Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597292

A Measurement System Analysis Approach for Hard-to-Repeat Events

Mahmoud Awad [a]; Tashi P. Erdmann [b]; Yassir Shanshal [c]; Bruce Barth [d]
[a] Alhosn University, Abu Dhabi, United Arab Emirates [b] Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA), Amsterdam, The Netherlands [c] Product Development, Case New Holland, Burr Ridge, Illinois [d] Powertrain Operations, Ford Motor Company, Dearborn, Michigan

Online Publication Date: 01 July 2009

To cite this Article Awad, Mahmoud, Erdmann, Tashi P., Shanshal, Yassir and Barth, Bruce(2009)'A Measurement System Analysis Approach for Hard-to-Repeat Events',Quality Engineering,21:3,300 — 305

To link to this Article: DOI: 10.1080/08982110902852344

URL: http://dx.doi.org/10.1080/08982110902852344

Quality Engineering, 21:300–305, 2009

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982110902852344

A Measurement System Analysis Approach for Hard-to-Repeat Events

Mahmoud Awad[1], **Tashi P. Erdmann**[2], **Yassir Shanshal**[3], **Bruce Barth**[4]

[1]Alhosn University, Abu Dhabi, United Arab Emirates
[2]Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA), Amsterdam, The Netherlands
[3]Product Development, Case New Holland, Burr Ridge, Illinois
[4]Powertrain Operations, Ford Motor Company, Dearborn, Michigan

ABSTRACT When measurements are nonreplicable, assessing the measurement system is a difficult task and cannot be done using the conventional gauge repeatability and reproducibility methods. This is due to the fact that the object measured is affected by the measurement and/or changes with time. This article describes an alternative method that can be used to study the repeatability and reproducibility of a measurement system in case the measured quantity varies over time. The essence of the method is that each part is measured simultaneously by multiple gauges.

KEYWORDS destructive measurement, gauge R&R, measurement system analysis, variance components

INTRODUCTION

Assessing the precision of a measurement system is a vital step that should be carried out before any design or process improvement effort. The method most commonly used to do this is a gauge repeatability and reproducibility (R&R) study, which aims to answer two main questions:

- How much of the total observed variability is due to real part-to-part variation and how much is due to random measurement error?
- What is the breakdown of the measurement variation? How much is due to repeatability versus reproducibility?

Repeatability is the extent to which measurement values are equal if measurements are repeated by the same appraiser, and *reproducibility* is the extent to which measurement values are equal if measurements are done by different appraisers.

In a standard gauge R&R study a number of appraisers measure a sample of parts several times. The results are analyzed using the random effects analysis of variance (ANOVA). The error variance represents the repeatability and the variance between appraisers the reproducibility.

There are many relevant situations in which the standard gauge R&R study described above is not applicable. For instance, if the true value of the measured characteristic of a particular part is not constant for each measurement, the error variance will not purely be caused by measurement error but partly by variation in the true value of that characteristic, and

Address correspondence to Mahmoud Awad, Industrial Engineering Department, Alhosn University, P.O. Box 38772, Abu Dhabi, 60453 United Arab Emirates. E-mail: mawad237@yahoo.com

therefore the measurement error will be overestimated. Or in case each part cannot be measured more than once by each operator, the error is confounded with the part-appraiser interaction effect. Consequently, the standard gauge R&R study as described by Automotive Industry Action Group (2002), assumes that for each part the true value of the measured quantity is constant over time, is not affected by the measurement, and can be measured at least twice by each appraiser under identical circumstances.

Unfortunately, practical situations often deviate from these assumptions. Measurements on a part for which the true value of the measured quantity changes over time or is affected by the measurement are called nonreplicable. Some examples of nonreplicable measurements in which the measured quantity changes over time are

1. The push or pull force needed to connect a plastic electrical connector.
2. The engine emissions or oil consumption using a radioactive tracer method.
3. The noise and vibration measurement during a certain duty cycle.
4. The strain measurement during crack propagation.
5. The brake disc temperature during operation.

Variables such as engine emissions or oil consumption, for example, physically change with every cycle, making a standard gauge R&R study fail. Oil consumption is traditionally measured by quantifying the amount of oil burned during combustion in terms of volume rate (gallon/h). Temperature, calibration, and surface finish are among other noise factors that contribute to oil consumption variability, which changes, from one cycle to another.

In these examples, incorrectly using a standard gauge R&R study produces overestimations of the gauge R&R% values, because part of the estimated measurement spread is not really measurement variation but variation in the true value of the measured quantity.

In the remainder of this article, the following definitions will be used:

- Parts are called stable if the true values of the measured characteristic do not change over time.
- Random measurement error is called homogeneous if its probability distribution does not depend on the part and the time.

If parts are stable and the random measurements error is homogeneous, a standard gauge R&R can be carried out. In the five examples given in the introduction, however, the parts are not stable.

With the presence of run-to-run variation over time, an alternative method is needed that does not require the measured quantity to be constant over time. De Mast and Trip (2005) describe methods to deal with this issue, such as using alternative parts, using an alternative measurement system, or modeling the variation in the parts. Other examples of gauge R&R studies for nonreplicable measurements are Phillips et al. (1997), and Bergeret et al. (2001).

This article proposes a method as an addition to the approaches listed in De Mast and Trip (2005), which can be used to handle the problem of instable parts under the assumption that the measurement errors are homogeneous and that it is possible to obtain simultaneous measurements by multiple gauges. In the next section we first define the standard gauge R&R study in more detail. Then in the subsequent section we introduce a method for instable parts. The method will be illustrated by an example involving measurements of brake disc temperature.

STANDARD GAUGE R&R STUDY

In a standard gauge R&R study I different parts are measured K times by each of J different appraisers (Burdick et al., 2005). The tth measurement of part i by appraiser j has value y_{ijt}. The measurement values are then modeled by the following random effects model:

$$y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijt} \qquad [1]$$

The constant μ is the overall mean of the measurements, α_i is the part effect, β_j is the appraiser effect, and $(\alpha\beta)_{ij}$ is the part-appraiser interaction. The effects α_i, β_j, $(\alpha\beta)_{ij}$, and ε_{ijt} are independent random variables that follow a normal distribution with mean 0 and variances σ_α^2, σ_β^2, $\sigma_{\alpha\beta}^2$, and σ^2, respectively. The model is analyzed using the random effects analysis of variance (ANOVA) method, and the variances of the effects are estimated by taking appropriate linear combinations of the mean squares (Montgomery, 2005). Repeatability is represented by the standard deviation of the error term (σ), reproducibility by the standard deviation of the appraiser effect plus the part-appraiser interaction ($\sqrt{\sigma_\beta^2 + \sigma_{\alpha\beta}^2}$), and part-to-part variation by the standard deviation of the part effect (σ_α). The gauge R&R% value is calculated as:

$$\text{Gauge R\&R\%} = \frac{\sqrt{\sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2}}{\sqrt{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2}} \qquad [2]$$

TABLE 1 Expected Mean Squares for the Mixed Model

Term	Expected mean square
α_i	$\sigma^2 + T\sigma_{\alpha\beta}^2 + J\sigma_{\alpha\gamma}^2 + JT\sigma_\alpha^2$
β_j	$\sigma^2 + T\sigma_{\alpha\beta}^2 + \left(IT\sum_{j=1}^{J}\beta_j^2\right)/(J-1)$
γ_t	$\sigma^2 + J\sigma_{\alpha\gamma}^2 + \left(IJ\sum_{t=1}^{T}\gamma_t^2\right)/(T-1)$
$(\alpha\beta)_{ij}$	$\sigma^2 + T\sigma_{\alpha\beta}^2$
$(\alpha\gamma)_{it}$	$\sigma^2 + J\sigma_{\alpha\gamma}^2$
$(\beta\gamma)_{jt}$	$\sigma^2 + \left(I\sum_{j=1}^{J}\sum_{t=1}^{T}(\beta\gamma)_{jt}^2\right)/((J-1)(T-1))$
ε_{ijt}	σ^2

GAUGE R&R STUDY FOR INSTABLE PARTS

Now we are ready to introduce the gauge R&R study for instable parts. The premises for the proposed method are

- The parts are instable.
- The random measurement error is homogeneous.
- It is possible to measure with two or more gauges simultaneously.
- There is no measurement variability which is caused by appraisers; that is, the measurement error is equal regardless of which appraiser operates the gauge.

The idea is to take a random sample of I parts that represent the population of parts and then measure each part simultaneously by J gauges at T different times. The measurement of part i by gauge j at time t is written as y_{ijt}, $i=1,\ldots,I$, $j=1,\ldots,J$, and $t=1,\ldots,T$. For the analysis the three-factor mixed model [3] is used, with *part* as a random factor and *gauge* and *time* as fixed factors. The gauges are treated as a fixed factor here but may also be treated as a random factor if one is interested in the effect of a large population of gauges of which the J gauges are only a small sample. Time is a fixed factor, because the moments in time that are used are not a random sample but a fixed sequence.

$$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + (\alpha\beta)_{ij} + (\alpha\gamma)_{it} + (\beta\gamma)_{jt} + \varepsilon_{ijt} \qquad [3]$$

Here β_j, γ_t, and $(\beta\gamma)_{jt}$ are fixed parameters that sum to zero over both j and t, representing gauge and time effects and their interaction. The parts effect α_i and its interactions with the other factors $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{it}$ are independent random variables that are each normally and independently distributed with mean equal to 0 and variances σ_α^2, $\sigma_{\alpha\beta}^2$, $\sigma_{\alpha\gamma}^2$, respectively. The error term has mean 0 and variance σ^2. The main gauge effect and gauge-time interaction are indicators of systematic differences between the gauges, and the variance of the error term and the part-gauge interaction indicate random measurement error. Repeatability is represented by σ and gauge reproducibility by $\sigma_{\alpha\beta}$.

The model is analyzed as a mixed model ANOVA. Montgomery (2005) provides rules for deriving expected mean squares in terms of the variance components and fixed effects parameters of mixed models. The expected mean squares for each term in model [3] are given in Table 1.

The variance components are estimated by equating the observed mean squares to their expected values and solving for each variance component.

BRAKE DISC TEMPERATURE EXAMPLE

We will illustrate the procedure with an example. The temperature of five different brake discs was measured over time by two different infrared temperature gauges, which collected temperature readings simultaneously. The study was part of a project aimed at resolving brake disc overheating at high speed on construction equipment. The measurements were conducted as follows:

1. The machine was started after at least one hour cool-down period.
2. The wheels of the machine were run freely at maximum speed for 8 minutes.
3. During the operation, two infrared guns were aimed at two different but adjacent points on the disc.

TABLE 2 Brake Disc Example Data

	Time	Gauge 1	Gauge 2
Part 1	1	35	31
	2	43	40
	3	51	48
	4	69	56
	5	91	83
	6	103	91
	7	117	103
	8	138	119
Part 2	1	57	51
	2	79	67
	3	125	98
	4	156	142
	5	202	187
	6	236	214
	7	237	214
	8	275	251
Part 3	1	85	84
	2	90	93
	3	95	99
	4	101	109
	5	123	131
	6	137	142
	7	156	165
	8	179	183
Part 4	1	101	102
	2	118	112
	3	103	107
	4	128	126
	5	149	151
	6	169	177
	7	194	200
	8	200	203
Part 5	1	103	101
	2	117	112
	3	105	107
	4	131	129
	5	153	149
	6	172	175
	7	195	199
	8	200	202

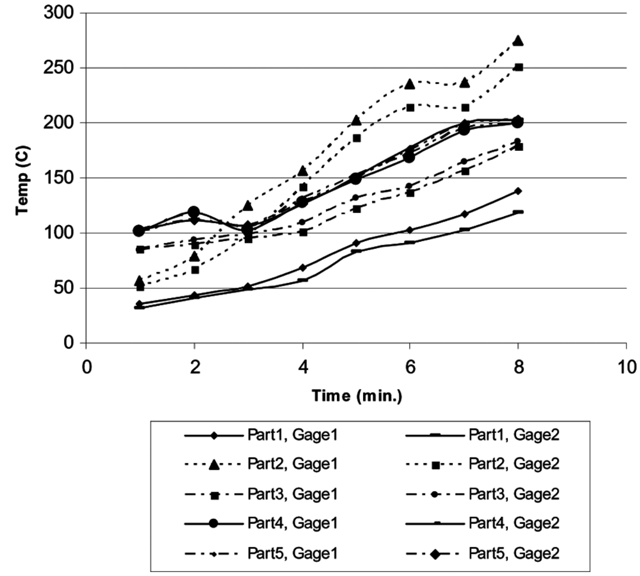

FIGURE 1 Brake disc temperature measurements.

4. The temperature was measured by both gauges simultaneously every minute.

The 8-minute period was selected so no permanent damage to the disc would occur.

Table 2 shows the resulting temperature data. Figure 1 is a graphical representation of the measurements.

The mean temperatures are significantly different for different times and parts, so the parts are instable. Variances were also compared and no statistically significant differences were found, supporting the assumption that the measurement error is homogeneous. Variability caused by appraisers is not an issue in this example, because operators have a negligible influence on the measurement outcomes. Considering all this, model [3] seems a suitable model to assess repeatability and gauge reproducibility of this measurement system.

TABLE 3 ANOVA for Model [3]

Source	DF	SS	MS	F	P
Part	4	72,037.6	18,052.9	16.34	0.000
Gauge	1	340.3	340.3	0.96	0.382
Time	7	140,478.1	20,104.0	26.25	0.000
Part × Gauge	4	1,415.0	353.7	23.42	0.000
Part × Time	28	21,446.9	766.0	50.70	0.000
Gauge × Time	7	29.2	4.2	0.28	0.958
Error	28	423.0	15.1		
Total	79	236,590.0			

TABLE 4 ANOVA for Model [4]

Source	DF	SS	MS	F	P
Part	4	72,207.6	18,051.9	16.31	0.000
Gauge	1	340.3	340.3	0.96	0.382
Time	7	140,728.1	20,104.0	26.25	0.000
Part × Gauge	4	1,415.0	353.7	27.38	0.000
Part × Time	28	21,446.9	766.0	59.29	0.000
Error	35	452.2	12.9		
Total	79	236,590.0			

The model is analyzed by three-factor mixed model ANOVA, and the results are shown in Table 3.

The F-test has different interpretations for fixed effects and random effects. For fixed effects it tests whether the fixed effect is different from zero, whereas for random effects it tests whether the random effect has positive variance. Recall that the gauge and time effects are fixed and the part effect is random.

Based on its P-value, the gauge-time interaction is not significantly different from zero. This is a pleasing result, because a gauge-time interaction would indicate a systematic measurement error depending on time. The gauge-time interaction will be assumed to equal zero and left out of the model, resulting in model [4].

$$y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t + (\alpha\beta)_{ij} + (\alpha\gamma)_{it} + \varepsilon_{ijt} \quad [4]$$

Table 4 shows the ANOVA results for model [4].

There is no evidence of a systematic difference between the gauges, because the main gauge effect is not significantly different from zero, but the random part-gauge interaction has a significant positive variance. This implies that for any given brake disc, on average one gauge will measure it differently from the other gauge. However, for some parts gauge 1 measures the higher temperature and

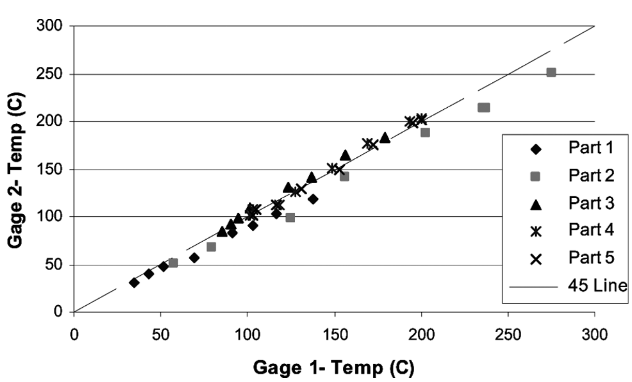

FIGURE 2 Scatterplots by part.

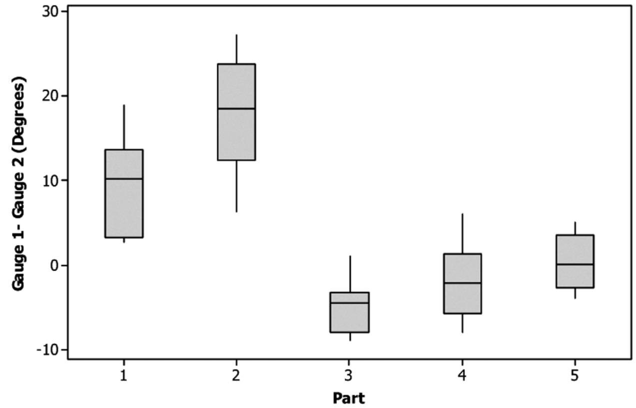

FIGURE 3 Box plot of gauge 1–gauge 2.

for some parts gauge 2 measures the higher temperature, averaging out over all parts. This might be explained by the way the gauges are placed on the brake disc. Possibly, the factor causing the differences in measurements of the two gauges is not really the gauge itself but rather the placement of the gauge on the brake disc.

The conclusions above can be graphically illustrated. In the scatterplot of the measurements of gauge 1 versus gauge 2 in Figure 2 the plotted points seem to fall around the 45° line on average, supporting the hypothesis that there is no systematic difference between the gauges. Some parts, however, are measured differently by the two gauges. The temperature of part 2, for example, is measured higher by gauge 1 than by gauge 2. This illustrates the conclusion that there is no main effect but a significant part-gauge interaction.

Figure 3 shows box plots of the paired differences between the simultaneous measurements of the gauges for each part. Because the temperature readings were taken simultaneously, the depicted differences in readings are due to the gauges only. When measuring parts 1 and 2, the measurements by

TABLE 5 Variance Contributions for the Disc Brake Example

Source	Variance	% Contribution
Part	1,059.07	71.03
Gauge		
Time		
Part × Gauge	42.60	2.86
Part × Time	376.52	25.25
Error	12.92	0.87
Total	1,491.11	100.00

TABLE 6 Gauge R&R

Source	Std. dev.	% Study var
Total Gauge R&R	7.45	19.30
Repeatability	3.59	9.31
Reproducibility	6.53	16.90
Total variation	38.61	100.00

gauge 1 were higher on average than those by gauge 2, whereas when measuring parts 3, 4, and 5, it was the other way round, once again backing the findings of no main effect but significant part-gauge interaction.

Table 5 shows the variance estimates of all terms along with the percentage contribution of variances relative to the total variance. Note that the main effects of gauge and time have no variance because they are fixed effects. Table 6 summarizes the standard deviations representing gauge repeatability σ and gauge reproducibility $\sigma_{\alpha\beta}$.

The measurement spread due to repeatability and reproducibility (part-gauge interaction) is 19.30% of the total observed variation, which is quite a substantial part. Of course this gauge R&R% only has practical meaning if the parts are a representative sample of all products that are normally used during production. Furthermore, the sample size of five parts used in this case study is rather small. It would be safer to compare the measurement spread to an estimate of the total variation based on historical data.

CONCLUSION

In practice sometimes the assumption that the measured characteristic of a part is constant over time does not hold. The standard gauge R&R methods cannot handle such systems. In this article a measurement systems analysis approach for such hard-to-repeat measurements is introduced. A new experimental setup to assess gauge repeatability and reproducibility is proposed along with analysis tools.

ABOUT THE AUTHOR

Mahmoud Awad received his BS degree in Mechanical Engineering from Jordan University of Science and Technology (Irbid, Jordan) and received his MS degree in Industrial Engineering from University of Jordan (Amman, Jordan). And his PhD in Industrial Engineering from Wayne State University (USA). He worked for many years at Ford Motor Company as a quality and reliability engineer. He also worked as a Master Black Belt at case New Holland and taught DFSS as part of his job. He is currently working as an associate professor at Industrial engineering department at Alhosn University, UAE.

Tashi Erdmann, MSc, is a consultant and researcher at the Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA). He works on a PhD project about Measurement Systems Analysis, and he teaches courses in Lean Six Sigma and coaches and evaluates project managers of quality improvement projects. In 2006 he obtained his MSc degree (cum laude) in Econometrics at the University of Amsterdam. After his education he worked as mathematics teacher for two years at Keera Pat International School in Bangkok, and then joined IBIS UvA in 2008

Yassir Shanshal received his BS & MS degrees in Mechanical Engineering from Southern Illinois University at Carbondale (Carbondale, Illinois) and MS in Product Development from the University of Detroit Mercy (Detroit, Michigan). He currently works as a six sigma director at Case New Holland. He is a certified 6 sigma master black belt level 2. He taught 6-Sigma black belt and DFSS at Ford Motor Company. He is also an adjunct professor at the University of Detroit Mercy. He is currently the 6-Sigma director at Case New Holland

Bruce Barth is a certified Six Sigma black belt at Ford Motor Company. He holds a Bachelor of Science degree in Metallurgy and a Master's of Science degree in Mechanical Engineering from University of Michigan. He is a senior member of ASQ, a Certified Quality Engineer, and a Certified Six Sigma Black-Belt.

REFERENCES

Automotive Industry Action Group. (2002). *Measurement Systems Analysis; Reference Manual* (3rd ed.). Detroit, MI: Automotive Industry Action Group.

Bergeret, F., Maubert, S., Sourd, P., Puel, F. (2001). Improving and applying destructive gauge capability. *Quality Engineering*, 14(1):59–66.

Burdick, R. K., Borror, C. M., Montgomery, D. C. (2003). A review of methods for measurement systems capability analysis. *Journal of Quality Technology*, 35(4):342–354.

De Mast, J., Trip, A. (2005). Gauge R&R studies for destructive measurements. *Journal of Quality Technology*, 37(1):40–49.

Montgomery, D. C. (2005). *Design and Analysis of Experiments* (6th ed.). New York: Wiley.

Phillips, A. R., Jeffries, R., Schneider, J., Frankoski, S. P. (1997). Using repeatability and reproducibility studies to evaluate a destructive test method. *Quality Engineering*, 10(2):283–290.