Gage Capability Case Study in a Quality Control Course

Mr. Shannon K. Sweeney, Pennsylvania State University, Erie

Mr. Shannon Sweeney is an Associate Professor of Engineering at Penn State Erie, The Behrend College. He received an A.S. in Drafting and Design Engineering Technology and a B.S. in Mechanical Engineering from West Virginia Institute of Technology in 1981 and 1985 respectively. Mr. Sweeney also received a M.S. in Mechanical Engineering in 1996 from Case Western Reserve University. He is certified as a Quality Engineer by the American Society for Quality and is a Licensed Professional Engineer in the Commonwealth of Pennsylvania. Mr. Sweeney began his career as a Quality Engineer for Lord Corporation. He worked there eleven years with the last eight years being as a Product Designer of vibration control products. He joined the faculty at Penn State Behrend in 1996 where he teaches a variety of engineering mechanics courses. His research interests are materials testing, industrial statistics, and vibration analysis.

Gage Capability Case Study in a Quality Control Course

Shannon Sweeney, P.E.
Associate Professor of Engineering
Penn State Erie, The Behrend College

Abstract

Gage capability studies have been widely used in industrial practices for over three decades. In that time, industry practices for studying gage capability have evolved substantially. The early practice of studying gage capability was called the Averages and Ranges Method, or Tabular Method. The later practice utilizes the Design of Experiment Method. Although the Tabular Method is becoming obsolete in industrial practices, it has an extensive history. Consequently, students in a Quality Control and Quality Improvement elective course for seniors are given an assignment to conduct a gage capability study by both the Tabular Method and the Design of Experiment Method. They use the same set of repeated measurement data for each method. The primary goals are for students to be able to conduct both methods and to compare the results for a test case. Secondary goals are to investigate the sources of variation in the measurement process and to seek improvements to the measurement process.

Introduction

Gage capability studies are necessary for any organization to evaluate variation in their measurement processes.[1] They have been used to evaluate variation in the measurement process for anything from linear dimensions taken by a micrometer to hardness of metals obtained by indentation after thermal processing[2] to imbalance of rotating components.[3] Gage capability studies are often required by industrial customers during their quality audits of suppliers and may also be a part of an organization’s ISO certification efforts. Although capability studies can be applied to both variables gaging and attribute gaging, this work will focus on variables gaging.

Consequently, gage capability is a topic in a Quality Control and Quality Improvement elective course at Penn State Behrend. Since the course is for seniors in many engineering and technology majors and is not a metrology course, the purpose is not to train on the use of any particular measurement system but to convey and quantify the concepts of variation, precision, and capability and to conduct the methods of study so that they can be used with a variety of measurement systems.

The heart of the early Tabular Method of gage capability study requires the familiar estimation of standard deviation from range values. Although the Tabular Method is simple to use, it has some disadvantages[4]. First, the range estimation of standard deviation is an approximation and is sometimes inefficient. Second, it is sometimes desirable to obtain confidence intervals on the sources of measurement variation[5], and that is not easily accomplished with the Tabular Method. Third, a gage capability study is truly a designed experiment so the principles of good experimental analysis should be applied. It is noteworthy the D. Montgomery, a leading author in the field of quality control, has removed the Tabular Method from recent editions of his textbook(s).

The later Design of Experiment Method applies those good principles of experimental analysis as it requires an Analysis of Variance (ANOVA) with two factors and random effects, and it is becoming the method of choice for progressive quality programs. It is important to note that, although the early Tabular Method is becoming obsolete, many gage capability studies done by the early Tabular Method are archived that way and are still presented to industrial customers during quality audits or to satisfy ISO certification requirements.

It is noteworthy that Wheeler[6] identifies shortcomings even with the later Design of Experiment method. He suggests alternatives to repair the deficiencies in what would be a third method of gage capability study. This is sometimes referred to as the Evaluating the Measurement Process (EMP) Method but it is not currently taught in the Quality Control and Quality Improvement course.

Each method seeks to partition the variance of the measurement process into the respective sources of measurement variation. The first major source is equipment variation, *EV*, often call repeatability (or lack thereof). The second major source is appraiser variation, *AV*, often called reproducibility (or lack thereof). Gage capability studies are sometimes referred to as gage R&R studies.

Equipment variation is often a function of the quality of the materials and the tolerances used in the making of the measurement system itself. Appraiser variation is often a function of the experience and/or training of the operators using the measurement system. Examples of efforts to reduce appraiser variation include clutch mechanisms in the thimbles of micrometers and touch probes on coordinate measuring machines.

Background

Students in the Penn State Behrend course are lectured on the definitions of accuracy and precision and are shown that gage capability studies quantify the precision of a measurement system and usually make no effort to quantify accuracy. Gage capability studies are an integral part of an organization's overall quality control efforts. Gage capability studies are separate from, but complement, other quality control efforts such as process control charts, process capability studies and gage calibration procedures (for accuracy).

Gage capability studies have been introduced in other upper education courses. M. Kozak[7] has students conduct gage capability studies in four activities. However, the activities are done in a metrology course where the focus is on measurement improvement and not on the methods of studying gage capability. D. Timmer and M. Gonzalez[8] introduce some innovative and novel pedagogical techniques for teaching gage capability. It is unclear what gage capability method is utilized, but presumably it is the Design of Experiment method.

Equipment variation and appraiser variation are determined separately in a gage capability study and are useful in identifying the sources of measurement variation. However, equipment variation and appraiser variation from a gage capability study can be combined to estimate the overall standard deviation of repeated measurements. With the assumed normal distribution of repeated measurements, the overall standard deviation can be multiplied by 6 to capture 99.7% of the distribution of repeated measurements or by 5.15 to capture 99.0% of the distribution of repeated measurements. This result is considered the precision P of the measurement system.

The precision P of the measurement system must be compared to the tolerance T, upper specification limit minus lower specification limit, of the feature being measured by the measurement system to determine if the measurement system is 'capable'. This is done by determining the precision-to-tolerance P/T ratio. The smaller the P/T ratio, the more capable a measurement system is. Guidelines[9] for the P/T ratio suggest that a P/T ratio less than 0.1 is excellent, between 0.1 and 0.2 is good, between 0.2 and 0.3 is fair, and greater that 0.3 is unacceptable.

It is emphasized to students that the general process of conducting a gage capability study has some controversies. The first major controversy involves the quality of operators used in the study. One school of thought is that the study should be conducted with only the most highly skilled operators that will use the measurement system. The other school of thought is that the study should be conducted with a mixture of skill levels. Another major controversy involves the quality of the feature to be subjected to repeated measurements. For example, one school of thought is that surfaces that need to be touched for linear dimensional measurements should be refined, such as lapped and honed like gage blocks. The other school of though is that those surfaces should typical of what is produced by the associated manufacturing processes, such as a surface roughness of 125 or even 250 microinches. This author is of the latter school of thought in both cases because such measurement systems won't always be used by the most skilled operators and they won't always be used on refined surfaces.

The Assignment

The repeated measurement data provided to the students was generated by a team of peers using a handheld digital caliper and measuring the width dimension, in inches, of steel Charpy impact specimens, or parts. The measurement system and a sample part are shown in Figure 1.

Figure 1. The Measurement System and Sample Part

The team included three operators, a recorder and a coordinator. Prior to the experiment, students were told how important it is to randomize the order of data collection and they were shown how to use a random number chart or generator to accomplish randomization for this experiment. The coordinator assured that the measurements were taken in the predetermined random order and that the measurements were taken at the same location on each part. Three operators (*o* = 3) made two measurement trials (*n* = 2) each on five parts (*p* = 5), not knowing which part was being measured, and the data is shown in Table 1. The recorder worked closely with the coordinator to assure proper data entry.

Table 1. Measurement Data

	Operator 1		Operator 2		Operator 3	
Part	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
1	0.3905	0.3900	0.3920	0.3915	0.3925	0.3935
2	0.3920	0.3895	0.3950	0.3940	0.3970	0.3940
3	0.3925	0.3925	0.3940	0.3940	0.3940	0.3960
4	0.3930	0.3920	0.3945	0.3940	0.3935	0.3935
5	0.3915	0.3915	0.3930	0.3930	0.3920	0.3925

A digital caliper was chosen as the measurement system due to its simplicity. The caliper is easily held and adjusted with one hand while the measured part is held in the other. The chosen model has a digital display to eliminate scale reading errors but it does not have a clutch in the thumbscrew or roll adjustment to assure that each part is clamped onto with uniform force. In fact, the thumbscrew is just a thumb knob for this caliper model. The Charpy impact specimens have a finely milled surface so they are a compromise of surface roughness. They are not lapped and honed but they are smoother than a typical machined surface that might be 125 microinches.

Tabular Method & Results

The assignment to do the gage capability study by the Tabular Method is made at an early point in the course, shortly after the range estimation of standard deviation has been used to describe process control and process capability. The results of the Tabular Method are shown in Table 2, which utilizes common procedure and nomenclature[4].

Table 2. Tabular Method of Gage Capability Study

	Operator 1		Operator 2		Operator 3	
Part	Average	Range	Average	Range	Average	Range
1	0.3903	0.0005	0.3918	0.0005	0.3930	0.0010
2	0.3908	0.0025	0.3945	0.0010	0.3955	0.0030
3	0.3925	0.0000	0.3940	0.0000	0.3950	0.0020
4	0.3925	0.0010	0.3943	0.0005	0.3935	0.0000
5	0.3915	0.0000	0.3930	0.0000	0.3923	0.0005
		$\bar{\bar{x}}_1 = 0.39150$		$\bar{\bar{x}}_2 = 0.39350$		$\bar{\bar{x}}_3 = 0.39385$
		$\bar{R}_1 = 0.00080$		$\bar{R}_2 = 0.00040$		$\bar{R}_3 = 0.00130$
	$UCL_{R,1}$	$= 0.00260$	$UCL_{R,2}$	$= 0.00130$	$UCL_{R,3}$	$= 0.00420$

Table 2 shows the respective averages and ranges for each operator for each part. The following calculations use the familiar estimation of standard deviation from range values. Together, Table 2 and the calculations demonstrate how within-operator ranges are used to determine equipment variation, σ_{EV}, and how operator-to-operator range(s) are used to determine appraiser variation, σ_{AV}. Table 2 also shows control limits to demonstrate that the measurement process is in a state of statistical control (i.e., none of the operators are having substantial difficulty using the gage).

$$R_{\bar{\bar{x}}} = \max(\bar{\bar{x}}_1, \bar{\bar{x}}_2, \bar{\bar{x}}_3) - \min(\bar{\bar{x}}_1, \bar{\bar{x}}_2, \bar{\bar{x}}_3) = 0.002350$$

$$\sigma_{AV} = \frac{R_{\bar{\bar{x}}}}{d_2} = \frac{0.002350}{1.693} = 0.001388 \qquad (d_2\text{=1.693 for } o = 3)$$

$$\bar{\bar{R}} = \frac{\bar{R}_1 + \bar{R}_2 + \bar{R}_3}{3} = 0.000833$$

$$\sigma_{EV} = \frac{\bar{\bar{R}}}{d_2} = \frac{0.000833}{1.128} = 0.000739 \qquad (d_2\text{=1.128 for } n = 2)$$

Design of Experiment Method and Results

The assignment to do the gage capability study by the Design of Experiment method is made later in the course, after an introduction to Design of Experiment techniques and after instruction on how to create an Analysis of Variance (ANOVA) table for a two-factor experiment. The resulting ANOVA for the measurement data is Table 3.

Table 3. ANOVA for Design of Experiment Method

Source	*SS*	*DOF*	*MS*
Oper.	3.215E-05	2	1.607E-05
Part	2.122E-05	4	5.304E-06
Part*Oper.	1.218E-05	8	1.523E-06
Error	1.162E-05	15	7.750E-07
Total	7.717E-05	29	

The following components-of-variation calculations demonstrate how the data from Table 3 is used to partition variance into its respective sources, equipment variation and appraiser variation. Equipment variation is related to the overall random error variance of the experiment and appraiser variation is the related to the combination of operator variance and operator/part interaction variance within the experiment.

$$\sigma_{EV}^2 = MS_E$$

$$\sigma_{EV} = \sqrt{MS_E} = \sqrt{7.750x10^{-7}} = 0.000880$$

$$\sigma_{AV}^2 = \sigma_O^2 + \sigma_{PO}^2$$

$$\sigma_O^2 = \frac{MS_O - MS_{PO}}{pn} = \frac{1.607x10^{-5} - 1.523x10^{-6}}{5 \cdot 2} = 1.455x10^{-6}$$

$$\sigma_{PO}^2 = \frac{MS_{PO} - MS_E}{n} = \frac{1.523x10^{-6} - 7.750x10^{-7}}{2} = 3.740x10^{-7}$$

$$\sigma_{AV} = \sqrt{\sigma_O^2 + \sigma_{PO}^2} = \sqrt{1.455x10^{-6} + 3.740x10^{-7}} = 0.001352$$

Comparisons & Improvement Opportunities

The results of the two methods are shown in Table 4 where $\sigma_{GAGE} = \sqrt{\sigma_{EV}^2 + \sigma_{AV}^2}$ and $P = 5.15\sigma_{GAGE}$.

Table 4. Comparison of Gage Capability Methods

	Method		%
	Tabular	Des. Exp.	Diff.
σ_{EV}	0.000739	0.000880	17.5
σ_{AV}	0.001388	0.001352	2.6
σ_{GAGE}	0.001572	0.001614	2.6
P	0.008098	0.008311	2.6

For this set of measurement data, both methods indicate much more appraiser variation than equipment variation. Both methods indicate about the same appraiser variation but the Design of Experiment method indicates a somewhat higher equipment variation. An argument[10] is made that the comparison of the sources of variation should be done with variances rather than standard deviations. With this argument, the Design of Experiment method says that 70.2% of the variation is due to appraiser(s) with only 29.8% due to equipment.

Appraiser variation should be the first focus when investigating improvement opportunities of the measurement system. A higher appraiser variation is expected for this experiment, given the simplicity of the measurement system and the lack of experience and training of the operators. When asked about improvement opportunities, students usually cite the thumb knob mechanism of the caliper. Without a clutch in the mechanism, operators may clamp onto a part with an inconsistent force. Proper training and practice should make all operators use the thumb knob similarly, applying about the same clamp force, and reducing appraiser variation.

Using published guidelines[9], this measurement system should currently be used to measure a tolerance of no greater than about 0.06 inches (1/16”). That means this caliper would currently have the precision to measure most as-cast, as-forged or as-molded features but would not have the precision to measure most machined features. That may not be a very popular assessment to the party responsible for selection and procurement of this caliper.

For this case, the two gage capability methods produce comparable results for appraiser variation and for equipment variation. However, Klaput and Plura[11] show that the two methods can produce substantially different results. This author has compared the two methods for only one other set of measurement data and found the methods to produce comparable results in that case also.

Learning Assessment

Since both assignments are rather lengthy, most student work is done out of class. With ample out-of-class time to complete the assignments, student scores are usually quite good. The first assignment is mid-semester homework and amounts to about 2% of the course grade. Students are encouraged to create a spreadsheet that can be used for future gage capability studies. The second assignment is often part of a take-home final exam and the assignment amounts to about 4% of the course grade. Students are reminded of what the results were from the first assignment. Students again are encouraged to create a spreadsheet that can be used for future work. Students are discouraged from using a statistical software package such as Minitab only so that, at this early learning stage of Designed Experiments, they can gain an appreciation for what an analysis of variance is doing for them. Further, it is not overly cumbersome to create a spreadsheet to get the ANOVA table for a two factor experiment with three and five levels respectively and with only two replicates.

Conclusion

Throughout both assignments, students experience measurement variation first-hand, possibly for the first time. They also experience how industrial practices and associated tools evolve over time. The students gain an appreciation for the effort that is necessary to properly conduct an experiment, including randomization of data collection. Further, they see that for this case, the two methods produce comparable results.

Bibliography

1. Senvar, O., Firat, O., Umit, S., An overview of capability evaluation of Measurement Systems and Gauge Repeatability and Reproducibility Studies, *International Journal of Metrology and Quality Engineering*, v 1, n 2, p 121-127, 2010; ISSN: 21076839; DOI: 10.1051/ijmqe/2010022
2. O'Neill, B., Gage Repeatability and Reproducibility in Rockwell Hardness Testers, *Advanced Materials and Processes*, v 168,n 11, 33-5, Nov.-Dec. 2010; ISSN: 0882-7958
3. Sweeney, S., Analysis of Two-dimensional Gage Repeatability and Reproducibility, *Quality Engineering*, Volume 19, 2007 - Issue 1, p 29-37; DOI: 10.1080/08982110601057641
4. Montgomery, D.C., *Introduction to Statistical Quality Control*, 4th ed., 2001, p 383, John Wiley & Sons, Inc., ISBN: 0-471-31648-2
5. Borror, C., Montgomery, D., Runger, G., Confidence intervals for variance components from gauge capability studies, *Quality and Reliability Engineering International*, v 13,n 6, p 361-369, Nov.-Dec. 1997; ISSN: 07488017; DOI: 10.1002/(SICI)1099-1638(199711/12)13:6<361

6. Wheeler, D., Problems with gauge R&R studies, *Annual Quality Congress Transactions*, v 46, p 179-185, 1992
7. Kozak, M., The creation and assessment of a gage repeatability and reproducibility study exercise in a metrology class, *Proceedings of the 118th ASEE Annual Conference and Exposition*, June 2011.
8. Timmer, D., Gonzalez, M., Teaching gage reproducibility and repeatability using the Mouse Factory, *Proceedings of the 120th ASEE Annual Conference and Exposition*, June 213.
9. Measurement Systems Analysis, Automotive Industry Action Group, Product Code MSA-4, 4th ed., p 78, June 2010
10. Ermer, D., Improved gage R&R measurement [quality control], *Quality Progress*, v 39, n 3, 77-9, March 2006; ISSN:0033-524X
11. Klaput, P., Plura, J., Influence of used method on the results of gage repeatability and reproducibility study, *Proceedings of the 21st International Conference on Metallurgy and Materials*, p 1756-1761, 2012; ISBN-13:9788087294314