

Three Ways to Analyze a Gage R&R Study

You have just completed a Gage R&R study on one of your critical to quality measurement systems. You want to find out how “good” that test method is. You have done everything right. You carefully selected the parts to reflect the range of production. You carefully selected the operators to do the testing and randomized the run order for the parts. You ensured that the operators didn’t know what part they were testing. Each operator tested each part the required number of times. Now, you are ready to analyze the results. What method do you use?



You can analyze the Gage R&R study using one of the following analysis techniques:

- Average and Range Method
- ANOVA
- EMP (Evaluating the Measurement Process)

All three techniques have been covered in detail in past publications. This publication compares the output from the three techniques and attempts to decide which is best. We will assume that we want to use the test method for process control.

In this issue:

- [The Data](#)
- [The Sources of Variation in a Gage R&R Study](#)
- [Average and Range Gage R&R Analysis](#)
- [ANOVA Gage R&R Analysis](#)
- [EMP Gage R&R Analysis](#)
- [Comparison of Results](#)
- [Summary](#)
- [Quick Links](#)

The Data



The data we will use is from the 4th edition of the Measurement Systems Analysis manual published by AIAG. In this Gage R&R study, there are three operators and ten parts. Each operator runs each part three times. The data are shown in Table 1.

For example, operator A ran part 1 three times with the following results: 0.29, 0.41, and 0.64. The data from this table are analyzed using each of the three Gage R&R analysis techniques using the *SPC for Excel software*. Before we start, we will quickly review the sources of variation in a Gage R&R study. There are four sources we primarily follow: repeatability, reproducibility, part, and total.

Table 1: Gage R&R Data

Op.	Part	1	2	3	4	5	6	7	8	9	10
	Trial										
A	1	0.29	-0.56	1.34	0.47	-0.8	0.02	0.59	-0.31	2.26	-1.36
A	2	0.41	-0.68	1.17	0.5	-0.92	-0.11	0.75	-0.2	1.99	-1.25
A	3	0.64	-0.58	1.27	0.64	-0.84	-0.21	0.66	-0.17	2.01	-1.31
B	1	0.08	-0.47	1.19	0.01	-0.56	-0.2	0.47	-0.63	1.8	-1.68
B	2	0.25	-1.22	0.94	1.03	-1.2	0.22	0.55	0.08	2.12	-1.62
B	3	0.07	-0.68	1.34	0.2	-1.28	0.06	0.83	-0.34	2.19	-1.5
C	1	0.04	-1.38	0.88	0.14	-1.46	-0.29	0.02	-0.46	1.77	-1.49
C	2	-0.11	-1.13	1.09	0.2	-1.07	-0.67	0.01	-0.56	1.45	-1.77
C	3	-0.15	-0.96	0.67	0.11	-1.45	-0.49	0.21	-0.49	1.87	-2.16

The Sources of Variation in a Gage R&R Study

The two major sources of variability that we are interested in a Gage R&R study are the repeatability and reproducibility.

1. *Repeatability* is the variation in the measurements obtained by one operator measuring the same item repeatedly. This is also called measurement or equipment variation.
2. *Reproducibility* is the variation of the measurement system caused by differences in the way operators perform the test. It is the variation in the average values obtained by several operators while measuring the same item and is sometimes called the appraiser variation.



The combined repeatability and reproducibility make up the Gage R&R variability. The third major source of variation is the *part variation*. This variation is a measure of how much the parts vary and should be representative of what occurs in production if you are using the measurement system to control the process.

The last major source of variation is the *total variation* – which is a measure of the variation in all the results.

The relationship between the total, part and measurement system variation is given by the equation below:

$$\sigma_t^2 = \sigma_p^2 + \sigma_{ms}^2$$

where the subscripts represent the source (t = total, p = part, and ms = measurement system). Note that this equality is based on variances. Remember that the variance is the square of the standard deviation (sigma).

Average and Range Gage R&R Analysis

This methodology has been around for many years and was, for much of that time, the preferred method for analyzing a Gage R&R study - mostly for the ease of calculation. That is no longer the case today. One of our [previous publications](#) laid out the calculations using the Average and Range method in detail.

The average and range method forms subgroups based on each operator-part combination (e.g., one subgroup is A-1 for operator A and part 1). The three trials from Table 1 make up that subgroup (0.29, 0.41, 0.64). Subgroup averages and ranges are calculated. Each operator's average and range is then calculated. The average range for the three operators is then found. In this example, the average range is $\bar{R} = 0.342$. To find the repeatability (called EV = equipment variation by AIAG), this average range is multiplied by a constant, K_1 , that depends on the number of trials. For 3 trials, $K_1 = 0.5908$. Thus,



$$EV = \bar{R}(K_1) = (0.342)(0.5908) = 0.202$$

A word of caution here. The value of EV does not represent a variance. It represents a standard deviation. This is the start of the problems associated with the average and range method.

The range in operator averages is then calculated. This is called \bar{X}_{DIFF} and is 0.445 in this example. This value is used in the following equation to find the reproducibility or the appraiser variability (AV).

$$AV = \sqrt{(\bar{X}_{DIFF} * K_2)^2 - (EV)^2 / nr} = 0.230$$

where K_2 is a constant that depends on the number of operators (0.5231 for three operators), n is the number of parts (10) and r is the number of trials (3). The value of AV for our example data is 0.230

The Gage R&R value is then found by combining the EV and AV results using the following equation:

$$\text{Gage R\&R} = \sqrt{EV^2 + AV^2} = 0.306$$

The part variation (PV) is found by determining the range in part values (R_p) and multiplying this range by a constant (K_3) that depends on the number of parts. For this example:

$$PV = R_p(K_3) = (3.511)(0.3146) = 1.105$$

Finally, the total variation (TV) is then found by the following equation:

$$TV = \sqrt{(\text{Gage R\&R})^2 + PV^2} = 1.146$$

Again, not that the above equation for TV is not the variance – but the *variation* represented by the standard deviation. We can use these results to determine the % of variation (NOT variance) due to each source of variation. The results are shown in Table 2.

Table 2: Average and Range Method Gage R&R Results

Measurement Unit Analysis	% Total Variation (TV)
Repeatability - Equipment Variation (EV)	
EV = 0.202	%EV = EV/TV = 17.61%
Reproducibility - Operator Variation (AV)	
AV = 0.230	%AV = AV/TV = 20.04%
Repeatability & Reproducibility (R & R)	
R&R = 0.306	% R&R = R&R/TV = 26.68%
Part Variation (PV)	
PV = 1.104	PV = PV/TV = 96.37%
Total Variation (TV)	
TV = 1.146	

The key result that most people look at is the % R&R. From Table 2, the % R&R has a value of 26.68%. Note that the values in the second column do not add to 100%.

So, what does this value of %R&R mean? The acceptance criteria from AIAG are given on page 78 of their measurement system analysis manual. The criteria given there are reproduced in Table 3 below.



Table 3: AIAG Gage R&R Acceptance Criteria*

% R&R	Decision	Comments
Under 10	Generally considered to be an acceptable measurements system.	Recommended, especially useful when trying to sort or classify parts or when tightened process control is required.
10 to 30	May be acceptable for some applications.	Decision should be based upon for example, importance of the application measurement, cost of measurement device, cost of rework or repair. Should be approved the customer.
Over 30	Considered to be unacceptable	Every effort should be made to improve the measurement system.

* from Measurement Systems Analysis, 4th Edition, 2010, AIAG

The manual does say that these criteria alone are not an acceptable practice for determining the acceptability of a test method. They are just guidelines. But, in reality, many people do just that. So, with our value of 26.88% for the % R&R, we would conclude that the test method may or may not be acceptable – it depends on the situation.

ANOVA Gage R&R Analysis

Analysis of variance (ANOVA) is a technique that examines what sources of variation have a significant impact on the results. This approach actually adds another source of variation to the mix: the operator*part interaction. This interaction is usually not significant so we will leave it out of this discussion. What ANOVA does is compare the variation in part and operator results to the repeatability of the test method.

The Gage R&R output for our example is shown below. Remember, there is no operator*part interaction so it was taken out of the ANOVA table below.

Table 4: ANOVA Gage R&R without Interaction Report

Source	df	SS	MS	F	p Value
Part	9	88.362	9.818	245.614	0.000
Operator	2	3.167	1.584	39.617	0.000
Repeatability	78	3.118	0.0400		
Total	89	94.647			



The first column is the source of variability. Operator here represents the reproducibility. The second column is the degrees of freedom associated with the source of variation. This is a measure of the amount of data present. The third column is the sum of squares. This is a measure of the variation in the data for that source.

The fourth column is the mean square associated with the source of variation. The mean square is the estimate of the variance for that source of variability (not necessarily by itself) based on the amount of data we have (the degrees of freedom). So, the mean square is the sum of squares divided by the degrees of freedom. We use the mean square information to estimate the variance of each source of variation – this is the key to analyzing the Gage R&R results.

The fifth column is the F value. This is the statistic that is calculated to determine if the source of variability is statistically significant. It is based on the ratio of two variances (or mean squares in this case). The last column is the p value – a value ≤ 0.05 is considered significant. So, both the parts and operator have a significant effect on the results.

With ANOVA, you determine the % of the total variance (not standard deviation) due to each source. The following equations can be used to calculate the variances when there is no operator*part interaction. The repeatability variance is simply the mean square of the repeatability source of variation.

$$\sigma_{\text{Repeatability}}^2 = \text{MS}_{\text{Repeatability}} = 0.04$$

The reproducibility comes from the mean square of the operators (with n = number of trials and p = number of parts):

$$\sigma_{\text{Operators}}^2 = \frac{MS_{\text{Operators}} - MS_{\text{Repeatability}}}{n * p} = \frac{1.584}{3 * 10} = 0.0515$$

The part variance comes from the mean square of the parts (where o = number of operators).

$$\sigma_{\text{Parts}}^2 = \frac{MS_{\text{Parts}} - MS_{\text{Repeatability}}}{o * n} = \frac{9.818}{9} = 1.086$$

The results are shown in Table 5. The calculations are covered in our [September 2012 publication](#).

Table 5: % Contribution to Variance by Source

Source	Variance	% Contr.
Total Gage R&R	0.0914	7.76%
Repeatability	0.0400	3.39%
Reproducibility	0.0515	4.37%
Part-to-Part	1.086	92.24%
Total Variation	1.178	100.00%

From this analysis, the % Gage R&R is 7.76%. The AIAG reference manual does include ANOVA as a way of analyzing a Gage R&R study. In fact, using these same data, the manual now says that the test method is acceptable since the % Gage R&R is below 10. What? How can it be one thing with the Average and Range method and another with the ANOVA? You probably already know the answer, but we will review it later. Next, we look at the EMP methodology.

EMP Gage R&R Analysis



Our last two publications took an in-depth look at the EMP methodology. The EMP methodology is similar to the ANOVA method in that it determines the variances due to the different sources of variation and determines the % contribution due to each source. Like the Average and Range method, it uses subgroups of data to determine the variance due to the various sources of variation. It does not take into account the operator*part interaction.

The approach, not surprisingly since it is Dr. Donald Wheeler's approach, includes the use of control charts. A range chart is made based on the subgroups composed of each operator-part combination. As long as the range chart is in statistical control, the repeatability can be estimated from the average range (using Dr. Wheeler's nomenclature):

$$\sigma_{pe}^2 = \left(\frac{\bar{R}}{d_2} \right)^2$$

where d_2 is a control chart constant that depends on subgroup size (the number of trials). The numerical results of the calculations are shown in the table below.

The range of operator averages is used to find the reproducibility using the following:

$$\sigma_o^2 = \left(\frac{R_o}{d_2^*}\right)^2 - \left(\frac{o}{(n)(o)(p)}\right)\sigma_{pe}^2$$

where R_o is the range of the operator averages, d_2^* is a bias correction factor that depends on the number of operators, n = number of trials, o = number of operators, and p = number of parts.

The combined R&R variance is the sum of the repeatability variance and the reproducibility variance:

$$\sigma_e^2 = \sigma_{pe}^2 + \sigma_o^2$$

The range of the p part averages is used to determine the product variance using the following:

$$\sigma_p^2 = \left(\frac{R_p}{d_2^*}\right)^2 - \left(\frac{p}{(n)(o)(p)}\right)\sigma_{pe}^2$$



The total variance is the sum of the product variance and the combined Gage R&R variance:

$$\sigma_x^2 = \sigma_p^2 + \sigma_e^2$$

The variances for each source of variation are shown in Table 6. The last column is the % of total variance due to each source of variation.

Table 6: Contribution to % Variance using EMP Methodology

<i>Component</i>	<i>Variance</i>	<i>% of Total</i>
Repeatability	0.0407	3.1%
Reproducibility	0.0531	4.1%
R&R	0.0938	7.2%
Product	1.216	92.8%
Total	1.310	

These results are very close to those obtained from the ANOVA Gage R&R methodology. Now, let's compare the results.

Comparison of Results

The results are compared in Table 7. The source is given in the first column. The average and range method results are given first. There has been an addition to the results for the Average and Range method. The first two columns under the Average and Range results are based on the calculations shown above - which use the standard deviation for the results. Those standard deviations were

squared to generate variances and then the % of Total Variance was calculated for the Average and Range method.

Table 7: Comparison of Results

Source	Average and Range				ANOVA		EMP	
	Variation	% of Total Variation	Variance	% of Total Variance	Variance	% of Total Variance	Variance	% of Total Variance
Repeatability	0.202	17.61%	0.04	3.11%	0.0400	3.39%	0.0407	3.1%
Reproducibility	0.23	20.04%	0.05	4.03%	0.0515	4.37%	0.0531	4.1%
R&R	0.306	26.68%	0.09	7.13%	0.0914	7.76%	0.0938	7.2%
Product	1.104	96.37%	1.22	92.80%	1.086	92.24%	1.216	92.8%
Total	1.146		1.31	100.00%	1.178	100.00%	1.310	100.00%

Note that the columns for variance and % of total variance are very close for all three methods. So, the three methods, when using the variance, generate very similar results.

Obviously, the Average and Range approach of using the standard deviation gives significantly different results. This is simply because the standard deviations are not additive.

$$\sigma_t \neq \sigma_p + \sigma_{ms}$$

So, the % of variation column does not sum to 100. This makes it much more difficult to interpret the results. Why AIAG continues to include the Average and Range approach in their manual is beyond me. At a minimum, all they have to do is to square the results to convert the results to variances. **Bottom line: do not use the Average and Range method.**



But the bigger question is how to interpret the results. The criteria given by AIAG are just guidelines to consider (see Table 3). But if you apply them directly to the variance, most of the % of variance range is unacceptable – anything over 30% is not acceptable.

Dr. Wheeler’s EMP approach uses a completely different method. He classifies the test method as a First, Second, Third or Fourth Class monitor based on the intraclass correlation coefficient (ρ), which is the ratio of the part variance to the total variance:

$$\rho = \frac{\sigma_p^2}{\sigma_x^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2} = 1 - \frac{\sigma_e^2}{\sigma_x^2}$$

The subscripts are as follows: x = total variance, p = part variance, e = measurement system variance. So the intraclass correlation coefficient is also equal to 1 minus the % of variance due to the measurement system (the % R&R).

Table 8 shows how Dr. Wheeler suggests the results be interpreted. The last column in the table was added to show the %R&R value and AIAG guidelines for acceptability.

Table 8: Interpreting the EMP Results

ρ	Type of Monitor	Reduction of Process Signal	Chance of Detecting ± 3 Std. Error Shift	Ability to Track Process Improvements	% R&R/AIAG Guideline
0.8 to 1.0	First Class	Less than 10%	More than 99% with Rule 1	Up to Cp ₈₀	0 to 20% Acceptable to Marginal
0.5 to 0.8	Second Class	From 10% to 30%	More than 88% with Rule 1	Up to Cp ₅₀	20 to 50% Marginal to Unacceptable
0.2 to 0.5	Third Class	From 30% to 55%	More than 91% with Rules 1, 2, 3 and 4	Up to Cp ₂₀	50% to 80% Unacceptable
0.0 to 0.2	Fourth Class	More than 55%	Rapidly Vanishing	Unable to Track	80% to 100% Unacceptable

Note: table adapted from EMP III Evaluating the Measurement System, by Donald J. Wheeler, Copyright 2006 SPC Press.



This table was described in detail in our [previous publication](#). Please refer to that publication for more information. The first column lists the value of the Intraclass Correlation Coefficient. The second column lists whether it is a First Class, Second Class, Third Class or Fourth Class monitor – with “First” being the best. The rest of the table (except the last column I added) gives information about how much a reduction in process signal there is, the chance of detecting a major shift, and the ability to track process improvements.

You can see from the table, a First Class Monitor has a % R&R range of 0 to 20%. Under the AIAG guideline, a test method is acceptable if the % R&R is 10% or less. It is marginal in the range of 10 to 30%. The % R&R for our example data is about 7%. So, it is First Class Monitor and acceptable under AIAG guidelines. This means that there is less than a 10% reduction in a process signal, there is a better than 99% chance of detecting a point beyond the control limits (Rule 1) and that the measurement system will be able to track process improvements up to Cp₈₀. Cp₈₀ is calculated based on specifications and marks the point from the measurement system will move from a first class to a second class monitor. Rules 2 to 4 refer to the zone tests. Wow, a lot more information than the guidelines in Table 3.

But, if the result was 15% R&R, it would be marginal under AIAG guidelines but still be a First Class Monitor. It appears to me that the AIAG guidelines are unduly restrictive. A Third Class monitor would be unacceptable under the AIAG guidelines but from the table above still can be used to track a process.

So, what should you do to analyze a Gage R&R study? Use the ANOVA or EMP method to analyze the Gage R&R study. They will give similar results for % of variance. The EMP method does have some control charting built in which gives it the edge to me (see our [last month's publication](#)). Then interpret

the results using Dr. Wheeler's approach in Table 8. Rate the test method as a First, Second, Third or Fourth Class monitor and then use the information in the table to understand what that means.

Summary

In our early publications, I said that a precise measurement system was one where the measurement system is in statistical control and the % of variance due to the measurement system was less than 10% of the total variance. This would make it a Class One Monitor under Dr. Wheeler's system. So, was I wrong? Well, I too was probably too restrictive. I do believe that critical test methods should be monitored on an on-going basis by running a control and analyzing the results using an individuals control chart. The objective should also be continuous reduction in the measurement system variability.

But for Gage R&R studies, use the ANOVA or EMP methodology and interpret the results as a First, Second, Third or Fourth class monitor. Forget about the Average and Range method.

Quick Links

[Visit our home page](#)

[SPC for Excel Software](#)

[SPC Training](#)

[SPC Consulting](#)

[SPC Knowledge Base](#)

[Ordering Information](#)

Thanks so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese
BPI Consulting, LLC