Quality Engineering, 22:21–29, 2010

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982110903344820

Assessment of a Measurement System Using Repeat Measurements of Failing Units

Ryan P. Browne

Department of Statistics and Actuarial Sciences, Business and Industrial Statistics Research Group, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT Inspection systems measure every part produced. If the measurement is outside of the inspection limits, then that part is measured again. To assess the quality of these measurement systems, traditionally gauge repeatability and reproducibility (R&R) studies are preformed. Instead of performing a gauge R&R study, we present a method of assessing these measurement systems with operational data from an inspection system. Using the inspection data, we provide a justification for the pooled variance of the measured values for each part that has two measurements. The bias and variance of this analysis of variance (ANOVA) estimator are derived using properties of the truncated normal distribution. We show that the ANOVA estimator has a relatively small bias and high efficiency when compared with the maximum likelihood estimator for most common values of γ or GRR%, which is the measurement system standard deviation divided by total inspection system standard deviation.

KEYWORDS gauge repeatability, inspection system, measurement system assessment, reproducibility study

INTRODUCTION

Many manufacturers require parts to pass an inspection system before being shipped. The purpose of the inspection is to prevent customers from receiving poor quality parts. The ideal system rejects each part with a true value outside of inspection limits, but due to measurement error, the actual system rejects parts with an observed or measured characteristic outside of these limits. Thus, an inspection system will reject some good parts and accept some bad parts. Accepting and rejecting the wrong parts can be costly, making it essential to verify or quantify the performance of the inspection system. Measurement variability explains why inspection limits are often tighter than the specification limits.

In general, an inspection system has two parts, a measurement system and an inspection protocol. The measurement system is the method or device used to measure the characteristic of interest. The inspection protocol is the set of decision rules for the inspection system. Figure 1 gives an example of a commonly used protocol, but there are many possibilities.

The performance of any inspection system is highly reliant on the measurement system used to measure the characteristic of interest. To assess

Address correspondence to Ryan P. Browne, Department of Statistics and Actuarial Sciences, Business and Industrial Statistics Research Group, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1. E-mail: rpbrowne@uwaterloo.ca

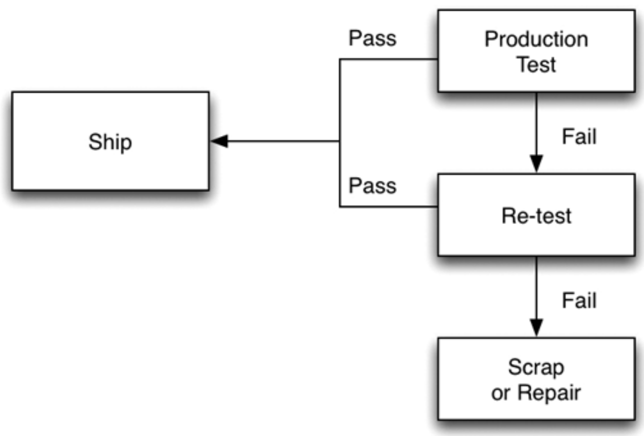

FIGURE 1 Typical inspection protocol.

a continuous measurement system, a gauge repeatability and reproducibility (R&R) is typically performed (see Automotive Industry Action Group 2002 and Burdick et al. 2003). Other pertinent references are Wheeler and Lyday (1984) and Montgomery (2000). Because these studies are conducted off-line, they can be costly and may not reflect the properties of the measurement system during production.

In some industries, such as integrated circuit or electronic device manufacturing, and in testing flammability of plastics and other chemically based products, it is common to use the protocol given in Figure 1. The inspection protocol starts with each part being measured. This first measurement is called the *production test*. The test result can be either pass or fail based on whether the measured value lies within the inspection limits or not. A pass allows the part to be shipped, whereas a failure means the part is retested. Commonly, this retest is carried out immediately, and if the part passes the second test it is shipped. Otherwise, it is sent to be repaired, scrapped, or set aside for investigation.

Using the inspection protocol shown in Figure 1, some parts are measured twice. As a result it is possible to carry out an assessment of the measurement system using data from the inspection system alone and avoid off-line studies such as a standard gauge R&R. The data from the inspection system have a special form because a part is measured a second time if and only if the first measurement falls outside the pass region, denoted B = (LIL, UIL), where (LIL) and (UIL) are the lower and upper inspection limits. The possible outcomes for any part are (PASS), (FAIL, PASS) and (FAIL, FAIL).

Suppose, for example, we have the results of an inspection system with limits (95, 110) for 100 parts. Of these, 17 have second measurements. The data are shown in two tables. Table 1 gives the production measurements and Table 2 gives the repeated measurements. The first measurements or production data average and standard deviation are 100.1 and 4.86, respectively.

Generally, any existing process operates well, so we assume that the majority of the observed measurements are within the inspection limits. Thus, there are typically a large number of first measurements and a relatively small number of second measurements.

We use the following notation. The production data, the first measurement from each of n_1 parts, is denoted by $\{y_{11}, y_{21}, \ldots, y_{n_1 1}\}$. For the retest data, we use S, a subset of $\{1, 2, \ldots, n_1\}$, to indicate all the parts that have the failed production test so that $y_{i1} \notin B$. Suppose there are n_2 such parts. The retest data, the second measurements, are denoted by $\{y_{i2},\ i \in S\}$.

To model an inspection system, we follow Burdick et al. (2003) and Doganaksoy (2000) by assuming that a normal random effects model [1] describes the observed characteristics. The model is

$$Y_{ij} = P_i + E_{ij} \qquad [1]$$

TABLE 1 100 Production Test Observations or 1st Measurements

103.6	100.2	107.6	97.4	92.4	96.1	97.3	102.1	95.2	101.6
96.8	105.8	100.9	101.6	105.5	107.6	112.9	104.2	104.3	91.9
105.5	96.0	92.9	101.1	92.6	94.9	97.7	98.8	105.0	104.2
105.3	104.4	99.5	103.1	101.5	93.8	101.6	99.4	101.2	98.9
100.6	105.9	103.9	98.3	99.5	98.0	98.1	97.3	100.9	93.9
96.5	97.8	98.8	100.3	99.1	93.6	107.1	85.7	107.2	101.5
100.1	97.9	107.8	99.8	104.0	99.3	96.8	95.8	103.1	100.4
112.2	97.8	95.3	97.5	101.5	99.1	107.9	111.5	89.5	91.9
93.8	101.6	99.2	98.1	99.8	103.9	101.2	103.1	102.4	93.3
95.6	96.9	97.3	94.5	104.1	98.6	104.4	98.3	105.8	100.6

TABLE 2 **From the 100 Production Test Observations, 17 2nd Measurements Where Taken**

Part #	1st	2nd	Part #	1st	2nd
5	92.4	91.3	71	112.2	111.8
17	112.9	111.1	78	111.5	110.8
20	91.9	92.2	79	89.5	88.8
23	92.9	93.3	80	91.9	91.1
25	92.6	94.1	81	93.8	95.4
26	94.9	94.2	90	93.3	90.8
36	93.8	92.4	94	94.5	93.6
50	93.9	92.9			
56	93.6	92.2			
58	85.7	84.6			

where P_i is a random variable representing the possible values for the true dimension of part i $(i = 1, \ldots, n)$ and E_{ij} is a random variable representing the error on each measurement $(j = 1,\ 2)$ for part i. We assume that the part effects P_i are independent and identically distributed normal random variables with mean μ and variance σ_p^2, the measurements errors E_{ij} are independent and identically distributed normal random variables with mean zero and variance σ_m^2, and P and E are mutually independent. The variance of Y_{i1}, called the *total variation*, is $\sigma_t^2 = \sigma_p^2 + \sigma_m^2$. By adopting model [1], we assume that μ, σ_p, and σ_m are constant over the time needed to conduct the investigation and that σ_m is constant across true part dimensions.

We also assume that the measurement system has no material operator effects. This contrasts with what is typically assumed in the literature (see Burdick et al. 2003), but examples with no or little operator error are common. For instance, in one example, piston diameters were inspected by an inline gauge with automated part handling. Using manufacturing jargon, with no operator effects σ_m captures measurement repeatability but not reproducibility.

Burdick et al. (2003) described a variety of metrics used to quantify measurement system quality or reliability. The metric we use for this article is the gauge repeatability labeled as %GRR or γ. In terms of the model parameters, γ is defined as

$$\gamma = \frac{\sigma_m}{\sigma_t}. \qquad [2]$$

The Automotive Industry Action Group (2002) classifies a measurement system as (1) acceptable, if $\gamma < 0.1$; (2) needs improvements, if $0.1 < \gamma < 0.3$; or (3) is not acceptable, if $\gamma > 0.3$. Any reasonable measurement system used for 100% inspection has $\gamma < 0.5$.

This article provides justification for using the ANOVA estimator when assessing the measurement system using repeated measurements from units that fail an inspection system. Note that our purpose is not to assess or try to optimize the inspection protocol. We assume the inspection protocol is described by Figure 1.

The ANOVA estimator is based on the pooled variance of the measured values for each part that has two measurements. We derive the bias and variance of the estimator using properties of the truncated normal distribution. We compare the root mean square errors of the ANOVA and maximum likelihood estimator (MLE). We show that the ANOVA estimator has a relatively small bias and high efficiency when compared to the maximum likelihood estimator for most common values of γ. Finally, we consider some other applications for this assessment method.

MEASUREMENT SYSTEM ANALYSIS WITH INSPECTION DATA

Two methods of analysis are presented and compared: analysis of variance (ANOVA) and maximum likelihood (ML). Although interest lies in estimating γ, two other parameters σ_t^2 and μ are unknown and need to be estimated. To estimate these additional parameters, the ANOVA procedure uses the first measurements—that is, the production data—only. In contrast, the ML procedure uses all the data.

ANOVA

A natural estimate of the measurement variation σ_m^2 is the average within-part variance from those parts with two measurements. We estimate σ_t^2 by the sample variance of all the first measurements. The ANOVA estimate for the %GRR, denoted as $\hat{\gamma}_a$, is

$$\hat{\gamma}_a = \sqrt{\frac{s_m^2}{s_1^2}} = \frac{s_m}{s_1} \qquad [3]$$

where $s_1^2 = \frac{1}{n_1 - 1}\sum_i^{n_1}(y_{i1} - \bar{y}_{.1})^2$ is the production data variance, $s_m^2 = \frac{1}{n_2}\sum_{i \in S}\sum_{j=1}^{2}(y_{ij} - \bar{y}_{i.})^2$ is the average variation within parts with two measurements, $\bar{y}_{i.} = \frac{1}{2}(y_{i1} + y_{i2})$ is the average for any part i with two measurements, and $\bar{y}_{\cdot 1} = \frac{1}{n_1}\sum_i^{n_1} y_{i1}$ is the production data average. In the corresponding estimator $\widetilde{\gamma}_a$, each y_{ij} is replaced with the corresponding random variable Y_{ij} from model [1]. Note that we use a circumflex (ˆ) to overscore a parameter to denote the estimate (a number) and an overscore tilde (˜) to denote the corresponding estimator (a random variable).

To find the expectation and variance of the estimator $\widetilde{\gamma}_a$, we note that the distribution of the second measurement, conditional on the first measurement, is given by

$$Y_{i2}|(Y_{i1} = y_{i1}) \sim N[\mu + (1-\gamma^2)(y_{i1} - \mu), \sigma_t^2\gamma^2(2-\gamma^2)] \quad [4]$$

and when a second measurement occurs, the first measurement is outside the inspection limits. This means that if there is a second measurement, the distribution of the first measurement, Y_{i1}, is a truncated $N(\mu, \sigma_i^2)$ such that $Y_{i1} \notin B$. By conditioning on the first measurements, we can determine the expectation of S_m^2 as

$$E[S_m^2] = \sigma_t^2\gamma^2\left[1 - \gamma^2\frac{\beta_1}{2}\right] \quad [5]$$

and the variance of S_m^2 is

$$var(S_m^2) = \frac{2\sigma_t^4\gamma^4}{n_2}\left[1 - \gamma^2\beta_1 + \frac{1}{8}\gamma^4(3\beta_1 - \beta_1^2 - \beta_3)\right] \approx \frac{2\sigma_t^4\gamma^4}{n_2}(1-\gamma^2\beta_1) \quad [6]$$

where β_i is defined in Eq. [A3] in the Appendix. The derivation of both variance and expectation are shown in the appendix. The variance in Eq. [6] can be approximated by the right-most term because the contribution of the term $\frac{1}{8}\gamma^4(3\beta_1 - \beta_1^2 - \beta_3)$ is small. This simplification shows that variance of S_m^2 is inflated by $(1-\gamma^2\beta_1)$ relative to measuring any part twice. Because the variation from two independent measurements on a part estimates the measurement variation $\sigma_m^2 = \sigma_t^2\gamma^2$ and has a χ_1^2 distribution, the variance of this estimator is $2\sigma_m^2 = 2\sigma_t^2\gamma^2$.

The covariance between, S_m^2 and S_1^2 is near 0 because $\gamma < 0.5$ and n_1 is large. We can apply the delta method (see Stuart and Ord 1998) to find the approximate expectation and variance of $\widetilde{\gamma}_a$. We get

$$E[\widetilde{\gamma}_a] \approx \gamma\sqrt{1 - \gamma^2\frac{\beta_1}{2}} \approx \gamma - \gamma^3\frac{\beta_1}{4}, \quad \text{and} \quad [7]$$

$$var(\widetilde{\gamma}_a) \approx \frac{\gamma^2}{2}\left\{\frac{1-\gamma^2\frac{\beta_1}{2}}{n_1 - 1} + \frac{1}{n_2}\left[\frac{1-\gamma^2\beta_1}{1-\gamma^2\frac{\beta_1}{2}}\right]\right\} \approx \frac{\gamma^2}{2}\left\{\frac{1}{n_1-1} + \frac{1}{n_2}\right\}\left[1 - \gamma^2\frac{\beta_1}{2}\right]. \quad [8]$$

Equation [7] shows that $\widetilde{\gamma}_a$ is biased. Figure 2 shows the bias, $\gamma^3\beta_1/4$, as a function of the standardized inspection limits $\left(\alpha_1 = \frac{LIL-\mu}{\sigma_t}, \alpha_2 = \frac{UIL-\mu}{\sigma_t}\right)$ and γ. Note that the vertical axis is $\alpha_2 - \alpha_1$. For example, if γ is 0.2 and $(-\alpha, \alpha) = (-2,2)$, then we find the point (−2, 4) on the left panel of Figure 2 to obtain $\beta_1 = -5$ and then on the right panel, we find the bias to be ≈0.01.

For one-sided inspection limits, β_1 can similarly be determined from the left panel of Figure 2 because β_1 is the same for limits of the form $(-k, k)$, $(-\infty, k)$, and $(-k, \infty)$. For example, to find β_1 when the standardized inspection limits are $(-\infty, 1.5)$, we look up the point $\alpha_1 = -1.5$ and $\alpha_2 - \alpha_1 = 2 \times 1.5 = 3$ on the left panel of Figure 2.

We analyze the example presented in the introduction using the above method. From Tables 1 and 2, we have $s_1 = 4.86$, $s_m = 0.851$, $n_1 = 100$, and $n_2 = 17$. Thus, the ANOVA estimate, $\widehat{\gamma}_a$, from [3] is 0.175. Using Table 1, we estimate β_1 for our example as −2.05. This estimate is useful in determining the approximate bias of the ANOVA estimator. In Figure 2, viewing the line along $\beta_1 = -2.05$ we can see how the bias for this estimator depends on γ. Additionally, from [8] the standard error of $\widehat{\gamma}$ can be approximated as 0.033.

Maximum Likelihood

The log-likelihood for the data is the sum of two log-likelihoods: $l_1(\mu, \sigma_t^2)$, the likelihood of the

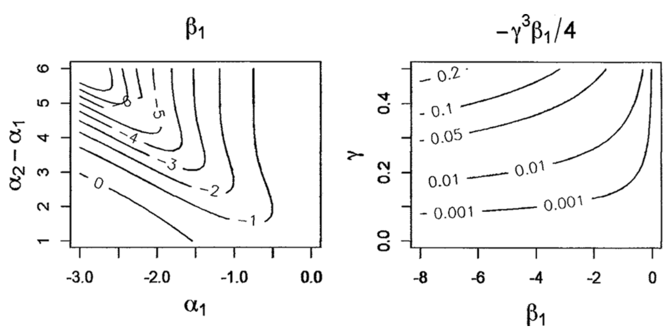

FIGURE 2 **The left panel gives β_1 for different values of the standardized inspection limits. The right panel is the bias of $\widetilde{\gamma}_a$ as a function of γ and β_1.**

production data, and $l_{21}(\mu, \sigma_t^2, \gamma | y_{i1}, i \in S_1)$, the likelihood of the retest data given the production data. The distribution of the production data is independent $N(\mu, \sigma_t^2)$. The distribution of Y_{i2} given $Y_{i1} = y_{i1}$ is given in [4]. Thus, the two log-likelihoods are

$$l_1(\mu, \sigma_t^2) = -\frac{n_1}{2}\log\sigma_t^2 - \frac{1}{2\sigma_t^2}\left[n_1 S_1^2 + n_1(\bar{y}_{.1} - \mu)^2\right] \quad \text{and} \qquad [9]$$

$$l_{21}(\gamma, \mu, \sigma_t^2 | y_{i1}, i \in S_1) = -\frac{n_2}{2}\log[\sigma_t^2\gamma^2(2-\gamma^2)] - \frac{1}{2}\frac{\sum_{i\in S}[y_{i2} - \mu - (1-\gamma^2)(y_{i1} - \mu)]^2}{\sigma_t^2\gamma^2(2-\gamma^2)} \qquad [10]$$

The complete log-likelihood for the inspection process is the sum of the two log-likelihoods [9] and [10]. To get the maximum likelihood estimates (MLEs) of μ, σ_t^2, and γ, we numerically maximize the complete log-likelihood.

Asymptotic standard errors for the estimators can be obtained from the inverted Fisher information matrix with the parameters replaced by their estimates. The Fisher information is obtained by taking the expectation of -1 times the log-likelihood second derivatives. The expectations depend on the distribution of the observations. For retest data, the distribution of the second measurements Y_{i2} is given in [4] and the first measurements Y_{i1} have a truncated $N(\mu, \sigma_t^2)$ with Y_{i1} outside the inspection limits. Corresponding to the two components of the log-likelihood, the Fisher information is the sum of the two matrices, J_1 and J_{21}. Taking derivatives and applying expectations, we obtain

$$J_1(\mu, \sigma_t^2, \rho) = n_1\begin{pmatrix} \frac{1}{\sigma_t^2} & 0 & 0 \\ 0 & \frac{1}{2\sigma_t^4} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \qquad [11]$$

$$J_{21}(\mu, \sigma_t^2, \gamma) = n_2\begin{bmatrix} \frac{\gamma^2}{\sigma_t^2(2-\gamma^2)} & 0 & -\frac{2\beta_0\gamma}{\sigma(2-\gamma^2)} \\ 0 & \frac{1}{2\sigma^4} & \frac{2(1-\gamma)(1+\gamma)}{\sigma^2\gamma(2-\gamma^2)} \\ -\frac{2\beta_0\gamma}{\sigma(2-\gamma^2)} & \frac{2(1-\gamma)(1+\gamma)}{\sigma^2\gamma(2-\gamma^2)} & 4\gamma^2(2-\gamma^2)(-\beta_1 - 1) + \frac{8}{\gamma^2(2-\gamma^2)^2} \end{bmatrix}. \qquad [12]$$

The Fisher information in [12] is obtained using the properties of conditional expectation and the moments given in the Appendix.

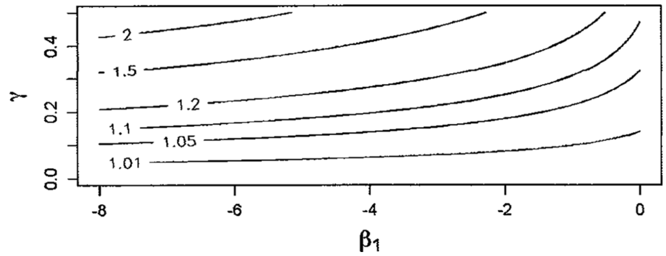

FIGURE 3 **The ratio $\sqrt{\text{MSE}(\tilde{\gamma}_a)/\text{MSE}(\tilde{\gamma}_{mle})}$ by γ and β_1.**

The asymptotic variance of the MLE for γ can be obtained by inverting the matrix $J_1 + J_{21}$. In general, to get a reasonable number of retests, we need a large number of production tests n_1. We can simplify the calculations if we let n_1 tend to infinity, so that the asymptotic variance of the maximum likelihood estimator for γ becomes

$$var(\tilde{\gamma}_{mle}) \approx \frac{1}{4n_2}\frac{\gamma^2(2-\gamma^2)^2}{[\gamma^2(2-\gamma^2)(-\beta_1 - 1) + 2]} \qquad [13]$$

This is also the variance we get by assuming that μ and σ_t are known.

Applying maximum likelihood to our example, we get estimates for μ, σ_t^2, and γ of (100.0, 24.03, 0.171), respectively. The standard error for the maximum likelihood estimator for γ is 0.0295.

In Figure 3, we compare the mean squared errors (MSE) of the MLE and ANOVA estimator as given by [13] and [8] as functions of γ and β_1. The figure shows, as expected, that the MLE is more efficient. However, when $\gamma \leq .1$ the two estimators are almost equivalent, and when $.1 \leq \gamma \leq .3$, the MLE is only slightly better than the ANOVA. When $\gamma \geq .3$, the ANOVA estimator performs poorly relative to the MLE. Using simulation we confirmed the results from Figure 3 and the MLE's unbiasedness.

Given the cost and complexity of finding the MLE, we recommend the ANOVA estimate in most situations. Also, the ANOVA estimates can be quickly calculated from the summary statistics generated by statistical software such as JMP (SAS Institute, Inc.) or *MINITAB* (Minitab, Inc.).

NONNORMALITY

The results we obtained were derived assuming normality of both the part and measurement distrbution. We present a small sensitivity analysis of the normality assumption on the part distribution.

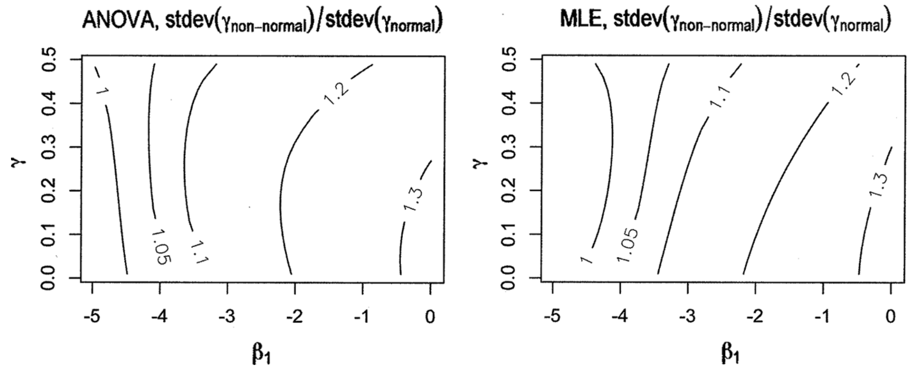

FIGURE 4 The ratio $\sqrt{Var(\tilde{\gamma}_{non-normal})/Var(\tilde{\gamma}_{normal})}$ of by the ANOVA (left panel) and the MLE (right panel) by γ and β_1.

Because the inspection system measures every part produced, we should be able to detect any nonnormality in the observed distribution and then change our analysis to compensate for this.

In the simulation, the part distribution was switched to a Student's t-distribution with 5 degrees of freedom (t_5). Because, the variance of a t_5 is 5/3, the random variables simulated from this distribution were scaled by $\sqrt{3/5}$. We compare the standard deviations from the simulation to the theoretical standard deviations we would obtain if the part distribution was normal. The smoothed results in Figure 4 suggest that the standard deviation does not increase dramatically if the part distribution is t_5. No significant changes were found in the bias of either estimator.

MODEL ASSESSMENT AND ASSESSMENT INTERVALS

This section gives suggestions for assessing the model [1] and applies them to the example. We can assess the overall normality (part plus measurement) from the production data. The normal quantile plot of the production data from Table 1 given in the left panel of Figure 5 shows no evidence to reject normality. Another check on normality comes from Eq. [4]. Because, Y_{i2} depends on Y_{i1} conditionality through the mean with the equal variance for each pair of measurements, we can regress y_{i2} onto y_{i1} and check the residuals for departures in normality. A normal quantile of the residuals for the data in Table 2 is shown in the right panel of Figure 5. Checking for constant variance in these residuals is the same as checking linearity in the measurement system. Linearity occurs if the measurement variation σ_m does not depend on the true dimension P_i of any part.

Using the methods discussed in the Measurement System Analysis with Inspection Data Section we can assess the measurement system without conducting a separate off-line study, but we need to specify assessment times. We could conduct analysis on a regular schedule, say, weekly/monthly, or when some desired precision is achieved. The precision can be achieved by specifying the number of

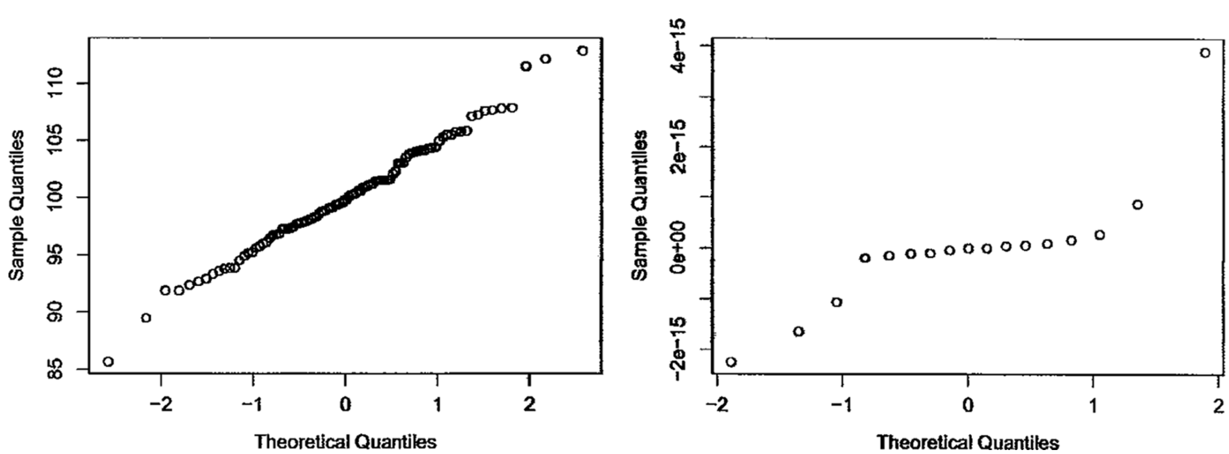

FIGURE 5 Normal quantile plots of the production data (left panel) and the residuals from regressing y_{i2} onto y_{i1} (right panel).

parts n_1 or the number of second measurements n_2 included in the study.

To illustrate, suppose we specify the desired precision for the ANOVA estimator. Using the ANOVA estimate from the example as the true values, Eq. [8], and assuming n_1 is large, we can obtain the number of second measurements, n_2, required to achieve the desired precision. If the desired standard error for $\hat{\gamma}$ is 0.01 and the standardized inspection limits are $A = (-1.05, 2.04)$ and $\gamma = 0.175$, then the number of second measurements needs to be at least 163.

To obtain standard errors in any inspection system use Figure 2 to calculate β_1. Then along with the approximate value of γ, input these values into [8].

DISCUSSION AND CONCLUSIONS

In this article we looked at the case where there is a single response. These methods can be extended to inspection systems with k independent characteristics, although the independence assumption is likely unreasonable. If all characteristics from a part are remeasured when a single characteristic fails, then, because of independence, some of the repeated measurements are equivalent to randomly selecting a part and measuring twice. For example, suppose there are two independent characteristics X and Y and during the production test, Y fails resulting in a second measurement for both X and Y. Then the two repeated measurements for X are equivalent to taking a random part and measuring X twice. Other parts will have second measurements because the production test for X failed. So, in this situation, the ANOVA estimator will be composed of two types of repeated measurements.

The assumption that the inspection system has k independent characteristics is very restrictive, but removing this assumption complicates things considerably. First we need to incorporate the dependency structure into the model [1]. Second, we note that with multiple measurments, we are apt to get continuous, ordinal, and binary characteristics determined by the same inspection system.

Using the methodology from the Measurement System Analysis with Inspection Data Section, we can also assess the measurement variation in situations where multiple gauges are used in parallel and we are trying to detect differences in variation among the gauges. This means σ_m now becomes σ_{mi} where i denotes the gauge. We assume the parts are randomly allocated to the gauges to ensure that over the long term the part variation is the same for each gauge. Here we assume that there are only two gauges but the methodology can be generalized.

If we did not have knowledge of the section mentioned above, we would use the production data or the first measurements to assess the differences in the parallel gauges. This method detects differences in σ_{m1} and σ_{m2} through the total variation $\sigma_{11}^2 = \sigma_p^2 + \sigma_{m1}^2$ and $\sigma_{12}^2 = \sigma_p^2 + \sigma_{m2}^2$. Detection of differences will be difficult if σ_p^2 is the dominant component of the variation. This is the typical situation. Thus, adding repeated measurements into this type of analysis will greatly improve the power to detect differences in the two measurement variation components when they are both small relative to the part variation. The analysis in the aforementioned section suggests that we can use the ratio of the ANOVA estimates for each gauge to compare σ_{m1} to σ_{m2i}.

Modifying the inspection protocol (see Figure 1) will change the results given in this article. One example brought up by a referee is that some inspection protocols allow two retests instead of just one. This modification will likely inflate the variance of the ANOVA estimator beyond what is tolerable. However, to apply the methodology presented, we could ignore any second retest. Presumably very few parts are measured three times, so not much information would be lost.

We present two ways to analyze a measurement system from inspection data. They enable to us to avoid off-line studies such as a standard gauge R&R. In addition, we showed that the ANOVA estimator, although biased, is comparable to the MLE if the gauge repeatability and reproducibility (denoted as γ within this article) is ≤ 0.1. If γ is within the interval (.1, .3), we suggest using the MLE but using the ANOVA estimator is not unreasonable. However, if $\gamma \geq .3$, the bias of the ANOVA estimator is substantial and the MLE is more efficient, but in most manufacturing situations, if $\gamma \geq .3$, the measurement system used for 100% inspection is so poor that estimator efficiency will not be a primary concern.

ACKNOWLEDGMENTS

Ryan Browne acknowledges the financial support of the National Science and Engineering Research Council of Canada (NSERC), Research in Motion

Limited, and the Mathematics of Information Technology and Complex Systems (MITACS). We thank the Editor (for the example in testing flammability of plastics and other chemically based products) and the referees for numerous suggestions that greatly improved the article.

APPENDIX—PROPERTIES OF S_m^2

Truncated Normal

The moment generating function (MGF) for $X \sim N(\mu, \sigma^2)$ (see Fisher 1931) that is truncated such that $X \epsilon [b_1, b_2]$ where $b_2 > b_1$ (denoted as truncated $N[\mu, \sigma^2; b_1, b_2]$) is

$$M(t; b_1, b_2) = e^{\mu t + \sigma^2 t^2/2} \frac{\left[\Phi\left(\frac{b_2-\mu-\sigma^2 t}{\sigma}\right) - \Phi\left(\frac{b_1-\mu-\sigma^2 t}{\sigma}\right)\right]}{\left[\Phi\left(\frac{b_2-\mu}{\sigma}\right) - \Phi\left(\frac{b_1-\mu}{\sigma}\right)\right]} \quad \text{[A1]}$$

where $\Phi(x)$ is the standard normal cumulative distribution function. Using the MGF the first four moments are

$$\begin{aligned}
E(X) &= \mu - \lambda_0 \sigma \\
E(X^2) &= \mu^2 - 2\lambda_0 \sigma \mu + (1-\lambda_1)\sigma^2 \\
E(X^3) &= \mu^3 - 3\lambda_0 \sigma \mu^2 + (3 - 3\lambda_1)\sigma^2 \mu + (-2\lambda_0 - \lambda_2)\sigma^3 \\
E(X^4) &= \mu^4 - 4\lambda_0 \sigma \mu^3 + (-6\lambda_1 + 6)\sigma^2 \mu^2 \\
&\quad + (-8\lambda_0 - 4\lambda_2)\sigma^3 \mu + (-3\lambda_1 + 3 - \lambda_3)\sigma^4
\end{aligned}$$

where $\alpha_1 = \frac{b_1-\mu}{\sigma_t}$, $\alpha_2 = \frac{b_2-\mu}{\sigma_t}$, for $i = 0, 1, 2, 3$; $\lambda_i = \frac{\alpha_2^i \phi(\alpha_2) - \alpha_1^i \phi(\alpha_1)}{\Phi(\alpha_2) - \Phi(\alpha_1)}$; and $\phi(x)$ is the standard normal probability density function.

If Y is truncated $N(\mu, \sigma^2)$, such that $Y \notin A = (a_1, a_2)$, where $a_2 > a_1$, then we can write $Y = uX_1 + (1-u)X_2$ where $X_1 \sim$ truncated $N(\mu, \sigma^2; -\infty, a_1)$, $X_2 \sim$ truncated $N(\mu, \sigma^2; a_2, \infty)$, and $u = \left[\Phi\left(\frac{a_1-\mu}{\sigma}\right)\right] / \left[1 - \Phi\left(\frac{a_2-\mu}{\sigma}\right) + \Phi\left(\frac{a_1-\mu}{\sigma}\right)\right]$ then the moment generating function for Y is

$$M_Y(t) = uM(t; -\infty, a_1) + (1-u)M(t; a_2, \infty) \quad \text{[A2]}$$

where $M(t; b_1, b_2)$ is given in [A1]. Thus, Y has the same moments as X with the exception that for $i = 0, 1, 2, 3$, λ_i is replaced with

$$\beta_i = u\lambda_i\left(-\infty, \frac{a_1-\mu}{\sigma}\right) + (1-u)\lambda_i\left(\frac{a_2-\mu}{\sigma}, \infty\right)$$
$$\text{where} \quad \lambda_i(z_1, z_2) = \frac{z_2^i \phi(z_2) - z_1^i \phi(z_1)}{\Phi(z_2) - \Phi(z_1)} \quad \text{[A3]}$$

Normal Moments

If $X \sim N(\mu, \sigma^2)$, then the first four moments (see Johnson and Kotz 1970) are

$$\begin{aligned}
E(X) &= \mu \\
E(X^2) &= \mu^2 + \sigma^2 \\
E(X^3) &= \mu^3 + 3\mu\sigma^2 \\
E(X^4) &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4
\end{aligned}$$

Expectation

We define

$$\begin{aligned}
S_m^2 &= \frac{1}{n_2}\sum_{i \epsilon S}\sum_{j=1}^{2}(Y_{ij} - \overline{Y}_i)^2 = \frac{1}{n_2}\sum_{i \epsilon S}\frac{1}{2}(Y_{i1} - Y_{i2})^2 \\
&= \frac{1}{n_2}\sum_{i \epsilon S} S_m^2. \quad \text{[A4]}
\end{aligned}$$

Each pair of measurements from different parts is independent, so we need to determine the properties of $\frac{1}{2}(Y_{i1} - Y_{i2})^2$. To simplify the calculations, we notice that we can define $Y_{ij} = X_{ij} + \mu$ where X_{ij} has the same distribution as Y_{ij} but with paramter $\mu = 0$. Then

$$\begin{aligned}
E[S_{im}] &= \frac{1}{2}E\left[(Y_{i1} - Y_{i2})^2\right] \\
&= \frac{1}{2}E\left[(X_{i1} + \mu - X_{i2} - \mu)^2\right] \\
&= \frac{1}{2}E\left[(X_{i1} - X_{i2})^2\right] \\
&= \frac{1}{2}E\left[E(X_{i1}^2 - 2X_{i1}X_{i2} + X_{i2}^2 \mid X_{i1})\right] \\
&= \frac{1}{2}E\left[X_{i1}^2 - 2X_{i1}E(X_{i2} \mid X_{i1}) + E(X_{i2}^2 \mid X_{i1})\right] \\
&= \frac{1}{2}E\{X_{i1}^2 - 2X_{i1}[(1-\gamma^2)X_{i1}] \\
&\quad + [(1-\gamma^2)X_{i1}]^2 + \sigma_t^2\gamma^2(2-\gamma^2)\} \\
&= \left[-\frac{1}{2} + \gamma^2 + \frac{1}{2}(1-\gamma^2)^2\right]E[X_{i1}^2] + \frac{1}{2}\sigma^2\gamma^2(2-\gamma^2)
\end{aligned}$$

$$= \left[-\frac{1}{2} + \gamma^2 + \frac{1}{2}(1-\gamma^2)^2 \right][\sigma^2] + \frac{1}{2}\sigma^2\gamma^2(2-\gamma^2)$$
$$= \sigma^2\gamma^2[1-\gamma^2\beta_1/2]$$

Thus, the expectation of S_{im}^2 is $\sigma^2\gamma^2[1-\gamma^2\beta_1/2]$.

Variance

To calculate the variance of S_m we use S_{im}, defined in [A4] to get

$$\mathrm{var}(S_m^2) = \left(\frac{1}{n_2}\right)^2 \sum_{i \epsilon S} \mathrm{var}(S_{im}^2) = \frac{1}{n_2}\mathrm{var}(S_{im}^2) \qquad [A5]$$

because the measurements on different parts are independent. If we use the identity

$$\mathrm{var}(S_{im}^2) = E\left[S_{im}^4\right] - E\left[S_{im}^2\right]^2 \qquad [A6]$$

we only need to calculate $E\left[S_{im}^2\right]$ to determine the variance.

$$E\left[S_{im}^4\right] = E\left[\left(\frac{1}{2}(Y_{i1}-Y_{i2})^2\right)^2\right]$$
$$= \frac{1}{4}E\left[X_{i1}^4 - 4X_{i1}^3X_{i2} + 6X_{i1}^2X_{i2}^2 - 4X_{i1}X_{i2}^3 + X_{i2}^4\right]$$
$$= \frac{1}{4}E[E(X_{i1}^4 - 4X_{i1}^3X_{i2} + 6X_{i1}^2X_{i2}^2$$
$$- 4X_{i1}X_{i2}^3 + X_{i2}^4 \mid X_{i1})]$$
$$= \frac{1}{4}E\left[X_{i1}^4 - 4X_{i1}^3E(X_{i2} \mid X_{i1}) + 6X_{i1}^2E(X_{i2}^2 \mid X_{i1})\right.$$
$$\left. -4X_{i1}E(X_{i2}^3 \mid X_{i1}) + E(X_{i2}^4 \mid X_{i1})\right]$$
$$\vdots$$
$$= \frac{1}{4}\left[-3 + 4\gamma^2 - 4(1-\gamma^2)^3 + (1-\gamma^2)^4\right.$$
$$\left. +6(1-\gamma^2)^2\right]E[X_{i1}^4] + \frac{1}{4}[6\sigma^2\gamma^2(2-\gamma^2)$$
$$+ 6(1-\gamma^2)^2\sigma^2\gamma^2(2-\gamma^2)$$
$$- 12(1-\gamma^2)\sigma^2\gamma^2(2-\gamma^2)]E[X_{i1}^4]$$
$$+ \frac{1}{4}3\sigma^4\gamma^4(2-\gamma^2)^2$$
$$\vdots$$
$$= \frac{1}{4}\sigma^4\gamma^4(3\gamma^4\beta_1 - \gamma^4\beta_2 - 12\beta_1\gamma^2 + 12)$$

Now, combining this result with the previous formula we have

$$\mathrm{var}(S_{im}^2) = E[S_{im}^4] - E[S_{im}^2]^2$$
$$= \frac{1}{4}\sigma^4\gamma^4(3\gamma^4\beta_1 - \gamma^4\beta_3 - 12\beta_1\gamma^2 + 12)$$
$$- [\sigma^2\gamma^2(1-\gamma^2\beta_1/2)]^2$$
$$\vdots$$
$$= \frac{1}{4}\sigma^4\gamma^4(3\gamma^4\beta_1 - \gamma^4\beta_3 - 8\beta_1\gamma^2 - \gamma^4\beta_1^2)$$
$$= 2\sigma^4\gamma^4\left(1 - \beta_1\gamma^2 - \frac{\gamma^4}{8}(\beta_3 - 3\beta_1 + \beta_1^2)\right)$$

and thus,

$$\mathrm{var}(S_{im}^2) = \frac{2\sigma^4\gamma^4}{n_2}\left(1 - \beta_1\gamma^2 - \frac{\gamma^4}{8}(\beta_3 - 3\beta_1 + \beta_1^2)\right) \qquad [A7]$$

ABOUT THE AUTHOR

Ryan Browne recently graduated with a Ph.D. in statistics from the University of Waterloo. Currently, he is doing a Biostatistics post doctoral at the NCIC Clinical Trials Group in Kingston, Ontario.

REFERENCES

Automotive Industry Action Group. (2002). *Measurement Systems Analysis,* 3rd ed. Southfield, MI: Chrysler, Ford, General Motors Supplier Quality Requirements Task Force.

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

Doganaksoy, N. (2000). Assessment of impact of measurement variability in the presence of multiple sources of product variability. *Quality Engineering*, 13:83–89.

Fisher, R. A. (1931). The truncated normal distribution. *British Association for Advancement of Science*, 5:xxxiii–xxxiv.

Johnson, N., Kotz, S. K. (1970). *Distributions in Statistics: Continuous Univariate Distributions*. Boston, MA: Houghton Mifflin.

Montgomery, D. C. (2000). *Introduction to Statistical Quality Control*. Hoboken, NJ: Wiley.

Stuart, A., Ord, J. K. (1998). *Kendalls Advanced Theory of Statistics*, Vol. 1, 6th ed. London: Arnold.

Wheeler, D. J., Lyday, R. (1984). *Evaluating the Measurement Process*. Knoxville, TN: SPC Press, Inc.