Evaluating Measurement System by Gauge Repeatability and Reproducibility

Bahar Sennaroglu
Department of Industrial Engineering
Marmara University
Istanbul, Turkey
bsennar@marmara.edu.tr

Ozlem Yurtsever
Vocational School of Technical Sciences
Marmara University
Istanbul, Turkey
ozlem.yurtsever@marmara.edu.tr

Abstract

In this paper, the measurement system of a pharmaceutical company is evaluated by Gauge Repeatability and Reproducibility study. The objectives are to identify inadequacies of the measurement system and provide guidelines for performance improvement. Measurement variability is evaluated by two components which are the variability due to the operators (reproducibility) and the variability due to the gauge itself (repeatability). In this study, designed experiments are used to isolate and estimate the components of variability in the measurement system. The data are analyzed by Analysis of Variance and gauge capability is evaluated by precision-to-tolerance ratio and the percent contribution of variance components to the total variability. The results are interpreted and suggestions are made to improve the performance of the measurement system.

Keywords
Gauge Capability, Repeatability, Reproducibility, Designed Experiments

1. Introduction

Tests and measurements are applied to samples taken from production lines in order to control whether the products meet the required performance characteristics and quality levels. It is also necessary to evaluate the performance of the measurement system to be able to rely on the test and measurement results. The measurement errors can cause both poor quality products to be delivered to the customers and good quality products to be rejected and not delivered. Therefore, it is important to take into account the measurement errors in the system to prevent the misclassification of products and misleading decisions.

In this paper, the measurement system of a pharmaceutical company is evaluated by Gauge Repeatability and Reproducibility (GR&R) study in order to identify inadequacies of the measurement system and provide guidelines for performance improvement. GR&R studies focus on quantifying the measurement errors (Pearn and Kotz, 2006). Reproducibility is defined as the variability due to different operators using the gauge (or in general, different conditions) and repeatability as reflecting the basic inherent precision of the gauge itself (Montgomery, 2005). The gauge is a measuring instrument. The capability of the gauge is assessed by its ability to repeat and reproduce measurements. A measurement system is repeatable if its variability is consistent and it is reproducible when different operators produce consistent results (Pyzdek and Keller, 2010).

Most studies using operators and tests employ two primary methods to calculate GR&R results which are the average and range method (a.k.a. the Automotive Industry Action Group method) and the Analysis of Variance

(ANOVA) method (Kappele and Raffaldi, 2005). The appropriate quality measures to use for the evaluation of gauge capability include precision-to-tolerance (P/T) ratio, signal-to-noise (S/N) ratio, and discrimination ratio (DR) (Al-Refaie and Bata, 2010). In the literature, different approaches are adapted in GR&R studies. Papananias et al. (2017) evaluate uncertainty of length and diameter measurements associated with versatile automated gauging using full factorial designs, ANOVA, and effect graphs. Aquila et al. (2018) evaluate the behavior of wind average speed in different wind energy-producing states using nested design and ANOVA for GR&R. Weaver et al. (2012) propose the Bayesian approach, which requires specifying a statistical model for the data and a prior distribution for the model parameters, to data analysis and show how to estimate variance components associated with the sources of variability and relevant functions of these using GR&R data together with prior information. Wang and Chien (2010) apply process-oriented basis representation method for a multivariate GR&R study to identify specific causes of production problems and map them into a basis matrix, and then analyze these patterns individually using a random factor experiment. Erdmann et al. (2010) apply a GR&R study in health care using ANOVA, P/T ratio and graphical analysis. In this study, designed experiments are used to isolate and estimate the components of variability in the measurement system. The data are analyzed by ANOVA and gauge capability is evaluated by P/T ratio and percent contribution of variance components to the total variability.

2. Gauge Repeatability and Reproducibility (GR&R) Study

The main purpose of a GR&R study is to determine how much of the total observed variability is due to the gauge (or instrument) so that the capability of the gauge can be assessed. The total observed measurement (x) can be defined by Eq.1.

$$x = x_{\text{Product}} + \varepsilon \tag{1}$$

where x_{Product} is the true value of the measurement and ε is the measurement error.

The variance of the total observed measurement (σ^2_{Total}) is defined by Eq.2, assuming that x and ε are normally and independently distributed random variables with means μ and 0 and variances $\sigma^2_{\text{Product}}$ and σ^2_{Gauge}, respectively (Montgomery, 2005).

$$\sigma^2_{\text{Total}} = \sigma^2_{\text{Product}} + \sigma^2_{\text{Gauge}} \tag{2}$$

Total variability includes both product variability and gauge variability. The product variability is the actual variation between parts produced by the process; it is also called the part-to-part variability. The variance of the measurement error (or the variance of the gauge) is defined by two components, repeatability (variability due to the gauge itself) and the reproducibility (variability due to the operators) of the gauge as in Eq.3.

$$\sigma^2_{\text{Measurement Error}} = \sigma^2_{\text{Gauge}} = \sigma^2_{\text{Repeatability}} + \sigma^2_{\text{Reproducibility}} \tag{3}$$

The experiment used to measure these two components is usually called a GR&R study (Montgomery, 2005).

After calculating the variance of the gauge, the capability of the measurement system can be assessed by the precision-to-tolerance ratio.

$$P/T = \frac{6\sigma_{\text{Gauge}}}{USL - LSL} = \frac{6\sigma_{\text{Gauge}}}{T} \tag{4}$$

where T is the tolerance, USL and LSL are the upper and lower specification limits, respectively. A gauge is judged capable if the P/T ratio is less than or equal to 0.1, and incapable if the P/T ratio is greater than 0.3 in which case the measurement system needs improvement.

In order to identify inadequacies in the measurement system, percent contribution to the total variability made by each variance component can be calculated by dividing each variance by the total variance and by multiplying 100.

3. Application and Results

The amount of active ingredient (mg) of a specific drug produced by the pharmaceutical company is an important quality characteristic and measured by operators. Since every measurement destroys the part due to destructive testing, the measurements cannot be repeated on the same part.

Kappele and Raffaldi (2010) suggest two alternatives for assessing repeatability with a destructive measurement system:

- Use a replacement nondestructive test that correlates with the results of the destructive test.
- Collect parts that are so similar in the property to be measured that it can be assumed they are the same part.

In this study, a sample of size six was taken from the same batch of production to provide uniformity within the sample so that it can be assumed that they are the same part; thus the variation in these identical parts from the same batch was assumed negligible in the data analysis.

Two operators were randomly selected for the GR&R study. The reference standard of the drug was prepared separately by each operator prior to the measurements using the gauge. The samples were collected from three different batches. The measurement data are given in Table 1. Since the reference standard of the drug was prepared separately by each operator, a two-stage nested design was used (Figure 1).

Table 1. The measurement data

Part (Batch) Number	**Operator 1**			**Operator 2**		
	Test 1	**Test 2**	**Test 3**	**Test 1**	**Test 2**	**Test 3**
1	489.2083	495.2939	488.2212	509.8210	503.8735	514.5253
2	488.3434	489.0278	488.4349	498.3963	492.5487	503.9633
3	494.8635	493.7041	491.0811	506.2688	507.3185	512.6657

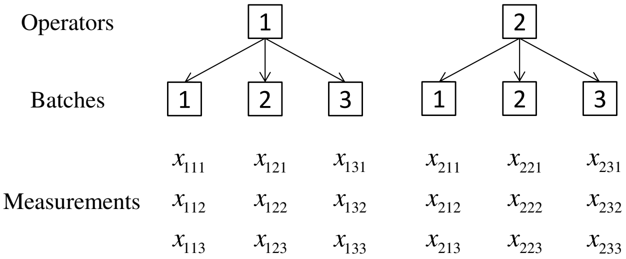

Figure 1. Two-stage nested design

The random effects model for the two-stage nested design is given in Eq.5.

$$x_{ijk} = \mu + O_j + P_{i(j)} + \varepsilon_{(ij)k} \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,o \\ k = 1,2,\ldots,n \end{cases} \tag{5}$$

where μ represents the overall mean, O_j random effects of different operators, $P_{i(j)}$ random effects of different parts nested under operator, and $\varepsilon_{(ij)k}$ random error. It is assumed that model parameters O_j, $P_{i(j)}$, and $\varepsilon_{(ij)k}$ are independent and normally distributed random variables with (mean, variance) $(0,\sigma_O^2)$, $(0,\sigma_{P(O)}^2)$ and $(0,\sigma_E^2)$, respectively. Therefore, the variance of the observed measurement is

$$V(x_{ijk}) = \sigma_O^2 + \sigma_{P(O)}^2 + \sigma_E^2 \tag{6}$$

and ANOVA is used to estimate the variance components. The variance component for the error term is gauge repeatability. The ANOVA procedure involves partitioning the total variability in the measurements into its components with the following sum of squares identity.

$$SS_{Total} = SS_O + SS_{P(O)} + SS_E \tag{7}$$

The summary of the procedure is given in Table 2.

Table 2. ANOVA table for the two-stage nested design and random effects model

Source	***SS* (Sum of Squares)**	***df* (Degrees of Freedom)**	***MS* (Mean Square)**	F_0
Operator	$SS_O = pn\sum_j (\bar{x}_{.j.} - \bar{x}_{...})^2$	$df_O = o - 1$	$MS_O = \frac{SS_O}{df_O}$	$\frac{MS_O}{MS_{P(O)}}$
Part (Operator)	$SS_{P(O)} = n\sum_i \sum_j (\bar{x}_{ij.} - \bar{x}_{.j.})^2$	$df_{P(O)} = o(p-1)$	$MS_{P(O)} = \frac{SS_{P(O)}}{df_{P(O)}}$	$\frac{MS_{P(O)}}{MS_E}$
Error	$SS_E = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{ij.})^2$	$df_E = po(n-1)$	$MS_E = \frac{SS_E}{df_E}$	
Total	$SS_{Total} = \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{...})^2$	$df_{Total} = pon - 1$		

The calculations are carried out using the experimental data in Table 1 and the ANOVA results are given in Table 3. Based on the ANOVA results, it is concluded that both the effect of operator and the effect of part nested under operator are statistically significant at 0.05 significance level (p-value<0.05).

Table 3. ANOVA for the nested design experiment

Source	***SS***	***df***	***MS***	F_0	***p-value***
Operator	956.344	1	956.344	14.445	0.019
Part (Operator)	264.827	4	66.207	4.346	0.021
Error	182.816	12	15.235		
Total	1403.988	17			

Using the ANOVA results, the estimates of components of gauge variability $\hat{\sigma}^2_{\text{Repeatability}}$ and $\hat{\sigma}^2_{\text{Reproducibility}}$, product variability $\hat{\sigma}^2_{\text{Product}}$, and total observed variability $\hat{\sigma}^2_{\text{Total}}$ are calculated by the formulas given in Table 4 (Minitab, 2010). If any variance component has a negative value, its value is set to zero. Percent contribution to the total variability made by each variance component is calculated by dividing each variance by the total variance and by multiplying 100.

Tablo 4. Variance estimates and percent contribution to total variability

Source	**Variance Estimate ($\hat{\sigma}^2$)**	$\hat{\sigma}^2$	**% Contribution**
Repeatability	$\hat{\sigma}^2_{\text{Repeatability}} = MS_E$	15.235	11.62
Reproducibility	$\hat{\sigma}^2_{\text{Reproducibility}} = \dfrac{MS_O - MS_{P(O)}}{pn}$	98.904	75.42
Gauge	$\hat{\sigma}^2_{\text{Gauge}} = \hat{\sigma}^2_{\text{Repeatability}} + \hat{\sigma}^2_{\text{Reproducibility}}$	114.139	87.04
Product	$\hat{\sigma}^2_{\text{Product}} = \dfrac{MS_{P(O)} - MS_E}{n}$	16.991	12.96
Total	$\hat{\sigma}^2_{\text{Total}} = \hat{\sigma}^2_{\text{Product}} + \hat{\sigma}^2_{\text{Gauge}}$	131.130	100.00

Precision-to-tolerance ratio of 1.28, which is higher than 0.3, indicates that the gauge is not capable.

$$T = USL - LSL = 525 - 475 = 50 \text{ mg}$$

$$P/T = \frac{6\hat{\sigma}_{\text{Gauge}}}{T} = \frac{6\sqrt{114.139}}{50} = 1.28$$

The results for percent contribution indicate that gauge variability accounts for 87.04% of the total variability and most of it is due to reproducibility component (75.42%). Therefore, the measurement system is inadequate and it is necessary to reduce the variability due to the operators.

A training program was prepared in order to eliminate the differences between operators. A new measurement procedure was prepared and it was decided that the reference standard of the drug be prepared by a single operator. After the operator training, a new experiment was conducted (Table 5). Since the reference standard was prepared by a single operator, the following random effects model for the factorial design, also called crossed design, was used for the analysis.

$$x_{ijk} = \mu + P_i + O_j + (PO)_{ij} + \varepsilon_{ijk} \begin{cases} i = 1,2,\ldots,p \\ j = 1,2,\ldots,o \\ k = 1,2,\ldots,n \end{cases} \tag{8}$$

where μ represents the overall mean, P_i random effects of different parts, O_j random effects of different operators, $(PO)_{ij}$ random effect of part-operator interaction and ε_{ijk} random error. It is assumed that model parameters P_i, O_j, $(PO)_{ij}$ and ε_{ijk} are independent and normally distributed random variables with (mean, variance) $(0, \sigma_P^2)$, $(0, \sigma_O^2)$, $(0, \sigma_{PO}^2)$ and $(0, \sigma_E^2)$, respectively.

Table 6 presents the formulas for the ANOVA and Table 7 its results. It is concluded from the ANOVA results that the model sources of variability (part, operator, and part-operator interaction) are nonsignificant (p-value>0.05). The formulas and calculations for variance estimates and percent contribution of variance components to total variability are given in Table 8 (Minitab, 2010).

Table 5. Measurement data after operator training

Part (Batch) Number	**Operator 1**			**Operator 2**		
	Test 1	**Test 2**	**Test 3**	**Test 1**	**Test 2**	**Test 3**
1	495.968	502.964	496.623	508.167	491.563	503.215
2	498.721	496.118	495.396	491.108	493.701	498.041
3	491.055	499.385	484.665	493.963	499.997	492.449

Table 6. ANOVA table for the crossed design and random effects model

Source	***SS***	***df***	***MS***	$\boldsymbol{F_0}$
Part	$SS_P = on\sum_{i}(\bar{x}_{i..} - \bar{x}_{...})^2$	$df_P = p - 1$	$MS_P = \dfrac{SS_P}{df_P}$	$\dfrac{MS_P}{MS_{PO}}$
Operator	$SS_O = pn\sum_{j}(\bar{x}_{.j.} - \bar{x}_{...})^2$	$df_O = o - 1$	$MS_O = \dfrac{SS_O}{df_O}$	$\dfrac{MS_O}{MS_{PO}}$
Part*Operator	$SS_{PO} = n\sum_{i}\sum_{j}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$	$df_{PO} = (p-1)(o-1)$	$MS_{PO} = \dfrac{SS_{PO}}{df_{PO}}$	$\dfrac{MS_{PO}}{MS_E}$
Error	$SS_E = \sum_{i}\sum_{j}\sum_{k}(x_{ijk} - \bar{x}_{ij.})^2$	$df_E = po(n-1)$	$MS_E = \dfrac{SS_E}{df_E}$	
Total	$SS_{Total} = \sum_{i}\sum_{j}\sum_{k}(x_{ijk} - \bar{x}_{...})^2$	$df_{Total} = pon - 1$		

Table 7. ANOVA for the crossed design experiment

Source of variation	***SS***	***df***	***MS***	$\boldsymbol{F_0}$	***p*-value**
Part	119.321	2	59.660	3.685	0.213
Operator	7.105	1	7.105	0.439	0.576
Part*Operator	32.383	2	16.192	0.560	0.585
Error	346.706	12	28.892		
Total	505.515	17			

Tablo 8. Variance estimates and percent contribution after operator training

Source	**Variance Estimate ($\hat{\boldsymbol{\sigma}}^2$)**	$\hat{\boldsymbol{\sigma}}^2$	**% Contribution**
Repeatability	$\hat{\sigma}^2_{\text{Repeatability}} = MS_E$	28.892	79.95
Reproducibility	$\hat{\sigma}^2_{\text{Reproducibility}} = \hat{\sigma}^2_O + \hat{\sigma}^2_{PO}$ $\hat{\sigma}^2_O = \dfrac{MS_O - MS_{PO}}{pn}$ $\hat{\sigma}^2_{PO} = \dfrac{MS_{PO} - MS_E}{n}$	0.000	0.00
Gauge	$\hat{\sigma}^2_{\text{Gauge}} = \hat{\sigma}^2_{\text{Repeatability}} + \hat{\sigma}^2_{\text{Reproducibility}}$	28.892	79.95
Product	$\hat{\sigma}^2_{\text{Product}} = \dfrac{MS_P - MS_{PO}}{on}$	7.245	20.05
Total	$\hat{\sigma}^2_{\text{Total}} = \hat{\sigma}^2_{\text{Product}} + \hat{\sigma}^2_{\text{Gauge}}$	36.137	100.00

Based on the variance estimate of the gauge, the capability of the measurement system is assessed by the precision-to-tolerance ratio.

$$P/T = \frac{6\hat{\sigma}_{\text{Gauge}}}{T} = \frac{6\sqrt{28.892}}{50} = 0.645$$

P/T ratio is higher than 0.3, therefore, the gauge is incapable.

The variability due to the operators (reproducibility) was eliminated but the measurement system needs improvement for the gauge itself due to 80% contribution of repeatability variance component to the total variability. Repeatability is the basic inherent precision of the gauge itself; therefore, the precision of the measuring instrument should be increased. Calibration, proper use and maintenance of the instrument or replacement are the solutions.

It should also be noted that repeatability component will be overestimated because of destructive measurement system. Although the test samples are taken from the same batch of production and the variation in these identical parts from the same batch is assumed negligible, there will be some variation that will cause repeatability be overestimated.

4. Conclusion

This paper presented the principles of a Gauge Repeatability and Reproducibility study and its application to the measurement system of a pharmaceutical company. A nested design and a crossed design were used for experiments in order to isolate and estimate the components of variability in the measurement system. The experimental data were analyzed by Analysis of Variance and gauge capability was evaluated by precision-to-tolerance ratio and percent contribution of variance components to the total variability.

The GR&R study provided guidelines for improving performance of the measurement system. The variability due to the operators (reproducibility) was eliminated through a new measurement procedure and operator training program. Calibration, proper use and maintenance of the instrument or replacement were suggested in order to reduce the variability due to the gauge itself (repeatability).

Acknowledgements

Authors acknowledge the support of the Marmara University Scientific Research Projects Committee (BAPKO) through the project FEN-D-120418-0175.

References

Al-Refaie, A., and Bata, N., Evaluating measurement and process capabilities by GR&R with four quality measures, *Measurement*, vol. 43, pp. 842-851, 2010.

Aquila, G., Peruchi, R. S., Junior, P. R., Rocha, L. C. S., de Queiroz, A. R., de Oliveira Pamplona, E., and Balestrassi, P. P., Analysis of the wind average speed in different Brazilian states using the nested GR&R measurement system, *Measurement*, vol. 115, pp. 217-222, 2018.

Erdmann, T. P., Does, R. J. M. M., and Bisgaard, S., Quality Quandaries: A Gage R&R Study in a Hospital, *Quality Engineering*, vol. 22, pp. 46-53, 2010.

Kappele, W. D., and Raffaldi, J., Select the Right Gage R&R Calculation Method, *Quality*, vol. 44, no. 4, pp. 38-41, 2005.

Kappele, W. D., and Raffaldi, J., GR&R for Destructive Measurement Systems, *Quality*, vol. 49, no. 5, pp. 32-34, 2010.

Minitab, *Minitab 16 Help Menu: Methods and Formulas-Measurement Systems Analysis*, Minitab Inc., Pennsylvania, 2010.

Montgomery, D. C., *Introduction to Statistical Quality Control*, 5th Edition, Wiley, New York, 2005.

Papananias, M., Fletcher, S., Longstaff, A. P., and Forbes, A. B., Uncertainty evaluation associated with versatile automated gauging influenced by process variations through design of experiments approach, *Precision Engineering*, vol. 49, pp. 440-455, 2017.

Pearn, W. L., and Kotz, S., *Encyclopedia and Handbook of Process Capability Indices*, World Scientific, Singapore, 2006.

Pyzdek, T., and Keller, P., *The Six Sigma Handbook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels*, 3rd Edition, McGraw Hill, New York, 2010.

Wang, F.-K., and Chien T.-W., Process-oriented basis representation for a multivariate gauge study, *Computers & Industrial Engineering*, vol. 58, no. 1, pp. 143-150, 2010.

Weaver, B. P., Hamada, M. S., Vardeman, S. B., and Wilson, A. G., A Bayesian approach to the analysis of gauge R&R data, *Quality Engineering*, vol. 24, no. 4, pp. 486-500, 2012.

Biographies

Bahar Sennaroglu is an Associate Professor in the Industrial Engineering Department of Faculty of Engineering of Marmara University. Dr. Sennaroglu received her MSc and PhD degrees in Industrial Engineering from Marmara University. Her research interests include design and analysis of experiments, decision techniques, forecasting techniques, project management, and production and quality systems.

Ozlem Yurtsever is a Ph.D candidate in Industrial Engineering at Marmara University, Istanbul. She completed her B.Sc. in Chemical Engineering in 2006 and M.Sc. in Engineering Management in 2008 both at Marmara University. She is currently a lecturer in Vocational School of Technical Sciences at the same university. Her research interests include sustainability, calculation of carbon emissions, green supply chain management and energy efficiency.