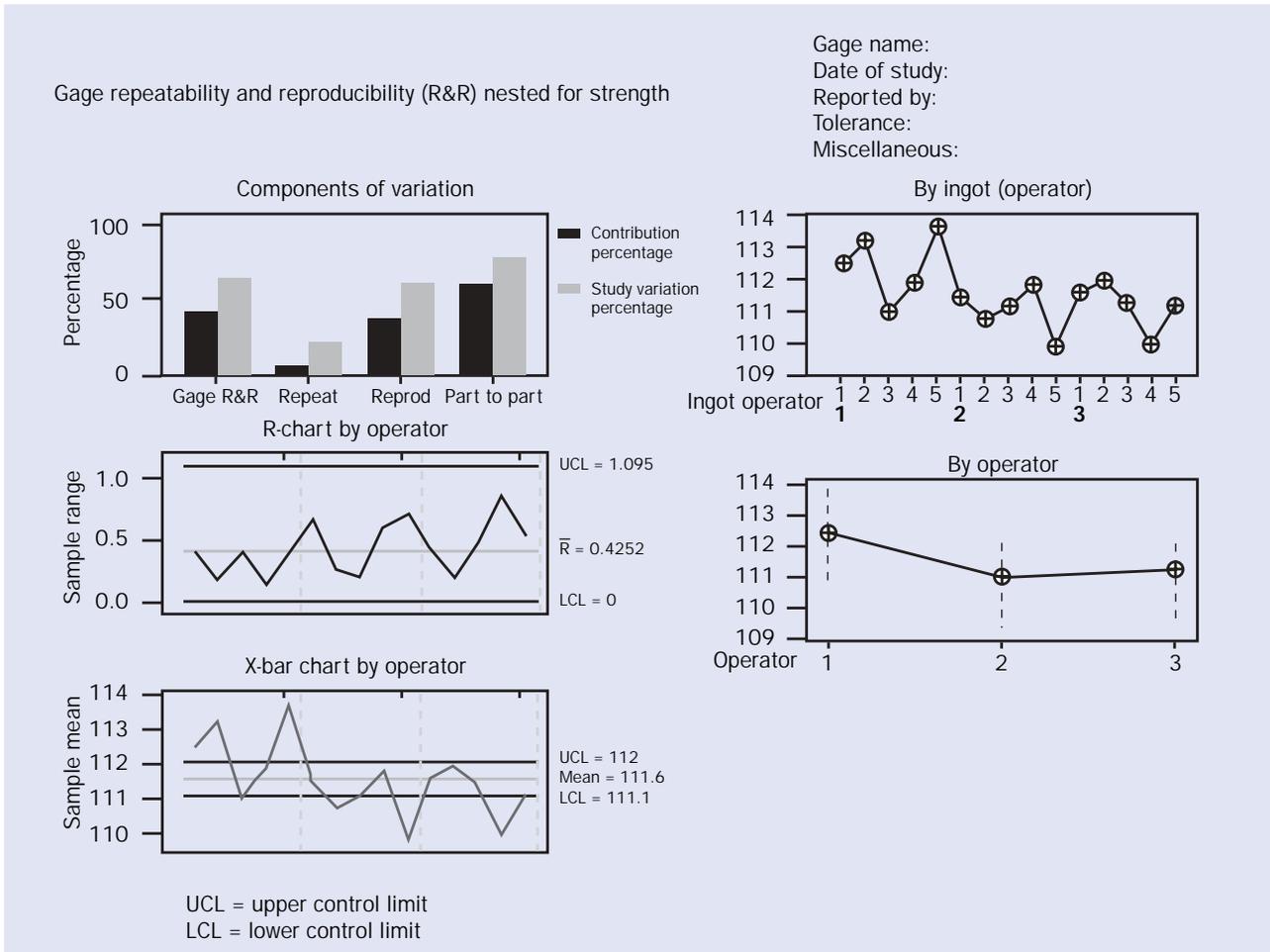
Table 2. Analysis of Variance Output

Source	Degrees of freedom (DF)	Sum of squares (SS)	Mean square (MS)	F-ratio (F)	P-value (P)
Operator	2	19.3012	9.65060	4.0675	0.04481
Ingot (operator)	12	28.4715	2.37262	40.3774	0.00000
Repeatability	30	1.7628	0.05876		
Total	44	49.5355			

fore, reject the null hypothesis $H_0: \sigma_{operator}^2 = 0$ in favor of $H_1: \sigma_{operator}^2 > 0$.

The ingot to ingot differences also contribute a relatively large amount of variability to the study (76.58%). It is certainly desirable to have the largest component of variation in the study be due to batch or ingot dif-

Figure 3. Graphical Results From Nested Gage R&R Analysis



ferences. This ingot-to-ingot effect (within operators) is also statistically significant at the $\alpha = 0.05$ level of significance. This result is important because it shows that even though the measurement system needs improvement, it is capable of detecting differences among ingots under each operator.

Table 1 also shows that the percentage study metric associated with the repeatability—what is not explained by the operator or ingots—is small relative to the others at 21.14%. With destructive R&R results, regardless whether you use a crossed or nested model, the repeatability estimate actually contains within-batch variability. This means the repeatability estimate is likely inflated to the extent that the homogenous batch assumption is violated. A measurement system found inadequate due mostly to repeatability error should raise questions about the homogenous batch assumption.

In essence, the information in the R&R table (see Table 1) tells us the ingot-to-ingot variation is larger than the combined repeatability and within-ingot variation. This supports the previous claim that specimens created from the same ingot are more homogeneous than those created from different ingots. Table 1 also indicates the operators are contributing almost as much variability to the study as the different ingots are.

Graphical Results

The R-chart in Figure 3 shows the level of variation exhibited within each ingot appears to be relatively constant. Since the range for the example using this scenario actually represents the combined variability from the repeatability of the measurement system and the within-ingot variability, this chart would be helpful in identifying whether certain operators have difficulty consistently preparing and testing specimens as well as identifying specific ingots that were not homogeneous.

From the \bar{X} chart in Figure 3 you can see the ingot averages vary much more than the control limits. This is a desirable result, as the control limits are based on the combined repeatability and within-ingot variations. It indicates the between-ingot differences will likely be detected over the repeatability error.

The chart also makes it apparent that the averages for operator one appear generally higher than those for operators two and three, but you cannot automatically lay blame on operator one. Instead, you should design an additional study to assess procedurally what the operators might do differently with regard to the important aspects of obtaining the measurement. For

example, you may wish to inspect how the test specimens are prepared or how they are fixtured in the testing device.

Remember, it is also possible the randomization procedure was ineffective in ensuring a representative sample of ingots was given to each operator and, by chance alone, operator one happened to get the three highest strength ingots out of the 15.

The “by ingot (operator)” and “by operator” graphs in Figure 3 also indicate the prevalence of operator one’s generally recording higher values for the five ingots measured, in comparison with operators two and three.

Experiment Design And Execution Are Important

By making use of an appropriate experimental design structure it is possible to assess the performance of a measurement system when destructive testing is required. It is necessary to use a nested approach when homogenous batch sizes are limited and each batch can only be tested multiple times by one operator.

When using an R&R approach to assess a destructive measurement system, the results are not as straightforward as those in a nondestructive case. Specifically, repeatability variation is indistinguishable from the within-batch variation. If a destructive measurement system is deemed unacceptable from a repeatability standpoint, the homogenous batch assumption should be questioned. As with other measurement system analyses, the design and execution of the experiment itself is instrumental in obtaining useful results.

REFERENCES

1. D.C. Montgomery, *Design and Analysis of Experiments*, fifth edition (New York: John Wiley & Sons, 2001).
2. G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building* (New York: John Wiley & Sons, 1978).

BIBLIOGRAPHY

Montgomery, D.C., and G.C. Runger, “Gauge Capability and Designed Experiments: Part I: Basic Methods,” *Quality Engineering*, Vol. 6, No. 1, July 1993. 

WHAT DO YOU THINK OF THIS ARTICLE? Please share your comments and thoughts with the editor by e-mailing

Blanton_Godfrey@ncsu.edu.

© 2002 American Society for Quality. Reprinted with permission.