

2015

Being Smart about Parts

Nathaniel Stevens

University of San Francisco, ntstevens@usfca.edu

S H. Steiner

R.J. MacKay

Follow this and additional works at: <http://repository.usfca.edu/math>

 Part of the [Mathematics Commons](#)

Recommended Citation

Stevens, N.T., Steiner, S.H., & MacKay, R.J. (2015). Being smart about parts. *Quality Progress*, Volume 48, Issue 3, 1 March 2015, Pages 32-37.

This Article is brought to you for free and open access by the College of Arts and Sciences at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Mathematics by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.

Being Smart About Parts

Performing gage R&R studies with **baseline information**

by Nathaniel T. Stevens, Stefan H. Steiner and R. Jock MacKay

IT WAS LUNCHTIME, I was hungry, and so of course the phone rang.

“Hey, it’s John from the foundry. How are you? We haven’t talked in a while.”

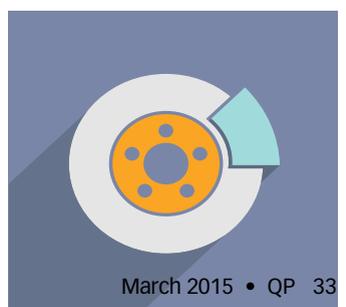
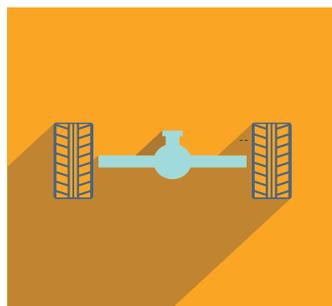
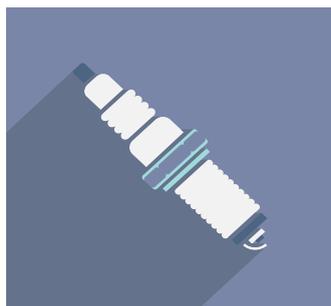
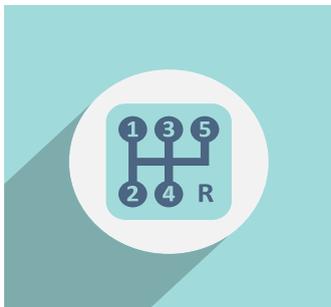
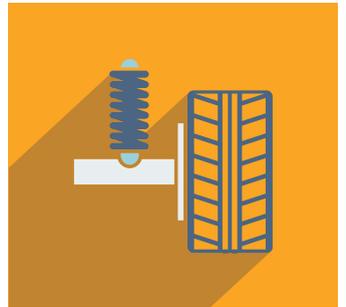
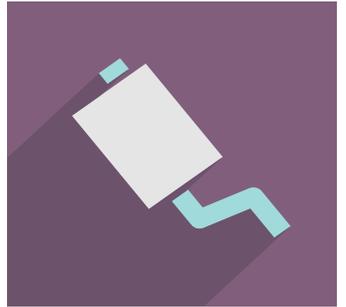
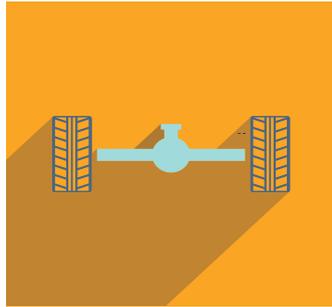
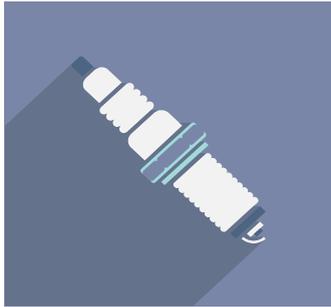
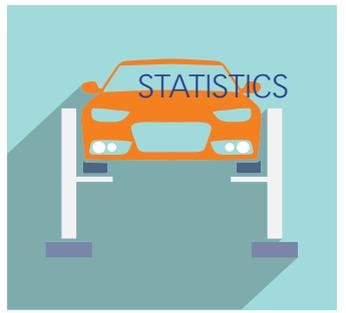
John is a process engineer in a casting and machining plant that is a supplier to automotive original equipment manufacturers. He is prone to encountering statistical emergencies.

“I’m good. How are you?” I asked.

“Fine,” he said. “I’ve got a little problem for you. I was reading QP the other day and found an article¹ that said we should do our gage R&Rs (repeatability and reproducibility studies) with only five parts instead of 10. So this morning, that’s what we did: We measured five parts, six times each.”

In 50 Words Or Less

- Gage repeatability and reproducibility (R&R) studies can be used to assess the adequacy of a measurement system.
- Using freely available production data can drastically improve the estimation of R&R and hence reduce the number of parts and measurements needed for the assessment study.



“How many operators?” I asked.

“Only one. This gage is automated so operators don’t have any effect,” he answered. “Anyway, I just did the analysis and the R&R was 34%. You remember R&R is supposed to be less than 30% for the gage to be acceptable. Last time we did the assessment in the standard way, and the gage was fine. So do you think we made a mistake using only five parts? Have you got any suggestions?”

Because I was missing my lunch, I decided to have a little fun. Hungry statisticians have an odd sense of humor.

“Why not do the study over? Maybe you’ll get lucky,” I suggested.

“But should I use five or 10 parts? If I use only five parts, I’ll get the same answer, won’t I? And what do you mean by lucky?” he asked suspiciously.

“Well, if you repeat the five-part study, the R&R could easily change by 10 percentage points in either direction. So if you’re lucky, the R&R in the new study will be less than 30%, and the gage is good to go,” I said, answering the last question first.

“Ten percent change? That can’t be right. You statisticians are always causing trouble. Why can’t you be more like an engineer and just give me a method that works? With my luck, I’ll do the study over and the gage will look even worse. Give me something certain,” John complained.

“Ten percentage point difference,” I corrected him. “Here’s an idea: Do the five-part study over, but carefully select two small castings and three large castings to measure. This will inflate the total variation and make the R&R smaller,” I suggested facetiously.

“We can’t do that. You’re supposed to select the parts to represent current production,” he replied.

“Just testing to see whether you’re paying attention. How about putting each part in the fixture, and just hit the switch to measure it six times. That should reduce the repeatability a lot,” I said.

“No way! You spent a long time in that measurement systems short course convincing us that the fixturing was part of the measurement system. I’d never get away with that. Stop fooling around and give me some useful answers. I’m busy, you know!” John said.

“OK. What is the gage used for?” I asked, knowing the fun was over.

“Process control. We measure every sixth part and

make adjustments occasionally. The data are recorded and plotted automatically. How can that help?” he questioned.

“Can you email me with the past 100 measurements, plus the data from this morning’s R&R study? I think we can improve your estimate of the R&R without any extra work for you. After I get the data, I’ll get back to you.”

“OK. I’ll send it right away,” John answered. “But I don’t understand what statistical trick you are trying to play. Don’t forget I might need to explain what you’re doing to my manager and our auditor. Can you get back to me this afternoon? Otherwise, I think we’ll repeat the entire study with 10 parts so the gage will be acceptable.”

“Fine,” I said, realizing that I had just agreed to missing lunch. “Always nice talking with you.”

Exchange of emails

John’s email arrived.

“Thanks for your help with this. Here is the data from that R&R study and the last 100 measurements as you requested. All measurements are taken from nominal, so some are negative. As always, we’re in a rush. We were supposed to get this done last week. Talk with you soon.” See Tables 1A and 1B for the data included in the email.

I replied shortly after in an email:

“Here’s a quick analysis. I think you’ll like the result. Remember R&R is the ratio of the repeatability variation divided by the total process variation. We estimate the repeatability variation from the five-part study as you did this morning. Here is part of the output using the gage R&R study (crossed) menu in Minitab² (see Table 2).

“So the estimate of the repeatability standard deviation is $\sqrt{0.06446}=0.254$. We estimate the overall process variability using the 100 parts to represent current production, just as you said on the phone. The standard deviation of these baseline measurements is 0.963. The estimated R&R is $\frac{0.254}{0.963} \times 100 = 26\%$ so the measurement system is acceptable, according to your standard. I assume you made no adjustments to the process while measuring the 100 parts.

I’ll send a follow-up report with more explanation later. I’m guessing that now you won’t have to miss your golf game (just a little joke).”

'You statisticians are **always causing trouble**. Why can't you be more like an engineer and **just give me a method that works?**'

Reporting out

A follow-up report, titled "Planning and Analysis of Gage R&R Studies With Baseline Information," was written. The executive summary and recommendations included:

- Standard gage R&R studies can be improved by including freely available production data.
- Gage studies produce an estimate of the true R&R subject to sampling error.
- We recommend using 100 baseline measurements to represent current production and an assessment plan with three parts and at least 10 repeated measurements.
- Training is available.

The report continued. "Here we assume that there are no operator effects for the measurement system being assessed."³

1. The gage repeatability γ is the ratio of two standard deviations. That is, $\gamma = \frac{\sigma_r}{\sigma_t}$ in which σ_r represents the variability in the output if the same part is measured repeatedly, and σ_t represents the variability in the output when several parts are measured once. The goal of the assessment study is to estimate γ by estimating σ_r and σ_t . The better we can estimate these two standard deviations, the better we can estimate the gage repeatability γ .
2. Suppose you have available baseline data from recent routine use of the measurement system. We

Repeatability and reproducibility study data / TABLE 1A

Part	Repeat number					
	1	2	3	4	5	6
1	-0.477	-0.5295	-0.091	-0.3401	-0.6264	-0.2063
2	-0.1639	-0.6838	-0.492	-0.3063	0.0164	-0.0462
3	-1.3168	-0.8991	-1.2699	-1.5244	-1.034	-1.0203
4	0.6726	0.9742	0.4388	0.6442	0.9991	0.7584
5	-0.5716	-0.5688	-1.3187	-0.5182	-0.5771	-1.0161

Baseline data / TABLE 1B

-0.6361	0.9101	0.8607	0.5828	-0.4783	-0.0749	-0.6322	0.2489	-0.8901	0.241
-0.2871	-0.1864	-0.0648	0.1072	0.4601	-0.9693	-0.4432	-0.6032	0.9359	-0.1582
0.3547	-0.5578	0.2224	-2.1455	-1.3263	0.1094	0.0774	-0.8912	-0.0025	1.0559
-1.1014	-0.0912	0.9839	1.2873	-0.9821	-0.5467	-0.1289	0.291	-0.9965	-0.8421
-0.7664	-1.3994	-1.3912	0.3893	2.0291	-1.08	-0.9206	1.9959	0.3852	-0.5285
0.8153	-0.3306	0.4611	0.1425	-1.9489	-1.1618	0.4667	-0.2229	0.7991	-2.3185
-1.486	-0.9235	0.7156	-0.1898	-1.4444	1.5989	-1.0595	-0.6119	-0.7437	-0.2346
-2.6045	1.8591	-1.1152	-1.0762	-0.324	-0.0645	0.2632	-0.9603	-0.2483	-0.1583
-1.9074	0.5709	0.5294	-0.2713	0.7938	-0.4963	1.0401	-2.7395	-0.4853	-1.2344
-0.1868	0.8008	-0.8258	0.4118	-0.5721	-0.2509	0.154	1.1644	-1.6386	1.3121

can get an estimate of σ_t by finding the standard deviation of these measurements, as in the example. We call this the simple method. This estimate is much better than that from the usual estimate from a standard five or 10-part plan.

3. We can do even better by combining the information on σ_t from the baseline and the standard plan, although the calculations become more obscure.^{4,5}
4. We use the repeated measurements on the parts to estimate σ_r . The five-part, six-repeated measurements plan gives the same information about σ_r as it would have had we measured one part 31 times. So, we can use fewer parts in the assessment plan when we use the baseline data to estimate σ_t . We recommend using two or three parts in the gage R&R study and measuring each part 15 or 10 times, respectively.
5. We combine the estimates of σ_t (from point No. 2 or No. 3 earlier in this list) with the estimate of σ_r (from point 4) to get the estimate of γ . If we use the simple method, we can substitute the estimate of σ_t as a known historical value in the options of the Minitab⁶ routine gage R&R study (crossed) to get the results.
6. The five-part study, as recommended by authors in the QP article you mentioned,⁷ has too few parts to produce a precise estimate of σ_t and hence of γ . For your data, the standard error of the estimate (a measure of the precision) is 0.106. This is why I said the estimate could easily vary by 10 percentage points if the five-part study was repeated. When the baseline data are included, the standard error of the estimate is reduced to 0.044.

7. There are substantial benefits to including baseline information in the analysis. Suppose we regard the Automotive Industry Action Group⁸ threshold of 0.3 as a decision limit. If we obtain a value of $\hat{\gamma}$ (we put a “hat” on γ to indicate we get an estimate, not the actual value of γ), which is less than 0.3, we will say the measurement system is acceptable. If $\hat{\gamma}$ is larger than 0.3, we will say the measurement system is unacceptable. The perfect plan would produce an estimate $\hat{\gamma}$ that is less than 0.3 whenever γ is less than 0.3 and greater than 0.3 when γ is greater than 0.3.

In line with the earlier QP article,⁹ we used simulation to investigate the probability of accepting the measurement system (that is, the probability that $\hat{\gamma} < 0.3$). We generated data from a baseline and various assessment plans for varying values of γ . For each γ , we generated 10,000 data sets, found $\hat{\gamma}$ and calculated the proportion of time that $\hat{\gamma} < 0.3$. We used three different plans with two methods of analysis for the plan that uses the baseline data:

- Ten parts, six repeated measurements per part.¹⁰
- Five parts, six repeated measurements per part.¹¹
- Three parts, 10 repeated measurements per part + 100 baseline measurements, estimate of σ_t based solely on the baseline (simple method).
- Three parts, 10 repeated measurements per part + 100 baseline measurements, estimate of σ_t using the baseline and assessment study data (maximum likelihood method).

Figure 1 shows the results. We see as the overall number of parts increases, the probability the measurement system is acceptable approaches that of the perfect plan. The best plan is one in which baseline data are included. We also note the superiority of the three-part plan with 100 baseline parts over the 10-part plan, which has twice the number of repeated measurements. Using the freely available baseline data means we can carry out a smaller and cheaper assessment plan with better results.

Figure 1 also demonstrates the difference between the simple and maximum likelihood methods in the plan using a baseline.¹²

The maximum likelihood method is better when γ is near the 0.3 threshold, but given the ease and good properties of the simple method, it is a viable alternative.

One-way ANOVA table / TABLE 2

Source	DF	SS	MS	F	P
Part	4	12.3801	3.09502	48.0119	0.000
Repeatability	25	1.6116	0.06446		
Total	29	13.9917			

ANOVA = analysis of variance DF = degrees of freedom

SS = sum of squares MS = mean square

F = (found variation of the group averages) /
(expected variation of the group averages)

P = p-value. The probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming the null hypothesis is true.

Use what's freely available

Many practitioners can sympathize with the problem John faced. When conducting a measurement system assessment study, we want as precise an estimate of the true R&R as possible so we can correctly decide whether the measurement system is acceptable. We are, however, faced with cost and time constraints that restrict the number of parts and repeated measurements that can be used in the study.¹³

By incorporating freely available production measurements (baseline data), we can reduce the number of parts in the study to two or three and still obtain a better estimate of the R&R than we would have otherwise.

To avoid bias, we must ensure the baseline data reflect the current manufacturing and measurement processes. In the analysis, we assume the process and measurement system is stable for the time interval that covers both the baseline data and the gage R&R study. To address this issue, we suggest checking for stability in the baseline data.

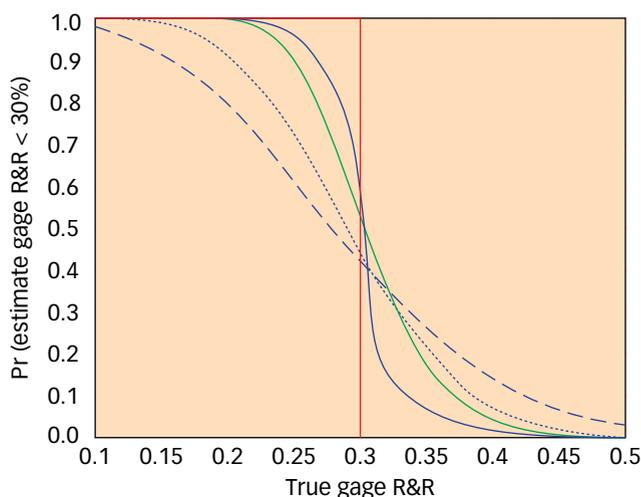
We have focused on estimating the R&R metric. If interest lies in estimating other measurement system criteria, such as the precision-to-tolerance ratio,¹⁴ which do not involve the overall variation σ_t , there is no value in the baseline data.

The example in this article uses 100 baseline measurements. In another article, Nathaniel T. Stevens and co-authors demonstrate that substantial improvements in precision can be realized with as few as 60 baseline measurements, and even better gains are realized for a larger number of parts. They also show that incorporating baseline information is beneficial when there are multiple operators.¹⁵ **QP**

REFERENCES AND NOTE

1. Victor M. Aguirre-Torres and Maria Teresa Lopez-Alvarez, "Parts of the Process," *Quality Progress*, June 2013, pp. 34-39.
2. Minitab Inc., Minitab Statistical Software, release 16, 2013.
3. Nathaniel T. Stevens, Stefan H. Steiner, Ryan P. Browne and R. Jock MacKay, "Gauge R&R Studies That Incorporate Baseline Information," *IIE Transactions*, 2013, Vol. 45, No. 11, pp. 1,166-1,175.
4. *Ibid.*
5. For more information on planning and analyzing a gage R&R study that incorporates baseline data, software with instructions to perform the calculations is available at www.bisrg.uwaterloo.ca.
6. Minitab Inc., Minitab Statistical Software, see reference 2.
7. Aguirre-Torres, "Parts of the Process," see reference 1.
8. Automotive Industry Action Group (AIAG), *Measurement Systems Analysis*, fourth edition, AIAG, 2010.
9. Aguirre-Torres, "Parts of the Process," see reference 1.
10. AIAG, *Measurement Systems Analysis*, see reference 8.
11. Aguirre-Torres, "Parts of the Process," see reference 1.
12. Stefan H. Steiner and R. Jock MacKay, *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes*, ASQ Quality Press, 2005.

Different R&R sizes with various study designs / FIGURE 1



Note: This figure represents the probability of accepting a measurement system by repeatability and reproducibility (R&R) size with various study designs.

- The dotted blue line represents 10 parts with six repeated measurements.
- The broken blue line represents five parts with six repeated measurements.
- The solid blue line represents three parts with 10 repeated measurements + 100 baseline parts, maximum likelihood method.
- The solid green line represents three parts with 10 repeated measurements + 100 baseline parts, simple method.
- The solid red line represents the perfect plan.

13. Aguirre-Torres, "Parts of the Process," see reference 1.

14. AIAG, *Measurement Systems Analysis*, see reference 8.

15. Stevens, "Gauge R&R Studies That Incorporate Baseline Information," see reference 3.



NATHANIEL T. STEVENS is a research assistant at the University of Waterloo in Ontario. He holds a doctorate in statistics from the University of Waterloo. He is an ASQ member.



STEFAN H. STEINER is a professor at the University of Waterloo. He holds a doctorate in management science/systems from McMaster University in Hamilton, Ontario. He is co-author of *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes* (ASQ Quality Press, 2005). He is an ASQ fellow.



R. JOCK MACKAY is a retired associate professor from University of Waterloo. He holds a doctorate in statistics from the University of Toronto. He is co-author of *Statistical Engineering: An Algorithm for Reducing Variation in Manufacturing Processes* (ASQ Quality Press, 2005). He is an ASQ member.