2013

A Comparison of Gauge Repeatability and Reproducibility Methods

Scott Stamm
Indiana State University

A COMPARISON OF GAUGE REPEATABILITY AND REPRODUCIBILITY METHODS

A dissertation

Presented to

The College of Graduate and Professional Studies

College of Technology

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Scott Stamm

May 2013

Keywords: Gauge R&R Study, Measurement Precision Error, Technology Management, Quality Systems

COMMITTEE MEMBERS

Committee Chair: Michael A. Hayden, Ph.D.

Professor, Department of Applied Engineering and Technology Management

Indiana State University

Committee Member: M. Affan Badar, Ph.D.

Chair, Department of Applied Engineering and Technology Management

Indiana State University

Committee Member: Marion D. Schafer, Ph.D.

Professor, Department of Applied Engineering and Technology Management

Indiana State University

ABSTRACT

Common in industrial applications is the need for estimates for measurement precision error. Measurement precision error is important because manufacturers make decisions about product acceptance or rejection based on product measurements. A frequent method of determining measurement precision error is the Gauge Repeatability and Reproducibility Study (GR&R Study). A typical GR&R Study determines estimates of repeatability error, reproducibility error as well as estimates of total measurement precision error and the part-to-part component. This dissertation compares three methods of performing GR&R studies on 10,080 simulated GR&R study data sets. The 10,080 simulations were derivations of 224 actual Gauge R&R studies. The three methods of analysis are Donald Wheeler's "Honest Gauge R&R Study," the Automotive Industry Action Group's Average and Range Method and the ANOVA Method. The study results were analyzed by ANOVA, Kruskal-Wallis and Pearson correlation. The analysis showed the three methods are different in their estimates of total Gauge R&R and the components of repeatability, reproducibility, and part-to-part. The analysis also estimated the pair-wise comparisons of the three methods and showed they are different from one another for total GR&R, repeatability, reproducibility and part-to-part. The correlation analysis showed Donald Wheeler's method to be correlated with both the Average and range method and the ANOVA method and the Average and range method to be correlated to the ANOVA method. For critical products the ANOVA method is recommended for Gauge R&R analysis, while for less critical products the Average and range method and Wheeler's "Honest Gauge R&R Study" approach are recommended.

ACKNOWLEDGMENTS

My deep appreciation goes to Dr. Michael Hayden for his wisdom, wise guidance, critique, and advice during the dissertation process. Additionally, my deep appreciation goes to the dissertation committee members; Dr. Michael Hayden, Dr. Affan Badar and Dr. Marion Schafer.

TABLE OF CONTENTS

COMMITTEE MEMBERS ii

ABSTRACT iii

ACKNOWLEDGMENTS iv

LIST OF TABLES viii

LIST OF FIGURES xiv

CHAPTER 1 THE PROBLEM AND ITS SETTING 1

- General Statement of the Problem 4
- Primary Research Question of the Study 4
- Statement of the Hypothesis 5
- Delimitations of the Study 12
- The Definition of Terms 13
- Assumptions 15

CHAPTER 2 A REVIEW OF RELATED LITERATURE 17

- Automotive Industry Action Group Methods of Gauge R&R 19
- The A&R Method in Detail 22
- The ANOVA Method in Detail 25
- EMP III Methods 35
- The EMP III Method in Detail 39
- Other Methods 42

Standards..44

U.S. Government Requirements 45

A Review of Related Literature-Summary 45

CHAPTER 3 METHOD OF INVESTIGATION 49

Overview.. 49

Description of the Subjects 50

Simulation to Achieve Sample Size...56

Research Design Procedures/Description of the Measurers Employed.............................62

CHAPTER 4 FINDINGS...65

Overview..65

Hypothesis One..68

Hypothesis Two 72

Hypothesis Three 75

Hypothesis Four...79

Hypothesis Five 82

Hypothesis Six 83

Hypothesis Seven...84

Hypothesis Eight..85

CHAPTER 5 SUMMARY AND DISCUSSION 87

Conclusions..87

Practical Significance..90

Further Study 93

REFERENCES 95

APPENDIX A: RESULTS FROM 224 ACTUAL (SEED) GAUGE R&R STUDIES 100

APPENDIX B: SAMPLE SIZE CALCULATION REPORTS..101

APPENDIX C: HISTOGRAM COMPARISONS OF SEED GR&R STUDY DATA AND SIMULATED GR&R STUDY DATA..106

APPENDIX D: MINITAB MACROS...119

APPENDIX E: HYPOTHESIS ONE THROUGH FOUR ANALYSIS WITH OPERATOR BY INTERACTION DATA REMOVED...124

APPENDIX F: COMPARISON OF SEED GAUGE R&R STUDY DATA AND SIMULATION GAUGE R&R STUDY DATA..134

APPENDIX G: VARIANCE COMPARISON OF EMP III, A&R AND ANOVA METHODS FOR SIMULATED DATA..136

LIST OF TABLES

Table 1.1 Study Hypothesis by Source of Variation and Analsyis Method 11

Table 1.2 EMP III Results Comparison between Wheeler (2009) and Study Spreadsheet...........16

Table 2.1 A&R Method Layout of Gauge R&R Study Data and Initial Calculations..................22

Table 2.2 ANOVA Method Layout of Gauge R&R Study Data and Initial Calculations.............26

Table 2.3 Squared Gauge R&R Measured Values and Their Summation....................................27

Table 2.4 Sum of Measured Values by Operator and Squares of the Sums of the Measured Values 28

Table 2.5 ANOVA Table for Gauge R&R Study 30

Table 2.6 Gauge R&R Study Results Table 31

Table 2.7 EMP III Method Layout of Gauge R&R Study Data and Initial Calculations..............39

Table 2.8 Comparison of the Three GR&R Method Examples in this Chapter and Minitab Results for the Same Data for A&R and ANOVA Methods 42

Table 3.1 Comparison of Actual (Seed) Gauge R&R Study Total Gauge Error...........................50

Table 3.2 Summary of Sample Size by Hypothesis...55

Table 3.3 Typical Gauge R&R Study Data Collection Table..62

Table 4.1 Comparison of Simulated Gauge R&R Study Total Gauge Error.................................66

Table 4.2 Hypothesis one, ANOVA Analysis of Total Measurement Precision Error 68

Table 4.3 Hypothesis one, Descriptive Statistics for the Three Analysis Methods.......................69

Table 4.4 Hypothesis one, Tukey's Pair-wise Comparisons EMP III Method of Total Measurement Precision Error 70

Table 4.5 Hypothesis one, Tukey's Pair-wise Comparisons A&R of Total Measurement Precision Error 70

Table 4.6 Hypothesis one, Kruskal-Wallis Analysis of Total Measurement Precision Error 71

Table 4.7 Hypothesis one, Kruskal-Wallis Multiple Comparisons for Total Measurement Precision Error 71

Table 4.8 Hypothesis two, ANOVA Analysis for Repeatability Measurement Precision Error 72

Table 4.9 Hypothesis two, Repeatability Descriptive Statistics 73

Table 4.10 Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability 73

Table 4.11 Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability 74

Table 4.12 Hypothesis two, Kruskal-Wallis Tests on Repeatability 74

Table 4.13 Hypothesis two, Kruskal-Wallis Multiple Comparisons for Repeatability 75

Table 4.14 Hypothesis three, ANOVA Analysis for Reproducibility Measurement Precision Error 76

Table 4.15 Hypothesis three, Reproducibility Descriptive Statistics 76

Table 4.16 Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility 77

Table 4.17 Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility 77

Table 4.18 Hypothesis three, Kruskal-Wallis Test on Reproducibility 78

Table 4.19 Hypothesis three, Kruskal-Wallis Multiple Comparisons for Reproducibility 78

Table 4.20 Hypothesis four, ANOVA Analysis for Part-to-part Measurement Precision Error 79

Table 4.21 Hypothesis four, Part-to-part Descriptive Statistics 80

Table 4.22 Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part 80

Table 4.23 Hypothesis four, Tukey's Pair-wise Comparison for Part-to-part 81

Table 4.24 Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part Portion of Measurement Precision Error 81

Table 4.25 Hypothesis four, Kruskal-Wallis Test on Part-to-part Portion of Measurement Precision Error 82

Table 4.26 Hypothesis five, Pearson r Correlation Values for Total Measurement Precision Error 83

Table 4.27 Hypothesis six, Pearson r Correlation Values for Repeatability 83

Table 4.28 Hypothesis seven, Pearson r Correlation Values for Reproducibility 84

Table 4.29 Hypothesis eight, Pearson r Correlation Values for Part-to-part 85

Table 5.1 Standard Deviation and Percent Tolerance for Example Calculations in Chapter 2 89

Table 5.2 Study Results for Total Measurement Precision Error (Total GRR)............................91

Table 5.3 Recommendations for GR&R methods under Varying Conditions 93

Table A1 Descriptive Statistics fir Total 224 Seed Gauge R&R Studies (in percentages) 100

Table D1 Example %GRRXRc Minitab Macro Output 121

Table D2 Example %GRRAVc Minitab Macro Output 122

Table E1 Hypothesis one, ANOVA Analysis of Total Measurement Precision Error with Operator by Part Interaction Data Removed...124

Table E2 Hypothesis one, Descriptive Statistics for the Three Analysis Methods with Operator by Part Interaction Data Removed...124

Table E3 Hypothesis one, Tukey's Pair-wise Comparisons EMP III Method of Total Measurement Precision Error with Operator by Part Interaction Data Removed 125

Table E4 Hypothesis one, Tukey's Pair-wise Comparisons A&R Method of Total Measurement Precision Error with Operator by Part Interaction Data Removed 125

Table E5 Hypothesis one, Kruskal-Wallis Analysis of Total Measurement Precision Error with Operator by Part Interaction Data Removed..125

Table E6 Hypothesis one, Kruskal-Wallis Multiple Comparisons for Total Measurement Precision Error with Operator by Part Interaction Data Removed 126

Table E7 Hypothesis two, ANOVA Analysis for Repeatability Measurement Precision Error with Operator by Part interaction Data Removed..126

Table E8 Hypothesis two, Repeatability Descriptive Statistics with Operator by Part Interaction Data Removed..127

Table E9 Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability with Operator by Part Interaction Data Removed 127

Table E10 Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability with Operator by Part Interaction Data Removed 127

Table E11 Hypothesis two, Kruskal-Wallis Test on Repeatability with Operator by Part Interaction Data Removed 128

Table E12 Hypothesis two, Kruskal-Wallis Multiple Comparisons for Repeatability with Operator by Part Interaction Data Removed..128

Table E13 Hypothesis three, ANOVA Analysis for Reproducibility Measurement Precision Error with Operator by Part Interaction Data Removed...129

Table E14 Hypothesis three, Reproducibility Descriptive Statistics with Operator by Part Interaction Data Removed 129

Table E15 Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility with Operator by Part Interaction Data Removed 130

Table E16 Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility with Operator by Part Interaction Data Removed 130

Table E17 Hypothesis three, Kruskal-Wallis Test on Reproducibility with Operator by Part Interaction Data Removed 130

Table E18 Hypothesis three, Kruskal-Wallis Multiple Comparisons for Reproducibility with Operator by Part Interaction Data Removed 131

Table E19 Hypothesis four, ANOVA Analysis for Part-to-part Measurement Precision Error with Operator by Part Interaction Data Removed 131

Table E20 Hypothesis four, Part-to-part Descriptive Statistics with Operator by Part Interaction Data Removed 132

Table E21 Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part with Operator by Part Interaction Data Removed 132

Table E22 Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part with Operator by Part Interaction Data Removed 132

Table E23 Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part with Operator by Part Interaction Data Removed 133

Table E24 Hypothesis four, Kruskal-Wallis test on Part-to-part Portion of Measurement Precision Error with Operator by Part interaction Data Removed 133

Table F1Comparison of Seed and Simulation Gauge R&R Study Data P-Values 134

Table G1 Hypothesis one, Tests for Equal Variance 136

Table G2 Hypothesis two, Tests for Equal Variance 138

Table G3 Hypothesis three, Tests for Equal Variance..140

Table G4 Hypothesis four, Tests for Equal Variance..142

LIST OF FIGURES

Figure 1.1. Measurement Demonstrating a Bias 2

Figure 1.2. Three Measurement Processes Demonstrating Differences in Precision 2

Figure 3.1. Histogram of A&R Analyzed Seed Data for Total Measurement Precision Error 57

Figure 3.2. Partial Diagram of Simulation After First Five Simulation Runs 59

Figure 3.3. Partial Diagram of Simulation After Ten Simulation Runs 59

Figure 3.4. Histogram of A&R Total Measurement Precision Error 61

Figure 5.1. Chart of 67 Randomly Selected Study Measurement Precision Error (Total GRR) Results by Method of Analysis 92

Figure C1. Seed GR&R Study EMP III Total Measurement Precision Error 107

Figure C2. Simulated GR&R Study EMP III Total Measurement Precision Error 107

Figure C3. Seed GR&R Study A&R Total Measurement Precision Error 108

Figure C4. Simulated GR&R Study A&R Total Measurement Precision Error 108

Figure C5. Seed GR&R Study ANOVA Total Measurement Precision Error 109

Figure C6. Simulated GR&R Study ANOVA Total Measurement Precision Error 109

Figure C7. Seed GR&R Study EMP III Repeatability Component of Measurement Precision Error 110

Figure C8. Simulated GR&R Study EMP III Repeatability Component of Measurement Precision Error 110

Figure C9. Seed GR&R Study A&R Repeatability Component of Measurement Precision Error 111

Figure C10. Simulated GR&R Study A&R Repeatability Component of Measurement Precision Error 111

Figure C11. Seed GR&R Study ANOVA Repeatability Component of Measurement Precision Error 112

Figure C12. Simulated GR&R Study ANOVA Repeatability Component of Measurement Precision Error 112

Figure C13. Seed GR&R Study EMP III Reproducibility Component of Measurement Precision Error 113

Figure C14. Simulated GR&R Study EMP III Reproducibility Component of Measurement Precision Error 113

Figure C15. Seed GR&R Study A&R Reproducibility Component of Measurement Precision Error 114

Figure C16. Simulated GR&R Study A&R Reproducibility Component of Measurement Precision Error 114

Figure C17. Seed GR&R Study ANOVA Reproducibility Component of Measurement Precision Error 115

Figure C18. Simulated GR&R Study ANOVA Repeatability Component of Measurement Precision Error 115

Figure C19. Seed GR&R Study EMP III Part-to-part Component of Measurement Precision Error 116

Figure C20. Simulated GR&R Study EMP III Part-to-part Component of Measurement Precision Error 116

Figure C21. Seed GR&R Study A&R Part-to-part Component of Measurement Precision Error 117
Figure C22. Simulated GR&R Study A&R Part-to-part Component of Measurement Precision Error 117
Figure C23. Seed GR&R Study ANOVA Part-to-part Component of Measurement Precision Error 118
Figure C24. Simulated GR&R Study ANOVA Part-to-part Component of Measurement Precision Error 118
Figure G1. Hypothesis one, Test for Equal Variance Total GR&R 137
Figure G2. Hypothesis two, Test for Equal variance Repeatability 139
Figure G3. Hypothesis three, Test for Equal Variance Reproducibility..........141
Figure G4. Hypothesis four, Test for Equal variance Part-to-part..........143

CHAPTER 1

THE PROBLEM AND ITS SETTING

In manufacturing environments, as noted by Montgomery and Runger (1993), measurement plays a significant role in helping firms improve quality. Knowles, Antony and Vickers (2000) emphasize that measurement is the cornerstone of decision making. Further, Knowles et al. (2000) point out that when businesses make decisions with measurement data, those decisions are predicated on the presumption that the data supporting the decision are reliable.

Examples of businesses using measurement information to make decisions about product conformance include accepting or rejecting a product or when to adjust a process. Product acceptance decisions are frequently made in manufacturing during in-process or final inspection activities where the measurement results are evaluated and compared with specifications. An example of using measurement results to adjust a process would be a process operator making off-set adjustments to a milling machine based on product measurement results. As mentioned by Knowles et al. (2000) previously, the assumption is that the measurement results are reliable.

Typically in industry, depending on the criticality of the measurement results, studies are performed to determine the accuracy and precision of a measurement system. Measurement accuracy is estimated by bias studies, for example, as described in the *Measurement Systems Analysis* manual (Measurement Systems Analysis, 2010). Bias is described as the difference from the true value being measured as compared to the measured value. Figure 1.1 presents a diagram showing bias in a measurement system. Measurement precision is commonly estimated by performing Gauge Repeatability and Reproducibility (Gauge R&R) studies, again as described by the *Measurement Systems Analysis* manual (Measurement Systems Analysis, 2010). Bias of a measurement system is how close the measurements are to the true value. The precision of a measurement system is how much variation is there in repeated measurements of the same object or dimension. Figure 1.2 demonstrates differences in precision in three different measurement systems. Note the differences in the width of the distribution of measured values. Other attributes of a measurement system as described by Measurement Systems Analysis (2010) are linearity and stability.

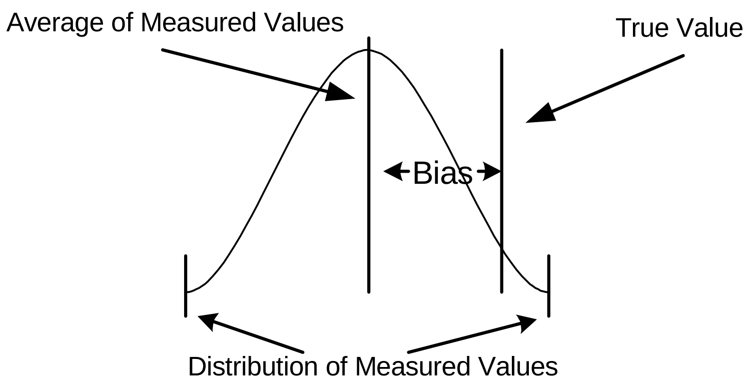

Figure 1.1. Measurement Demonstrating a Bias.

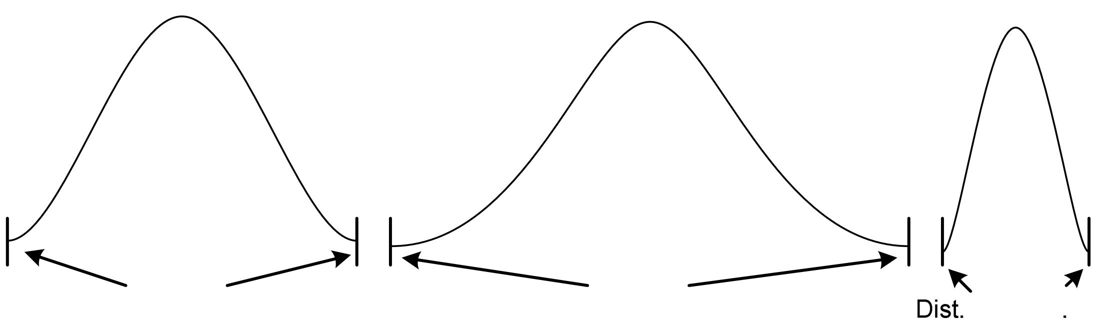

Figure 1.2. Three Measurement Processes Demonstrating Differences in Precision.

One standard method of assessing measurement precision error is the Automotive Industry Action Group (AIAG) method as described by Measurement Systems Analysis (2010). Here the measurement precision error is estimated by performing a study, for example three operators measuring ten parts each three times for a total of 90 measurements. After the measurements are complete, and the results calculated, the components of variation are broken down into total repeatability and reproducibility components as well as the part-to-part component of measurement precision error. Also typical is a comparison of the total variation to the dimensional product tolerance, the precision to tolerance ratio. Here the measurement precision error is compared to the tolerance width to determine how much of the tolerance band is consumed by measurement precision error. In the *Measurement Systems Analysis* manual an average and range method- hereafter referred to as the A&R method- as well as an ANOVA method- hereafter referred to as the ANOVA method- is described for estimating measurement precision error (Measurement Systems Analysis, 2010).

One critic of the AIAG methods of measurement systems analysis is Donald Wheeler. Wheeler, a consulting statistician, contends the AIAG methods significantly overstate the measurement precision error of a measurement system (Wheeler, 2006). While Wheeler's *Evaluating the Measurement Process III* (EMP III) method uses the same method of collecting the measurement study data (typically, three operators, ten parts and three replications of measurement), his method of analyzing the data is different in the way it calculates the percent of measurement precision error and its components. Collectively the measurement precision error estimation methods mentioned above are known as Gauge Repeatability and Reproducibility (GR&R) studies.

General Statement of the Problem

The significance of the different methods of estimating measurement precision error is considerable because the results of a measurement study are used to make decisions about the adequacy of a measurement process. For example, a less expensive prospective measurement system may be needlessly rejected in favor of a more expensive system based on an inflated estimate of measurement precision error. In addition, decisions are made about adjustments to manufacturing process tolerances as a result of measurement precision error to ensure only acceptable product is passed- defined as guard banding. If the measurement precision error is indeed overstated by the A&R and ANOVA methods, manufacturers may be overly conservative in their compensations for measurement precision errors and thus might be scrapping or needlessly reworking conforming product.

This study attempts to address this problem by examining multiple Gauge R&R studies on various typical industrial measurement devices and comparing the results under the EMP III, A&R, and ANOVA methods to determine if the EMP III method reports lower measurement precision error estimates. This comparison relates to Technology Management because it defines how officials of a firm manage the product inspection process. The thesis of this study is that managers of the measurement process do not know the extent to which the methods are different.

Primary Research Question of the Study

The overall research question is: Do the methods to estimate measurement precision error produce the same results?

Statement of the Hypotheses

This study addresses four sources of measurement precision error identified by Gauge R&R studies. These sources of error are: total measurement precision error, repeatability component of measurement precision error, reproducibility component of measurement precision error and the part-to-part component of measurement precision error. These four sources of measurement precision error are estimated by three methods of Gauge R&R analysis. Both the differences among these methods and relationship/correlations are investigated. This investigation includes eight hypotheses as described in the following paragraphs.

A first hypothesis of the study addresses the averages of measurement precision error. In other words, the measurement precision error could be minor or a relatively low percentage of the tolerance range. The measurement precision error could also be major or a large percentage of the tolerance range. The AIAG guidelines suggest a measurement precision error of less than ten percent of the tolerance range is considered acceptable, a measurement precision error between ten and thirty percent is acceptable for some applications and above thirty percent is unacceptable (Measurement Systems Analysis, 2010). This first hypothesis tests if the EMP III, A&R, and ANOVA methods are different in their average total measurement precision error estimate of the same set of data. The research question is: Do the three methods (EMP III, A&R and ANOVA) estimate total measurement precision error equally?

HO_1: The EMP III, A&R, and ANOVA methods estimate average total measurement precision error equally. That is, $\mu_{EMP\ III} = \mu_{A\&R} = \mu_{ANOVA}$

HA_1: At least one of the EMP III, A&R, and ANOVA methods, estimates average total measurement precision error differently than the others. That is, $\mu_{EMP\ III \neq A\&R}$, $\mu_{A\&R \neq} \mu_{ANOVA}$, or $\mu_{EMP\ III \neq} \mu_{ANOVA}$.

The second and third hypotheses relate to two components of measurement precision error attributed to a measurement system and how consistently these components are estimated. Measurement precision error is typically divided into repeatability and reproducibility error. Repeatability error is how consistent an operator of a measurement system is across multiple measurements of the same objects- the within operator error. Repeatability precision error component is covered in Hypothesis two. The research question is: Do the three methods (EMP III, A&R and ANOVA) estimate repeatability measurement precision error equally?

HO_2: The repeatability component of measurement precision error is the same among the EMP III method, the A&R method, and the ANOVA method of measurement precision error assessment. That is, $\mu_{\text{EMP III-Repeatability}} = \mu_{\text{A\&R-Repeatability}} = \mu_{\text{ANOVA-Repeatability}}$.

HA_2: The repeatability component of measurement precision error is not the same among the EMP III method, the A&R method, and the ANOVA method of measurement precision error assessment. That is, $\mu_{\text{EMP III-Repeatability}} \neq \mu_{\text{A\&R-Repeatability}}$, or $\mu_{\text{A\&R-Repeatability}} \neq \mu_{\text{ANOVA-Repeatability}}$, or $\mu_{\text{EMP III-Repeatability}} \neq \mu_{\text{ANOVA-Repeatability}}$.

The third hypothesis addresses the Reproducibility component of measurement precision error estimation. Reproducibility error is how consistent operators are among themselves when measuring the same objects, described as operator-to-operator error. This hypothesis tests whether or not the three systems are consistent in their estimates of the reproducibility component of measurement precision error. The research question is: Do the three methods (EMP III, A&R and ANOVA) estimate reproducibility measurement precision error equally?

HO_3: The reproducibility component of measurement precision error is the same among the EMP III method, A&R method, and ANOVA method of measurement precision error assessment. That is, $\mu_{\text{EMP III-Reproducibility}} = \mu_{\text{A\&R-Reproducibility}} = \mu_{\text{ANOVA-Reproducibility}}$.

HA_3: The reproducibility component of measurement precision error is not the same among the EMP III method, A&R method, and ANOVA method of measurement precision error assessment. That is, $\mu_{\text{EMP III-Reproducibility}} \neq \mu_{\text{A\&R-Reproducibility}}$, $\mu_{\text{A\&R-Reproducibility}} \neq \mu_{\text{ANOVA-Reproducibility}}$, or $\mu_{\text{EMP III-Reproducibility}} \neq \mu_{\text{ANOVA-Reproducibility}}$.

A fourth hypothesis relates to the estimation of the Part-to-part variation component of the EMP III, A&R, and ANOVA methods of Gauge R&R. Part-to-part variation in a Gauge R&R study is comprised of the actual differences in the parts used in the study for the measurement of interest. If the three methods of estimating measurement precision error are consistent, they should have similar estimates of part-to-part variation. The research question is: Do the three methods (EMP III, A&R and ANOVA) estimate part-to-part measurement precision error equally?

HO_4: The average of the EMP III method, A&R method, and the ANOVA method for estimating part-to-part measurement precision error are equal. That is, $\mu_{\text{EMP III-Part-to-part}} = \mu_{\text{A\&R-Part-to-part}} = \mu_{\text{ANOVA-Part-to-part}}$.

HA_4: The average of the EMP III method, A&R method, and the ANOVA method for estimating part-to-part measurement precision error are different. That is, $\mu_{\text{EMP III-Part-to-part}} \neq \mu_{\text{A\&R-Part-to-part}}$, $\mu_{\text{A\&R-Part-to-part}} \neq \mu_{\text{ANOVA-Part-to-part}}$, or $\mu_{\text{EMP III-Part-to-part}} \neq \mu_{\text{ANOVA-Part-to-part}}$.

A fifth hypothesis of the study concerns the relationship among the three measurement precision error estimate methods when compared two at a time. The research question is: Is there a relationship of the estimate of total measurement precision error among the three methods (EMP III, A&R, and ANOVA) when compared two at a time?

HO_5: There is no relationship between the EMP III method and A&R method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III /A\&R}} = 0$

There is no relationship between the EMP III method and ANOVA method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III/ANOVA}} = 0$

There is no relationship between the A&R method and ANOVA method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{A\&R/ANOVA}} = 0$

HA_5: There is a relationship between the EMP III method and A&R method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III/A\&R}} \neq 0$

There is a relationship between the EMP III method and ANOVA method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III/ANOVA}} \neq 0$

There is a relationship between the A&R method and ANOVA method of total measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{A\&R/ANOVA}} \neq 0$

A sixth hypothesis relates to the repeatability portion of measurement precision error estimation. The research question is: Is there a relationship of the estimate of total repeatability precision error among the three methods (EMP III, A&R, and ANOVA) when compared two at a time?

HO_6: There is no relationship between the EMP III method and the A&R method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III Repeatability/A\&R-Repeatability}} = 0$.

There is no relationship between the EMP III method and ANOVA method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III Repeatability/ANOVA-Repeatability}} = 0$.

There is no relationship between the A&R method and ANOVA method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{A\&R Repeatability/ANOVA Repeatability}}=0$.

HA$_6$: There is a relationship between the EMP III method and A&R method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III-Repeatability/A\&R-Repeatability}} \neq 0$.

There is a relationship between the EMP III method and ANOVA method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III Repeatability/ANOVA-Repeatability}}\neq 0$.

There is a relationship between the A&R method and ANOVA method of repeatability measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{A\&R Repeatability/ANOVA Repeatability}}\neq 0$.

A seventh hypothesis questions whether there is a relationship among the three methods (EMP III, A&R, and ANOVA) when compared two at a time in their respective estimates of reproducibility error. The research question is: Is there a relationship of the estimate of total reproducibility precision error among the three methods (EMP III, A&R, and ANOVA) when compared two at a time?

HO$_7$: There is no relationship between the EMP III method and A&R method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{\text{EMP III-Reproducibility/A\&R-Reproducibility}} =0$.

There is no relationship between the EMP III method and ANOVA method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{EMP\ III\ Reproducibility/ANOVA\text{-}Reproducibility}=0$.

There is no relationship between the A&R method and ANOVA method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{A\&R\ Reproducibility/ANOVA\ Reproducibility}=0$.

HA_7: There is a relationship between the EMP III method and A&R method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{EMP\ III\ Reproducibility/A\&R\text{-}Reproducibility}\neq 0$.

There is a relationship between the EMP III method and ANOVA method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{EMP\ III\ Reproducibility/ANOVA\text{-}Reproducibility}\neq 0$.

There is a relationship between the A&R method and ANOVA method of reproducibility measurement precision error estimation across Gauge R&R data. That is, $\rho_{A\&R\ Reproducibility/ANOVA\text{-}Reproducibility}\neq 0$.

The eighth hypothesis of the study seeks to determine if there is a relationship among the three methods of estimating the part-to-part portion of measurement precision error. The research question is: Is there a relationship of the estimate of total part-to-part precision error among the three methods (EMP III, A&R, and ANOVA) when compared two at a time?

HO_8: There is no relationship between the EMP III method and A&R method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{EMP\ III\text{-}Part\text{-}to\text{-}part/A\&R\text{-}Part\text{-}to\text{-}part}=0$.

There is no relationship between the EMP III method and ANOVA method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{\text{EMP III-Part-to-part/ANOVA-Part-to-part}}=0$.

There is no relationship between the A&R method and ANOVA method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{\text{A\&R Part-to-part/ANOVA Part-to-part}}=0$.

HA_8: There is a relationship between the EMP III method and A&R method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{\text{EMP III-Part-to-part/A\&R-Part-to-part}}\neq 0$.

There is a relationship between the EMP III method and ANOVA method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{\text{EMP III-Part-to-part/ANOVA-Part-to-part}}\neq 0$.

There is a relationship between the A&R method and ANOVA method of Part-to-part variation estimation across Gauge R&R data. That is, $\rho_{\text{A\&R Part-to-part/ANOVA Part-to-part}}\neq 0$.

Table 1.1 provides a review of the eight hypotheses and the four sources of variation as well as the comparison types for each combination.

Table 1.1

Study Hypotheses by Source of Variation and Analysis Method

		Source of Variation			
		Total GR&R	**Repeatability Component**	**Reproducibility Component**	**Part-to-Part Component**
Method	A&R	Difference (Hypothesis 1) Correlation (Hypothesis 5)	Difference (Hypothesis 2) Correlation (Hypothesis 6)	Difference (Hypothesis 3) Correlation (Hypothesis 7)	Difference (Hypothesis 4) Correlation (Hypothesis 8)

Table 1.1 *(continued)*

		Source of Variation			
		Total GR&R	**Repeatability Component**	**Reproducibility Component**	**Part-to-Part Component**
Method	ANOVA	Difference (Hypothesis 1) Correlation (Hypothesis 5)	Difference (Hypothesis 2) Correlation (Hypothesis 6)	Difference (Hypothesis 3) Correlation (Hypothesis 7)	Difference (Hypothesis 4) Correlation (Hypothesis 8)
	EMP III	Difference (Hypothesis 1) Correlation (Hypothesis 5)	Difference (Hypothesis 2) Correlation (Hypothesis 6)	Difference (Hypothesis 3) Correlation (Hypothesis 7)	Difference (Hypothesis 4) Correlation (Hypothesis 8)

Delimitations of the Study

The type of Measurement Systems Analysis described so far for this study is known as a Variable Gauge R&R Study. This study is of a variable non-destructive measurement system. This characteristic means the data provided by the study is numerical on a continuous scale. This numerical data type is contrasted with attribute type data, where the data provided by the measurement system is categorical - good/bad or yes/no or pass/fail - type data. Another delimitation of the study is that it is limited to non-destructive type measurements that utilize a Crossed Gauge R&R Study data analysis. Non-destructive type measurements do not alter the sample as part of the measurement process. Some measurement systems employ destructive type measurements, where the sample studied in the measurement process is altered or destroyed. An example of the latter type of analysis is a tensile test. Destructive type measurement systems are analyzed through nested data analysis, which are outside the scope of this study.

Full measurement systems analysis can include checks of the bias, precision, linearity and stability of a measurement system as described in Measurement Systems Analysis (2010). This study is limited to an examination of the precision and part variation contributions of a measurement system study. Typically the components of measurement precision error are subdivided into the categories of total error, repeatability error, reproducibility error, and part-to-part, Measurement Systems Analysis (2010). A further delimitation of the study is it is limited to the EMP III method, A&R method, and the ANOVA method. The Range method as well as other methods of measurement uncertainty estimation is not included in this study.

The Definition of Terms

A&R method- see Average and range method.

ANOME-(Analysis of Main Effects) - A statistical technique that determines if the main effect contributors in a study are statistically different (Wheeler, 2003).

ANOVA (Analysis of Variance) - A basic statistical technique for analyzing experimental data that subdivides the total variation into component parts associated with specific sources of variation (Omdahl, 1997).

ANOVA Method- In the context of this dissertation the ANOVA Method is one of three GR&R analysis methods utilized for analysis of the GR&R study data (signified ANOVA method).

ANOMR- (Analysis of Mean Range) - A statistical technique that demonstrates from the range of data for a factor contributing to a study is statistically different from another factor (Wheeler, 2006).

Attribute Data – A form of qualitative data. Numerical information representing the frequency of occurrence within some discrete category. For example, 16 bad, 250 good. Also called go/no go information (Omdahl, 1997).

Average and range method- The average and range method ($\overline{X}$ & R) with respect to Gauge R&R studies is an approach that provides an estimate of both repeatability and reproducibility for a measurement system. This method utilizes calculation techniques from statistical process control (SPC), (Measurement Systems Analysis, 2010). In the context of this dissertation the Average and range method (signified A&R method) is one of three analysis methods utilized for GR&R analysis of the study data.

Bias (In-Accuracy) - The difference between the observed average of measurements and the reference value. A systematic error component of the measurement system (Measurement Systems Analysis, 2010).

EMP III- In the context of this dissertation EMP III is Donald Wheeler's "Honest Gauge R&R Study" (Wheeler, 2006). It is one of three GR&R analysis methods used in this study (signified EMP III method).

Linearity- The change in bias over the normal operating range of the instrument, (Measurement Systems Analysis, 2010). In other words, the bias and precision of the measurement system over the full operating range of the instrument.

Measurement System- The collection of hardware, software, procedures and methods, human effort, environmental conditions, associated device, and the objects that are measured for the purpose of producing a measurement (Standard Guide for Measurement Systems Analysis, 2010).

Precision- Closeness of repeated readings to each other. A random error component of the measurement system (Measurement Systems Analysis, 2010).

Repeatability- Variation in measurements obtained with one measuring instrument when used several times by an appraiser while measuring identical characteristics on the same part (Measurement Systems Analysis, 2010).

Reproducibility- Variation in the average of the measurements made by different appraisers using the same gauge when measuring a characteristic on one part (Measurement Systems Analysis, 2010).

Stability (gage) - Absence of a change, drift, or erratic behavior in bias over a period of time (Standard Guide for Measurement Systems Analysis, 2010).

Variable Data- Quantitative data, where measurements are used for analysis. Examples include diameter of a bearing journal in millimeters, or torque of a fastener in Newton-meters (Omdahl, 1997).

Assumptions

An assumption of the study is that the underlying distribution from which the parts used in a Gauge R&R study is normally and independently distributed. The normality assumptions are typical in measurement systems analysis as described in the *Measurement Systems Analysis* manual (Measurement Systems Analysis, 2010). A final assumption is that the typical structure of the Gauge R&R study is followed for the data used in the study. For example, two or three operators measuring five or ten parts, two or three times each. All of the Gauge R&R studies included in this study include three operators, measuring ten parts; three times each for a total of 90 measurements each. A presumption of the study that is not extensively tested is that the software used for the A&R and ANOVA methods is accurately estimating the true measurement precision error of the Gauge R&R study data and its repeatability, reproducibility and part-to-part subcomponents. This is because it is commercial off-the-shelf software. The software was checked against the Chapter 2 manually calculated results for the A&R and ANOVA methods and the results were within a reasonable level of rounding error. The comparison is in Table 2.8. The Microsoft Excel spreadsheet used for the EMP III method was verified against the example in Wheeler (2009) and compared favorably. Table 1.2 shows the results of the comparison.

Table 1.2

EMP III Results Comparison between Wheeler (2009) and Study Spreadsheet

Component of Variation	Wheeler (2009) Example	Study Spreadsheet
Repeatability (EV)	3.783	3.78250591
Reproducibility (AV)	4.296	4.296196957
Combined Repeatability and Reproducibility (GRR)	5.724	5.724042213
Part Variation (PV)	23.483	23.48270758
Total Variation (TV)	24.171	24.1702754
Repeatability Proportion	0.0245	0.024490409
Reproducibility Proportion	0.0316	0.03159404
Combined Proportion	0.0561	0.056084
Part Proportion	0.9438	0.943916

CHAPTER 2

A REVIEW OF RELATED LITERATURE

One of the earliest journal articles addressing the precision of measuring devices in a manufacturing environment was Frank Grubbs' *On Estimating Precision of Measuring Instruments and Product Variability*, written in 1948 (Grubbs, 1948). Grubbs (1948) describes a measurement or observed value as being comprised of an absolute value of the characteristic being measured and an error of measurement component. Additionally, Grubbs (1948) describes the importance of understanding measurement error. For example, excessive measurement error can call into question the accuracy and usefulness of the reported results from a measurement process. Grubbs (1948) was also one of the first to describe measurement precision error in terms of reproducibility. Grubbs (1948) describes assumptions surrounding measurement error determination that are current today, such as the concept that there is no correlation between the errors of measurement and the values being measured over the limited range of measurements. Additionally, Grubbs (1948) suggests that errors of measurements are normally distributed and that measured values and the measurement error components are independent of one another.

Grubbs (1948) describes methods of partitioning measurement errors between measurement instruments and the actual measured values. What appears to be missing when compared to more modern work on the subject is recognition of the various operators' contributions to measurement precision error. Current theory partitions measurement precision error estimates into reproducibility (operator-to-operator) error component and repeatability (within operator) error component. Grubbs (1948) does not explicitly identify an acceptability criterion for measurement precision error, but does explain that measurement precision error should be appreciably smaller than the process variation of the characteristic being measured. Additionally, he points out that the relationship of measurement precision error to the characteristic being measured depends on the purpose of the measurements and the cost of ensuring measurement precision error is small compared to the characteristic being measured.

A second early article was written by Jack Gantt of General Electric Company in 1959 (Gantt, 1959). This article is one of the earliest commentaries that approaches the modern methods of Gauge R&R studies and defines the “gray area” of measurement precision error near the tolerance limits of a process. The gray area is the region around a dimensional tolerance limit that is created due to measurement error. Within the gray area, a product could either be accepted or rejected due to the variation caused by measurement precision error. Gantt (1959) also points out that in order to contend with the gray area, manufacturers must use process limits that are tighter than blueprint tolerances to avoid false rejection of parts. Gantt (1959) is also one of the first to mention that a Gauge R&R error percentage of 10% of the tolerance range is an acceptable level of measurement precision error, although he does not provide a justification for this percentage. A final important contribution by Gantt (1959) is commentary on determining if the gauge discrimination is adequate. As other authors such as Wheeler (2006) and Ermer (2006) point out, the discrimination of the gauge plays an important role in Gauge R&R studies. Wheeler and Lyday (1989) indicate that inadequate measurement discrimination will contribute to improper reporting of the variation present in a process.

Another early article was authored by Robert Traver also of General Electric Company in the American Society for Quality Control’s 1962 Annual Convention Transactions (Traver, 1962). Traver (1962) describes the importance of accurate measurement and the tendency for users of measurement data to accept numbers as accurate, without question. Traver (1962) takes Grubbs' (1959) work on Gauge R&R studies further and explains the preparations and steps for Gauge R&R studies, many of which are standard today. Traver (1962) is also one of the first to include measurement precision error as a percentage of the product tolerance among the methods of reporting the results of a Gauge R&R study. In addition to Gauge R&R, Traver (1962) discusses gauge bias and stability. Like Gantt (1959), Traver (1962) describes the gray area around product tolerances and points out that the greater the measurement precision error, the greater the probability of rejecting conforming parts and accepting non-conforming parts.

Automotive Industry Action Group Methods of Gauge R&R

The Automotive Industry Action Group (AIAG) is a non-profit industry organization formed to serve the interests of the automotive industry (AIAG History Highlights, n.d.). In 1990, AIAG published their *Measurement Systems Analysis Reference Manual*, which was authored by representatives of Chrysler Corporation, Ford Motor Company, General Motors Corporation and the American Society for Quality Control's (ASQC, now ASQ) Supplier Requirements Task Force (Measurement Systems Analysis, 1990).

The book's original purpose was to aid automotive industry suppliers in meeting the combined requirements of Chrysler, Ford and General Motors for measurement systems credibility (Measurement Systems Analysis, 1990). The reference manual today, currently in its fourth edition, has become the standard guide for performing measurement system analysis as noted by Knowles et al. (2000) and van den Heuvel (2000). The manual covers the many aspects of measurement systems analysis, including gauge accuracy, gauge stability, gauge linearity, as well as gauge reproducibility and repeatability (Measurement Systems Analysis, 2010). The manual also covers attribute measurement system analysis.

Specific to estimating measurement precision error, the manual describes methods of estimation that include calculations of measurement precision error standard deviation and the percentage of the process variation consumed by the measurement precision error. In addition, the percentage of the tolerance consumed by the measurement precision error (precision-to-tolerance ratio) is calculated as well as an indication of the number of distinct categories of discrimination the measurement system can resolve. The manual includes acceptance criteria for these results and guidance for performing a Gauge R&R study. Also included are interpretations of the various charts that are generated by the various methods and suggestions for what mitigations are available if a measurement system study fails the acceptance criteria (Measurement Systems Analysis, 2010).

With respect to measurement precision error estimation (Total GR&R), the manual covers three methods of analysis; the range method, the A&R method, and the ANOVA method. The range method only provides information on total gauge error not decomposed additionally into sub categories of repeatability, reproducibility and part-to-part (Measurement Systems Analysis, 2010). The range method described in the manual involves the collection of data by identifying a set of production parts to be measured by multiple operators once each. The example in the manual used five parts and two operators measuring the parts. Like the name implies, the range method is based on the differences in ranges from repeated measurements of the same parts by multiple operators. The average range (average difference among repeat measurements by multiple operators) is multiplied by an adjustment factor based on sample size, then multiplied by a 99% factor and then taken as a percentage of process variation or dimensional tolerance. The answer is expressed as a percentage that estimates the percentage of the measurement precision error contribution to process variation or process tolerance. As noted in the manual, the range method is intended for a quick approximation of measurement precision error (Measurement Systems Analysis, 2010).

The second method described in the Measurement Systems Analysis (2010) manual of estimating measurement precision error is the A&R method. This method is able to estimate the repeatability, reproducibility and part-to-part contributions of the measurement precision error as well as the total measurement precision error. In this method, multiple production parts are selected and measured by multiple operators multiple times. In the example provided in the manual, ten parts are measured by three operators, two times each. This process collects information on operator-to-operator measurement precision error as well as within-operator measurement precision error. In this method, the data is analyzed through average and range charts, which are typical charts used in statistical process control. The manual provides information on interpretation of the charts and calculation of the equipment variation (EV) component and appraiser variation (AV) component of the measurement precision error. The inherent variation between the parts, or part variation (PV) used in the study is also estimated. The EV component is the repeatability or within-operator contribution to variation, the AV component is the reproducibility or operator-to-operator contribution to variation. Additionally, these components are estimates of the respective standard deviations of the measurement components. Like the range method these estimates are then reported as a percentage of the process variation or the process tolerance (Measurement Systems Analysis, 2010).

The A&R Method in Detail

The operator in a Gauge R&R study is the person making or orchestrating the measurement of the part. This person is also known as the appraiser or inspector. Repeatability is the within-operator variation or the variation in measurement results from the same operator making repeated measurements of the same part. It is also known as Equipment Variation (EV). Reproducibility is the operator-to-operator variation, or the variation from different operators measuring the same parts, also known as Appraiser Variation (AV). Part-to-part variation is the component of variation in a Gauge R&R study attributed to the differences in the parts used in the study (PV). In other words, the parts measured in a Gauge R&R study will not be identical; part-to-part variation is a measure of the inherent differences in these parts. A typical Gauge R&R study has multiple operators measuring multiple parts multiple times to achieve the repeatability and reproducibility estimates (Measurement Systems Analysis, 2010).

The steps to the A&R method of Gauge R&R calculation are shown below and are linked to Table 2.1 Gauge R&R data. The equations are from Measurement Systems Analysis (2010). For purposes of demonstration this example is shorter than a typical ten-part Gauge R&R study where the ten parts are measured three times each by three operators. This example is comprised of five parts, measured by three operators, two times each; however, the methodology is the same regardless of the number of parts, operators and measurements. Footnotes in Table 2.1 are referenced in the steps.

Table 2.1

A&R Method Layout of Gauge R&R Study Data and Initial Calculations

	Part Number					
Operator 1	1	2	3	4	5	Average
Meas. #1	0.460	0.445	0.450	0.449	0.449	0.4506

Table 2.1 *(continued)*

	Part Number						
Operator 1	1	2	3	4	5	Average	
Meas. #2	0.460	0.446	0.451	0.449	0.449	0.4510	
Average ($\overline{X}$)	0.460	0.4455	0.4505	0.449	0.449	0.4508	$\overline{X}$ Op1- <=Overall[5]
Range (R)	0	0.001	0.001	0	0	0.0004	$\overline{R}$ Op1 Avg. <=Range[1]
Operator 2						Average	
Meas. #1	0.461	0.446	0.454	0.451	0.450	0.4524	
Meas. #2	0.461	0.446	0.454	0.450	0.450	0.4522	
Average ($\overline{X}$)	0.461	0.446	0.454	0.4505	0.450	0.4523	$\overline{X}$ Op2- <=Overall
Range (R)	0	0	0	0.001	0	0.0002	$\overline{R}$ Op2 Avg. <=Range[2]
Operator 3						Average	
Meas. #1	0.460	0.445	0.451	0.452	0.452	0.4520	
Meas. #2	0.461	0.447	0.452	0.451	0.452	0.4526	
Average ($\overline{X}$)	0.4605	0.446	0.4515	0.4515	0.452	0.4523	$\overline{X}$ Op3 <=Overall[4]
Range (R)	0.001	0.002	0.001	0.001	0.000	0.0010	$\overline{R}$ Op3 Avg. <=Range[3]
Part Average	0.461	0.446	0.452	0.450	0.450	0.0147	Part Range <=(R_p)

1. Calculate the overall average range, $\overline{\overline{R}}$:

$$\overline{\overline{R}} = \left(\left[\overline{R}_{Op1}{}^{1} = 0.0004\right] + \left[\overline{R}_{Op2}{}^{2} = 0.0002\right] + \left[\overline{R}_{Op3}{}^{3} = 0.0010\right]\right) / \left[No.ofOperators = 3\right] = 0.000533$$

2. Determine $\overline{X}_{Diff.}$: $\left[MaxOverall\,Op\overline{X}^{\,4} = 0.4523\right] - \left[MinOverall\,Op\overline{X}^{\,5} = 0.4508\right] = \overline{X}_{Diff} = 0.0015$

3. Determine factor D_4. For 2 trials, D_4=3.27 from Measurement Systems Analysis (2010).

4. Determine UCL_R= $\left[\overline{\overline{R}} = 0.000533\right] \times \left[D_4 = 3.27\right] = 0.001743$

5. Determine K_1 factor. For 2 trials, K_1=0.8862 from Measurement Systems Analysis (2010).

6. Calculate repeatability (EV):

Repeatability (EV) = $\overline{\overline{R}} \times K_1 = 0.000533 \times 0.8862 = 0.000472$

7. Determine K_2 factor. For 3 operators, K_2=0.5231 from Measurement Systems Analysis (2010).

8. Calculate Reproducibility (AV):

Reproducibility (AV) = $\sqrt{\left(\overline{X}_{Diff} \times K_2\right)^2 - \left(EV^2 /(nr)\right)}$ n= no. of parts, r= no. of trials.

$$= \sqrt{(0.0015 \times 0.5231)^2 - \left(0.000472^2 / 10\right)} = 0.000770$$

9. Calculate Gauge R&R:

$$\text{GRR}=\sqrt{EV^2 + AV^2} = \sqrt{0.000472^2 + 0.000770^2} = 0.000903$$

10. Determine K_3 factor. For 5 parts, K_3= 0.4030 from Measurement Systems Analysis (2010).

11. Calculate Part-to-part variation (PV):

$$PV = R_p \times K_3 = 0.0147 \times 0.4030 = 0.00592$$

12. Calculate total variation (TV):

$$TV = \sqrt{GRR^2 + PV^2} = \sqrt{0.000903^2 + 0.00592^2} = 0.00599$$

13. Calculate the percent contribution for repeatability (EV):

$$\%EV = 100[EV / TV] = 100[0.000472 / 0.00599] = 7.88\%$$

14. Calculate the percent contribution for Reproducibility (AV):

$$\%AV = 100[AV / TV] = 100[0.000770 / 0.00599] = 12.85\%$$

15. Calculate the percent contribution for Gauge R&R (GRR):

$$\%GRR = 100[GRR / TV] = 100[0.000903 / 0.00599] = 15.08\%$$

16. Calculate the percent contribution for Part-to-part (PV):

$$\%PV = 100\left[PV / TV\right] = 100\left[0.00592 / 0.00599\right] = 98.83\%$$

The third version of estimating the measurement precision error described in Measurement Systems Analysis (2010) is the ANOVA method. ANOVA analysis is a standard statistical technique for analyzing sources of variability and is employed in this method of analysis of measurement precision error. The manual suggests this version is the preferred method of analysis when a computer is available for the calculations. This preference is because the ANOVA method further breaks down the reproducibility component of variation into the interaction between parts and operators. The manual also suggests the ANOVA method is more accurate than the A&R method, but does not go into the detail behind this assertion (Measurement Systems Analysis, 2010). Like the A&R method, the ANOVA method prescribes performing the study by selecting multiple production parts and measuring a feature with multiple operators multiple times. The example in the manual shows three operators measuring ten parts three times each. Similar to the A&R method the components of variation are broken down into equipment variation (EV), appraiser variation (AV) and part variation (PV). In addition, an interaction component of variation between operators and parts is identified if present. The percentage contribution of the components of each type of variation is calculated (Measurement Systems Analysis, 2010).

The ANOVA Method in Detail

The steps to ANOVA method of Gauge R&R analysis are described in the following paragraphs and tables. This example describes the ANOVA analysis for five parts measured by three operators two times each. For comparison, this Gauge R&R data is the same as the data used for the A&R example previously shown in this chapter. The ANOVA method calculations and equations are from Measurement Systems Analysis (2010). Table 2.2 shows the basic Gauge R&R measurement results. Footnotes in Table 2.2 are keyed to the steps below.

Table 2.2

ANOVA Method Layout of Gauge R&R Study Data and Initial Calculations. Footnotes keyed to steps below

	Part Number						
Operator 1	1	2	3	4	5	Average	Meas. Sum by Operator[1]
Meas. #1	0.460	0.445	0.450	0.449	0.449	0.4506	4.508
Meas. #2	0.460	0.446	0.451	0.449	0.449	0.4510	
Operator 2						Average	Meas. Sum by Operator[1]
Meas. #1	0.461	0.446	0.454	0.451	0.450	0.4524	4.523
Meas. #2	0.461	0.446	0.454	0.450	0.450	0.4522	
Operator 3						Average	Meas. Sum by Operator[1]
Meas. #1	0.460	0.445	0.451	0.452	0.452	0.4520	4.523
Meas. #2	0.461	0.447	0.452	0.451	0.452	0.4526	
Sum by Part	2.763	2.675	2.712	2.702	2.702		
Sum by Part Squared[2]	7.63416	7.15562	7.35494	7.30080	7.30080		

1. Develop the ANOVA table by calculating the sum of squares total. This task is done by squaring each of the measured values and summing them. Table 2.3 shows a table of the squared measured values from Table 2.2 and their summation. The sum of squared values is 6.124426.

Table 2.3

Squared Gauge R&R Measured Values and Their Summation. The values in Table 2.3 are the square of the same values in Table 2.2

	Parts				
Operator	**1**	**2**	**3**	**4**	**5**
1	0.2116	0.198025	0.2025	0.201601	0.201601
	0.2116	0.198916	0.203401	0.201601	0.201601
2	0.212521	0.198916	0.206116	0.203401	0.2025
	0.212521	0.198916	0.206116	0.2025	0.2025
3	0.2116	0.198025	0.203401	0.204304	0.204304
	0.212521	0.199809	0.204304	0.203401	0.204304
		Sum of Squares=	6.124426		

2. The measurements from each operator are summed then squared. The summation of all operator measured values is 13.554. This total squared is 183.7109. Total Sum of Squares (TSS) is then calculated from the results of step 1 and 2.

$$TSS = \sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{m=1}^{r}\left(x_{ijm}^{2}\right) - \frac{x_{\dots}^{2}}{nkr} = 6.124426 - \frac{183.7109}{30} = 0.000729$$

where, n= 5 (number of parts), k= 3 (number of operators), r= 2 (replicate meas.)

3. The operator (appraiser) sum of squares is calculated next by squaring each operator's total measurement value[1]. These values are 4.508, 4.523 and 4.523, respectively, for each operator

and squared values for each operator are 20.322, 20.4575 and 20.4575, respectively. The sum of these operator squared values is 61.237122. The operator sum of squares is then calculated:

$$SS_A = \sum_{j=1}^{k}\left(\frac{x_{.j.}^2}{nr}\right) - \frac{x_{...}^2}{nkr} = \left(\frac{61.237122}{10}\right) - \frac{183.7109}{30} = 0.00001553$$

where, n= 5 (number of parts), k= 3 (number of operators), r= 2 (replicate meas.)

4. The part-to-part sum of squares is calculated by summing and squaring the total measurements for each part[2]. The summed and squared value for each part is 7.63419, 7.155625, 7.354944, 7.300804, and 7.300804, respectively, for parts one through five. The sum of these values is 36.74634. The Part-to-part sum of squares is next calculated:

$$SS_P = \sum_{i-1}^{n}\left(\frac{x_{i..}^2}{kr}\right) - \frac{x_{...}^2}{nkr} = \frac{36.74634}{6} - \frac{183.7109}{30} = 0.0006933$$

where, n= 5 (number of parts), k= 3 (number of operators), r= 2 (replicate meas.)

5. The operator-by-part interaction term is next calculated. First, each operator's replicate measurements for each part are totaled. These values are represented in the upper half of Table 2.4 below. The values in the lower half of Table 2.4 are the square of the values in the upper half of the table.

Table 2.4

Sum of Measured Values by Operator (upper half of table) and Squares of the Sums of the Measured Values (lower half of table

	Part Measurement Totals				
Operator	**1**	**2**	**3**	**4**	**5**
1	0.92	0.891	0.901	0.898	0.898
2	0.922	0.892	0.908	0.901	0.9

Table 2.4 *(continued)*

	Part Measurement Totals				
Operator	**1**	**2**	**3**	**4**	**5**
3	0.921	0.892	0.903	0.903	0.904

	Part Measurements Totals Squared				
Operator	**1**	**2**	**3**	**4**	**5**
1	0.8464	0.793881	0.811801	0.806404	0.806404
2	0.850084	0.795664	0.824464	0.811801	0.81
3	0.848241	0.795664	0.815409	0.815409	0.817216

Sum of Squared Values= 12.248842

The operator-by-part interaction sum of squares is then calculated:

$$SS_{AP} = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\frac{x_{ij.}^2}{r}\right) - \sum_{i=1}^{n}\left(\frac{x_{i..}^2}{kr}\right) - \sum_{j=1}^{k}\left(\frac{x_{.j.}^2}{nr}\right) + \frac{x_{...}^2}{nkr} =$$

$$= \frac{12.248842}{2} - \frac{36.74634}{6} - \frac{61.237122}{10} + \frac{183.7109}{30} = 0.0000154$$

where, n= 5 (number of parts), k= 3 (number of operators), r= 2 (replicate meas.)

The equipment (EV) or repeatability measurement precision error is determined from the results of the previous calculations:

$$SS_e = TSS - \left[SS_A + SS_P + SS_{AP}\right] =$$

$$= 0.000729 - \left[0.00001553 + 0.0006933 + 0.0000154\right] = 0.0000048$$

6. The ANOVA table is created by summarizing the previous calculations in tabular form. Table 2.5 is the applicable ANOVA table. According to Measurement Systems Analysis (2010), the F statistic for the interaction term is the only F value calculated. This calculation is to determine if the interaction is significant. The F critical value for 0.05 significance and eight numerator

degrees of freedom and 15 denominator degrees of freedom is 2.64 (Devore, 2004). The F calculated value of 6.03 exceeding F critical value of 2.64 indicates that the interaction of operator-by-part is significant in this Gauge R&R study. The P-value of this statistic can be estimated by referencing Devore (2004). For eight numerator degrees of freedom and 15 denominator degrees of freedom the F calculated value of 6.03 falls between an alpha value of 0.01 and 0.001. This means the P-value is less than 0.01. When the interaction is not significant in the ANOVA method the results are rolled into the Equipment source of variation (Measurement Systems Analysis, 2010).

Table 2.5

ANOVA Table for Gauge R&R Study

Source	**Degrees of Freedom**	**Sum of Squares (SS)**	**Mean Square (MS)**	**F**
Appraiser	2 (k-1)	0.00001553	0.00000777	
Part	4 (n-1)	0.0006933	0.00017333	
Appraiser X Part	8 (n-1)(k-1)	0.0000154	0.00000193	6.03
Equipment	15 (r-1)	0.0000048	0.00000032	
Total	29 (nkr-1)	0.000729		

7. From the ANOVA table and the equations below from Measurement Systems Analysis (2010), the Gauge R&R table in Table 2.6 is constructed. The % Total Variation column in Table 2.6 is comparable to the final results of the EMP III and A&R examples in this chapter.

$$\tau^2=MS_e=0.00000032$$

$$\gamma^2 = \frac{MS_{AP} - MS_e}{r} = \frac{0.00000193 - 0.00000032}{2} = 0.00000081$$

$$\omega^2 = \frac{MS_A - MS_{AP}}{nr} = \frac{0.00000777 - 0.00000193}{10} = 0.000000584$$

$$\sigma^2 = \frac{MS_P - MS_{AP}}{kr} = \frac{0.00017333 - 0.00000193}{6} = 0.0000286$$

Table 2.6

Gauge R&R Study Results Table

Estimate of Variance (σ^2)	**Standard Deviation (σ)**	**6(σ)**	**% Total Variation**	**% Contribution**
τ^2= 0.00000032 (Equipment, EV, Repeatability)	0.00056569	EV= 0.003394	10.27	1.06
Reproducibility (ω^2+ γ^2)= 0.000001394	0.00118068	0.00708408	21.44	4.59
ω^2= 0.000000584 (Appraiser, AV)	0.00076420	AV=0.00459	13.87	1.93
γ^2= 0.00000081 (Interaction, Operator X Part)	0.0009	Int= 0.0054	16.35	2.67
GRR Total (τ^2+ ω^2+ γ^2)= 0.000001714	0.0013092	GRR= 0.0078552	23.78	5.65

Table 2.6 *(continued)*

Estimate of Variance (σ^2)	**Standard Deviation (σ)**	**6(σ)**	**% Total Variation**	**% Contribution**
Part-to-part (PV) σ^2= 0.0000286	0.005347897	0.032087381	96.77	94.35
Total Variation (GRR+σ^2)= 0.000030314	0.005505815	0.033034891	100.00	100.00

Barrentine (2003) also describes methods for performing Gauge R&R studies that correspond with the methods described by the Measurement Systems Analysis (2010). Barrentine (2003) describes two methods; the first is similar to the A&R method described in Measurement Systems Analysis (2010). Barrentine describes this method as the General Motors Long Form. Like Measurement Systems Analysis (2010), Barrentine (2003) prescribes making multiple measurements from multiple production parts multiple times. In his example he demonstrates measurements from ten parts, made by two operators with two measurements each. Barrentine's calculations produce measurement precision error standard deviations for equipment variation (EV) and appraiser variation (AV) and total measurement precision error (Barrentine, 2003). Barrentine (2003) departs from the Measurement Systems Analysis (2010) method by discouraging taking the total Gauge R&R measurement precision error estimate as a percentage of the part tolerance-the precision-to-tolerance ratio. Barrentine (2003) suggests that most tolerances are determined arbitrarily, and thus the precision to tolerance ratio is not an accurate assessment of measurement precision error. A second concept Barrentine (2003) emphasizes is that the measurement precision error estimate should not be taken as a percentage of the part variation included in the study, but rather as a percentage of an estimate of the process variation. He explains that parts selected for a Gauge R&R study represent the range of the measurement system and thus are not large enough to represent the typical variation of the process. He suggests a better assessment is the measurement precision error as a percentage of the process variation (Barrentine, 2003).

Some authors argue the ANOVA approach to the Gauge R&R study is superior to the A&R method. Antony, Knowles and Roberts (1998) and Kazerouni (2009) suggest the ANOVA method is more accurate than the A&R method in the presence of operator and part interaction in a Gauge R&R study. This suggestion of increased accuracy is because the A&R method underestimates the reproducibility component of interaction between the operators and parts if it is present in the study (Antony et al., 1998). Operator-by-part interaction means some or all operators participating in a Gauge R&R study measured some parts differently in the study; for example, there would be more variation when measuring the smaller parts in the study than the larger parts in the study. Antony et al. (1998) suggest using an ANOVA analysis of the Gauge R&R data to determine if operator-by-part interaction is significant and if so, to use the ANOVA analysis. If not significant Antony et al. (1998) suggest pooling the interaction variance with the error variance. Antony et al. (1998) summarize their concern by suggesting that if the operator-by-part interaction is present the ANOVA method will identify this fact and provide Gauge R&R users an avenue to investigate that would be unknown with the A&R method.

Another advantage of the ANOVA method of Gauge R&R study is described by Burdick, Borror, and Montgomery (2003). Burdick et al. (2003) suggest that an advantage of the ANOVA method is the ability to calculate confidence intervals on the results of the Gauge R&R study.

Burdick et al. (2003) also point out that the A&R method of Gauge R&R study is only comparable with two-factor ANOVA design of a Gauge R&R study. Burdick et al. (2003) suggest that a Gauge R&R study is in reality a designed experiment and analyzing such with the ANOVA methods facilitates more sophisticated designs and analysis types. For example, an analysis could be done by adding another source of variation to the study, such as multiple measuring instruments.

At least one author recently disagreed with the superiority of the ANOVA method for Gauge R&R analysis however. Osma (2011) describes research on Gauge R&R studies in an automotive application in which the A&R method and ANOVA method were compared in three studies. The three studies employed the conventional design of a Gauge R&R study with three operators, ten parts and three measurements per operator for a total of 90 measurements (Osma, 2011). In the first study, the A&R method and the ANOVA method agreed that the measurement system had unacceptable measurement precision error. In the second study, the results between the A&R method and the ANOVA method differed. The A&R method found the measurement system acceptable, while the ANOVA method found the measurement system unacceptable (Osma, 2011). A diagnosis of the ANOVA residuals for the second study-a common ANOVA check for validity-revealed that the residuals were not normally distributed and therefore, Osma (2011) concluded that the ANOVA results were not valid and the results of the A&R method were more reliable. In the third study, the Gauge R&R results calculated both by the A&R method and the ANOVA method were acceptable by the AIAG criteria, but showed dramatically different results (Osma, 2011). The third study ANOVA residuals were also analyzed and again showed non-normality and so again Osma (2011) concluded the A&R method to be more accurate. In Osma (2011) the acceptance or rejection of a Gauge R&R study was based on the Measurement Systems Analysis (2010) criteria, which provides a range of acceptability.

One concern with Osma's conclusions that he does not address is whether the ANOVA method, with its residual analysis, might be telling the researcher something about the data that the A&R method is not, and thus explain why one method is showing acceptability while the other is not. That is, just because the A&R method is not speaking to the non-normality of the residuals does not mean it is automatically a more accurate or reliable method.

EMP III Methods

Wheeler (2006) describes alternate methods of estimating the measurement precision error of a measuring device. Known as EMP methods (for Evaluating the Measurement Process), his methods include a Short EMP Study, a Basic EMP Study, a Two Factor EMP Study and a version of a traditional Gauge R&R Study (Wheeler, 2006). In Wheeler's Short EMP study, only one operator measures multiple parts with one instrument and in this process only the precision and the part variation are identified (Wheeler, 2006). Wheeler (2006) also introduces the Interclass Correlation Coefficient, which is the ratio of the estimated product variance to the total variance of the product measurement. The significance of the Interclass Correlation Coefficient is that it represents the proportion of variation that is attributed to the product, while one minus the Interclass Correlation Coefficient is the proportion of the variation attributed to measurement precision error (Wheeler, 2006). In some ways, the Short EMP study is similar to the range method of measurement precision error estimation described by Measurement Systems Analysis (2010), in that it is a quick method that does not break down the measurement precision error estimate into reproducibility and repeatability components.

Wheeler's Basic EMP Study captures not only part variation in a measurement study, but also what he calls the "nuisance factors" of test-retest error (precision error) and biases due to operators (Wheeler, 2006). The structure of the Basic EMP Study is similar to a Gauge R&R study in that multiple operators measure multiple parts multiple times. In Wheeler's Basic EMP Study, an average and range chart is created similar to the average and range chart created in the AIAG method. Wheeler (2006) encourages interpretation of the average chart differently in that the reader compares the similarity of the average chart by operator. In addition, Wheeler (2006) takes the analysis further by creating an Analysis of Means (ANOME) chart to statistically determine any bias differences among the operators. Further, Wheeler (2006) advocates creating a mean range chart (ANOMR) to determine any statistical differences in the test-retest rates of the operators. Wheeler (2006) summarizes the Basic EMP Study as a measurement precision error study that is used to check the bias and test-retest error of a single nuisance component, such as an operator or a measurement instrument (Wheeler, 2006).

Wheeler (2006) describes a Two-Factor EMP Study that captures two-factor influence in a measurement precision error study; for example, different operators, making multiple measurements with different instruments of a set of products. Similar to the Basic EMP Study, the Two-Factor EMP Study results in the creation of an average and range chart and the interpretation is very similar to the Basic EMP Study (Wheeler, 2006). If the researcher sees differences in the operator and instrument results in the charts, ANOME charts for operator or instrument can be created to determine if the differences are statistically significant (Wheeler, 2006). Similarly, if the test-retest values appear different on the range chart, ANOMR charts can be created to determine statistical differences in the precision estimates of operators and instruments (Wheeler, 2006).

Finally, Wheeler (2006) also addresses the traditional Gauge R&R study. He describes a Gauge R&R methodology known as an "Honest Gauge R&R Study". Wheeler (2006) takes issue with how the traditional AIAG Gauge R&R method adds standard deviation values, which is considered mathematically incorrect. The data collection phase of Wheeler's method of a Gauge R&R study is identical to a traditional Gauge R&R study (Wheeler, 2006). This similarity means multiple operators, measure multiple parts, multiple times. Wheeler's method is different from a traditional AIAG Gauge R&R study in that the components of measurement precision error sum to the total amount of variation (Wheeler, 2006). Wheeler's "Honest Gauge R&R Study" is the method utilized in this study and is designated the EMP III method in this study.

Another difference in Wheeler's teaching is the criteria for an acceptable Gauge R&R study. The AIAG method provides guidelines on acceptability; for example, under 10% measurement precision error is acceptable, 10% to 30% may be acceptable depending on the application and over 30% error is unacceptable (Measurement Systems Analysis, 2010). Wheeler is critical of the AIAG guidelines because no rationale is provided for how the numbers were derived or why they are appropriate (Wheeler, 2006). Wheeler's method of determining the acceptability of the measurement precision error is based on the Interclass Correlation Statistic and the class of monitor (Wheeler, 2006). For example, according to Wheeler (2006), an Interclass Correlation Statistic greater than .80 is a first class monitor and has a better then 99% chance of detecting a three standard error shift within ten subgroups of data collection when it occurs. Wheeler (2006) has similar rules for second, third and fourth class monitors based on decreasing Interclass Correlation Statistics and resulting in lowered chances of detecting shifts in a process. Wheeler (2006) leaves it up to the reader to determine which class of monitor is applicable for a particular measurement application.

Wheeler is not alone in his criticism of the AIAG method of Gauge R&R study. Ermer (2006) also finds errors with the AIAG method. He explains that the first error is with the calculation of the part-to-part variation component of the total measurement precision error, in that a correction factor that should be used is not employed (Ermer, 2006). Interestingly, Barrentine (2003) avoids this problem by not including part variation (PV) in his book. Technically, part variation is not needed for the total Gauge R&R statistic, as it is the total variation contribution by the parts in the study, not the measurement precision error, although most authors include this value in Gauge R&R study reports. Ermer (2006) also agrees with Wheeler (2006) that the final variation ratios are calculated by summing standard deviations rather than correctly summing variances. Ermer (2006) provides the correct calculations and notes that due to the errors, the AIAG method exaggerates the contributions of the components of repeatability, reproducibility, and part-to-part measurement precision error as well as total measurement precision error.

Knowles et al. (2000) also criticize the AIAG method of measurement precision error estimation. Like Barrentine (2003), Knowles et al. (2000) indicate the AIAG method of relying on the precision-to-tolerance ratio, as an acceptance criterion is weak, because tolerances can be arbitrarily set. Knowles et al. (2000) agree with Wheeler (2006) regarding the incorrectness of summing standard deviations rather than variances when reporting the proportion results of the Gauge R&R study. Knowles et al. (2000) also highlight the same error pointed out by Ermer (2006) regarding the part variation calculation in the AIAG method. Even though these errors are documented in the literature, they are not corrected in the fourth edition of the AIAG measurement systems analysis manual.

The EMP III Method in Detail

The steps to the EMP III method of "Honest Gauge R&R" are below and are footnoted to Table 2.7. The equations are from Wheeler (2006). The Gauge R&R measurement data is the same as used in the previous A&R and ANOVA examples in this chapter.

Table 2.7

EMP III Method Layout of Gauge R&R Study Data and Initial Calculations. Foot Notes keyed to steps below

	Part Number						
Operator 1	**1**	**2**	**3**	**4**	**5**		
Meas. #1	0.460	0.445	0.450	0.449	0.449		
Meas. #2	0.460	0.446	0.451	0.449	0.449		
Average ($\bar{X}$)[1]	0.46	0.4455	0.4505	0.449	0.449	0.4508	($\bar{\bar{X}}$) Grand <=Avg[4]
Range (R)[2]	0	0.001	0.001	0	0	0.0004	<= $\bar{R}$ Avg. Range
Operator 2							
Meas. #1	0.461	0.446	0.454	0.451	0.450		
Meas. #2	0.461	0.446	0.454	0.45	0.450		
Average ($\bar{X}$)[1]	0.461	0.446	0.454	0.4505	0.450	0.4523	($\bar{\bar{X}}$) Grand <=Avg.[4]
Range(R)[2]	0	0	0	0.001	0	0.0002	<= $\bar{R}$ Avg. Range
Operator 3							
Meas. #1	0.460	0.445	0.451	0.452	0.452		
Meas. #2	0.461	0.447	0.452	0.451	0.452		
Average ($\bar{X}$)[1]	0.461	0.446	0.452	0.452	0.452	0.4523	($\bar{\bar{X}}$) Grand <=Avg.[4]
Range (R)[2]	0.001	0.002	0.001	0.001	0.000	0.0010	<= $\bar{R}$ Avg. Range
Part Average[5]	0.461	0.446	0.452	0.450	0.450	0.0147	Part Range <=(R_p)[6]

Table 2.7 *(continued)*

Average of operator part ranges[3]	0.000533

1. Each operator's average is calculated for each part measurement[1].

2. The range of each operator's measurements is calculated (R)[2].

3. The average of all operator part ranges is calculated[3].

4. Determine the appropriate d_2 value (Bias correction factor) from look-up table A.1 (Wheeler, 2006). d_2 value=1.128. (d_2 value for n=2 measurement repetitions from table A.1).

5. Calculate the EMP III repeatability standard deviation (EV) by dividing the overall average range (from step 3) by the d_2 bias correction factor (from step 4), 0.000533/1.128= 0.000473.

6. The average of each operator's average measurements is calculated (Grand Average $\overline{\overline{X}}$)[4].

7. The range of each operator's Grand Average measurements is calculated. Range of Grand Average measurements is: 0.4523-0.4508= 0.0015, (R_o).

8. Determine the appropriate d_2* value; d_2* value=1.906 for n=3 measurement repetitions, K=1 from table A.2 (Wheeler, 2006).

9. Calculate EMP III reproducibility standard deviation (AV) from (Wheeler, 2006):

$$AV = \tilde{\sigma}_o = \sqrt{\left[\frac{R_o}{d_2 *}\right]^2 - \frac{o}{npo}\tilde{\sigma}_{pe}^{\ 2}} = \sqrt{\left[\frac{0.0015}{1.906}\right]^2 - \frac{3}{2 \cdot 5 \cdot 3}0.000473^2} = 0.000773$$

where, AV= Appraiser Variation

R_o= range of operator averages

d_2*=look up value from Table A.2 in Wheeler (2006).

o=number of operators in the study

p=number of parts in the study

n=number of measurement replications

$\tilde{\sigma}_{pe}$ = repeatability standard deviation from step #5 above.

10. Calculate the combined repeatability and reproducibility standard deviation:

$$GRR = \tilde{\sigma}_e = \sqrt{EV^2 + AV^2} = \sqrt{0.000473^2 + 0.000773^2} = 0.000906$$

11. Calculate the average measured value of each part[5].

12. Calculate the range of the average part measurements, 0.0147 from Table 2.7 above[6].

13. Determine the d_2* value based on the number of parts:

d_2* value=2.477, d_2* value for n=5 parts, K=1 from table A.2 (Wheeler, 2006).

14. Calculate the part-to-part standard deviation estimate (PV):

$$PV = \tilde{\sigma}_p = \frac{R_p}{d_2*} = \frac{0.0147}{2.477} = 0.00593$$

where, R_p= (Average range of part values) 0.0147 from step #12 above.

d_2* from step #13 above.

15. Calculate total variation (TV) by summing the EV, AV, and PV standard deviation estimates:

$$TV = \sqrt{EV^2 + AV^2 + PV^2} = \sqrt{0.000473^2 + 0.000773^2 + 0.00593^2} = 0.00600$$

16. Calculate the percent contribution for repeatability (EV):

$$EV\% = 100\frac{EV^2}{TV^2} = \frac{0.000473^2}{0.00600^2} = 0.621\%$$

17. Calculate the percent contribution for Reproducibility (AV):

$$AV\% = 100\frac{AV^2}{TV^2} = \frac{0.000773^2}{0.00600^2} = 1.66\%$$

18. Calculate the percent contribution for Total Gauge R&R (GRR):

$$GRR\% = 100\frac{GRR^2}{TV^2} = \frac{0.000906^2}{0.00600^2} = 2.28\%$$

19. Calculate the percent contribution for Part-to-part (PV):

$$PV\% = 100\frac{PV^2}{TV^2} = \frac{0.00593^2}{0.00600^2} = 97.68\%$$

Table 2.8 provides a comparison of the results of the three example methods demonstrated in this chapter (EMP III, A&R, and ANOVA). The three methods all used the same Gauge R&R study data for the analysis.

Table 2.8

Comparison of the Three GR&R Method Examples in this Chapter and Minitab Results for the Same Data for A&R and ANOVA Methods

Category	**A&R**	**A&R by Minitab**	**ANOVA**	**ANOVA by Minitab**	**EMP III**
Repeatability	7.88%	7.75%	10.27%	10.50%	0.621%
Reproducibility	12.85%	12.90%	21.44%	20.99%	1.66%
Total GRR	15.08%	15.05%	23.78%	23.47%	2.88%
Part-to-part	98.83%	98.86%	96.77%	97.21%	97.68%

Other Methods

The prior literature reviewed in this chapter addresses the approach industry has taken to address the issue of measurement precision error estimation. Other disciplines have also addressed the issue. Perhaps the most well-documented of these methods arises from the scientific community. From a physical science and metrological standpoint measurement precision error is addressed as *measurement uncertainty*. The U. S. Department of Commerce's, National Institute of Standards and Technology (NIST) have published *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*, also known as Technical Note 1297. The authors of the guideline, Barry Taylor and Chris Kuyatt provide information on determining and reporting the uncertainty of measurement as well as definitions of terms (Taylor & Kuyatt, 1994). In this NIST publication the standard deviation of the measurement is termed *standard uncertainty* and is classified by one of two methods, Type A or Type B uncertainty (Taylor & Kuyatt, 1994). Type A measurement uncertainty is determined by statistical means such as calculating the mean and standard deviation of a set of measurements, using the method of least squares to fit a curve to data or performing an ANOVA analysis (Taylor & Kuyatt, 1994). Type B evaluation of measurement precision error uses non-statistical means such as, scientific judgment, experience or general knowledge of the behavior of the measurement process, manufacturer's specifications, and uncertainties taken from reference books (Taylor & Kuyatt, 1994).

Other aspects of the NIST approach include recognition of contributors to measurement precision error as random or systematic (Taylor & Kuyatt, 1994). According to Taylor and Kuyatt (1994), random components to measurement precision error are random effects, while systematic uncertainty is inherent to the measurement process. Hughes and Hase (2010) provide clearer insight to random and systematic measurement errors. Systematic measurement errors influence the accuracy of the measurement result, while random errors influence the precision of the measured result (Hughes & Hase, 2010). Another contribution of the NIST publication is the concept of the uncertainty budget (Taylor & Kuyatt, 1994). In an uncertainty budget, as many contributors to measurement precision error that can be identified are listed and uncertainty estimates for each are summed to determine the entirety of measurement precision error for a measurement process (Taylor & Kuyatt, 1994). Examples of components of measurement precision error include repeated observations, calibration, random effects, systematic effects, and environmental influences (Taylor & Kuyatt, 1994). The uncertainty budget components also include information on whether the estimates were based on Type A or Type B evaluations (Taylor & Kuyatt, 1994).

Standards

International standards such as ISO 9001, *Quality management systems- Requirements* frequently address measurement instrument calibration but do not address measurement uncertainty or specify methods to determine measurement precision error (ISO 9001, 2008). This omission is also true of ISO 13485, *Medical devices- Quality management systems- Requirements for regulatory purposes* (ISO 13485, 2003). Two exceptions, however, are ISO/TS 16949 and ISO 17025. ISO/TS 16949, *Quality management systems- particular requirements for the application of ISO 9001:2008 for automotive production and relevant service part organizations* does require measurement systems analysis (ISO/TS 16949, 2009). The requirement, however, does not specify a particular method; rather it leaves that determination to the contract between the manufacturer and the automotive supplier (ISO/TS 16949, 2009). ISO 17025, *General requirements for the competence of testing and calibration laboratories*, specifies calibration of measurement equipment and requires estimation of measurement uncertainty (ISO 17025, 2005). ISO 17025 does not specify a particular method of determining uncertainty of measurement (ISO 17025, 2005).

U.S. Government Requirements

In some cases, the U.S. Government, by regulation, requires estimates for the uncertainty of measurement. For example, the U.S. Environmental Protection Agency, in 40 CFR Part 58, Appendix A, requires the use of precision and bias estimates in measurement data. The specific application is monitoring of environmental air quality, and the measurement precision error is based on a coefficient of variation calculation (Camalier, Eberly, Miller & Papp, 2007). Another U.S. Government agency, the Nuclear Regulatory Commission, requires estimates of bias and measurement precision error in regulation 10 CFR 74.45 (10 CFR 74.45, 2012). This regulation requires appropriate statistical methodologies for determining contributors to measurement uncertainty, but does not specify a particular method (10 CFR 74.45, 2012). Three U.S. Food and Drug Administration (FDA) regulations-21 CFR 210, 21 CFR 211 for the manufacture of pharmaceuticals and 21 CFR 820 for the manufacture of medical devices-require calibration of instrumentation and gauging used in the manufacture of pharmaceuticals and medical devices and also require identification of limits for accuracy and precision of these devices and medicines. However, they do not specify a method for determining accuracy and precision estimates (21 CFR 820, 2011), (21 CFR 210, 2011, (21 CFR 211, 2011). While the FDA does specify calibration for measurement instrumentation for biologics (primarily blood collection and storage) in regulation 21 CFR 606, *Current Good Manufacturing Practice for Blood and Blood Components,* the requirement stops short of requiring an estimation of uncertainty in the measurement results (21 CFR 606, 2011).

A Review of Related Literature-Summary

References to measurement error estimation date back to the middle half of the twentieth century with articles by Grubbs (1948), Gantt (1959) and Traver (1962) providing the underpinnings of the methods in use today. In the late 1980s, the Automotive Industry Action Group (AIAG) under the guidance of the "Big Three" automotive companies and what was then the Supplier Requirements Task Force of the American Society for Quality Control developed and published the *Measurement Systems Analysis Reference Manual*, which currently is in its fourth edition and is the presumed standard method of Gauge R&R study. This manual describes three methods of Gauge R&R study as well as acceptance criteria for the methods. The first method described is the range method, which is a quick way of determining the overall measurement precision error in a measurement process. A second method described in Measurement Systems Analysis (2010) is the A&R method. This method breaks down the measurement precision error into repeatability, reproducibility, and part-to-part components. Finally, Measurement Systems Analysis (2010) describes the ANOVA method, which is a standard statistical analysis technique. The advantage of the ANOVA technique is it further breaks down the reproducibility error estimate into operator-by-part interactions. Many authors, such as Kazerouni (2009) and Antony et al. (1998), suggest the ANOVA method is the superior method of Gauge R&R study analysis.

Other methods of measurement precision error estimation are described by Wheeler (2006). Wheeler's EMP methods include a Short EMP study, a Basic EMP study, a Two-Factor EMP Study and his version of a traditional Gauge R&R study. Wheeler (2006) is critical of the AIAG method of Gauge R&R study due to the addition of standard deviations rather than variances, which is considered mathematically incorrect. Wheeler is not alone in this criticism as Ermer (2006) and Knowles et al. (2000) agree. Other criticisms of the AIAG method by these authors include the failure to rationalize the measurement precision error acceptance criteria and failure to adequately use a correction factor when estimating the part variation component of a measurement precision error study.

The scientific and metrology communities recognize measurement precision error, typically called *measurement uncertainty* in these circles. They support the use of statistical (Type A) estimates and non-statistical (Type B) estimates of measurement uncertainty (Taylor & Kuyatt, 1994). Statistical methods typically include calculating the mean and standard deviation of a set of measurements, using the method of least squares to fit a curve to data, or performing an ANOVA analysis (Taylor & Kuyatt, 1994). Non-statistical methods typically include scientific judgment, experience, or general knowledge of the behavior of the measurement process, manufacturer's specifications, and uncertainties taken from reference books (Taylor & Kuyatt, 1994). Another aspect of the scientific and metrological approach is the development of an uncertainty budget that lists all of the possible sources contributing to the uncertainty of a measurement. These sources can include such items as calibration, repeated measurement, random effects, systematic effects, and environmental effects (Taylor & Kuyatt, 1994). This approach appears more comprehensive than the typical Gauge R&R and can mix both Type A and Type B estimates within the same uncertainty budget.

Another aspect of measurement precision error estimation is from international quality and related standards. International quality standards such as ISO 9001 and ISO 13485, while requiring calibration of measurement instrumentation, do not mention requirements for estimates of measurement precision error. However, two international standards, ISO 17025 and ISO/TS 16949, require not only instrument calibration but also an estimate of measurement uncertainty and measurement systems analysis, respectively. These two standards stop short of actually specifying a method of estimating measurement precision error, however.

U. S. Government regulation, like the international standards, is somewhat uneven in the requirements for estimating measurement uncertainty. Some regulations, like 21 CFR 606 for biologics, only require instrument calibration (21 CFR 606, 2011). Other regulations like 21 CFR 210, 21 CFR 211 and 21 CFR 820 for pharmaceuticals and medical devices, require instrument calibration and also establishment of limits for accuracy and precision of the devices (21 CFR 210, 2011), (21 CFR 211, 2011), (21 CFR 820, 2011). These regulations do not, however, specify a method for establishing these limits. Regulations 10 CFR 74.45 and 40 CFR 58, Appendix A, require estimates of measurement uncertainty but also stop short of specifying a method (10 CFR 74.45, 2012), (Camalier et al., 2007).

CHAPTER 3

METHOD OF INVESTIGATION

Overview

The methods of estimating measurement precision error, including the EMP III, A&R, and ANOVA methods estimate the standard deviation of the measurement process. The standard deviation estimation includes the total measurement precision error, the contribution due to repeatability, and the contribution due to reproducibility. In addition, the standard deviations of the measurement of the parts used in the study are also calculated as the part-to-part variation. Indirectly these methods also estimate the variance, as it is the square of the standard deviation (Devore, 2004). Standard deviation and variance are considered in each of these methods because they are primary measures of variation in a process and therefore are included in typical Gauge R&R analysis methods (Devore, 2004). The EMP III, A&R, and ANOVA methods also determine the percent contribution of each of the components- total measurement precision error, repeatability, reproducibility and part-to-part- to measurement precision error.

The goal of the study is to compare the estimates of the total and components of measurement precision error for numerous Gauge R&R data sets via the EMP III, A&R, and ANOVA methods to answer the research questions. To derive the data sets to be used in the study, actual Gauge R&R studies from a Midwestern U.S.-based medical device manufacturer that machines metal orthopedic implants (artificial knees, hips, shoulders etc.) was utilized as a seed or base to generate additional data sets. The results for the EMP III, A&R, and ANOVA, analysis of the 224 seed Gauge R&R studies for total Gauge R&R measurement precision error is shown in Table 3.1 delineated by AIAG acceptance categories. The AIAG acceptance criterion for total measurement precision error (total Gauge R&R) is as follows (Measurement Systems Analysis, 2010):

- Greater than 30% is unacceptable
- Between 10% and 30% is acceptable based on rationalization
- Below 10% is acceptable.

Table 3.1

Comparison of Actual (Seed) Gauge R&R Study Total Gauge Error

Total Gauge R&R Percent Category	**EMP III**	**A&R**	**ANOVA**
< 10%	143	91	85
10% -30%	31	51	55
>30%	50	82	84
Total	224	224	224

Description of the Subjects

To determine sample size for the study, 224 actual Gauge R&R studies were provided by a medical device manufacturer located in the Midwestern U.S., for a sample size basis. These variable gauge studies all involved, ten parts, with three operators measuring the parts three times each for a total of 90 measurements each. The provided studies involved results from manual gauges such as calipers, micrometers and height gauges as well as measurements from CMM and surface measuring equipment.

Software for performing Gauge R&R studies generally falls into two categories; software that is a Microsoft Excel spreadsheet add-in, and statistical analysis software. Examples of the Excel spreadsheet add-in include QI Macros and SPC for Excel. Examples of statistical software that include Gauge R&R studies are Minitab and JMP. In some cases, the Excel add-ins and the statistical software will include the AIAG A&R and ANOVA methods as well as the EMP III method. Most of these software packages, particularly the Excel add-in methods, require entering the Gauge R&R data into a template format in which a single Gauge R&R study is analyzed.

For each of the 224 provided Gauge R&R data sets, an analysis was conducted using the EMP III method, the A&R method and the ANOVA method. Minitab statistical software was used to calculate the A&R and ANOVA methods and a Microsoft Excel spreadsheet was used to calculate the EMP III method based on Wheeler (2006). To calculate the high volume of Gauge R&R studies needed for this study a Microsoft Excel spreadsheet was created for the EMP III process and macros were written to automate the Minitab Gauge R&R analysis for the A&R and ANOVA methods. The spreadsheet was tested against the EMP III example in Wheeler (2009) to verify the results were calculated correctly. The Minitab macros automated, but did not alter the Gauge R&R methodology. The results showed the expected differences among the EMP III and A&R and ANOVA methods calculating the total Gauge R&R, as well as the repeatability and reproducibility components and the part-to-part contribution. These results are presented in Appendix A. These 224 studies were also seed data for simulating additional Gauge R&R study data.

Hypothesis one utilized a one-way ANOVA analysis at three levels to compare the total measurement precision error (total Gauge R&R) values for the three methods (EMP III, A&R, and ANOVA). Note that for Hypotheses one through four, the differences are analyzed by ANOVA. The ANOVA analysis comparing the three methods in these four hypotheses is not the same ANOVA analysis used to analyze the GR&R data. To estimate the number of samples needed for the study's Total GR&R analysis, the total Gauge R&R results from the 224 seed Gauge R&R studies calculated under all three methods were compared. The maximum difference for total Gauge R&R among these three methods for the 224 data sets calculated by the three different methods was 15.40 with a standard deviation of 34.87. The Minitab sample size calculation for one-way ANOVA with three levels, with 0.90 power, with 0.05 alpha error requires a minimum of 131 samples for each method.

Hypothesis two is similar to Hypothesis one, except it compares the differences among the repeatability component of the measurement precision error. The comparison is done with an ANOVA analysis and compares the three methods, EMP III, A&R, and ANOVA. As before, the 224 seed GR&R studies were used to estimate the needed number of samples for the analysis. For repeatability, the maximum difference among these methods was 15.87 with a standard deviation of 28.46. Minitab statistical software was used to calculate a sample size for this difference and standard deviation assuming a one-way, three-level ANOVA test, with 0.90 power and 0.05 alpha error, which resulted in a minimum of 83 samples for each method.

Hypothesis three compares the differences among the three methods for the reproducibility component of a Gauge R&R. As before, the 224 seed Gauge R&R studies were used to estimate the sample sizes needed for this analysis. For reproducibility, the maximum difference among the methods for the 224 data sets was 9.17 with a standard deviation of 22.95.

Again, assuming a one-way, three-level ANOVA test with a 22.95 standard deviation and the ability to detect a difference of 9.17, Minitab determines the sample size to be at least 160 samples for each method.

The fourth hypothesis relates to the differences in the part-to-part component estimated by a Gauge R&R study. This hypothesis tests the assumption that the three methods (EMP III, A&R, and ANOVA) calculate the part-to-part variation the same way. To determine the sample size for this test, the three analysis methods were used to determine the part-to-part variation for the 224 seed Gauge R&R data sets. The maximum difference in means among these three methods was 6.41 with a standard deviation of 30.30. Minitab's sample size calculation for these parameters with 0.90 power and 0.05 alpha error was a minimum of 567 samples each.

Hypothesis five studies the relationship between the methods of estimating total Gauge R&R measurement precision error compared two at a time. Utilizing the 224 actual Gauge R&R data provided and comparing all possible correlations (EMP III vs. A&R, EMP III vs. ANOVA, and A&R vs. ANOVA) the maximum difference in these correlation relationships was 0.0540. For total Gauge R&R measurement precision error, van Belle (2008) demonstrates calculating sample size for correlation studies, and for a difference of 0.0540, a power value of 0.90, and an alpha error of 0.05 requires a sample size of 7,175.05 which is rounded up to 7176 each.

Hypothesis six studies the relationship between the methods of estimating the repeatability portion of measurement precision error compared two at a time. Utilizing the 224 actual Gauge R&R data provided and calculating these relationships shows a maximum difference of 0.0580 in correlation between the different methods (EMP III vs. A&R, EMP III vs. ANOVA, and A&R vs. ANOVA) methods for repeatability. For the repeatability relationships, van Belle (2008) demonstrates calculating sample size for correlation studies and for a difference of 0.0580, a power value of 0.90, and an alpha error of 0.05 requires a sample size of 6,221.10 which is rounded up to 6,222 samples each.

Hypothesis seven studies the relationship between the methods of estimating the reproducibility portion of measurement precision error when compared two at a time. Utilizing the 224 actual Gauge R&R data provided and calculating these relationships shows a maximum difference of 0.0730 in correlation between the (EMP III vs. A&R, EMP III vs. ANOVA, and A&R vs. ANOVA) methods for reproducibility. For the reproducibility component of precision error, using van Belle's (2008) method and calculating sample size for correlation studies for a difference of 0.0730, a power value of 0.90, and an alpha error of 0.05 requires a sample size of 3,929.92 which is rounded up to 3930 samples each.

Hypothesis eight studies the relationship between the methods of estimating the part-to-part component of variation measurement precision error when compared two at a time. Utilizing the 224 actual Gauge R&R data provided and calculating these relationships shows a maximum difference in correlation r value of 0.0860 for the part-to-part correlations. For the part-to-part variation, van Belle (2008) demonstrates calculating sample size for correlation studies and for a difference of 0.0860, a power value of 0.90, and an alpha error of 0.05 requires a sample size of 2,826.21 which is rounded up to 2827 samples each. Under the same conditions a difference of 0.05 requires 8,400 samples and a 0.04 difference requires 13,125 samples. In a study of 223 simulations Koehler, Brown and Haneuse (2009) reported the sample size of the simulations, a sample of size 10,000 for a simulation was in the ninety-fifth percentile of their study. A sample size of 10,080 by van Belle's (2008) method will detect a correlation difference of approximately of 0.0456 and is evenly divisible by the number of seed Gauge R&R studies. This is in the mid range between a correlation of no difference and 0.0860 and is in the range of typical simulation sample sizes as noted by Koehler et al., (2009) and was used as the sample size for this study.

A summary of the sample sizes for the eight hypotheses in the study is shown Table 3.2. The maximum sample size needed for the correlation study will be used for all analysis. Appendix B provides the Minitab sample size calculation output for Hypotheses one through four and the sample size calculations for Hypotheses five through eight.

Table 3.2

Summary of Sample Sizes by Hypothesis

Hypothesis	**Analysis Method**	**Standard Deviation**	**Difference**	**Calculated Sample Size Minimum**
#1 Total Gauge R&R Difference	One way, three level, ANOVA and Kruskal-Wallis	34.87	15.40	131
#2 Repeatability Difference	One way, three level, ANOVA and Kruskal-Wallis	28.46	15.87	83
#3 Reproducibility Difference	One way, three level, ANOVA and Kruskal-Wallis	22.95	9.17	160
#4 Part-to-part Difference	One way, three level, ANOVA and Kruskal-Wallis	30.30	6.41	567
#5 Total GR&R Correlation	Pearson Correlation	N/A	0.0540	7,176
#6 Repeatability Correlation	Pearson Correlation	N/A	0.0580	6,222
#7 Reproducibility Correlation	Pearson Correlation	N/A	0.0730	3,930
#8 Part-to-part Correlation	Pearson Correlation	N/A	0.0456	10,080

The tests will be performed at the 5% significance level because no prior knowledge of the results exists and a 5% significance level is a middle point between lower and higher conventional α levels. As Manderscheidt (1965) points out, selection of a significance level is a balancing act between Type I and Type II errors for a given sample size. Manderscheidt (1965) also stresses if the null hypothesis is firmly believed, on the basis of past experience, it would not likely be rejected and a small Type I error could be employed (e.g. Type I- α error of .01). On the other hand, if a null hypothesis is highly doubtful, a larger Type I error could be used (e.g. Type I- α error of .10) (Manderscheidt, 1965). The reasoning for the Type II error of .10 is the same, meaning no previous knowledge is available concerning the expectations of the hypothesis so a middle approach is taken.

Simulation to Achieve Sample Size

The normality of the distributions of the seed data when analyzed by all three methods was checked and the distributions failed these normality tests. Figure 3.1 is a typical example of the seed data distribution for total measurement precision error when analyzed by the A&R method. In addition, a successful identification of the seed data distributions could not be made with Minitab statistical software. That is, the seed GR&R distributions did not match any known common distributions from those identifiable with Minitab statistical software. Consequently, it was not possible to simulate additional data via computer by sampling from a known distribution. To simulate additional data a single value for each set of seed data was altered multiple times to create deviations of each of the seed data sets. The value selected for alteration was randomly identified and was stratified so each of the measurement sub-sets within the seed data sets was equally represented. The alteration was based on varying standard deviations which is a method for which simulations are based (Bhattacharya & Raj, 2004) and (Norman, 2005). The standard deviation values selected were based on preliminary testing to best match the seed data distributions to the simulation data distributions.

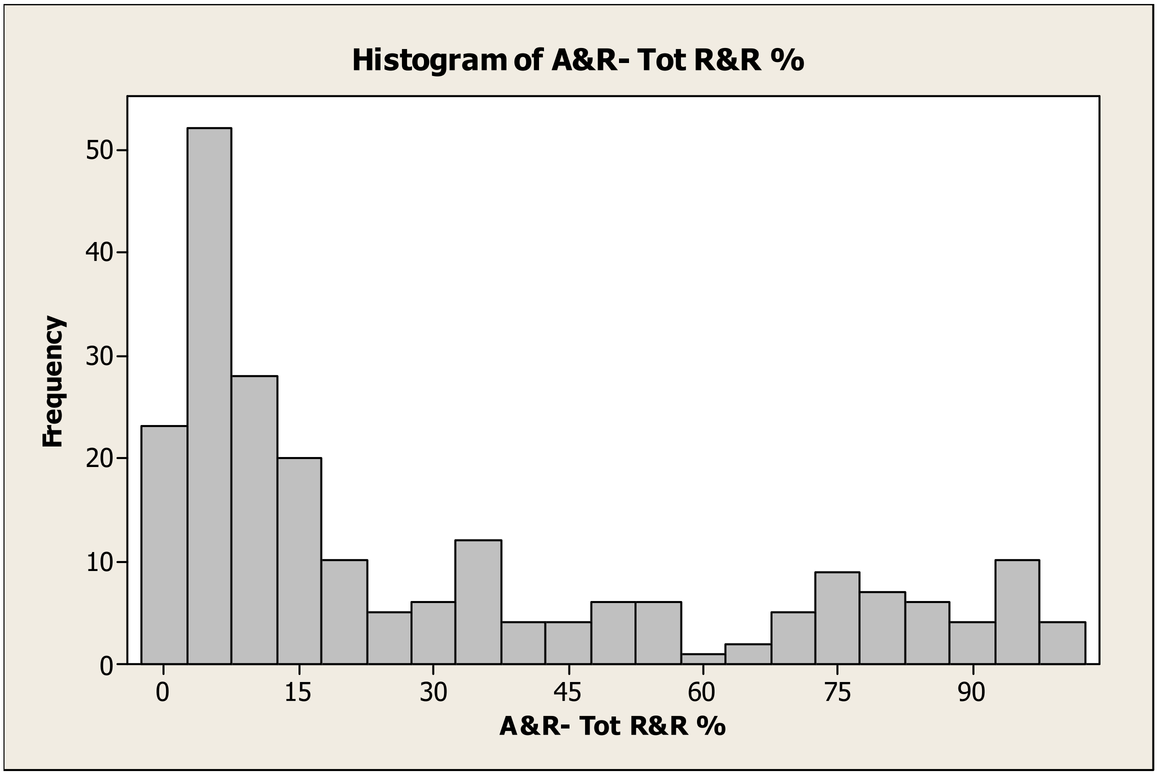

Figure 3.1. Histogram of A&R Analyzed Seed Data for Total Measurement Precision Error.

The required sample size to meet the minimum requirements of the study is 10,080 Gauge R&R data sets. To achieve the sample size, 10,080 data sets of 90 data points will be simulated based on the 224 actual Gauge R&R studies from the previously noted a U.S. medical device manufacturer. Forty-five data sets will be simulated from each of the 224 actual Gauge R&R studies as described in the following steps to achieve the 10,080 data sets. Each data set for the actual data and the simulated data is comprised of a total of 90 measurements, 30 measurements from each of three operators. The 90 values in each GR&R study is composed of nine subsets of ten measurements; three subsets from each operator. That is, the 30 measurements from each operator are comprised of three measurements each of ten parts. It is important to note the purpose of the simulation and the samples are to generate many different

Gauge R&R data sets for the comparison of the three methods of calculating Gauge R&R study results.

Simulation Steps

1. Simulate 45 versions of the first seed Gauge R&R data set.

a). Replace one randomly selected value from the first subset of ten measurements for operator one. The replacement value was randomly selected from 1,000 data points generated from Minitab software. The 1,000 generated data points were from a normal distribution with the same mean as operator one's 30 data points from the seed Gauge R&R study. However, the standard deviation of the distribution of the 1,000 data points for this first simulation set was a standard deviation seventy-five percent of the actual standard deviation for operator one's 30 actual data points. This restriction will create a new data set where one of the first ten measurements is a replaced value and the remaining 89 measurements are the actual data from Gauge R&R study one. Thus, one simulated data set (Gauge R&R study) has been created.

b). For the second simulation, repeat step a) above with a standard deviation 1.0 times operator one's standard deviation for the thirty measurements for the first seed Gauge R&R study. A new set of 1,000 normally generated points was created and, like step a), the same value randomly selected from step one is replaced. The other members of the data set will be the same as the actual study number one data set. At this point, a second data set (Gauge R&R study) has been created.

c). Repeat step a) above three more times with three different standard deviation values; 1.25, 1.5, and 2.0 times the actual standard deviation values. Each of these values will create a new 1,000 data point distribution, from which one value is randomly selected and replaced in the original seed Gauge R&R study one.

The above steps a) through c) will create five, 90 point Gauge R&R study data sets, where the one value from the first ten measurements was selected and replaced and the remaining 89 data points are the seed Gauge R&R study one data. Figure 3.2 shows a partial diagram of the simulation at this point in the process. At this point, five simulated data sets (Gauge R&R studies) have been created.

Sim. No.	Operator 1, First Set of 10 Measurements	Operator 1, Second Set of Readings
1	<==One value replaced w/ .75 Std. Dev. Value==>	<==Original values from Gauge R&R #1==>
2	<==One value replaced w/ 1.00 Std. Dev. Value==>	<==Original values from Gauge R&R #1==>
3	<==One value replaced w/ 1.25 Std. Dev. Value==>	<==Original values from Gauge R&R #1==>
4	<==One value replaced w/ 1.50 Std. Dev. Value==>	<==Original values from Gauge R&R #1==>
5	<==One value replaced w/ 2 Std. Dev. Value==>	<==Original values from Gauge R&R #1==>

Figure 3.2. Partial Diagram of Simulation after First Five Simulation Runs.

d). Steps a) through c) above are repeated in the same manner for the second set of measurements for operator one based on the seed Gauge R&R one study data. This process created a second group of five simulated data sets based on Gauge R&R study one for operator one's second set of measurements. These five sets will revert to the original data for the first set of ten measurements for operator one. Figure 3.3 shows a partial diagram of the simulation at this part in the process where a total of ten simulations have been created.

Sim. No.	Operator 1, First Set of 10 Measurements	Operator 1, Second Set of Readings
6	<==Original values from Gauge R&R #1==>	<==One value replaced w/ .75 Std. Dev. Value==>
7	<==Original values from Gauge R&R #1==>	<==One value replaced w/ 1.00 Std. Dev. Value==>
8	<==Original values from Gauge R&R #1==>	<==One value replaced w/ 1.25 Std. Dev. Value==>
9	<==Original values from Gauge R&R #1==>	<==One value replaced w/ 1.50 Std. Dev. Value==>
10	<==Original values from Gauge R&R #1==>	<==One value replaced w/ 2 Std. Dev. Value==>

Figure 3.3. Partial Diagram of Simulation after Ten Simulation Runs.

e). Steps a) through c) above are repeated in the same manner for the third set of measurements for operator one from the actual Gauge R&R study one data. This process will create a third group of five simulated data sets based on Gauge R&R study one. At this point in the process 15 simulated data sets have been created.

f). Steps a) through e) are repeated in like manner for operators two and three for Gauge R&R data set one. When this is process is complete, 45 simulated data sets have been created based on seed Gauge R&R study one. The simulated portion of the data sets retains the pattern of one value replaced in each of the 45 data sets. Each of the sets of five varies the standard deviation of one value to simulate different Gauge R&R study results.

2. Complete the simulation process for each of the remaining 223 seed Gauge R&R data sets. The process described in steps a) through f) above is completed for the remaining 223 seed Gauge R&R data sets, producing 10,080 simulated Gauge R&R data sets.

The simulation process can be summarized as follows: From 224 seed Gauge R&R studies, each containing 90 measurements in nine subgroups, 45 simulations were created from each; using five simulations for each of the nine subgroups. The five simulations were created by altering one value within the subgroup. The alteration was done by taking the average of the operator’s three subgroup measurements but with a different standard deviation value, five different standard deviation values in total. Within each of the final 10,080 simulations only one value was altered, the remaining 89 were directly from the seed Gauge R&R data.

The challenge with the simulation was to select values from a normal distribution, the seed Gauge R&R measurement data is predominately normally distributed, and end with a distribution of data after analysis that represents the analyzed seed data as shown in the histogram in Figure 3.1. The difficulty was altering the original seed data enough to be unique, yet not so different that a distributional shape vastly changed from the analyzed seed data distribution resulted. Figure 3.4 is the comparable simulation histogram to the Figure 3.1.

histogram of the seed data. The complete histogram comparisons for all three analysis methods (EMP III, A&R and ANOVA) between the seed data and the simulation data is in Appendix C. Appendix F contains statistical test comparisons of the seed and simulation data for all three analysis methods. A listing of the Minitab macros used in the simulation process is provided in Appendix D.

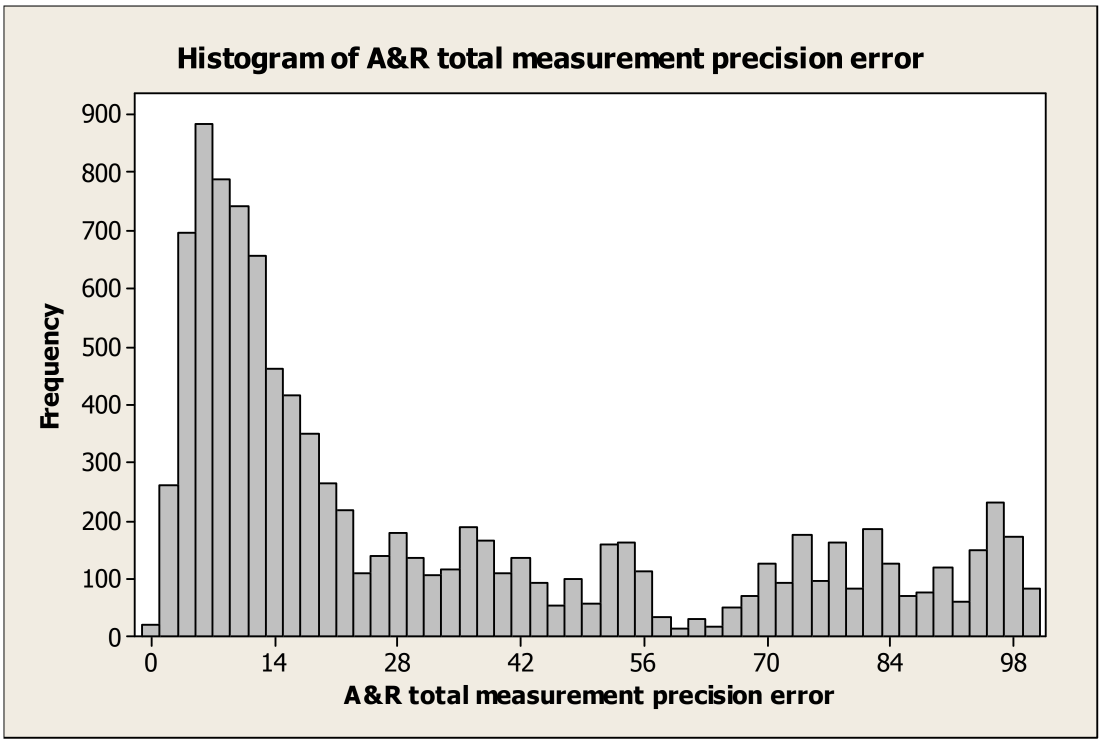

Figure 3.4. Histogram of A&R Total Measurement Precision Error, Compare with the Distributional Shape of Figure 3.1.

Table 3.3 provides a diagram of a typical Gauge R&R study layout. Note that this figure displays an abbreviated example compared to the Gauge R&R study data sets used in this study, abbreviated in that it contains only five parts measured two times each by three operators. It is intended to show an example of the typical Gauge R&R study data structure layout.

Table 3.3

Typical Gauge R&R Study Data Collection Table Indicating Three Operators Measuring Five Parts Two Times Each

	Part Number						
Operator 1	**1**	**2**	**3**	**4**	**5**		
Meas. #1	0.460	0.445	0.450	0.449	0.449		
Meas. #2	0.460	0.446	0.451	0.449	0.449		
Average ($\bar{X}$)	0.46	0.4455	0.4505	0.449	0.449	0.4508	($\bar{\bar{X}}$) Grand Avg.
Range (R)	0	0.001	0.001	0	0	0.0004	($\bar{R}$) Avg. Range
Operator 2							
Meas. #1	0.461	0.446	0.454	0.451	0.450		
Meas. #2	0.461	0.446	0.454	0.450	0.45		
Average ($\bar{X}$)	0.461	0.446	0.454	0.4505	0.450	0.4523	($\bar{\bar{X}}$) Grand Avg.
Range (R)	0	0	0	0.001	0	0.0002	($\bar{R}$) Avg. Range
Operator 2							
Meas. #1	0.460	0.445	0.451	0.452	0.452		
Meas. #2	0.461	0.447	0.452	0.451	0.452		
Average ($\bar{X}$)	0.461	0.446	0.452	0.452	0.452	0.4523	($\bar{\bar{X}}$) Grand Avg.
Range (R)	0.001	0.002	0.001	0.001	0.000	0.001	($\bar{R}$) Avg. Range

Research Design Procedures/Description of the Measurers Employed

Hypothesis One

Hypothesis one compares the average total measurement precision error estimates (Total Gauge R&R) among EMP III method, the A&R method, and the ANOVA method for the 10,080 data sets. The p-value results of the comparisons for all Hypotheses one through eight was done at the 5% significance level with Minitab statistical software and are reported and summarized in

Chapter 4. A nonparametric Kruskal-Wallis test is also included for Hypotheses one through four because the normality assumption of the data for the ANOVA tests was not met.

Hypothesis Two

Hypothesis two compares the average repeatability component of measurement precision error for the EMP III method, the A&R method, and the ANOVA method. The three-way comparison is done for all 10,080 data sets. The methods of comparison are the ANOVA and Kruskal-Wallis tests.

Hypothesis Three

Hypothesis three compares the average reproducibility component of measurement precision error for the EMP III method, the A&R method and the ANOVA method for the 10,080 data sets. The methods of comparison are the ANOVA and Kruskal-Wallis tests.

Hypothesis Four

The fourth hypothesis compares the average part-to-part component of the total measurement precision error. The three-way comparison is among the EMP III method, the A&R method, and the ANOVA method. A one-way, three-level ANOVA test is employed to determine the extent of the differences along with the nonparametric equivalent, the Kruskal-Wallis test

Hypothesis Five

Hypothesis five is a correlation study between the EMP III method, the A&R method and the ANOVA method each compared two at a time for the total measurement precision error (Total Gauge R&R). The Pearson r statistic is the output from the correlation studies for the Gauge R&R data sets. The r statistic is a measure of how strongly related the two methods are in the observed sample data sets (Devore, 2004). A value of zero indicates no correlation; a value of 1.0 indicates perfect positive correlation while -1.0 indicates a perfect negative correlation (Devore, 2004).

Hypothesis Six

Hypothesis six is a correlation study for the repeatability component of measurement precision error between the EMP III method, the A&R method, and the ANOVA method each compared two at a time. The Pearson r statistic is the output from the correlation studies for the repeatability component of the Gauge R&R data sets.

Hypothesis Seven

Hypothesis seven is a correlation study for the reproducibility component of measurement precision error between the EMP III method, the A&R method and the ANOVA method each compared two at a time. The Pearson r statistic is the output from the correlation studies for the reproducibility portion.

Hypothesis Eight

Hypothesis eight addresses the relationship of the part-to-part portion of measurement precision error between the EMP III method, the A&R method, and the ANOVA method each compared two at a time. The Pearson r statistic is the output from the correlation studies for this set of comparisons.

CHAPTER 4

FINDINGS

Chapter 3 described the method of investigation to the problem and hypotheses. This chapter describes the findings of the investigation as well as the techniques used in the analysis of the data.

Overview

Using 224 actual Gauge R&R studies as a simulation basis, 10,080 Gauge R&R studies were simulated. All of these Gauge R&R studies simulated three operators, measuring ten parts, three times each, for a total of 90 measurements per data set. The simulations were created by the method described in Chapter 3.

Data from each of the 10,080 simulated Gauge R&R studies was then analyzed by three methods of Gauge R&R study analysis; the EMP III method, the A&R method and the ANOVA method. The EMP III analysis was conducted with the aid of a Microsoft Excel spreadsheet following the method described in *EMP III using imperfect data* (Wheeler, 2006). The A&R and ANOVA methods were calculated with the assistance of Minitab statistical software, which follows the AIAG methodology.

Prior to analysis of the data the two assumptions described in Chapter 3 were verified. The first assumption is that the distributions from which the simulated samples are drawn are normally distributed. After the 10,080 simulations were completed, the distribution from which each set of samples was drawn was checked for normality. The distributions were part of the simulation process. The normality testing was done at the 0.05 level of significance with the Anderson-Darling test for normality. The tests were conducted with the aid of Minitab statistical software. Of the 10,080 simulation runs, 474 failed the normality assumption. These 474 simulations were re-run until the data from which the samples were drawn passed the Anderson-Darling normality test at the 0.05 level of significance

The second assumption of the study is that the simulations are made up of three simulated operators, measuring ten parts each, three times, for a total of 90 measurements per simulation. Each of the 10,080 simulations were verified to contain 90 data points.

Table 4.1 shows a comparison of the simulation data broken down by analysis method and how many simulations fell into each of the AIAG acceptance criteria categories. Compare this table with Table 3.1, which compared the seed Gauge R&R study data, from which the simulations were based.

Table 4.1

Comparison of Simulated Gauge R&R Study Total Gauge Error, Total Number of the Simulations falling into Each AIAG Acceptance Category by Analysis Method

Total Gauge R&R Error Percent AIAG Category	**EMP III**	**A&R**	**ANOVA**
< 10%	6,346	3,009	1,406
10% -30%	1,392	3,241	3,992
>30%	2,342	3,830	4,682
Total	10,080	10,080	10,080

The data analysis techniques were ANOVA, Kruskal-Wallis and Pearson correlation. ANOVA and Kruskal-Walls were used for analysis related to Hypotheses one through four. Correlation was used for analysis of Hypotheses five through eight. Minitab statistical software was used for all the ANOVA, Kruskal-Wallis and correlation calculations in this chapter.

An anomaly did occur during one simulation run. During the run of Minitab macro %RAND for seed Gauge R&R study GRR 41-1, for operator three the standard deviation value was too small for Minitab (1.06^{-17}). A value of 1.06^{-11}, which was the smallest standard deviation value Minitab allowed, was substituted. This adjustment affected the simulation run numbers 2821-2835, although not adversely, because the goal of the simulations was to create Gauge R&R data to be analyzed by three methods.

The data compared in the ANOVA analysis for Hypotheses one through four did not meet an assumption of ANOVA analysis that the data is normally distributed. This fact was evidenced by performing normality tests on the total measurement precision error (Total Gauge R&R) results, the Repeatability results, the Reproducibility results and the Part-to-part results for each of the three analysis methods (EMP III, A&R, and ANOVA). Both Anderson-Darling and Ryan-Joiner normality tests were performed. All of the Anderson-Darling tests for all conditions had a P-value < 0.005 and all Ryan-Joiner normality tests for all conditions had a P-value below 0.010. These values indicate that the analysis data did not meet the ANOVA normality assumption. As Good and Hardin (2009) note, a statistical test is most powerful when it meets the assumptions on which it is based. Consequently, for Hypotheses one through four an additional nonparametric statistical analysis was performed. The analysis added was the Kruskal-Wallis test of medians. As noted by Subrahmaniam, Subrahmaniam & Messeri (1975), minor departures from normality can be negligible in ANOVA; however, what is considered minor is subject to interpretation. Consequently, the ANOVA for Hypotheses one through four is included. The addition of the nonparametric analysis for these hypotheses does not change the hypotheses; they remain as stated in Chapter 1. Additionally, the sample sizes for Hypotheses one through four, although calculated for ANOVA tests, is appropriate for Kruskal-Wallis testing. Sheskin (2004) notes that the asymptotic relative efficiency of the Kruskal-Wallis test is 0.955 compared to the ANOVA test. The maximum number of samples needed for the ANOVA tests from Table 3.2 is 566. Applying the reciprocal of the Kruskal-Wallis asymptotic relative efficiency factor (1/0.955) to the 566 samples indicates a minimum 593 samples are required for the Kruskal-Wallis samples. The 10,080 actual samples used in the study are well beyond the 593 sample requirement.

Hypothesis One

Hypothesis one tests the assumption that all of the three methods of Gauge R&R study analysis are the same for the 10,080 simulated Gauge R&R studies for total measurement precision error (Total GR&R). Table 4.2 provides the results of this analysis. Table 4.3 provides the descriptive statistics of the three methods of analysis of the simulation data for Hypothesis one.

Table 4.2

Hypothesis one, ANOVA Analysis of Total Measurement Precision Error

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor (EMP III, A&R and ANOVA method)	2	2031754	1015877	1116.80	0.000

Table 4.2 *(continued)*

Source	DF	SS	MS	F	P
Error	30237	27504597	910		
Total	30239	29536351			
S= 30.16 R-Sq= 6.88% R-Sq (adj)= 6.87%					

Table 4.3

Hypothesis one, Descriptive Statistics for the Three Analysis Methods

Level	N	Mean	St Dev.
EMP III	10,080	19.69	29.06
A&R	10,080	32.43	30.28
ANOVA	10,080	39.49	31.11

Table 4.2 shows that overall, the methods of calculating total measurement precision error (Total Gauge R&R) by the three methods are different, but the ANOVA analysis does not provide pair-wise comparisons of the methods. Pair-wise comparisons of the methods utilizing t-Tests inflate the family-wise error rate beyond the specified 0.05 significance levels (Devore, 2004).   To preserve the 0.05 level of significance and provide pair-wise comparisons, Tukey's method (Devore, 2004) was used. Table 4.4 provides the results of this analysis by comparing the EMP III estimate of total measurement precision error with the A&R method and the EMP III method with the ANOVA method. Table 4.5 compares the A&R method and the ANOVA method. As long as zero is not in the difference range of the comparison, the comparisons are statistically different.

Table 4.4

Hypothesis one, Tukey's Pair-wise Comparisons EMP III Method of Total Measurement Precision Error

EMP III % Total measurement precision error subtracted from:			
	Lower	Center	Upper
A&R method	11.75	12.75	13.74
ANOVA method	18.81	19.81	20.80

Table 4.5

Hypothesis one, Tukey's Pair-wise Comparisons A&R Method of Total Measurement Precision Error

A&R % Study Variation subtracted from:			
	Lower	Center	Upper
ANOVA method	6.07	7.06	8.05

Because the total study variation data failed the normality assumption for ANOVA, a nonparametric statistical test was also employed for Hypothesis one. The Kruskal-Wallis test is used for comparing the medians of more than two independent nonparametric data sets (Conover, 1999). Table 4.6 shows the Minitab results of the Kruskal-Wallis test.

Table 4.6

Hypothesis one, Kruskal-Wallis Analysis of Total Measurement Precision Error

Source	**N**	**Median**	**Ave. Rank**	**Z**
EMP III Method	10080	3.085	10171.3	-69.71
A&R Method	10080	17.568	16383.4	17.79
ANOVA Method	10080	27.188	18806.7	51.92
Overall	30240		15120.5	
H = 5248.23 DF = 2 P = 0.000				
H = 5248.24 DF = 2 P = 0.000 (adjusted for ties)				

Like the ANOVA analysis, the results of the Kruskal-Wallis test indicates the methods are not the same for estimating total measurement precision error, but do not indicate which pair-wise comparisons are different. A further Kruskal-Wallis comparison was done comparing the pair-wise sets. Table 4.7 shows the Minitab results of this comparison. As with the Tukey comparison for Hypothesis one, the results indicate that all of the comparisons are statistically different.

Table 4.7

Hypothesis one, Kruskal-Wallis Multiple Comparisons for Total Measurement Precision Error

Groups	**Z vs. Critical Value**	**P-Value**
EMP III vs. A&R	19.7037 ≥ 1.834	0
EMP III vs. ANOVA	70.2266 ≥ 1.834	0

Table 4.7 *(continued)*

Groups	**Z vs. Critical Value**	**P-Value**
A&R vs. ANOVA	50.5194 ≥ 1.834	0

Hypothesis Two

Hypothesis two compares the average differences for the repeatability portions of the measurement precision error for the simulated data. ANOVA analysis was done to examine the data. For the repeatability portion, the null hypothesis is that there is no difference in the repeatability portion of the Gauge R&R measurement precision error for the three methods of analysis. The alternate hypothesis is that there is a difference in the repeatability portion estimates of measurement precision error by the three methods. Table 4.8 shows the ANOVA results for repeatability for the 10,080 simulated Gauge R&R studies. Table 4.9 shows the repeatability statistical results.

Table 4.8

Hypothesis two, ANOVA Analysis for Repeatability Measurement Precision Error

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor (EMP III, A&R and ANOVA method)	2	2276577	1138288	2151.28	0.000
Error	30237	15999008	529		
Total	30239	18275585			
S= 23.00 R-Sq=12.46% R-Sq (adj)= 12.45%					

Table 4.9

Hypothesis two, Repeatability Descriptive Statistics

Level	**N**	**Mean**	**St Dev.**
EMP III Repeatability	10,080	12.21	19.17
A&R Repeatability	10,080	25.51	23.90
ANOVA Repeatability	10,080	33.21	25.47

Similar to Hypothesis one, pair-wise comparisons were made of the repeatability results using Tukey's method (Devore, 2004) to determine if there were statistical differences among the various combinations of results. The results of the Tukey's method (Devore, 2004) are the differences in the means and for the repeatability results are shown in Tables 4.10 and 4.11. If zeros are not in the range of the mean differences indicated, there is a statistical difference in the comparison.

Table 4.10

Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability

EMP III Repeatability Method subtracted from:			
	Lower	Center	Upper
A&R Repeatability	12.55	13.31	14.06
ANOVA Repeatability	20.25	21.01	21.76

Table 4.11

Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability

A&R- Repeatability subtracted from:			
	Lower	Center	Upper
ANOVA Repeatability	6.94	7.70	8.46

Like Hypothesis one, the repeatability results did not meet the normality assumption for an ANOVA test. Consequently a Kruskal-Wallis test was also performed on the repeatability data. Table 4.12 shows the results of this comparison. As shown, the results indicate the three methods are not equivalent based on the Kruskal-Wallis median test in their estimates of the repeatability portion of measurement precision error.

Table 4.12

Hypothesis two, Kruskal-Wallis Test on Repeatability

Source	**N**	**Median**	**Ave. Rank**	**Z**
ANOVA Method	10080	24.284	19488.6	61.53
A&R Method	10080	14.517	16575.6	20.50
EMP III Method	10080	2.105	9297.3	-82.02
Overall	30240		15120.5	
H = 7289.15 DF = 2 P = 0.000				
H = 7289.15 DF = 2 P = 0.000 (adjusted for ties)				

In this case the Kruskal-Wallis test demonstrates that the methods are different, but does not perform pair-wise comparisons to assist in determining if all pair-wise comparisons are different or not. To determine whether pair-wise comparisons are different, an additional Kruskal-Wallis test for the repeatability pair-wise comparisons was performed. Table 4.13 shows the results of this testing. This test shows that none of the median values for the three methods are the same.

Table 4.13

Hypothesis two, Kruskal-Wallis Multiple Comparisons for Repeatability

Groups	**Z vs. Critical Value**	**P-Value**
EMP III Repeatability vs. ANOVA-Repeatability	82.8799 ≥ 1.834	0
EMP III Repeatability vs. A&R-Repeatability	59.1898 ≥ 1.834	0
A&R Repeatability vs. ANOVA Repeatability	23.6902 ≥ 1.834	0

Hypothesis Three

Hypothesis three compares the average differences for the reproducibility portion of the Gauge R&R measurement precision error for the simulated data. ANOVA analysis was performed for the analysis of the data. For the reproducibility portion, the null hypothesis is that there is no difference in the reproducibility portion of the Gauge R&R measurement precision error for the three methods of analysis. The alternate hypothesis is that there is a difference in the reproducibility portion estimates of measurement precision error by the three methods. Table 4.14 shows the ANOVA results for reproducibility for the 10,080 simulated Gauge R&R studies. Table 4.15 shows the repeatability statistical results. The results indicate that the data do not support that the analysis methods calculate reproducibility error the same.

Table 4.14

Hypothesis three, ANOVA Analysis for Reproducibility Measurement Precision Error

Source	DF	SS	MS	F	P
Factor (EMP III, A&R and ANOVA method)	2	498239	249119	578.86	0.000
Error	30237	13012883	430		
Total	30239	13511122			
S= 20.75 R-Sq= 3.69% R-Sq (adj)= 3.68%					

Table 4.15

Hypothesis three, Reproducibility Descriptive Statistics

Level	N	Mean	St Dev.
EMP III Reproducibility	10,080	7.48	16.92
A&R Reproducibility	10,080	15.93	22.20
ANOVA Reproducibility	10,080	16.24	22.63

As with Hypotheses one and two, it takes further analysis to determine if the levels in an ANOVA analysis are statistically different. As before, Tukey's method (Devore, 2004) is employed for pair-wise comparisons. Tables 4.16 and 4.17 provide these comparisons for the reproducibility portion of the Gauge R&R simulation data. Because zero is in the range for the ANOVA reproducibility/A&R reproducibility comparison in Table 4.16, there is no statistical difference in these two methods of estimating the reproducibility portion of measurement precision error when analyzed by Tukey's method.

Table 4.16

Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility

ANOVA Reproducibility subtracted from:			
	Lower	Center	Upper
EMP III Reproducibility	-9.45	-8.76	-8.08
A&R Reproducibility	-1.00	-0.31	0.37

Table 4.17

Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility

EMP III Reproducibility subtracted from:			
	Lower	Center	Upper
A&R Reproducibility	7.77	8.45	9.13

As with the repeatability portion, the reproducibility portion did not meet the normality assumption for an ANOVA statistical test. Consequently, an additional Kruskal-Wallis test of medians was employed. The results of this additional testing are shown in Table 4.18 and indicate that there is a statistical difference in the medians of the three methods of Gauge R&R study analysis for the reproducibility portion.

Table 4.18

Hypothesis three, Kruskal-Wallis Test on Reproducibility

Source	**N**	**Median**	**Ave. Rank**	**Z**
ANOVA Method	10080	5.0878	16158.3	14.62
A&R Method	10080	5.5354	17539.7	34.08
EMP III Method	10080	0.3078	11663.5	-48.69
Overall	30240		15120.5	
H = 2497.30 DF = 2 P = 0.000				
H = 2509.17 DF = 2 P = 0.000 (adjusted for ties)				

The results of the Kruskal-Wallis test for the reproducibility portion indicates that there is a difference among the three methods, but do not explore the pair-wise comparisons of the three methods. An additional Kruskal-Wallis pair-wise comparison was performed with the results presented in Table 4.19. The results indicate that all three methods differ in their estimates of the reproducibility portion of a Gauge R&R study unlike the Tukey pair- wise comparison for A&R reproducibility and ANOVA reproducibility.

Table 4.19

Hypothesis three, Kruskal-Wallis Multiple Comparisons for Reproducibility

Groups	**Z vs. Critical Value**	**P-Value**
EMP III Reproducibility vs. A&R-Reproducibility	47.9005 ≥ 1.834	0
EMP III Reproducibility vs. ANOVA-Reproducibility	36.6403 ≥1.834	0

Table 4.19 *(continued)*

Groups	**Z vs. Critical Value**	**P-Value**
A&R Reproducibility vs. ANOVA-Reproducibility	$11.2602 \geq 1.834$	0

Hypothesis Four

The fourth hypothesis compares the part-to-part portion of the simulation data by the three Gauge R&R measurement precision error analysis methods. As described in Chapter 3, this analysis was performed by ANOVA. Table 4.20 compares the 10,080 simulation Gauge R&R data sets for the average part-to-part proportion of measurement precision error. Table 4.21 provides the descriptive statistics for the part-to-part proportion of the data.

Table 4.20

Hypothesis four, ANOVA Analysis for Part-to-part Measurement Precision Error

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor (EMP III, A&R and ANOVA method)	2	255543	127771	181.48	0.000
Error	30237	21289003	704		
Total	30239	21544546			
S= 26.53 R-Sq=1.19% R-Sq (adj)= 1.18%					

Table 4.21

Hypothesis four, Part-to-part Descriptive Statistics

Level	**N**	**Mean**	**St Dev.**
EMP III Part-to-part	10,080	80.31	29.06
A&R Part-to-part	10,080	87.03	21.39
ANOVA Part-to-part	10,080	81.62	28.47

Hypothesis four was also further analyzed utilizing Tukey's method (Devore, 2004) to determine if the pair-wise comparisons were significant similar to the results in Hypotheses one through three. Tukey's method (Devore, 2004) results for Hypothesis four are shown in Tables 4.22 and 4.23. The differences in means are presented in these tables. If zero is not contained in the lower-to-upper interval, there is a statistical difference in the means of the comparison.

Table 4.22

Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part

ANOVA-Part-to-part subtracted from:			
	Lower	Center	Upper
EMP III Part-to-part	-2.18	-1.31	-0.43
A&R Part-to-part	4.53	5.41	6.28

Table 4.23

Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part

EMP III-Part-to-part subtracted from:			
	Lower	Center	Upper
A&R- Part-to-part	5.84	6.72	7.59

The part-to-part data did not meet the normality assumption of ANOVA. To continue the analysis under this condition, a nonparametric Kruskal-Wallis test of the part-to-part medians was conducted. The results of this analysis are shown in Table 4.24. The results of this test show the three methods estimate the median part-to-part study variation differently.

Table 4.24

Hypothesis four, Kruskal-Wallis Test on Part-to-part Portion of Measurement Precision Error

Source	**N**	**Median**	**Ave. Rank**	**Z**
EMP III Part-to- part	10080	96.92	14438.9	-9.60
A&R Part-to-part	10080	98.44	16955.3	25.84
ANOVA Part-to-part	10080	96.23	13967.3	-16.24
Overall	30240		15120.5	
H = 682.61 DF = 2 P = 0.000				
H = 682.61 DF = 2 P = 0.000 (adjusted for ties)				

The Table 4.24 Kruskal-Wallis analysis of the part-to-part medians based on the three analysis methods of the Gauge R&R data does not test the pair-wise comparisons of the three methods (EMP III, A&R, and ANOVA). The additional analysis results are shown in Table

4.25. The results indicate the three analysis methods are all different when compared to one another in pairs.

Table 4.25

Hypothesis four, Kruskal-Wallis Multiple Comparisons for Part-to-part

Groups	**Z vs. Critical Value**	**P-Value**
A&R Part-to-part vs. ANOVA Part-to-part	$24.2990 \geq 1.834$	0
EMP III Part-to-part vs. A&R Part-to-part	$50.3097 \geq 1.834$	0
EMP III Part-to-part vs. ANOVA Part-to-part	$3.851 \geq 1.834$	0.0001

Hypothesis Five

Hypothesis five examines the correlation between the total Gauge R&R analysis methods for the simulation study results when compared two at a time. This examination includes comparisons of EMP III to A&R analysis methods, the EMP III to ANOVA analysis methods and the A&R to ANOVA analysis methods, all for total Gauge R&R measurement precision error. Table 4.26 provides the results of the comparisons in Pearson r correlation statistics. The r statistic is a measure of how strongly related the two methods are in the observed sample data sets (Devore, 2004). A value of zero indicates no correlation; a value of 1.0 indicates perfect positive correlation while -1.0 indicates a perfect negative correlation (Devore, 2004).

Table 4.26

Hypothesis five, Pearson r Correlation Values for Total Measurement Precision Error

Hypothesis	**Comparison Description**	**Pearson r Correlation Statistic**	**P-Value**
5	EMP III total measurement precision error to A&R total measurement precision error	0.972	0.00
5	EMP III total measurement precision error to ANOVA total measurement precision error	0.930	0.00
5	A&R total measurement precision error to ANOVA total measurement precision error	0.969	0.00

Hypothesis Six

Hypothesis six examines the correlation between the repeatability portions of Gauge R&R measurement precision error for the simulation study results compared two at a time. Included are the comparisons for EMP III to A&R analysis method results, the EMP III to ANOVA analysis method results and the A&R to ANOVA analysis method results, all for the repeatability component of measurement precision error. Table 4.27 provides the comparisons in Pearson r correlation statistics.

Table 4.27

Hypothesis six, Pearson r Correlation Values for Repeatability

Hypothesis	**Comparison Description**	**Pearson r Correlation Statistic**	**P-Value**
6	EMP III repeatability measurement precision error to A&R repeatability measurement precision error	0.961	0.00

Table 4.27 *(continued)*

Hypothesis	**Comparison Description**	**Pearson r Correlation Statistic**	**P-Value**
6	EMP III repeatability measurement precision error to ANOVA repeatability measurement precision error	0.910	0.00
6	A&R repeatability measurement precision error to ANOVA repeatability measurement precision error	0.968	0.00

Hypothesis Seven

Hypothesis seven determines the correlation between the reproducibility portions of Gauge R&R measurement precision error for the simulation study results when the results are compared two at a time. Included are the comparisons for EMP III to A&R analysis method results, the EMP III to ANOVA analysis method results and the A&R to ANOVA analysis method results for reproducibility. Table 4.28 provides the comparisons in Pearson r correlation statistics.

Table 4.28

Hypothesis seven, Pearson r Correlation Values for Reproducibility

Hypothesis	**Comparison Description**	**Pearson r Correlation Statistic**	**P-Value**
7	EMP III reproducibility measurement precision error to A&R reproducibility measurement precision error	0.948	0.00
7	EMP III reproducibility measurement precision error to ANOVA reproducibility measurement precision error	0.875	0.00

Table 4.28 *(continued)*

Hypothesis	Comparison Description	Pearson r Correlation Statistic	P-Value
7	A&R reproducibility measurement precision error to ANOVA reproducibility measurement precision error	0.925	0.00

Hypothesis Eight

Hypothesis eight calculates the correlation between the part-to-part portions of Gauge R&R measurement precision error for the simulation study results when compared two at a time. Included are the comparisons for EMP III to A&R analysis method results, the EMP III to ANOVA analysis method results and the A&R to ANOVA analysis methods results for the part-to-part component of measurement precision error. Table 4.29 provides the comparisons in Pearson r correlation statistics.

Table 4.29

Hypothesis eight, Pearson r Correlation Values for Part-to-part

Hypothesis	Comparison Description	Pearson r Correlation Statistic	P-Value
8	EMP III Part-to-part measurement precision error to A&R Part-to-part measurement precision error	0.988	0.00
8	EMP III Part-to-part measurement precision error to ANOVA Part-to-part measurement precision error	0.907	0.00
8	A&R Part-to-part measurement precision error to ANOVA Part-to-part measurement precision error	0.902	0.00

Appendix G contains the statistical test results of comparing the variances of the three methods (EMP III, A&R, and ANOVA) to demonstrate a complete analysis. Comparing all three analysis methods for each of the components of variation suggests their variances are all different from one another.

CHAPTER 5

SUMMARY AND DISCUSSION

In the previous chapters, the research questions have been described, the hypotheses identified, the data simulated, the data analyzed and the results reported. This chapter discusses the implications of the study results from the standpoint of Technology Management.

Conclusions

The research question described in Chapter 1 is: Do the methods to estimate measurement precision error produce the same results? The results of the study suggest that the three methods do not estimate measurement precision error the same way. Understanding the differences in the methods of GR&R study analysis is important to technology managers because acceptance of a gauge is dependent on the inherent precision error of the gauge itself and the method of measurement precision error estimation. In other words, a gauge assessed by the EMP III method may be judged appropriate, while the same gauge assessed by the A&R or ANOVA method may be judged inappropriate with regard to measurement precision error. In addition, the three methods differ in their estimates of the repeatability, reproducibility and part-to-part precision error estimates. The study showed high correlation between the various methods of estimating total measurement precision error and its components. This correlation suggests the differences tend to be biased between the methods of estimating measurement precision error and its components.

The results are important and useful to the technology manager supervising the measurement process in a manufacturing organization because they demonstrate that the consumer of Gauge R&R study results should know which method of GR&R analysis is being used. Differentiation between the methods is important because as the study suggests, the results can be different. The differences between the EMP III method and A&R method and between the EMP III method and the ANOVA method were consistent with the literature authored by Wheeler (2006), Ermer (2006), and Knowles et al. (2000) due to the method of summing the standard deviation. What was not previously noted in the literature is the difference between the A&R and ANOVA methods, which the study suggests is significant.

Additionally, the technology manager needs to be aware of which method is being used to estimate measurement precision error if the results are going to be used to base adjustments to manufacturing tolerances (guard banding). The differences in Gauge R&R study methods could result in different adjustments to manufacturing tolerances. For example, a gauge assessed by the A&R method may have little adjustment to the manufacturing tolerances, while the same gauge assessed by the ANOVA method may have larger adjustments made to the manufacturing tolerances. For demonstration purposes, Table 5.1 shows the standard deviation estimates and percent tolerance for the example calculations described in Chapter 2 for the A&R and ANOVA methods. A tolerance of 0.010 was selected for the Table 5.1 example results. The EMP III method is combined with the A&R method in Table 5.1 because the standard deviation for the EMP III method and the A&R method are the same but diverge when summarizing the information as shown by the example calculations in Chapter 2. The percent tolerance column shown in Table 5.1 is the amount of the 0.010 tolerance consumed by the total measurement precision error. Note the difference in the percent tolerance between the A&R and ANOVA methods.

Table 5.1

Standard Deviation and Percent Tolerance for Example Calculations in Chapter 2

Method	**Standard Deviation Estimate for Total GR&R (From Chapter 2 example data)**	**Six Times Standard Deviation Estimate**	**Percent Tolerance (Six Times Standard Deviation Estimate as Percent of 0.010 Tolerance)**
EMP III and A&R	0.000903	0.005418	54.18%
ANOVA	0.0013092	0.0078552	78.55%

The ANOVA method of Gauge R&R study analysis is perplexing. Some authors such as Antony et al. (1998) and Kazerouni (2009) suggest the ANOVA method is the more accurate method in the presence of operator-by-part interaction in a Gauge R&R study. In this study, 2,774 of the 10,080 simulated Gauge R&R data sets, or 27.5%, showed operator-by-part interaction. This interaction would have gone undetected by the A&R and EMP III methods and indicates an additional source of variation that could assist technology managers in examining and reducing measurement precision error. Conversely, in the study, the ANOVA method had the highest estimates of total measurement precision error and the highest repeatability component of measurement precision error for both mean and median of the methods in the study.

To determine if the operator-by-part interaction contributed to the ANOVA method having the highest estimates in total measurement precision error the 2,774 simulations with operator-by-part interaction were removed and the analysis for Hypotheses one through four repeated. The results of the ANOVA, Tukey and Kruskal-Wallis tests agreed with the results for the full data for total measurement precision error, the repeatability component of measurement precision error and the part-to-part component of measurement precision error. The results differed for the reproducibility component for measurement precision error with regard to the Tukey pair-wise comparison for the ANOVA and A&R methods. The full data study found no difference between the ANOVA and A&R methods, while the reduced study eliminating the operator-by-part interaction measurement data found the two methods of estimating reproducibility measurement precision error were statistically different at the 0.05 level of significance. The Kruskal-Wallis comparisons for both overall and pair-wise for the reproducibility component were the same for both the full and reduced data sets. The reproducibility results showed the methods were different at the 0.05 level of significance. Complete results of the reduced comparison are shown in Appendix E.

Practical Significance

Table 5.2 shows the median and mean difference between the EMP III method and the A&R and ANOVA methods is at least 14.48 and 12.74, respectively for total measurement precision error. These differences are enough to move a Gauge R&R study from one AIAG acceptance category to another. In other words, the same Gauge R&R study analyzed by the EMP III method and accepted by the AIAG criteria could be unacceptable by the same criteria if analyzed by the A&R or ANOVA method. This apparent conflict means that depending on which analysis method is selected, a gauge could be acceptable or unacceptable. Industry is best served by ensuring the method of analysis is specified when reporting measurement precision error study results. Table 5.2 also shows that the mean and median differences between the

A&R and ANOVA methods for total measurement precision error could move the conclusions of a Gauge R&R study from one AIAG acceptance category to another.

Table 5.2

Study Results for Total Measurement Precision Error (Total GRR)

Level	Median	Median Difference	Mean	Mean Difference
EMP III	3.085	EMP III vs. A&R: 14.483	19.69	EMP III vs. A&R: 12.74
A&R	17.568	A&R vs. ANOVA: 9.62	32.43	A&R vs. ANOVA: 7.06
ANOVA	27.188	EMP III vs. ANOVA: 24.103	39.49	EMP III vs. ANOVA: 19.80

Figure 5.1 is a line chart of the total measurement precision error results from 67 randomly selected study results from the 10,080 simulations in this study. The intent of the chart is to demonstrate that the differences between the three methods are consistent. That is, the EMP III method tends to estimate total measurement precision error the lowest of the three methods, the A&R method next and the ANOVA method provides the highest measurement precision error estimates. Sixty-seven points were chosen to plot because this was the maximum number able to plot and provide a legible chart.

.

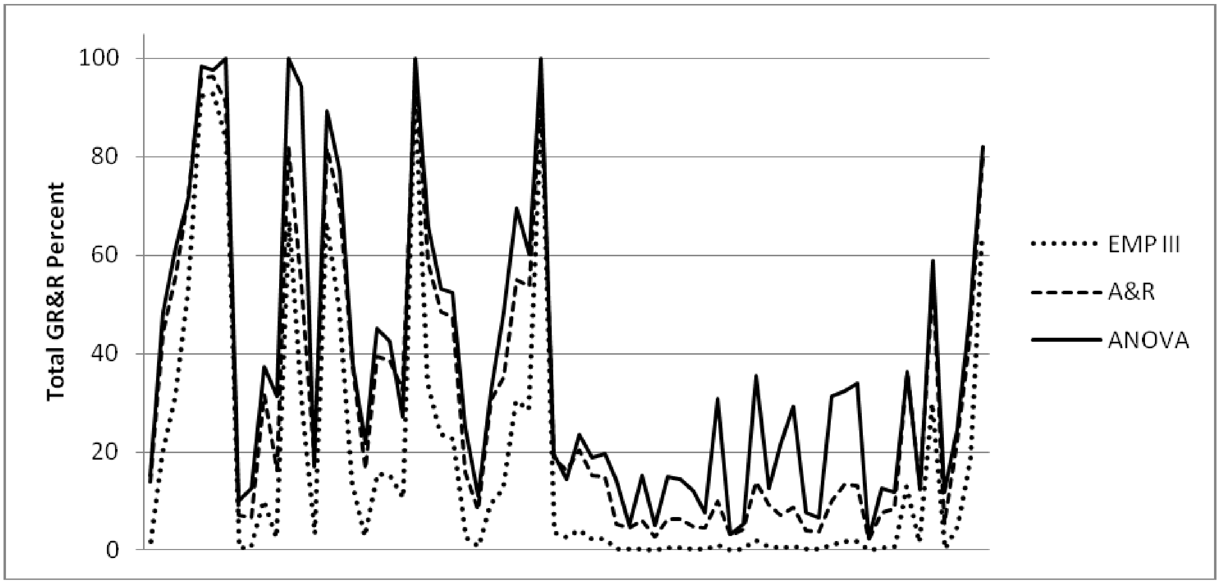

Figure 5.1. Chart of 67 randomly selected study measurement precision error (Total GRR) results by method of analysis.

Table 5.3 contains recommendations for which Gauge R&R method might be best suited under varying conditions. The product criticality column refers to the overall criticality of the firm's products. For example, is the firm producing high-risk medical devices, or less critical components? The EMP III and ANOVA methods might be better suited to less critical components where the accuracy and expense of the ANOVA method is not necessary.

The complexity of analysis column in Table 5.3 refers to the difficulty in performing the EMP III, A&R and ANOVA GR&R analysis. The EMP III and A&R methods lend themselves to manual and spreadsheet calculations and are available as Microsoft Excel spreadsheet add-ins. The ANOVA method is complicated and probably best suited for computers as noted by Measurement Systems Analysis (2010). It is recognized that some firms may not have the resources to purchase expensive computer software for performing GR&R studies.

The fourth column in Table 5.3 addresses which method of GR&R analysis is best suited if a more extensive GR&R study is needed to determine the root cause of high total measurement precision error or of the repeatability or reproducibility components of variation. The ANOVA method is best because it breaks down the reproducibility measurement precision error component into operator and operator by part interactions. Antony, Knowles and Roberts (1998) and Kazerouni (2009) suggest that the ANOVA method is more accurate than the A&R method in the presence of operator and part interaction in a Gauge R&R study. Additionally, the ANOVA method lends itself to adding other sources of variation, such as multiple gauges to assist in determining the root cause of high measurement precision error.

Table 5.3

Recommendations for GR&R Methods under Varying Conditions

Analysis Method	**Product Criticality**	**Complexity of Performing Analysis**	**Secondary Analysis for Deeper Understanding**	**Confidence Intervals Needed for Analysis Results**	**Guard Banding**
EMP III	Low-Medium	Low, Manual or Computer	Not well suited	Not Possible	Best suited for low to medium critical products
A&R	Low-Medium	Low, Manual or Computer	Not well suited	Not Possible	Best suited for low to medium critical products
ANOVA	High	High, Computer Recom-mended	Well suited	Possible	Best suited for high critical products

Further Study

A further study of the Gauge R&R method of measurement precision error estimation could include a survey of which of the three methods (EMP III, A&R, ANOVA) is used most frequently and if particular industry segments favor one method over another. Another potential survey area would be how pervasive the AIAG acceptance criterion use is in industry and among industry segments and is it an appropriate and valid criterion? Other potential survey areas could include which method of denominator selection is most frequently used in calculating the GR&R results (GR&R study data, historical process data, or dimensional tolerance) and how frequently and in what manner guard banding is employed to adjust dimensional tolerances to compensate for measurement precision errors.

Additional non-survey research could include a comparison of Gauge R&R study data by gauge type. For example, such research might take the form of a comparison of gauges that are highly operator skill dependent, such as micrometers and calipers, versus gauges that are less operator skill dependent, such as coordinate measurement machines (CMM).

If the study were repeated, it could include Gauge R&R studies from other medical device manufacturers and other industries. In addition, if enough Gauge R&R studies could be secured from multiple sources, the studies could be sub-divided and analyzed by gauge type and also the overall distribution of the Gauge R&R study results could be identified and characterized.

REFERENCES

10 CFR 74.45 Measurements and measurement control. (2012). Retrieved from http://www.nrc.gov/reading-rm/doc-collections/cfr/part074/part074-0045.html

21 CFR 210 Current good manufacturing practice in manufacturing, processing, packing, or holding of drugs (2011). Retrieved from http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=210&showFR=1&utm_source=fdaSearch&utm_medium=website&utm_term=21%20CFR%20210&utm_content=1

21 CFR 211 Current good manufacturing practices for finished pharmaceuticals (2011). Retrieved from http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=211&utm_source=fdaSearch&utm_medium=website&utm_term=21%20CFR%20211&utm_content=1

21 CFR 606 Current good manufacturing practice for blood and blood components. (2011). Retrieved from: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=606&utm_source=fdaSearch&utm_medium=website&utm_term=21%20CFR%20606&utm_content=1

21 CFR 820 Quality system regulation. (2011). Retrieved from http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=820&utmsource=fdaSearch&utm_medium=website&utm_term=21%20CFR%20820&utm_content=1

AIAG history highlights. (n.d.). Retrieved from: http://www.aiag.org/staticcontent/press/files/HistoryHighlights.pdf

Antony, J., Knowles, G. & Roberts, P. (1998). Gauge capability analysis: Classical versus ANOVA. *Quality Assurance*, *6*(3), 173-181.

Barrentine, L. B. (2003). *Concepts for R&R studies* (2nd ed.). Milwaukee, WI: American Society for Quality.

Bhattacharya, S. & Raj, R. A. (2004). Performance evaluation of multi-sensor data fusion technique for test range application. *Sadhana*, *29*(Part 2), 237-247.

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.

Camalier, L., Eberly, S., Miller, J. & Papp, M. (2007). *Guideline on the meaning and use of precision and bias data required by 40 CFR Part 58, appendix A.* Retrieved from http://www.epa.gov/ttnamti1/files/ambient/monitorstrat/precursor/07workshopmeaning.pdf

Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: John Wiley & Sons, Inc.

Devore, J. L. (2004). *Probability and statistics for engineering and the sciences* (6th ed.). Belmont, CA: Brooks/Cole-Thomson.

Ermer, D. S. (2006, March). Improved gage R&R measurement studies. *Quality Progress, 39*, 77-79.

Ermer, D. S. (2006, May). Appraiser variation in gage R&R measurement. *Quality Progress, 39*, 75-78.

Excel 2007 (Version 2007) [Computer software] Redmond, WA: Microsoft Corporation.

Gantt, J. S. (1959, March 9). Let's take the guesswork out of inspection! *American Machinist*, 117-122.

Good, P.I. & Hardin, J. W. (2009). *Common errors in statistics (3rd ed.).* Hoboken, NJ: John Wiley and Sons.

Grubbs, F. E. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association, 43*(242), 243-264.

Hughes, I.G. & Hase, T. P. A. (2010). *Measurements and their uncertainties*. New York: Oxford University Press.

ISO 9001 (2008). Geneva, Switzerland: International Organization for Standardization.

ISO 13485 (2003). Geneva, Switzerland: International Organization for Standardization.

ISO/TS 16949 (2009). Geneva, Switzerland: International Organization for Standardization.

ISO 17025 (2005). Geneva, Switzerland: International Organization for Standardization.

JMP (Version 10) [Computer Software]. Cary, NC: SAS Institute, Inc.

Kazerouni, A. M. (2009). Design and analysis of gauge R&R studies: Making decisions based on ANOVA method. *Proceedings of World Academy of Science, Engineering and Technology: Vol. 40*. (pp. 31-35).

Knowles, G., Antony, J. & Vickers, G. (2000). A practical methodology for analyzing and improving the measurement system. *Quality Assurance, 8*, 59-75.

Koehler, E., Brown, E., & Haneuse, S.J. (2009). On the assessment of monte carlo error in simulation-based statistical analysis. *The American Statistician, 63*(2), 155-162.

Kotz, S. & Johnson, N. L. (Eds.). (1983). *Encyclopedia of statistical sciences, Volume 4*. New York: John Wiley and Sons.

Manderscheidt, L. V. (1965). Significance levels-0.05, 0.01, or? *Journal of Farm Economics 47*, (5), 1381-1385.

Measurement systems analysis Reference Manual (1990). Southfield, MI: Automotive Industry Action Group (AIAG).

Measurement systems analysis- MSA (4th ed.) (2010). Southfield, MI: Automotive Industry Action Group (AIAG).

Minitab Statistical Software (Version 16) [Computer software]. State College, PA: Minitab Inc.

Montgomery, D.C. & Runger, G. C. (1993). Gauge capability and designed experiments. Part I: Basic methods. *Quality Engineering, 6*(1), 115-135.

Norman, G. R. (2005). The relation between the minimally important difference and patient benefit. *COPD: Journal of Chronic Obstructive Pulmonary Disease, 2*, 69-73.

Omdahl, T. *Quality dictionary*. West Terre Haute, IN: Quality Council of Indiana.

Osma, A. (2011). An assessment of the robustness of gauge repeatability and reproducibility analysis in automotive components. *Proceedings of the Institution of Mechanical Engineers, Journal of Automotive Engineering, Part D, 255*, 895-912. doi: 10.1177/0954407011401504

QI Macros [Computer Software]. Denver, CO: KnowWare.

Sheskin, D. J. (2004). *Handbook of parametric and nonparametric statistical procedures (3rd ed.*). New York: Chapman & Hall/CRC.

SPC for Excel [Computer Software]. Cypress, TX: BPI Consulting, LLC.

Standard guide for measurement systems analysis (MSA) - E2728. (2010). West Conshohocken, PA: ASTM International.

Subrahmaniam, K., Subrahmaniam, K., & Messeri, J. Y. (1975). On the robustness of some tests of significance in sampling form a compound normal population. *Journal of the American Statistical Association 70*, (350), 435-438.

Taylor, B. N. & Kuyatt, C. E. (1994). *Guidelines for evaluating and expressing the uncertainty of NIST measurement results* (NIST Technical Note 1297). Retrieved from http://www.nist.gov/pml/pubs/tn1297/index.cfm

Traver, R. W. (1962). *Measuring equipment repeatability- the rubber ruler*. Transactions of the American Society for Quality Control, Milwaukee, WI.

van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). Hoboken, N.J.: John Wiley and Sons, Inc.

van den Heuvel, E. R. (2000, May). *Gage R&R studies for non standard situations*. Paper presented at the ASQ 54th Annual Quality Conference Proceedings, Indianapolis, IN.

Wheeler, D. J. (2009). *An honest gauge R&R study*. Paper presented at the 2006 ASQ/ASA Fall Technical Conference. Retrieved from: http://www.spcpress.com/pdf/DJW189.pdf

Wheeler, D J. (2006). *EMP III using imperfect data*. Knoxville, TN: SPC Press.

Wheeler, D. J. (2003). *Range based analysis of means*. Knoxville, TN: SPC Press.

Wheeler, D. J., & Lyday, R. W. (1989). *EMP: Evaluating the measurement process (2nd ed.).* Knoxville, TN: SPC Press.

APPENDIX A: RESULTS FROM 224 ACTUAL (SEED) GAUGE R&R STUDIES

Table A1

Descriptive Statistics for Total 224 Seed Gauge R&R Studies (in percentages)

Source of Variation	Statistic	EMP III	A&R	ANOVA	Max-Min Difference in Average	Maximum Std. Dev.
Repeatability	Average	11.55	23.38	27.42	15.87	
	Std Dev	19.15	24.76	28.46		28.46
Reproducibility	Average	7.43	15.24	16.60	9.17	
	Std Dev	17.01	22.60	22.95		22.95
Total GR&R	Average	18.98	30.31	34.38	15.40	
	Std Dev	29.19	31.36	34.87		34.87
Part-to-part	Average	81.02	87.43	81.58	6.41	
	Std Dev	29.19	21.45	30.30		30.30

APPENDIX B, SAMPLE SIZE CALCULATION REPORTS

POWER AND SAMPLE SIZE- HYPOTHESIS ONE (TOTAL GAUGE R&R)

One-way ANOVA

Alpha = 0.05 Assumed standard deviation = 34.87

Factors: 1 Number of levels: 3

Maximum Difference	Sample Size	Target Power	Actual Power
15.4	131	0.9	0.900567

The sample size is for each level.

Minitab Statistical Software

SAMPLE SIZE- CORRELATION- HYPOTHESIS TWO (REPEATABILITY)

One-way ANOVA

Alpha = 0.05 Assumed standard deviation = 28.46

Factors: 1 Number of levels: 3

Maximum Difference	Sample Size	Target Power	Actual Power
15.87	83	0.9	0.902226

The sample size is for each level.

Minitab Statistical Software

POWER AND SAMPLE SIZE – HYPOTHESIS THREE (REPRODUCIBILITY)

One-way ANOVA

Alpha = 0.05 Assumed standard deviation = 22.95

Factors: 1 Number of levels: 3

	Maximum Sample	Target	
Difference	Size	Power	Actual Power
9.17	160	0.9	0.900911

The sample size is for each level.

Minitab Statistical Software

POWER AND SAMPLE SIZE – HYPOTHESIS FOUR (PART- TO-PART)

One-way ANOVA

Alpha = 0.05 Assumed standard deviation = 30.3

Factors: 1 Number of levels: 3

	Maximum Sample	Target	
Difference	Size	Power	Actual Power
6.41	567	0.9	0.900274

The sample size is for each level.

Minitab Statistical Software

SAMPLE SIZE- CORRELATION- HYPOTHESIS FIVE (TOTAL GR&R CORRELATION)

EMP III vs. A&R Seed Data Correlation= 0.969

EMP III vs. ANOVA Seed Data Correlation= 0.915

A&R vs. ANOVA Seed Data Correlation= 0.965

Maximum difference among the above comparisons: 0.0540

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.0540}{1-0.0540} = \frac{1}{2}\ln\frac{1.0540}{0.9460} = \frac{1}{2}0.1081 = 0.0541$$

Where, ρ=estimate of population correlation

Δ= desired detection in population correlation

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.0541)^2} = 7175.05$$ round up to 7176

van Belle (2008), 71-72, 29-30.

SAMPLE SIZE- CORRELATION- HYPOTHESIS SIX (REPEATABILITY CORRELATION)

EMP III Repeatability vs. A&R Repeatability Seed Data Correlation= 0.961

EMP III Repeatability vs. ANOVA Repeatability Seed Data Correlation= 0.910

A&R Repeatability vs. ANOVA Repeatability Seed Data Correlation= 0.968

Maximum difference among the above comparisons: 0.0580

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.0580}{1-0.0580} = \frac{1}{2}\ln\frac{1.0580}{0.9420} = \frac{1}{2}0.1161 = 0.0581$$

Where, ρ=estimate of population correlation

Δ= desired detection in population correlation

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.0581)^2} = 6221.10$$ round up to 6222

van Belle (2008), 71-72, 29-30.

SAMPLE SIZE- CORRELATION- HYPOTHESIS SEVEN (REPRODUCIBILITY CORRELATION)

EMP III Reproducibility vs. A&R Reproducibility Seed Data Correlation= 0.948

EMP III Reproducibility vs. ANOVA Reproducibility Seed Data Correlation= 0.875

A&R Reproducibility vs. ANOVA Reproducibility Seed Data Correlation= 0.925

Maximum difference among the above comparisons: 0.0730

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.0730}{1-0.0730} = \frac{1}{2}\ln\frac{1.0730}{0.9270} = \frac{1}{2}0.1463 = 0.0731$$

Where, ρ=estimate of population correlation

Δ= desired detection in population correlation

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.0731)^2} = 3929.92$$ round up to 3930

van Belle (2008), 71-72, 29-30.

SAMPLE SIZE- CORRELATION- HYPOTHESIS EIGHT (PART-TO-PART)

EMP III Part-to-part vs. A&R Part-to-part Seed Data Correlation= 0.988

EMP III Part-to-part vs. ANOVA Pat-to-part Seed Data Correlation= 0.907

A&R Part-to-part vs. ANOVA Part-to-part Seed Data Correlation= 0.902

Maximum difference among the above comparisons: 0.0860

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.0860}{1-0.0860} = \frac{1}{2}\ln\frac{1.0860}{0.9140} = \frac{1}{2}.0.1724 = 0.0862$$

Where, ρ=estimate of population correlation

Δ= desired detection in population correlation

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.0862)^2} = 2826.21$$
round up to 2827

0.05 Difference

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.05}{1-0.05} = \frac{1}{2}\ln\frac{1.05}{0.95} = \frac{1}{2}0.1001 = 0.05$$

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.05)^2} = 8400.0$$

0.04 Difference

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.04}{1-0.04} = \frac{1}{2}\ln\frac{1.04}{0.96} = \frac{1}{2}0.080 = 0.04$$

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.04)^2} = 13{,}125.0$$

0.0456 Difference

$$\Delta = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \frac{1}{2}\ln\frac{1+0.0456}{1-0.0456} = \frac{1}{2}\ln\frac{1.0456}{0.9544} = \frac{1}{2}0.0913 = 0.0456$$

$$n = \frac{21}{\Delta^2} = \frac{21}{(0.0456)^2} = 10{,}099.26$$
round up to 10,100

van Belle (2008), 71-72, 29-30.

APPENDIX C: HISTOGRAM COMPARISONS OF SEED GR&R STUDY DATA AND SIMULATED GR&R STUDY DATA

NOTE: Appendix C starts on the following page so top to bottom comparisons can be made for the same Gauge R&R components.

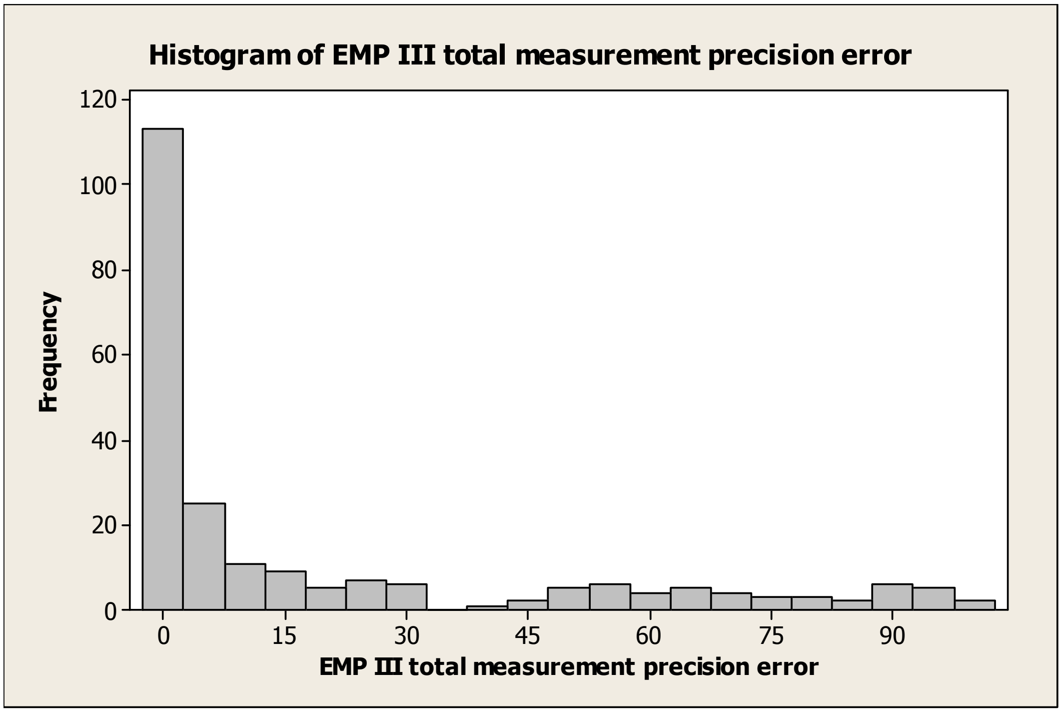

Figure C1. Seed GR&R Study EMP III Total Measurement Precision Error.

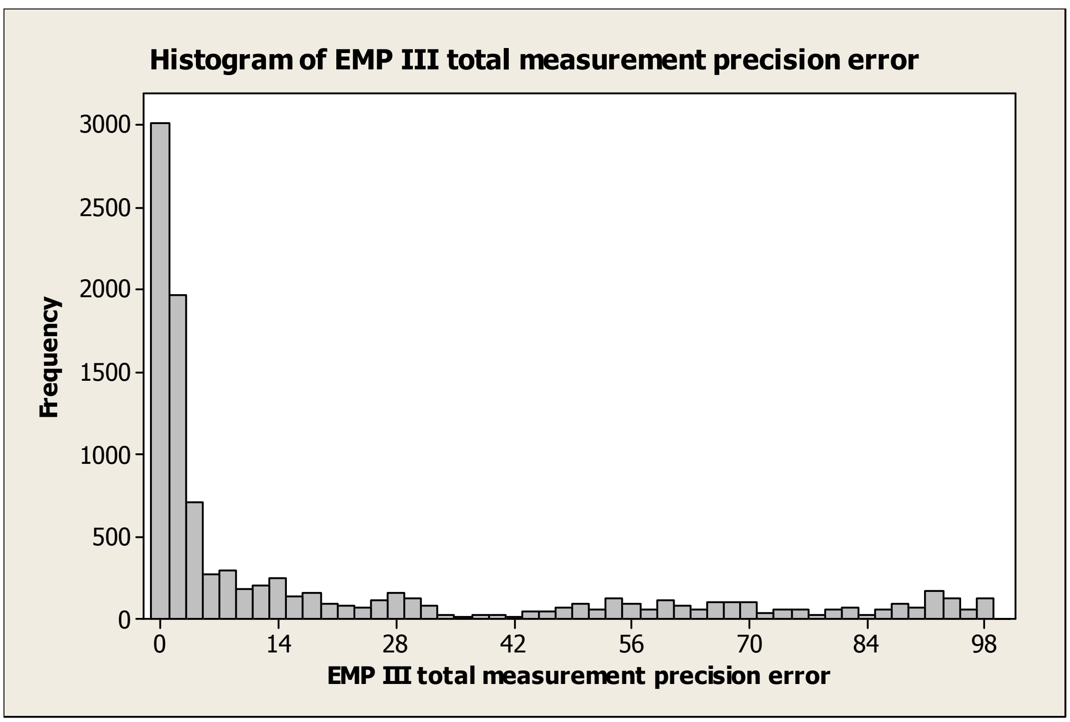

Figure C2. Simulated GR&R Study EMP III Total Measurement Precision Error.

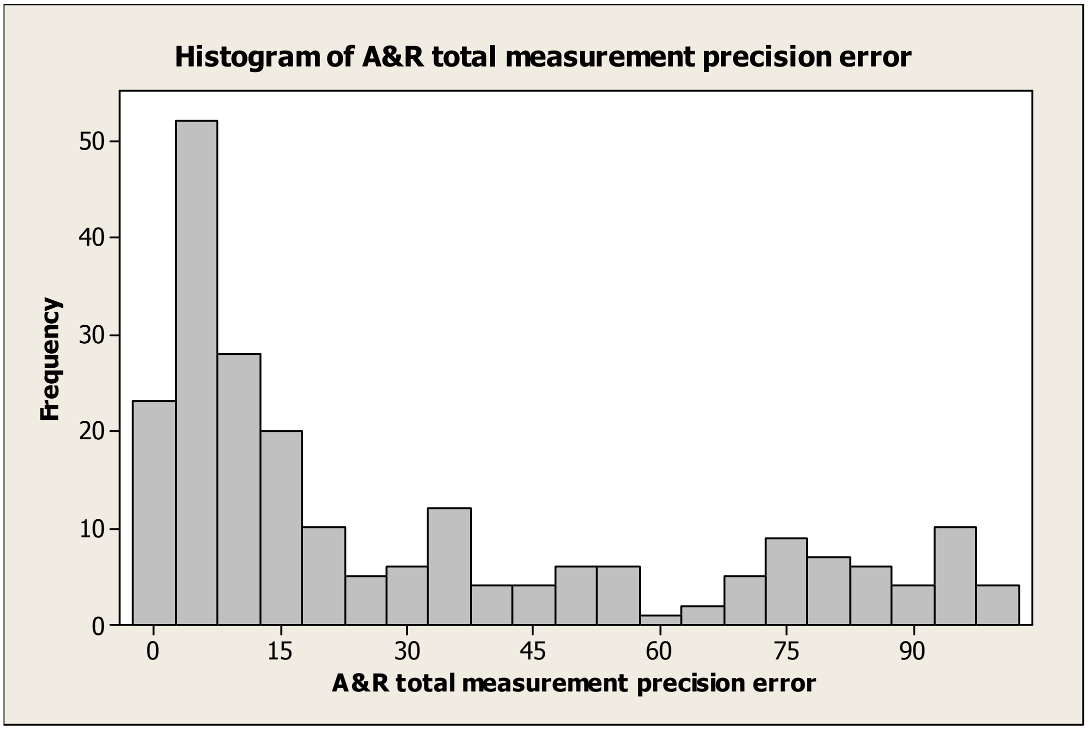

Figure C3. Seed GR&R study A&R Total Measurement Precision Error.

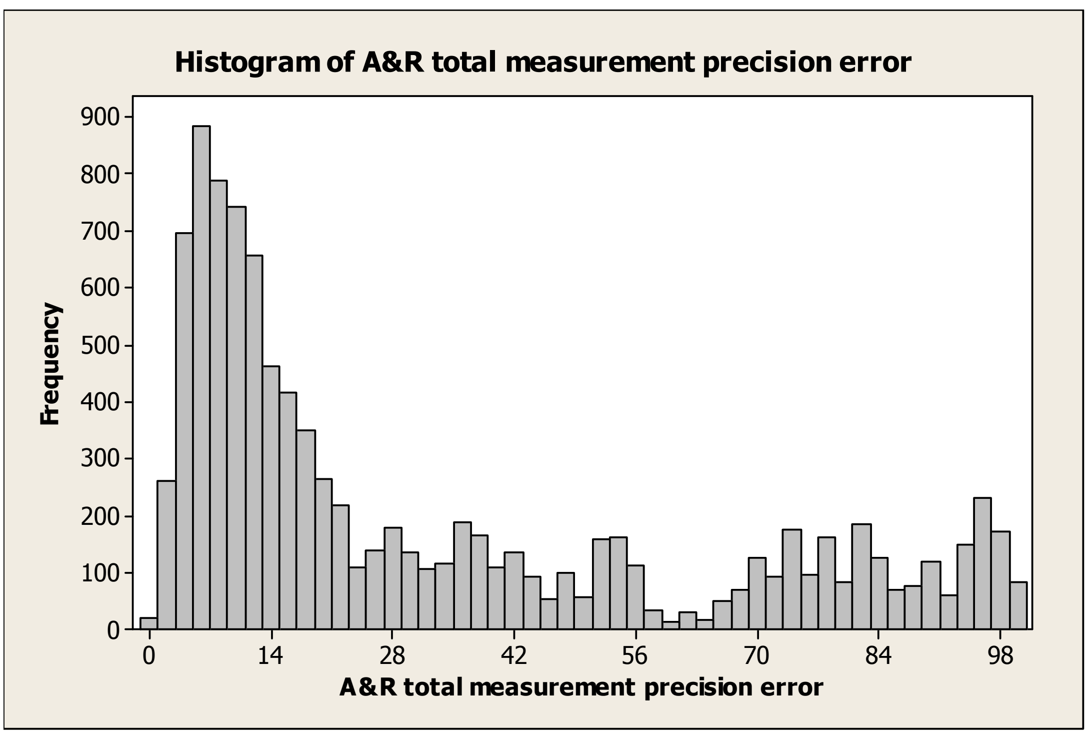

Figure C4. Simulated GR&R Study A&R Total Measurement Precision Error.

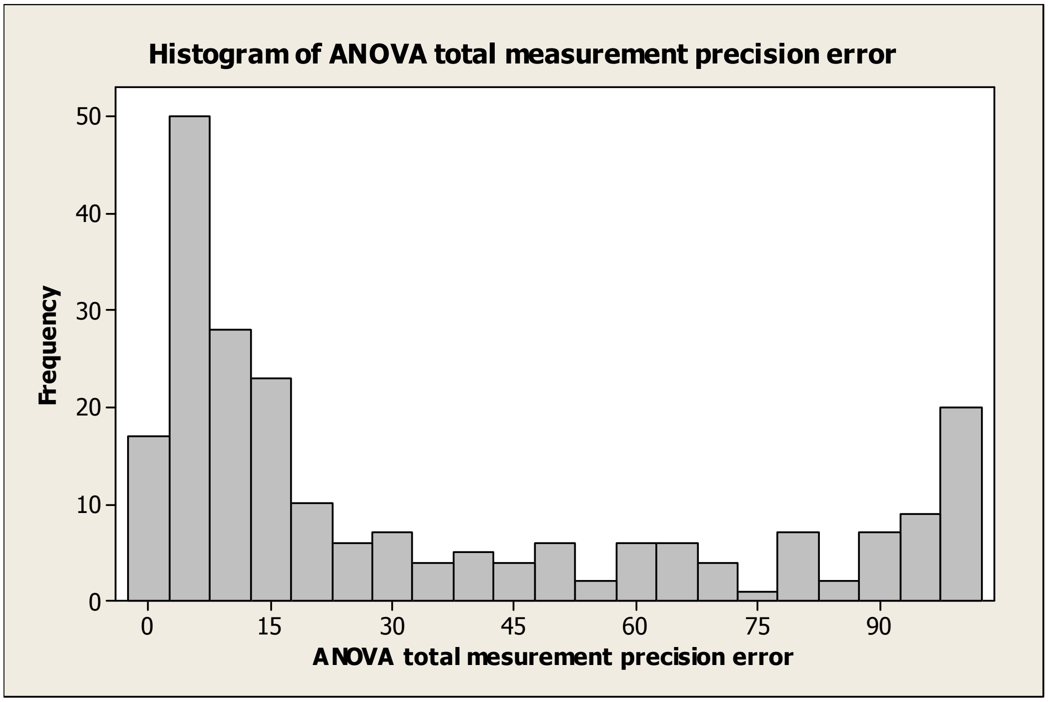

Figure C5. Seed GR&R Study ANOVA Total Measurement Precision Error.

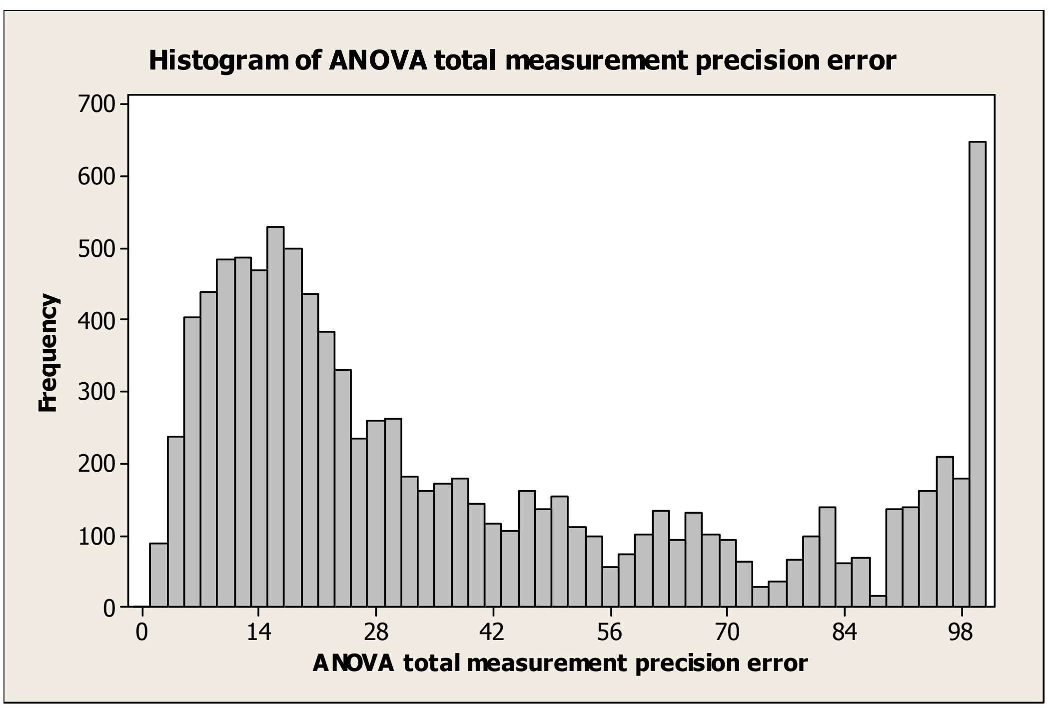

Figure C6. Simulated GR&R Study ANOVA Total Measurement Precision Error.

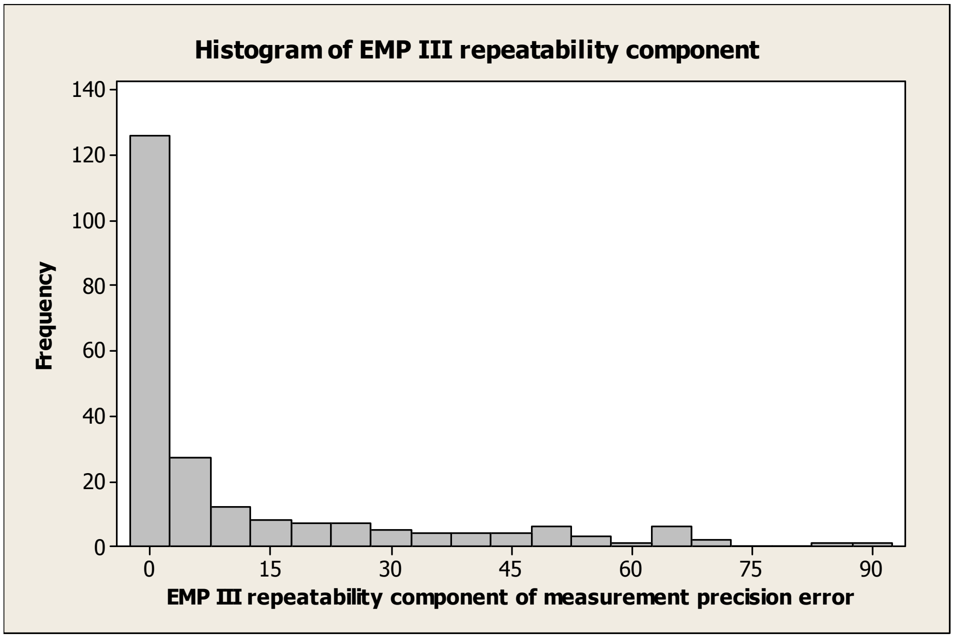

Figure C7. Seed GR&R study EMP III Repeatability Component of Measurement Precision Error.

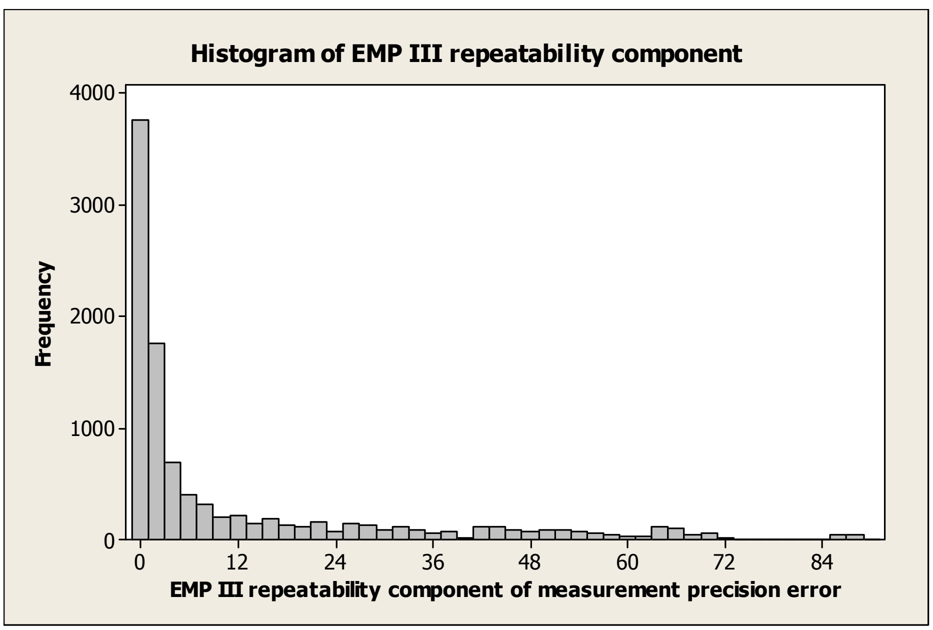

Figure C8. Simulated GR&R Study EMP III Repeatability Component of Measurement Precision Error.

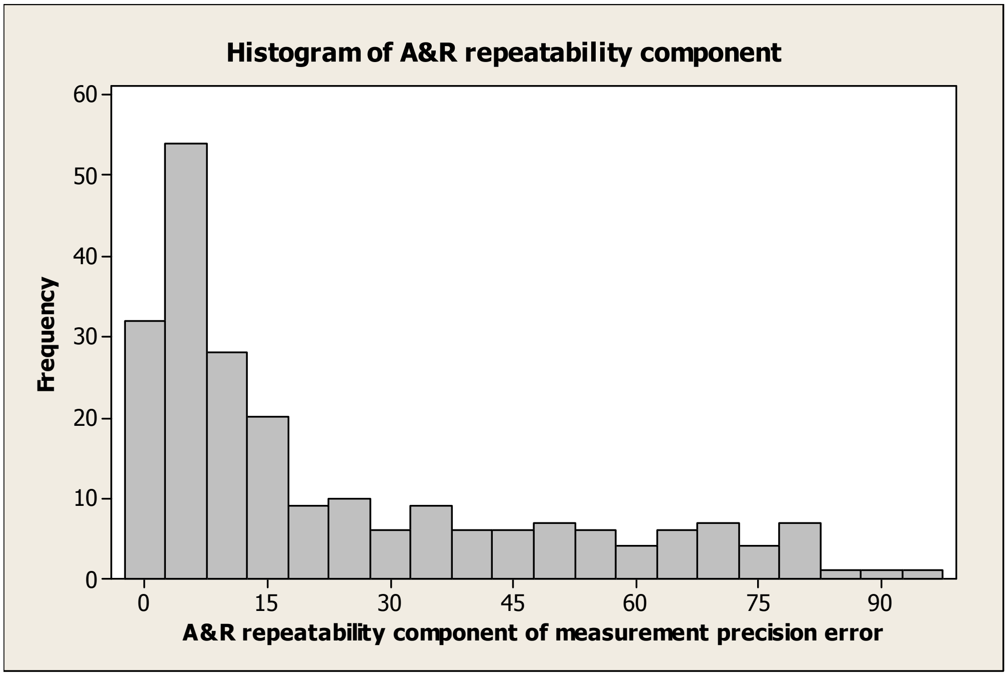

Figure C9. Seed GR&R Study A&R Repeatability Component of Measurement Precision Error.

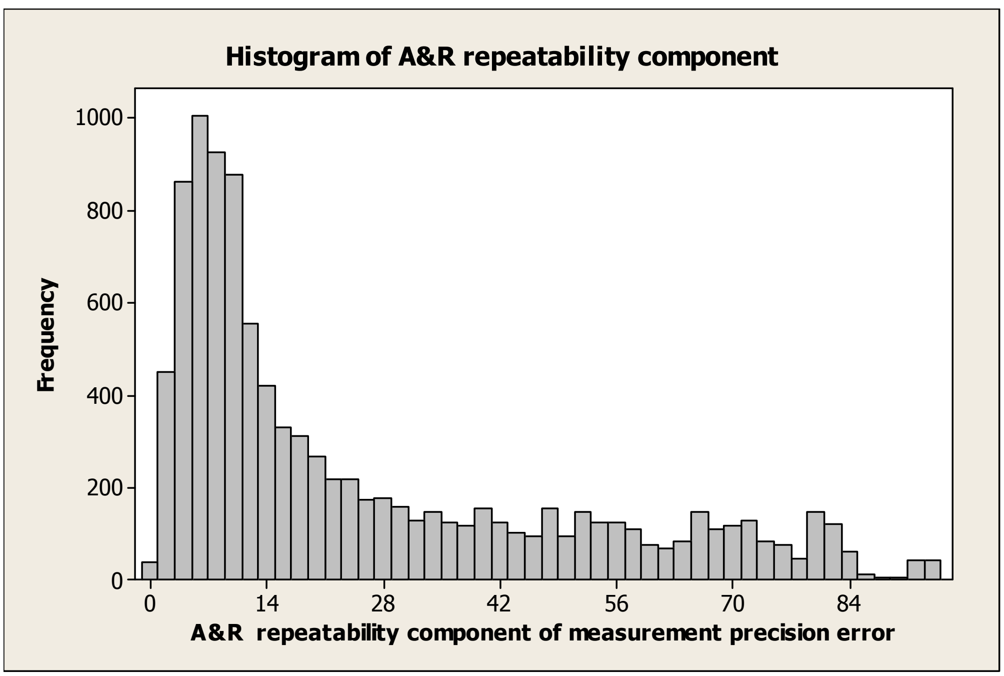

Figure C10. Simulated GR&R Study A&R Repeatability Component of Measurement Precision Error.

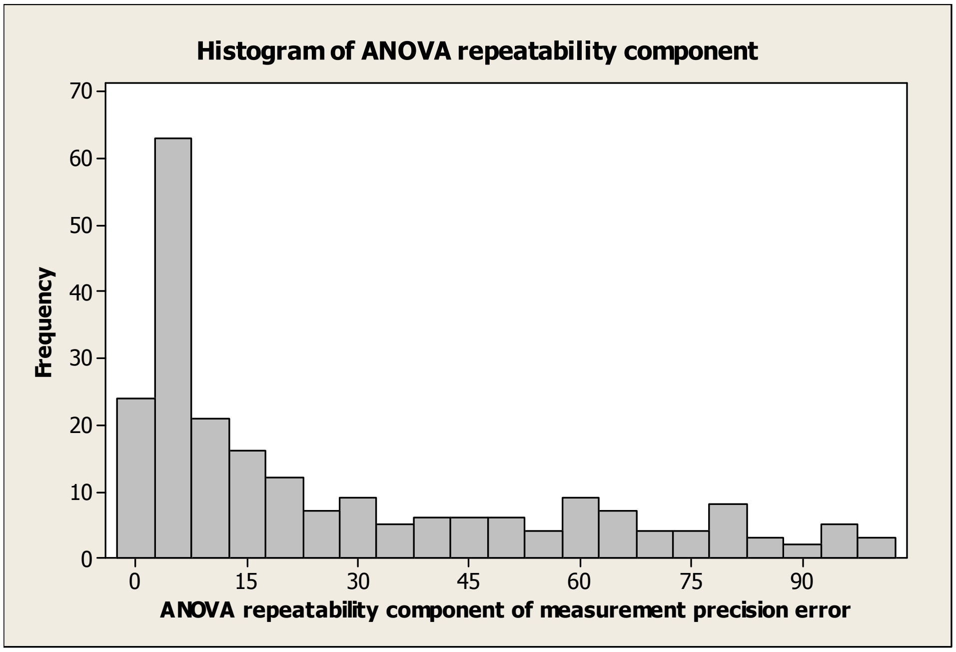

Figure C11. Seed GR&R Study ANOVA Repeatability Component of Measurement Precision Error.

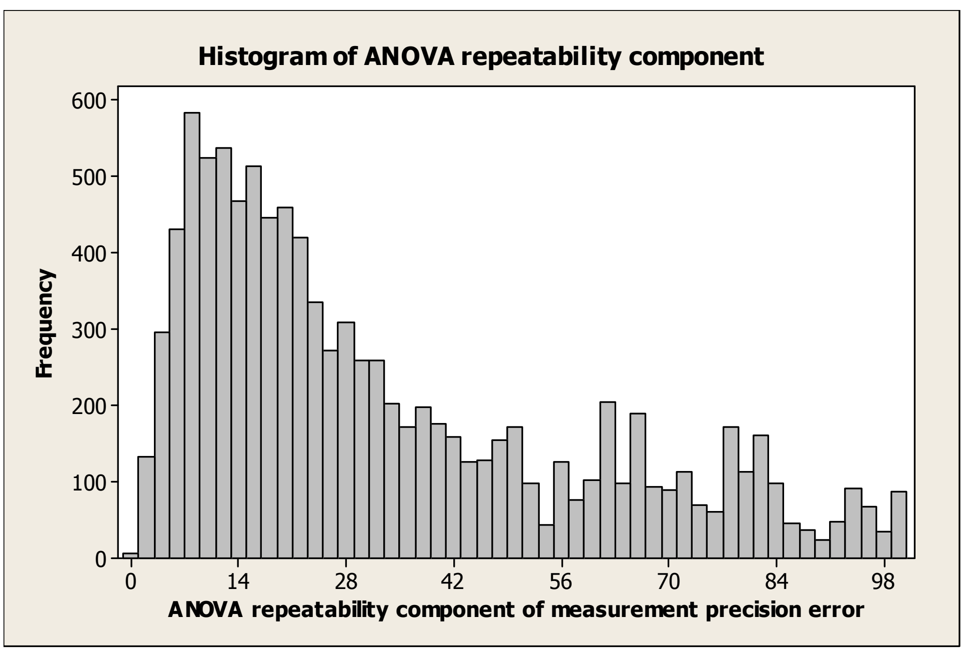

Figure C12. Simulated GR&R Study ANOVA Repeatability Component of Measurement Precision Error.

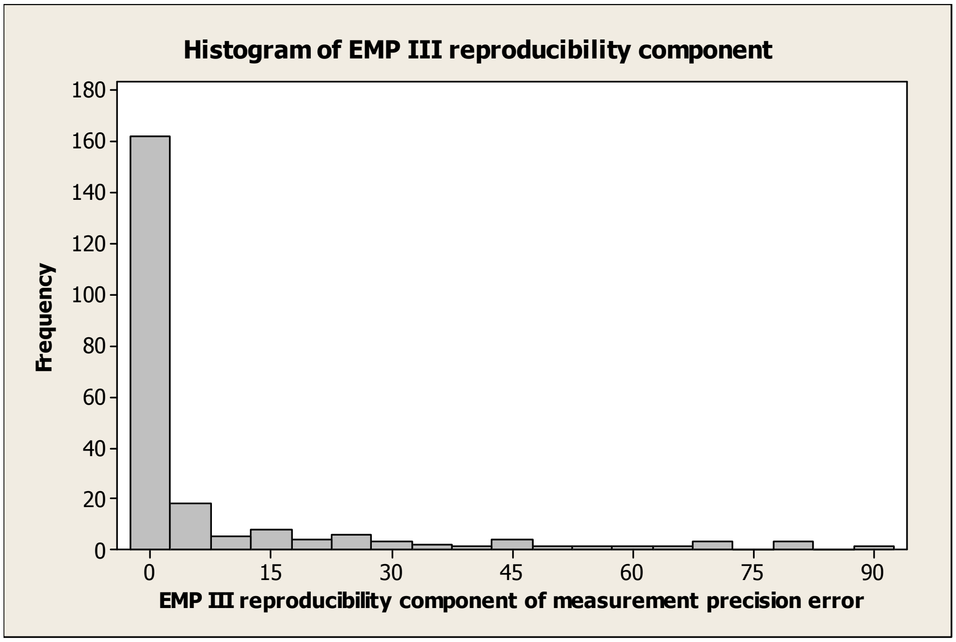

Figure C13. Seed GR&R Study EMP III Reproducibility Component of Measurement Precision Error.

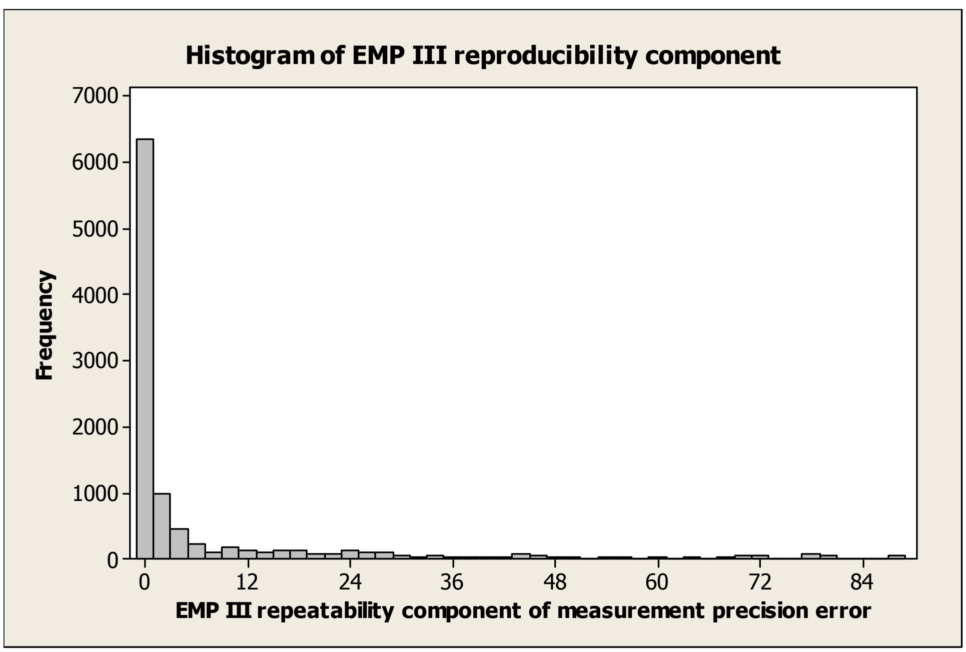

Figure C14. Simulated GR&R Study EMP III Reproducibility Component of Measurement Precision Error.

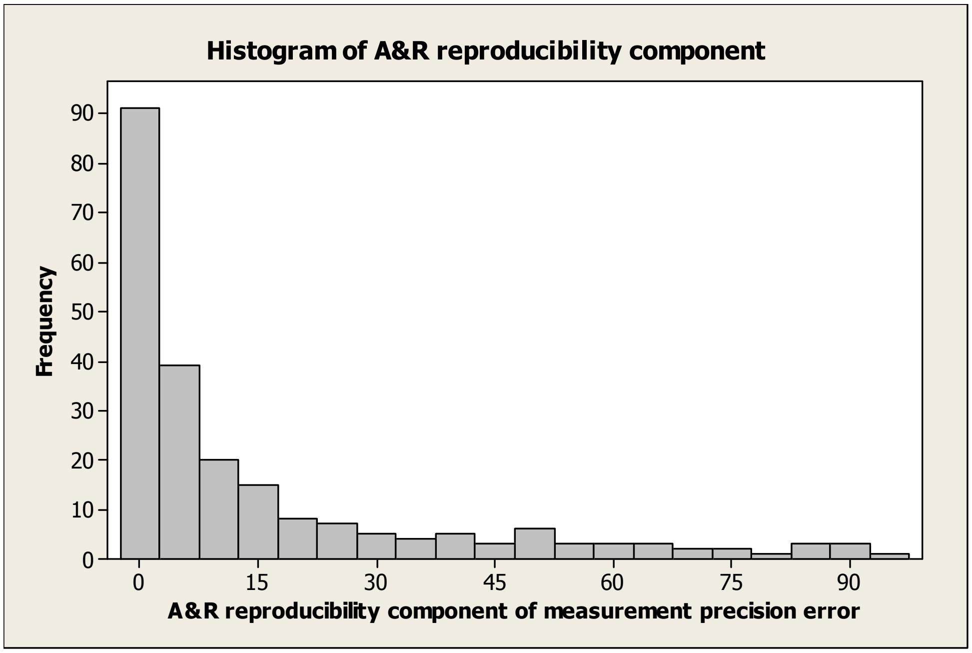

Figure C15. Seed GR&R Study A&R Reproducibility Component of Measurement Precision Error.

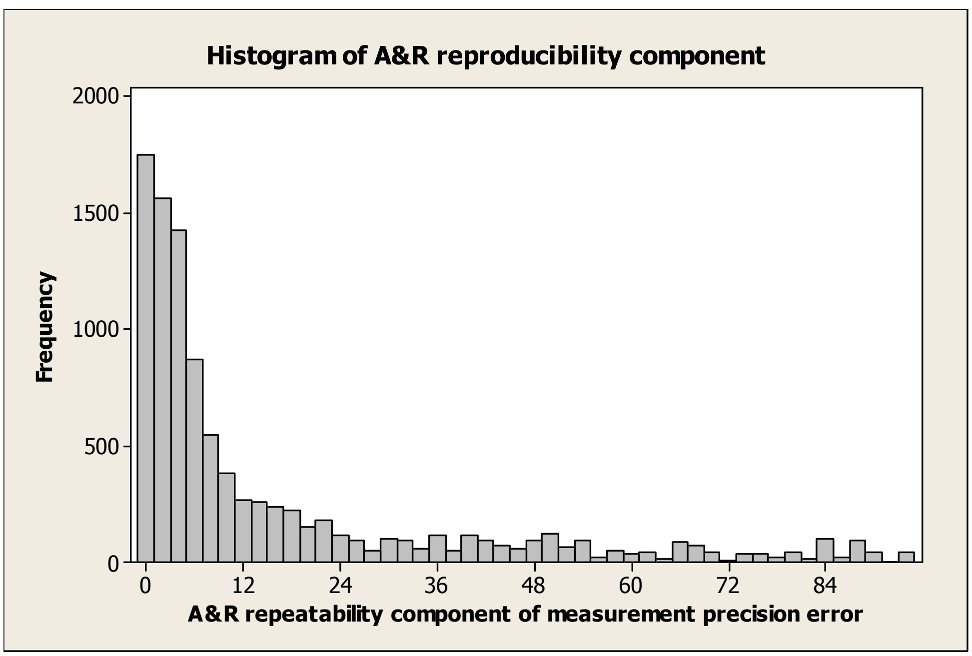

Figure C16. Simulated GR&R Study A&R Reproducibility Component of Measurement Precision Error.

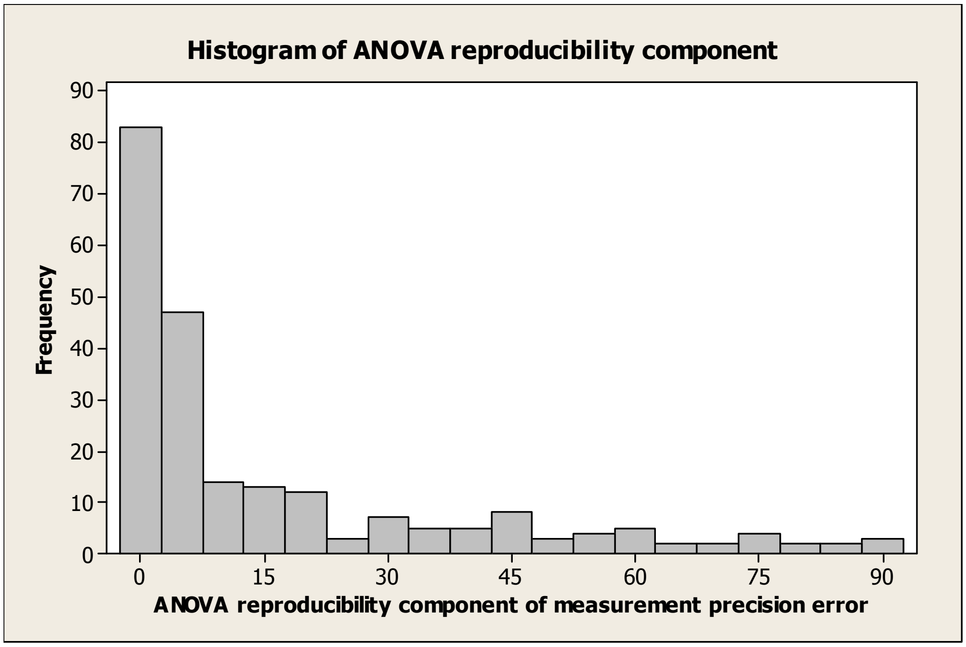

Figure C17. Seed GR&R Study ANOVA Reproducibility Component of Measurement Precision Error.

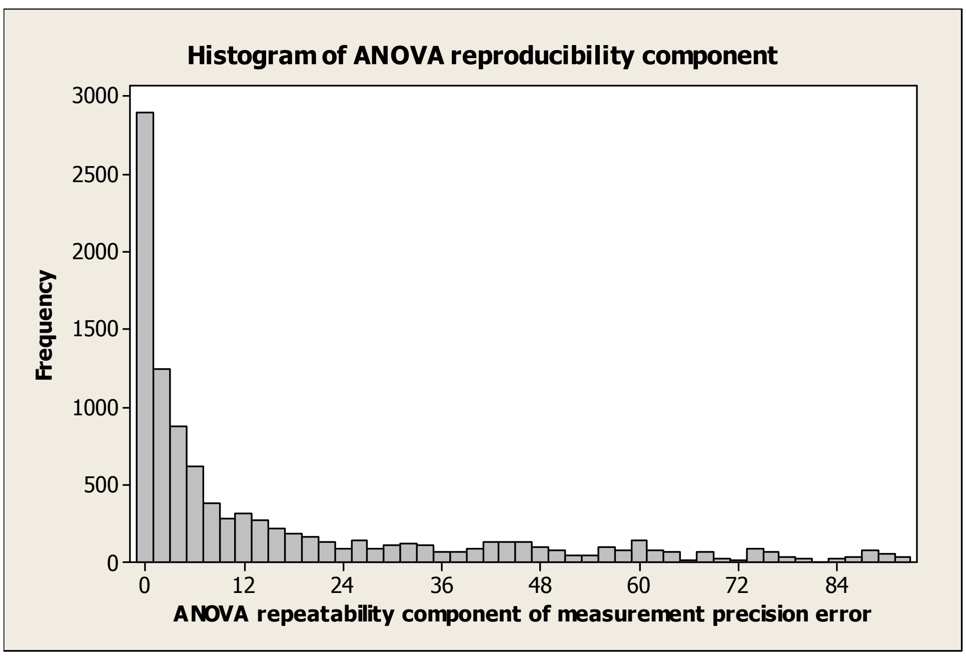

Figure C18. Simulated GR&R Study ANOVA Reproducibility Component of Measurement Precision Error.

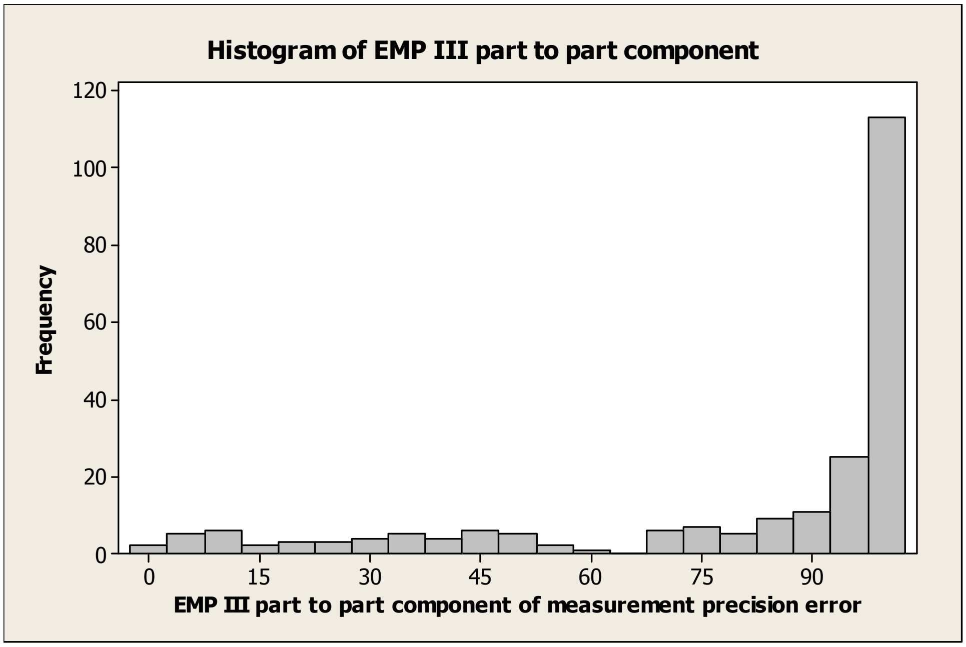

Figure C19. Seed GR&R Study EMP III Part-to-part Component of Measurement Precision Error.

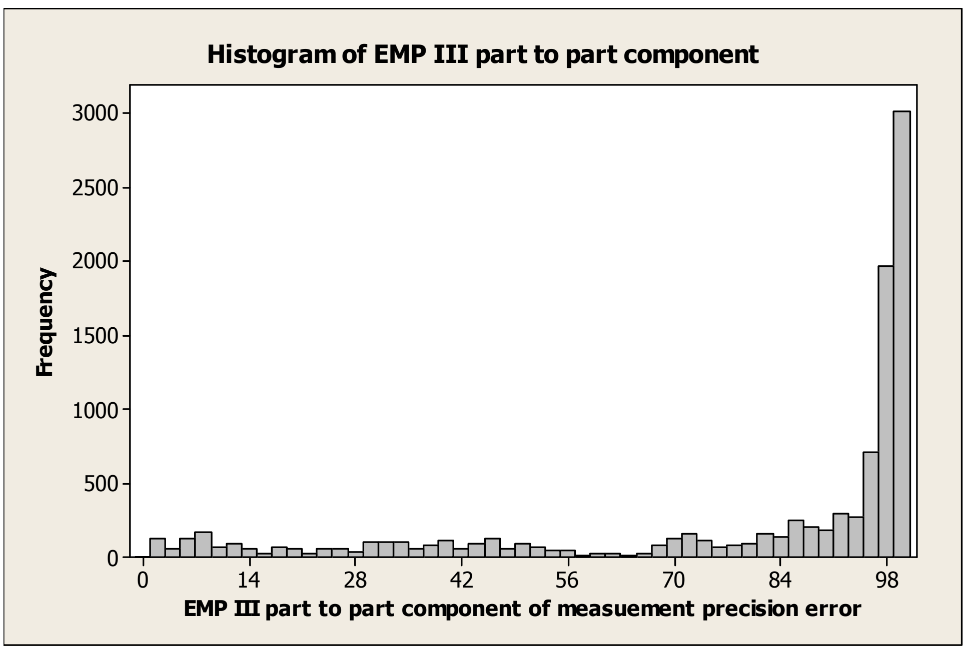

Figure C20. Simulated GR&R Study EMP III Part- to-part Component of Measurement Precision Error.

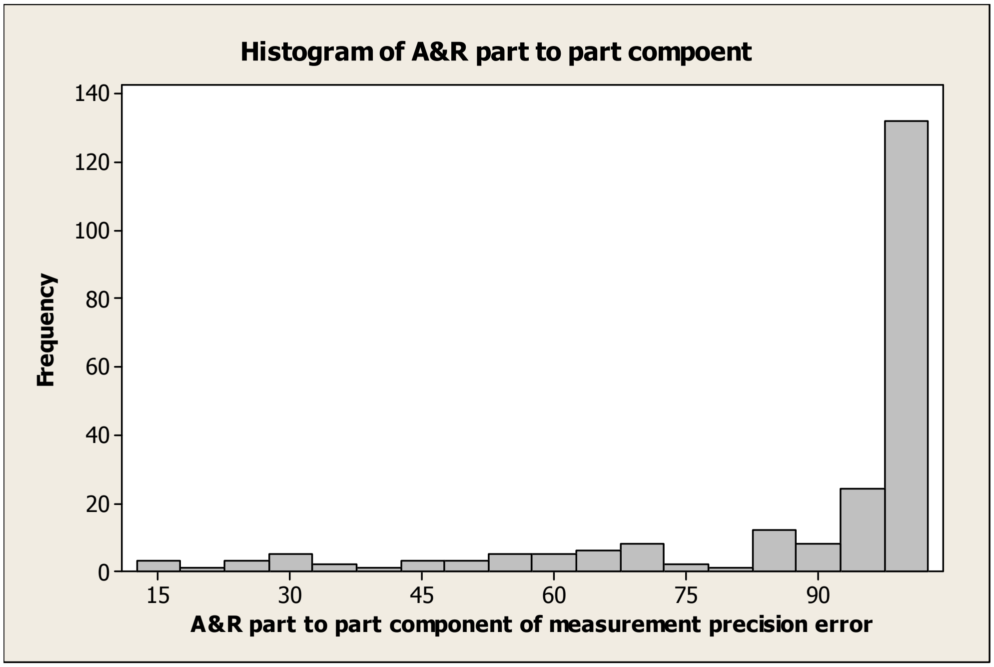

Figure C21. Seed GR&R Study A&R Part-to-part Component of Measurement Precision Error

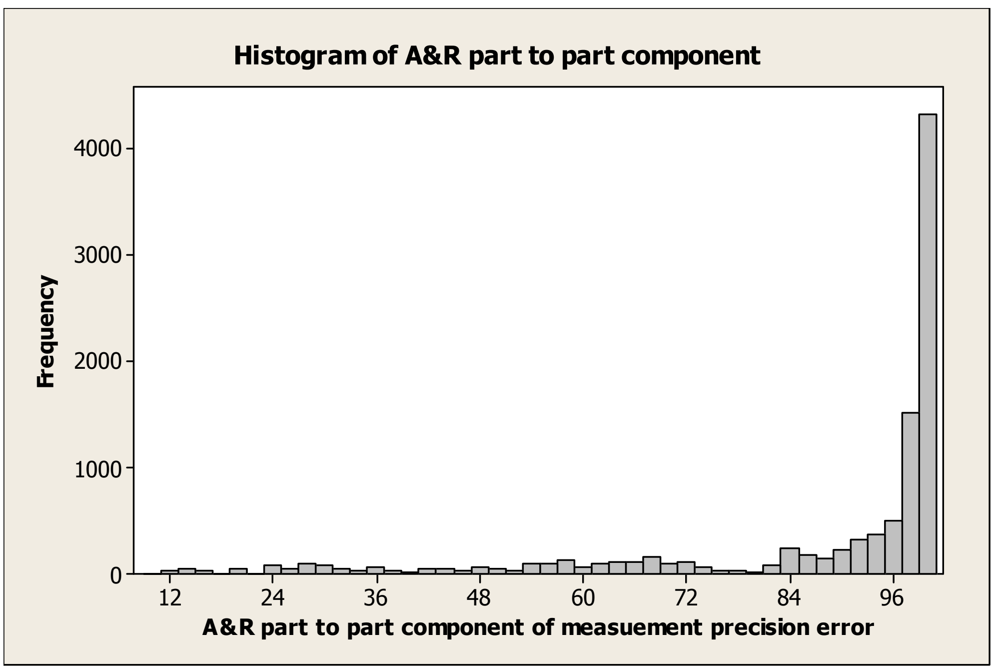

Figure C22. Simulated GR&R Study A&R Part-to-part Component of Measurement Precision Error.

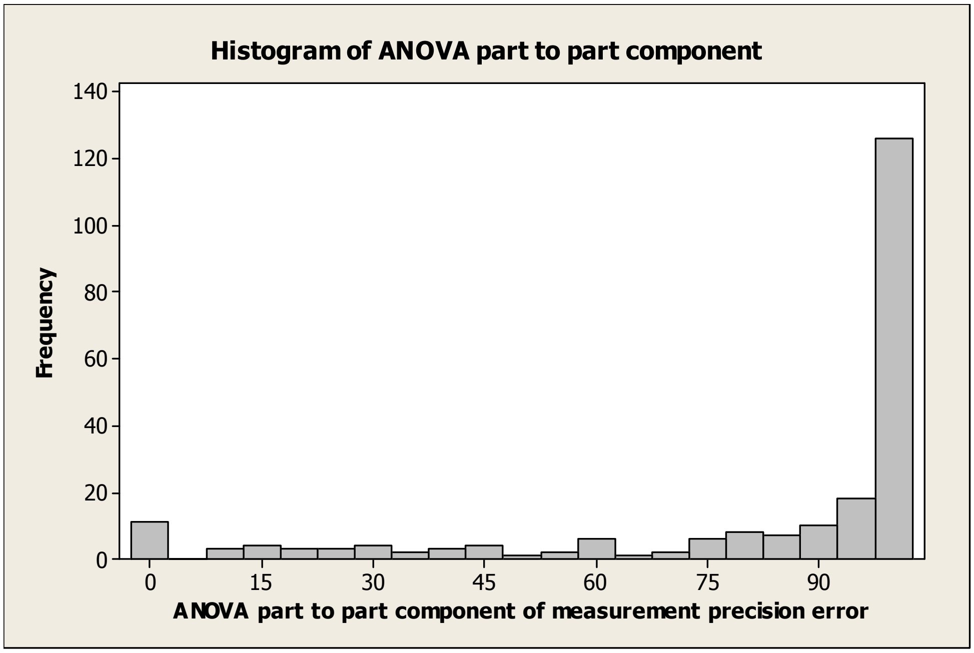

Figure C23. Seed GR&R Study ANOVA Part-to-part Component of Measurement Precision Error.

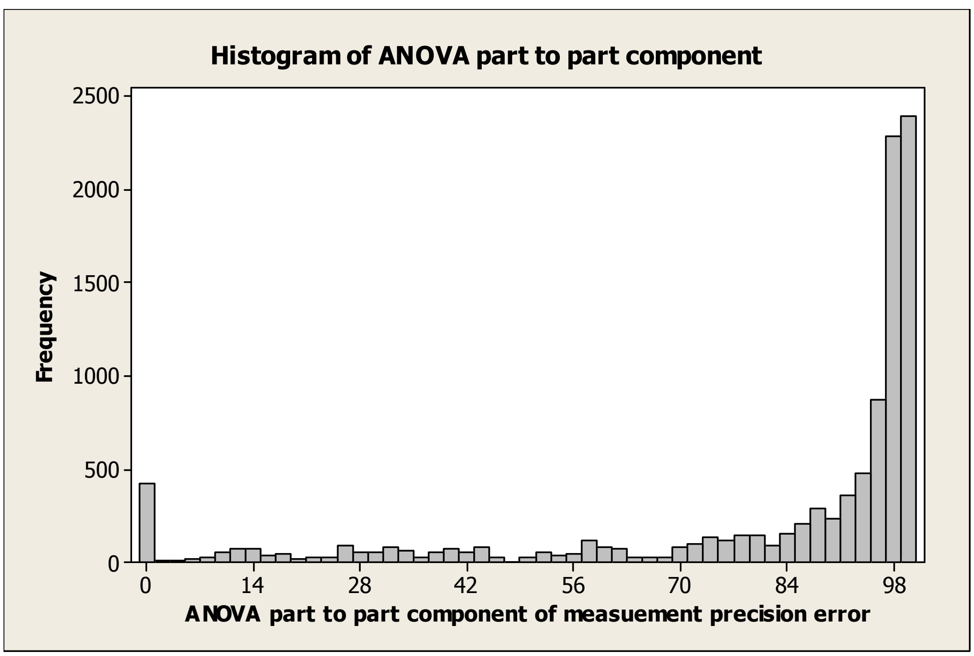

Figure C24. Simulated GR&R Study ANOVA Part-to-part Component of Measurement Precision Error.

APPENDIX D: MINITAB MACROS

% RAND X1 S1 X2 S2 X3 S3 c1-c45 c50-c94

This Minitab macro takes the input from one seed Gauge R&R study and outputs 45 sets of 1,000 normally distributed data points. The macro then randomly selects 10 values from each of the 45 sets of 1,000 normally distributed data points. One value is then randomly selected from the 10 values and used to substitute an actual value in a seed Gauge R&R study. In total, 45 simulations are created from each of the seed data sets.

- Inputs X1, X2, and X3 are the actual three-operator average measurement values from the seed Gauge R&R study. (X1=average for operator one, X2=average for operator two, X3=for operator three).
- Inputs S1, S2, S3 are the actual three-operator standard deviation measurement values from the seed Gauge R&R study. (S1=standard deviation for operator one, S2=standard deviation for operator two, S3=standard deviation for operator three).
- c1-c45 are the designated output columns for the 1,000 normally distributed data points for which the simulation output values will be randomly selected.
- Variables c50-c94 are the 45 sets of output of randomly selected ten member values. The first 15 are derived from the operator #1 input values, the next 15 are derived from the operator #2 input values and the final 15 are derived from the third operator input values.

The actual derivation in the 1,000 data point sets is a change in the standard deviation values; the actual average values were used to create all of the random normal distributions. The 10 randomly selected output values are set up in groups of five. The first value in each set of five was a normal distribution at .75 of the input standard deviation value; the second distribution in the set of five was 1.0 times the input standard deviation value; the next was 1.25, then 1.5 and finally 2.0 times the input standard deviation value. This pattern was repeated within the %RAND macro nine times, three times for each operator, representing the operators' first, second and third measurement sessions. Thus, the total output from the macro in columns c50-c94 was 45 measurement sets each comprising of 10 data points.

%NORMPLOTSTOREA c1-c45;

STORE c106.

This macro checks the normality of the columns of data in the arguments (c1-c45 in this example) and places the P-value of the Anderson-Darling normality test in the STORE column (c106 in this example). The P-value results are stacked in the STORE column (45, P-Values in the example). Note that this macro was modified from Minitab macro %NORMPLOTSTORE to accept multiple columns of data input and multiple P-Value outputs. This macro was used to test the assumption that the data input to the simulations were normally distributed.

%GRRXRc c1 c2 c3-c227 c230 c231 c232 c233 c234 c235

This Minitab macro inputs Gauge R&R measurement data (simulated data) and outputs the results of the AIAG A&R Gauge Study results. The inputs are c1 (operator column), c2 (Part number column) and c3-c227 (the raw Gauge R&R measurement data; in this example 225, sets of measurement data). The six columns of output are stored in columns c230 through c235 in this example.

The output results are stacked one on top of another. Table A4.1 below shows an example of the six columns of data output. Note that the column headings in the table and the left most column are not included in the macro data output. They are included here to describe the data output.

Table D1

Example % GRRXRc Minitab Macro Output

Source	**Data Column Heading**	**Variance Component**	**% Contri-bution of Var. Comp**	**Std. Dev. (SD)**	**Study Variation (6*SD)**	**% Study Variation (%SV)**
Tot. GR&R	SD.75-226	0.0000055	7.31	0.0023452	0.0140711	27.03
Repeatability	SD.75-226	0.0000052	6.93	0.0022844	0.0137063	26.33
Reproducibility	SD.75-226	0.0000003	0.37	0.0005305	0.0031832	6.12
Part-To-part	SD.75-226	0.0000698	92.69	0.0083528	0.0501169	96.28
Total Variation	SD.75-226	0.0000753	100.00	0.0086758	0.0520547	100.00

%GRRAVc c1 c2 c3-c227 c230 c231 c232 c233 c234 c235

This Minitab macro inputs Gauge R&R measurement data (simulated data) and outputs the results of the AIAG ANOVA Gauge R&R study results. The inputs are c1 (operator column), c2 (Part number column) and c3-c227 (the raw Gauge R&R measurement data; in this example, 225 sets of measurement data). The six columns of output are stored in columns c230 through c235 in this example.

The output results are stacked one on top of another. Table A4.2 below shows an example of the six columns of data output. Note that the column headings in the table and the left most column are not included in the macro data output. They are included here to describe the data output. Also note that in some cases, there are seven rows of data output if the operator-

by-part interaction is significant. The example in Table A4.2 shows the case when the operator- by-part interaction is significant. In the case where the operator-by-part interaction is not significant, this row is absent from the output.

Table D2

Example % GRRAVc Minitab Macro Output

Source	**Data Column Heading**	**Variance Component**	**% Contri-bution of Var. Comp**	**Std. Dev. (SD)**	**Study Variation (6*SD)**	**% Study Variation (%SV)**
Tot. GR&R	SD.75-231	0.0000369	35.867	0.0062898	0.0377385	59.889
Repeatability	SD.75-231	0.0000224	20.346	0.0047372	0.0284234	45.106
Reproducibility	SD.75-231	0.0000171	15.521	0.0041376	0.0248255	39.397
Operator	SD.75-231	0.0000000	0.000	0.0000000	0.0000000	0.000
Oper by Part	SD.75-231	0.0000171	15.521	0.0041376	0.0248255	39.397
Part-to-part	SD.75-231	0.0000707	64.133	0.0084107	0.0504640	80.083
Total Variation	SD.75-231	0.0001103	100.00	0.0105024	0.0630144	100.00

```
%KrusMC c6-c8;
unstacked.
```

This Minitab macro performs pair-wise comparisons for median data sets. Unlike the previous macros it was not written or adapted for this study. The macro was downloaded from Minitab (http://www.minitab.com/en-US/support/macros/default.aspx?id=27). The macro compares the pair-wise population medians from a set of data. The inputs (c6-c8 in the above example) are the columns of data to be compared. In this study there were three columns, one each for EMP III method, the A&R method and the ANOVA method.

APPENDIX E, HYPOTHESIS ONE THROUGH FOUR ANALYSES WITH OPERATOR BY PART INTERACTION DATA REMOVED

Hypothesis One

Table E1

Hypothesis one, ANOVA Analysis of Total Measurement Precision Error with Operator by Part Interaction Data Removed

Source	DF	SS	MS	F	P
Factor (EMP III, A&R and ANOVA method)	2	1358348	679174	774.97	0.000
Error	21915	19206153	876		
Total	21917	20564501			
S= 29.60 R-Sq= 6.61% R-Sq (adj)= 6.60%					

Table E2

Hypothesis one, Descriptive Statistics for the Three Analysis Methods with Operator by Part Interaction Data Removed

Level	N	Mean	St Dev.
EMP III	7,306	17.87	28.78
A&R	7,306	29.91	29.88
ANOVA	7,306	36.94	30.13

Table E3

Hypothesis one, Tukey's Pair-wise Comparisons EMP III Method of Total Measurement Precision Error with Operator by Part Interaction Data Removed

A&R % Study Variation subtracted from:			
	Lower	Center	Upper
ANOVA method	5.88	7.03	8.18
EMP III method	-13.18	-12.04	-10.89

Table E4

Hypothesis one, Tukey's Pair-wise Comparisons A&R Method of Total Measurement Precision Error with Operator by Part Interaction Data Removed

ANOVA % Study Variation subtracted from:			
	Lower	Center	Upper
EMP III method	-20.21	-19.07	-17.92

Table E5

Hypothesis one, Kruskal-Wallis Analysis of Total Measurement Precision Error with Operator by Part Interaction Data Removed

Source	**N**	**Median**	**Ave. Rank**	**Z**
EMP III Method	7,308	2.258	7082.6	-64.14
A&R Method	7,308	15.022	11907.3	15.68
ANOVA Method	7,308	24.431	13888.6	48.46
Overall	21,918		10959.5	

H = 4472.47 DF = 2 P = 0.000
H = 4472.48 DF = 2 P = 0.000 (adjusted for ties)

Table E6

Hypothesis one, Kruskal-Wallis Multiple Comparisons for Total Measurement Precision Error with Operator by Part Interaction Data Removed

Groups	Z vs. Critical Value	P-Value
EMP III vs. ANOVA	65.0122 ≥ 1.834	0
EMP III vs. A&R	46.0861 ≥ 1.834	0
A&R vs. ANOVA	18.9261 ≥ 1.834	0

Hypothesis Two

Table E7

Hypothesis two, ANOVA analysis for repeatability measurement precision error with Operator by Part Interaction Data Removed

Source	DF	SS	MS	F	P
Factor (EMP III, A&R and ANOVA method)	2	1667780	833890	1574.62	0.000
Error	21915	11605755	530		
Total	21917	13273535			
S= 23.01 R-Sq=12.56% R-Sq (adj)= 12.56%					

Table E8

Hypothesis two, Repeatability Descriptive Statistics with Operator by Part Interaction Data Removed

Level	N	Mean	St Dev.
A&R Repeatability	7,306	24.16	23.73
ANOVA Repeatability	7,306	32.68	25.56
EMP III Repeatability	7,306	11.45	19.29

Table E9

Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability with Operator by Part Interaction Data Removed

A&R Repeatability subtracted from:			
	Lower	Center	Upper
ANOVA Repeatability	7.64	8.53	9.42
EMP III Repeatability	-13.59	-12.70	-11.81

Table E10

Hypothesis two, Tukey's Pair-wise Comparisons for Repeatability with Operator by Part Interaction Data Removed

ANOVA Repeatability subtracted from:			
	Lower	Center	Upper
EMP III Repeatability	-22.12	-21.23	-20.34

Table E11

Hypothesis two, Kruskal-Wallis Test on Repeatability with Operator by Part Interaction Data Removed

Source	**N**	**Median**	**Ave. Rank**	**Z**
A&R Method	7,306	12.628	11938.0	16.19
ANOVA Method	7,306	23.274	14364.0	56.33
EMP III Method	7,306	1.593	6576.5	-72.52
Overall	21,918		10959.5	
H = 5795.74 DF = 2 P = 0.000				
H = 5795.74 DF = 2 P = 0.000 (adjusted for ties)				

Table E12

Hypothesis two, Kruskal-Wallis Multiple Comparisons for Repeatability with Operator by Part Interaction Data Removed

Groups	**Z vs. Critical Value**	**P-Value**
EMP III Repeatability vs. ANOVA Repeatability	74.3885 ≥ 1.834	0
EMP III Repeatability vs. A&R Repeatability	51.2145 ≥ 1.834	0
A&R Repeatability vs. ANOVA Repeatability	23.1741 ≥ 1.834	0

Hypothesis Three

Table E13

Hypothesis three, ANOVA Analysis for Reproducibility Measurement Precision Error with Operator by Part Interaction Data Removed

Source	DF	SS	MS	F	P
Factor (EMP III, A&R and ANOVA method)	2	206623	103311	273.79	0.000
Error	21915	8269511	377		
Total	21917	8476134			
S= 19.43 R-Sq= 2.44% R-Sq (adj)= 2.43%					

Table E14

Hypothesis three, Reproducibility Descriptive Statistics with Operator by Part Interaction Data Removed

Level	N	Mean	St Dev.
A&R Reproducibility	7,306	13.75	21.26
ANOVA Reproducibility	7,306	11.56	20.42
EMP III Reproducibility	7,306	6.42	16.22

Table E15

Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility with Operator by Part Interaction Data Removed

A&R Reproducibility subtracted from:			
	Lower	Center	Upper
ANOVA Reproducibility	-2.95	-2.19	-1.44
EMP III Reproducibility	-8.08	-7.33	-6.57

Table E16

Hypothesis three, Tukey's Pair-wise Comparisons for Reproducibility with Operator by Part Interaction Data Removed

ANOVA Reproducibility subtracted from:			
	Lower	Center	Upper
EMP III Reproducibility	-5.89	-5.13	-4.38

Table E17

Hypothesis three, Kruskal-Wallis Test on Reproducibility with Operator by Part Interaction Data Removed

Source	**N**	**Median**	**Ave. Rank**	**Z**
EMP III Method	7,306	0.2119	8794.8	-35.82
A&R Method	7.306	4.5925	13310.7	38.90
ANOVA Method	7,306	2,3107	10773.1	-3.08
Overall	21,918		10959.5	
H = 1870.32 DF = 2 P = 0.000				
H = 1887.11 DF = 2 P = 0.000 (adjusted for ties)				

Table E18

Hypothesis three, Kruskal-Wallis Multiple Comparisons for Reproducibility with Operator by Part Interaction Data Removed

Groups	**Z vs. Critical Value**	**P-Value**
EMP III Reproducibility vs. A&R Reproducibility	43.33025 ≥ 1.834	0
A&R Reproducibility vs. ANOVA Reproducibility	24.3482 ≥1.834	0
EMP III Reproducibility vs. ANOVA Reproducibility	18.9820 ≥ 1.834	0

Hypothesis Four

Table E19

Hypothesis four, ANOVA Analysis for Part-to-part Measurement Precision Error with Operator by Part Interaction Data Removed

Source	**DF**	**SS**	**MS**	**F**	**P**
Factor (EMP III, A&R and ANOVA method)	2	141367	70683	103.02	0.000
Error	21915	15035671	686		
Total	21917	15177037			
S= 26.19 R-Sq=0.93% R-Sq (adj)= 0.92%					

Table E20

Hypothesis four, Part-to-part Descriptive Statistics with Operator by Part Interaction Data Removed

Level	**N**	**Mean**	**St Dev.**
EMP III Part-to-part	7,306	82.13	28.78
A&R Part-to-part	7,306	88.04	21.47
ANOVA Part-to-part	7,306	83.42	27.73

Table E21

Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part with Operator by Part Interaction Data Removed

A&R Part-to-part subtracted from:			
	Lower	Center	Upper
ANOVA Part-to-part	-5.63	-4.62	-3.61
EMP III Part-to-part	-6.93	-5.92	-4.90

Table E22

Hypothesis four, Tukey's Pair-wise Comparisons for Part-to-part with Operator by Part Interaction Data Removed

ANOVA Part-to-part subtracted from:			
	Lower	Center	Upper
EMP III Part-to-part	-2.31	-1.30	-0.28

Table E23

Hypothesis four, Kruskal-Wallis Test on Part-to-part Portion of Measurement Precision Error with Operator by Part Interaction Data Removed

Source	**N**	**Median**	**Ave. Rank**	**Z**
EMP III Part-to- part	7,306	97.74	10531.9	-7.07
A&R Part-to-part	7,306	98.87	12448.0	24.63
ANOVA Part-to-part	7,306	96.97	9898.6	-17.55
Overall	21,918		10959.5	
H = 643.09 DF = 2 P = 0.000				
H = 643.09 DF = 2 P = 0.000 (adjusted for ties)				

Table E24

Hypothesis four, Kruskal-Wallis Multiple Comparisons for Part-to-part with Operator by Part Interaction Data Removed

Groups	**Z vs. Critical Value**	**P-Value**
A&R Part-to-part vs. ANOVA Part-to-part	$24.3526 \geq 1.834$	0
EMP III Part-to-part vs. A&R Part-to-part	$18.3028 \geq 1.834$	0
EMP III Part-to-part vs. ANOVA Part-to-part	$6.0498 \geq 1.834$	0

APPENDIX F, COMPARSION OF SEED GAUGE R&R STUDY DATA AND SIMULATION GAUGE R&R STUDY DATA

Table F1

Comparison of Seed and Simulation Gauge R&R Study Data P-Values

Test Condition	Equal Variance (Levene's test)	Mann-Whitney (adjusted for ties)	Moods Median	t-Test
Total GR&R				
EMP III Seed vs. EMP III Sim	0.900	0.0044	0.137	0.720
A&R Seed vs. A&R Sim	0.635	0.0043	0.137	0.299
ANOVA Seed vs. ANOVA Sim	0.114	0.0000	0.003	0.015
Repeatability				
EMP III Seed vs. EMP III Sim	0.809	0.0012	0.105	0.613
A&R Seed vs. A&R Sim	0.698	0.0012	0.105	0.188
ANOVA Seed vs. ANOVA Sim	0.126	0.0000	0.000	0.001
Reproducibility				
EMP III Seed vs. EMP III Sim	1.000	0.0254	0.079	0.966
A&R Seed vs. A&R Sim	0.837	0.0246	0.079	0.647
ANOVA Seed vs. ANOVA Sim	0.912	0.4028	0.893	0.817
Part-to-part				
EMP III Seed vs. EMP III Sim	0.900	0.0044	0.137	0.720
A&R Seed vs. A&R Sim	0.905	0.0043	0.137	0.780
ANOVA Seed vs. ANOVA Sim	0.595	0.0000	0.007	0.984

Table F1 *(continued)*

Note: Both Seed and Simulation data for all conditions failed tests for normality. Levene's test is a test for equal variance for any continuous distribution. Mann-Whitney and Moods Median are nonparametric tests.

APPENDIX G, VARIANCE COMPARISON OF EMP III, A&R AND ANOVA METHODS FOR SIMULATION DATA

Table G1

Hypothesis one, Tests for Equal Variances

95% Bonferonni Confidence Intervals for Total GR&R Standard Deviations

Method	N	Lower	Standard Deviation	Upper
EMP III	10080	28.5762	29.0584	29.5563
A&R	10080	29.7746	30.2771	30.7959
ANOVA	10080	30.5934	31.1096	31.6427

Bartlett’s Test (Normal Distribution) Test Statistic= 47.31, P-Value= 0.000

Levene’s Test (Any Continuous Distribution) Test Statistic= 170.84, P-Value= 0.000

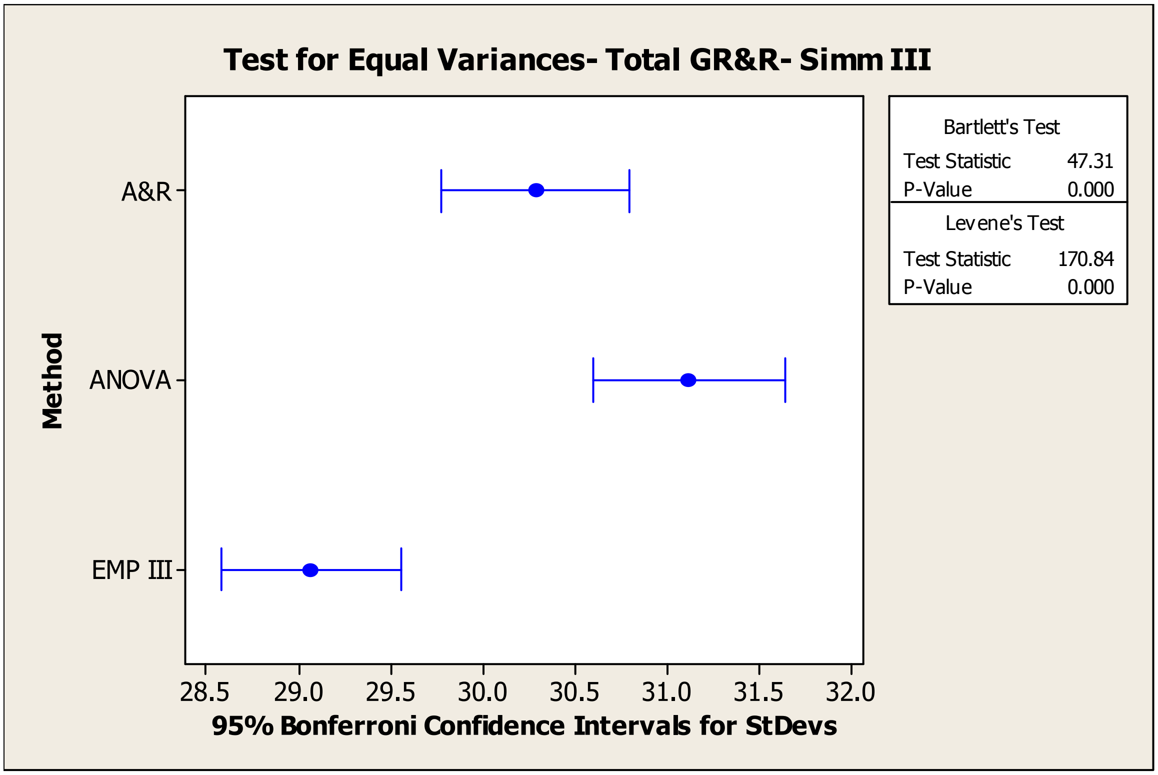

Figure G1. Hypothesis one, Test for Equal Variance Total GR&R.

Table G2

Hypothesis two, Tests for Equal Variances

95% Bonferonni Confidence Intervals for Repeatability Standard Deviations				
Method	N	Lower	Standard Deviation	Upper
EMP III	10080	18.8514	19.1696	19.4980
A&R	10080	23.5059	23.9025	24.3121
ANOVA	10080	25.0442	25.4668	25.9032
Bartlett’s Test (Normal Distribution) Test Statistic= 849.43, P-Value= 0.000				
Levene’s Test (Any Continuous Distribution) Test Statistic= 566.93, P-Value= 0.000				

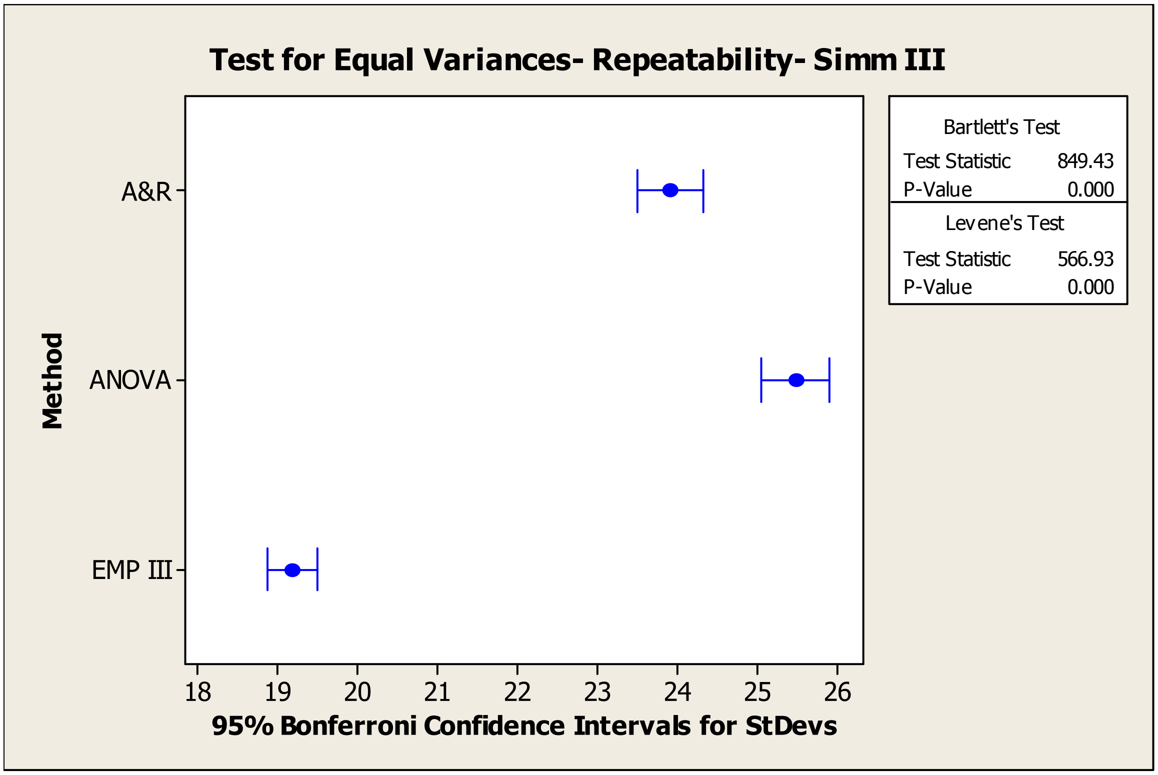

Figure G2. Hypothesis two, Test for Equal Variance Repeatability.

Table G3

Hypothesis three, Tests for Equal Variances

95% Bonferonni Confidence Intervals for Reproducibility Standard Deviations

Method	N	Lower	Standard Deviation	Upper
EMP III	10080	16.6366	16.9174	17.2073
A&R	10080	21.8288	22.1971	22.5775
ANOVA	10080	22.2558	22.6313	23.0191

Bartlett’s Test (Normal Distribution) Test Statistic= 993.78, P-Value= 0.000

Levene’s Test (Any Continuous Distribution) Test Statistic= 457.89, P-Value= 0.000

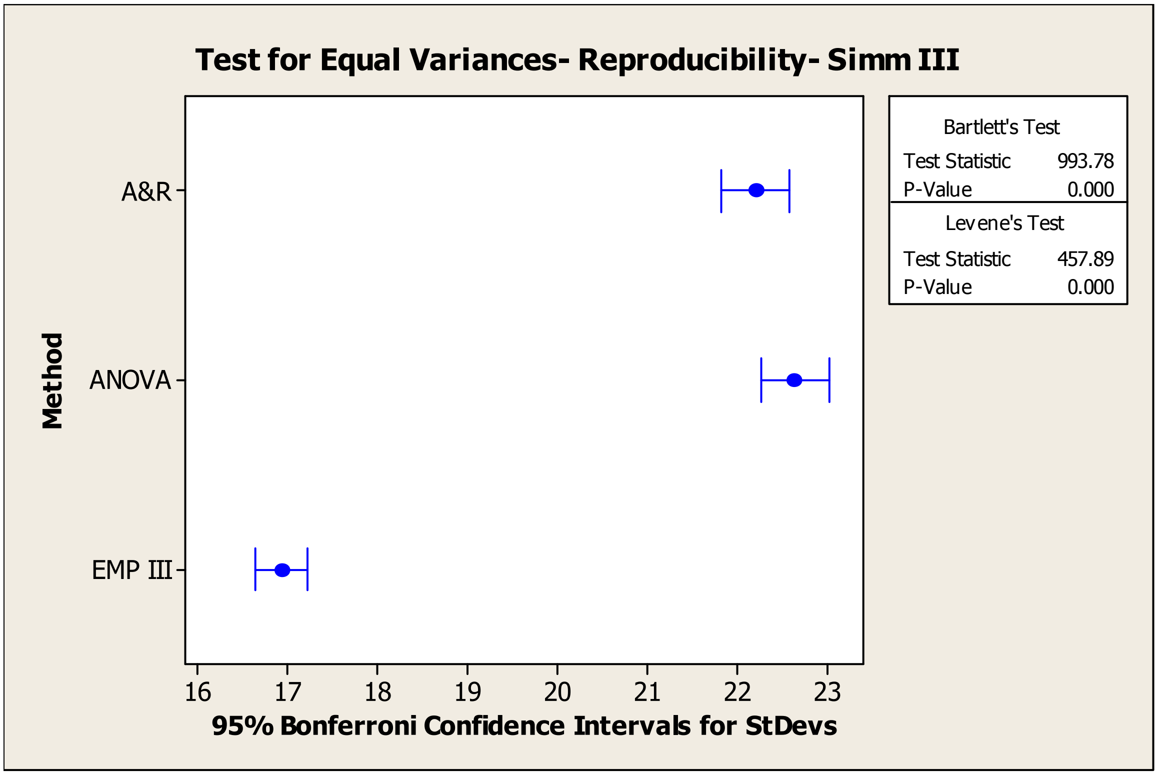

Figure G3. Hypothesis three, Test for Equal Variance Reproducibility.

Table G4

Hypothesis four, Tests for Equal Variances

95% Bonferonni Confidence Intervals for Part-to-part Standard Deviations				
Method	N	Lower	Standard Deviation	Upper
EMP III	10080	28.5762	29.0584	29.5563
A&R	10080	21.0308	21.3857	21.7521
ANOVA	10080	27.9964	28.4689	28.9567

Bartlett’s Test (Normal Distribution) Test Statistic1098.15, P-Value= 0.000

Levene’s Test (Any Continuous Distribution) Test Statistic= 163.37, P-Value= 0.000

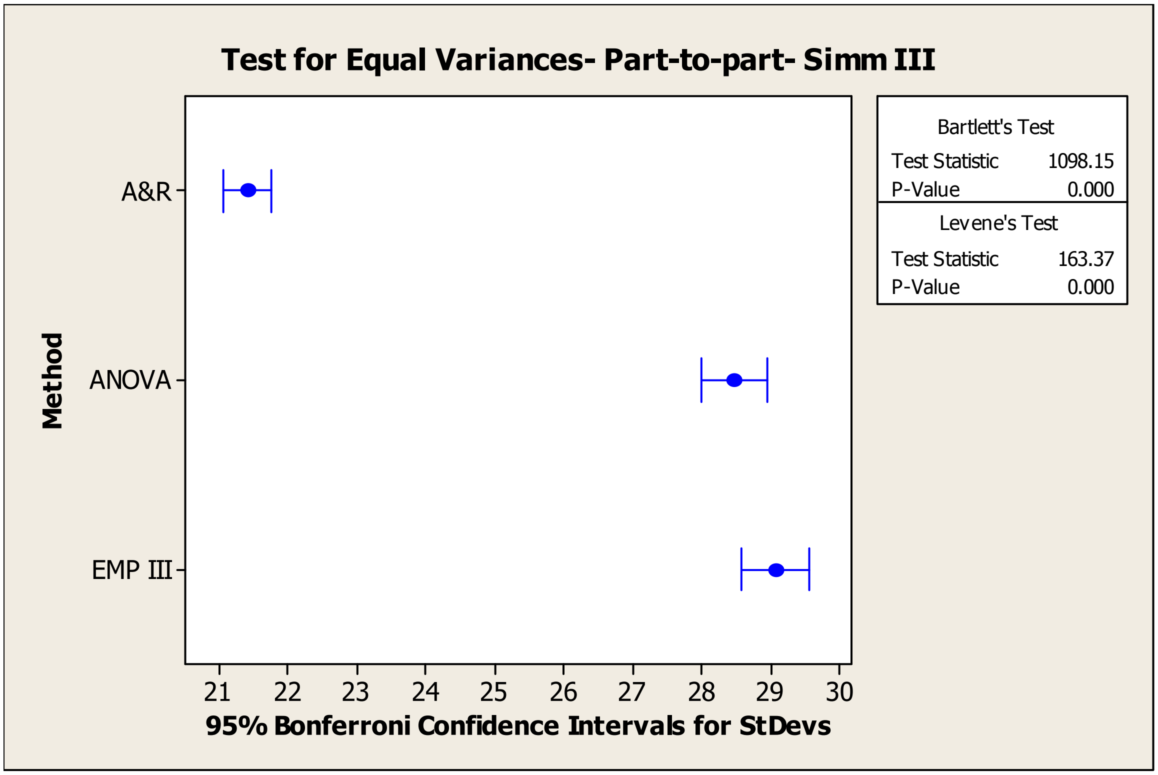

Figure G4. Hypothesis four, Test for Equal Variance Part-to-part.