

The impact of the Weibull distribution on the performance of the single-factor ANOVA modelGray Black^{a*}, Derek Ard^b, James Smith^c and Tim Schibik^d^a*School of Business, University of Sothern Indiana, Indiana, USA*^b*National Laboratories, Oak Ridge, Tennessee, USA*^c*Department of Industrial Engineering, Tennessee Technical University, Tennessee, USA*^d*School of Business, University of Sothern Indiana, Indiana, USA***ARTICLE INFO****ABSTRACT***Article history:*

Received 1 January 2010

Received in revised form

8 June 2010

Accepted 9 June 2010

Available online 9 June 2010

Keywords:

ANOVA robustness

Weibull

ANOVA normality assumption

Shape parameter

Scale parameter

Normality violation

This paper conducts a simulation study of the effects of violating the ANOVA normality assumption in the presence of Weibull data. Twelve specific Weibull distributions, characterizing the life data of a variety of real-world products and systems, are investigated. Confidence intervals on test significance and power are generated and compared against intervals from normally distributed data. The ANOVA procedure is found to be robust in the majority of cases. Furthermore, a designed experiment is conducted to isolate the effects of the Weibull shape and scale parameters within the preceding study. The shape parameter is found to have a significant effect on significance and power, whereas the scale parameter does not have a significant effect at the target $\alpha = 0.05$ test significance level.

© 2010 Growing Science Ltd. All rights reserved.

1. Introduction

Many physical systems have been observed to generate data that follow the Weibull distribution. Examples include physical models related to the reliability and life cycle analysis of product or machine failure times, ball bearing lives and environmental concepts such as wind speed. The Weibull distribution consists primarily of two parameters, shape and scale. These parameters provide the distribution with flexibility to model systems in which the number of events (e.g., failures) increases with time (e.g., bearing wear), decreases with time (e.g., some semiconductor failures) or remains constant with time (e.g., failures caused by external shocks to the system). A third Weibull parameter, location, is often not utilized in such reliability/life cycle applications since $t = 0$ (i.e., the origin) is a realistic starting point for the process.

The Analysis of Variance (ANOVA) method could conceivably be very useful to isolate effects and to determine the significance of sources of variation within a Weibull process. However, ANOVA is based on a set of assumptions that, at least theoretically, must be met to have confidence in the results. One of these assumptions is that the data, or error term, must follow a normal distribution. Although previous research has investigated the effect of normality violations for various distributions, the effect of Weibull shape and scale parameters has not been previously studied. Thus, this paper will explore the effects of various Weibull distributions on the robustness of the one-way fixed effects ANOVA method to determine whether the power or significance of the test is compromised. Moreover, a designed experiment will be conducted to isolate the effects of the shape

* Corresponding author. Tel: +01-812- 461-5385

E-mail addresses: gblack@usi.edu (G. Black)

and scale parameters to assess their individual significance. The next section will review background literature associated with this research.

2. Literature Review

Glass, et al. (1972) reported three distinct violations of the ANOVA assumptions pertaining to model errors: (1) non-normality, (2) unequal variances, and (3) non-independence. The first violation occurs when treatment errors conform to a distribution other than the normal distribution. The second violation exists when treatment variances are unequal. The third violation occurs when data correlation exists. This review will focus on violations of the first assumption, normality, in the presence of Weibull-distributed data. Previous research into the effects of violating the normality assumption has been both theoretical and empirical in nature. David and Johnson (1951) and Srivastava (1959) conducted theoretical studies on the effects of error non-normality upon the power of ANOVA. Four moments of a function representing a non-normal distribution were investigated. The moments were fitted to a frequency distribution curve using probability estimation, thus allowing the examination of power. Srivastava (1959) used the first four terms of the Edgeworth series to represent distributions of non-normal data. Non-normality of the data was defined in terms of skewness and kurtosis (peakedness). Numerical examples were provided to capture the effects of non-normality on the power of the F-test. In each of these studies, the populations considered were moderately non-normal and the resultant effects on power were slight. Smith (1966) conducted an inquiry into both the confidence and the power of the one-way fixed effects ANOVA model under the condition of non-normality. Monte Carlo simulation was used to determine the effects of various non-normal distributions on power. Distributions considered included variations of the L-shaped, U-shaped, J-shaped and bell-shaped curves. With the exception of the U-shaped distribution, the conclusions supported those of David and Johnson (1951) and Srivastava (1959) for moderately non-normal distributions. However, for cases of extreme non-normality as in the L-shaped and U-shaped populations, the power and the confidence of the test were found to be low. Games and Lucas (1966) investigated the effects of non-normality on the power of the fixed effects ANOVA model. Their empirical study provided a comparison of the power curves generated from a series of non-normal populations with the power curves generated from the respective normally transformed populations. Six populations of non-normal data were considered. One of the six populations was a discrete approximation of the normal curve; the remaining populations varied in degrees of kurtosis and skewness. The non-normal populations were labeled as slight, moderate, and extreme. The normal transformation procedures of log, square root and reciprocal were used. It was found that deviation from the normal power curve was insignificant for slight and moderate departures from normality. Moreover, it was found that transforming non-normal data increased the deviation from the theoretical power curve. Furthermore, significant deviation from the normal power curve was found in populations with a high degree of leptokurtosis. Donaldson (1968) studied the effects of non-normality and unequal variance on the power of the F-Test. An empirical investigation found that when within-treatment variance was equal, the power curves corresponding to the exponential and lognormal distributions were conservative. The non-normal power curves approached those of the normal distribution as the number of treatments and/or the number of replications per treatment increased. Driscoll (1990) presented a 'bootstrapping' procedure as an alternative to ANOVA for cases involving non-normality. Driscoll (1996) also conducted an empirical study on the effects of the gamma family of distributions with respect to Type I error for ANOVA. Multiple gamma parameter values were considered. Simulation was used to generate empirical Type I error values for each case which, in turn, were compared to the nominal significance level of $\alpha = 0.05$. It was found that the exponential distribution (i.e., a special case of the gamma) resulted in the largest deviation from the nominal α value. However, this discrepancy was only approximately 0.3%. Thus, it was concluded that the use of ANOVA was substantiated for most gamma distributions. Harwell, et al (1992) conducted a series of Monte Carlo studies on the robustness of the One-Way and Two-Way ANOVA, Welch and Kruskal-Wallis tests in terms of test significance and power under equal and

unequal sample sizes and variances. The ANOVA F-test was found to be robust overall, although the test significance did show sensitivity to unequal variances. To model non-normality, general variables were used for skew and kurtosis. However, the specific effects of the Weibull shape and scale parameters were not explicitly considered. Lix et al. (1996) provided a quantitative review of alternatives to the one-way ANOVA test. This study was prompted by the presence of variance heterogeneity and non-normality in educational and psychological data that may frequently invalidate the use of ANOVA. This paper offered recommendations to applied researchers on the use of various parametric and nonparametric alternatives to the F test under assumption violation conditions. Meta-analytic techniques were used to summarize the statistical robustness literature on the Type I error properties of the Brown-Forsythe (Brown & Forsythe, 1974), James (1951) second-order, Kruskal-Wallis (Kruskal & Wallis, 1952) and Welch (1951) tests. Mendes (2007) investigated the effects of non-normality on the Type III error rates for the ANOVA F-test and its three common parametric counterparts, namely the Welch test, Brown-Forsythe test and the Alexander-Govern test. These tests were compared in terms of Type III error rates across a variety of distributions, effect sizes and sample sizes. Results indicated that Type III error was affected by effect size and sample size, but not by the shape of the distribution. Although the paper studied one particular Weibull distribution, the $W(1.5, 1)$, the effects of varying shape and scale parameters were not investigated. Li Li (2007) developed a Modified ANOVA F-test based on the Kaplan-Meier and Satterthwaite methods and the Box-Cox transformation to compensate for censored data, heteroscedasticity and non-normality. Subsequently, it compared the Modified ANOVA to the Welch test to examine test power and significance under such cases. Results indicated the Modified ANOVA and Welch tests performed similarly for equal sample sizes, whereas the Modified ANOVA was generally more robust than the Welch. The paper examined the Weibull distribution as a function of two variables, the percentage of censored data and the difference of the standard deviations. However, the Weibull shape and scale parameters were not explicitly examined and, thus, no conclusions can be inferred regarding their effects.

In summary, various non-normal distributions have been examined using both theoretical and empirical approaches. ANOVA was found to be fairly robust to slight and moderate departures from normality. Only in cases of extreme non-normality was the power or significance of the test affected to a significant extent. Although a few sources considered the Weibull distribution, none of these sources explicitly studied the effects and statistical significance of Weibull shape and scale parameters. As shown in Table 3 and as discussed in Section 4.1, many real-world products have reliability/life cycle features which follow the Weibull distribution with predictable shape and scale parameter values. Thus, a study of the effect and significance of the shape and scale parameters is warranted and will be conducted in this paper to determine the impact on the significance and power of the one-way ANOVA test.

3. Model

This section will discuss the logic underlying the *Java* simulation model used to obtain the experimental results.

A three-step process will be used to generate an ANOVA data point y_{ij} from the Weibull distribution: (1) sample a uniform random number $r = U[0,1]$, (2) compute the Weibull inverse cumulative distribution value x , and (3) compute the corresponding ANOVA data value y_{ij} . These steps will be explained in more detail in the following paragraphs. Recall that the Weibull cumulative probability function is defined as:

$$F(x) = 1 - \exp[-(x/\delta)^\gamma], \quad (1)$$

where δ is the scale parameter and γ is the shape parameter. The shape parameter provides the distribution with flexibility to attain a variety of shapes ranging from an exponential curve (when $\gamma =$

1) to a bell-shaped curve (when γ becomes large). The scale parameter, or characteristic life, affects the elongation of the curve. To obtain a random Weibull value x , the Weibull cumulative probability function is rewritten, substituting the random variable $r = U[0,1]$ for $f(x)$, as follows:

$$x = -\delta [\ln(1-r)]^{1/\gamma}. \quad (2)$$

By repeatedly sampling Weibull data values x , an ANOVA table can be populated with individual y_{ij} observations using the following expression:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad (3)$$

where y_{ij} is the response value corresponding to the j^{th} replication of treatment i ($i = 1$ to a , $j = 1$ to n), μ is the overall average across all treatments, τ_i is the effect of treatment i ($\mu_i - \mu$) and ε_{ij} is the error term. The normality assumption inherent within ANOVA is that the error term ε_{ij} is normally distributed with a mean of zero and a constant variance of σ^2 . If the actual process follows a Weibull distribution, the mean of the error term ε_{ij} will differ from zero and the population variance σ^2 will be defined as follows:

$$\sigma^2 = \delta^2 \Gamma[1+(2/\gamma)] - \delta^2 [\Gamma(1+1/\gamma)]^2. \quad (4)$$

Violations of the normality assumption under Weibull data can be investigated by setting $\varepsilon_{ij} = x$ and expressing τ_i as a function of σ , thus yielding a y_{ij} value for each observation. Specific details regarding the experimental design will be discussed in the next section. After y_{ij} values have been generated, the ANOVA table shown in Table 1 can be populated.

Table 1
Data for one-way fixed-effects ANOVA model

| Treatment | Observations |
|-----------|------------------------------------|
| 1 | $y_{11} \ y_{12} \ \dots \ y_{1n}$ |
| 2 | $y_{21} \ y_{22} \ \dots \ y_{2n}$ |
| : | : : : : : |
| a | $y_{a1} \ y_{a2} \ \dots \ y_{an}$ |

Subsequently, the ANOVA analysis can be performed by decomposing the total variability in the data (SS_{TOTAL}) into explained and unexplained categories (SS_{TREAT} and SS_{ERROR} , respectively). The mean square values (MS_{TREAT} and MS_{ERROR}) and the F-ratio can then be computed (Montgomery, 2001) and the ANOVA hypothesis in Eq. 5 can be tested.

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a \quad (5)$$

$$H_1: \text{at least one } \mu_i \neq 0$$

The null hypothesis H_0 conjectures that all treatment means μ_i are equal (i.e., all treatment effects $\tau_i = 0$). The alternative hypothesis H_1 states that at least one of the treatment means differs. If the computed F-ratio $> F_{\alpha, a-1, a(n-1)}$ critical value, the null hypothesis is rejected. By replicating this ANOVA study using Weibull data and various treatment effects τ_i , we can tally the proportion of replications in which H_0 was correctly rejected (i.e., in cases where at least one $\tau_i \neq 0$) or was

incorrectly rejected (i.e., in cases where all $\tau_i = 0$). The first scenario provides an empirical estimate $1-\beta'$ of the power of the test under Weibull data ($\beta = \text{Type II error}$). The second scenario provides an empirical estimate α' of the significance level of the test ($\alpha = \text{Type I error}$) under Weibull data. Repeating this procedure multiple times, confidence intervals can be constructed on the significance level and power using standard confidence interval formulas, such as those found in Montgomery and Runger (2010). In turn, these confidence intervals (based on Weibull data) can be compared to the corresponding confidence intervals from normally distributed data (using common random numbers) to determine whether the validity of the ANOVA test has been significantly compromised. The next section will provide further details on the experimental design.

4. Experimentation

This section is comprised of two subsections. In Subsection 4.1, the experimental design will be presented. In Subsection 4.2, the experimental results will be discussed and analyzed.

4.1 Experimental Design

As stated, the research objectives are two-fold:

1. To investigate the robustness of the one-way fixed effects ANOVA test, relative to power and significance, in the case when the constituent data violates the underlying normality assumption by following the Weibull distribution
2. To isolate the effects of the Weibull shape and scale parameters to determine which, if any, parameters have a significant effect on power and significance

To consider the first objective, results collected from tests utilizing Weibull data (experimental group) will be compared to the respective baseline results using normal data (control group). The study consists of five major cases as shown in Table 2.

Table 2

Experimental cases (treatment effect settings)

| Case | τ_1 | τ_2 | τ_3 | Response |
|------|---------------|--------------|--------------|-------------------------------|
| 1 | 0 | 0 | 0 | 4.2 Significance (α) |
| 2 | -0.5σ | 0 | 0.5σ | |
| 3 | -0.75σ | 0 | 0.75σ | Power |
| 4 | $-\sigma$ | 0 | σ | $(1-\beta)$ |
| 5 | -0.5σ | -0.5σ | σ | |

Each case contains different settings for the effect τ_i of three random treatments, expressed as multiples of the population standard deviation σ . In turn, σ is computed as the square root of Equation 4 for the Weibull data (or $\sigma = 1$ for the baseline Normal data). In Case 1, all effects are zero. Thus, this case will be used to derive empirical estimates of the significance level (since H_0 should not be rejected). In Cases 2-5, two or more effects are non-zero. Thus, these cases will be used to derive empirical estimates of the power (since H_0 should be rejected). Although the specific coefficient values in Table 2 may appear somewhat arbitrary, the rationale is rather simple. If the

differences between the treatment means (i.e., effects) were very small, then any ANOVA study (i.e., based either upon normal or Weibull data) would have difficulty in detecting those differences. If the differences between the treatment means were very large, then any ANOVA study would easily be able to detect those differences. Thus, it is those “intermediate” cases which are of most interest. The most convenient way to quantify those cases (i.e., differences) is as a multiple of the population standard deviation σ . Each of the five cases in Table 2 will be studied under 13 different sub-cases as shown in Table 3. Twelve of the 13 sub-cases (# 2-13) correspond to specific Weibull distributions (i.e., shape and scale parameter values). The first sub-case corresponds to the baseline normal distribution against which Weibull results will be compared. These specific Weibull distributions were obtained from an online database (Barringer & Associates, Inc., 2001) listing specific Weibull distributions for real-world failure) characterizing life data for various types of machinery, components, instrumentation, static equipment and service liquids. Thus, the entire study consists of $5 * 13 = 65$ sub-cases. Within each major case, confidence intervals will be constructed and examined for each sub-case. Specifically, the 12 Weibull confidence intervals will be compared to the baseline (normal) results to determine if significant differences exist in the significance level (Case 1) or power (Cases 2-5) of the ANOVA test. Common random numbers will be used across the 65 sub-cases.

Table 3
Experimental distributions (Sub-cases)

| Sub-case | Distribution | Product | Shape Parameter (γ) | Scale Parameter (δ) |
|----------|--------------|-----------------------|------------------------------|------------------------------|
| 1 | Normal | Baseline distribution | n/a | n/a |
| 2 | Weibull | Controller | 0.7 | 100 |
| 3 | Weibull | Pressure indicators | 1 | 100 |
| 4 | Weibull | Coolants | 1.1 | 15 |
| 5 | Weibull | Boilers | 1.2 | 50 |
| 6 | Weibull | Roller Bearings | 1.3 | 50 |
| 7 | Weibull | Valves | 1.4 | 40 |
| 8 | Weibull | Circuit Breakers | 1.5 | 100 |
| 9 | Weibull | Steam Turbine | 1.7 | 65 |
| 10 | Weibull | Liner | 1.8 | 50 |
| 11 | Weibull | Diaphragm Couplings | 2 | 300 |
| 12 | Weibull | Gear Couplings | 2.5 | 75 |
| 13 | Weibull | Metal diaphragm | 3 | 65 |

The procedure used to construct a confidence interval for each sub-case is as follows:

1. Using Monte Carlo simulation, sample a random observation y_{ij} from the Weibull distribution for Sub-cases 2-13 using Eqs. 2, 3 and 4, or from the normal distribution using $\varepsilon_{ij} = N(\mu = 0, \sigma^2 = 1)$ for Sub-case 1. Set $\mu = 0$ in Eq. 3 for all distributions with no loss of generality. Perform this step 30 times to populate the ANOVA table with 30 values (i.e., $a = 3$ treatments * $n = 10$ replicants).

2. Employ the ANOVA procedure on the populated ANOVA table to make a “Reject H_0 ” or “Fail to Reject H_0 ” decision on the hypothesis (Eq. 5). Perform this step 3000 times (to achieve a good compromise between robustness and computer run time). In Case 1, the proportion of replications in which H_0 is (incorrectly) rejected gives an empirical estimate of the test significance (i.e., Type I error) α . In Cases 2-5, the proportion of replications in which H_0 is (correctly) rejected gives an empirical estimate of the test power $1-\beta$. A nominal α value of 0.05 will be used for the critical region.
3. Repeat Steps 1-2 ten times to obtain ten estimated values for α' or $1-\beta'$. Construct a 95% confidence interval on the true α or $1-\beta$ using these ten values.

Using these confidence intervals, each of the 12 Weibull distributions can be compared to the baseline (normal) results in each of the five major cases to determine whether the Weibull distribution had a significant effect on ANOVA significance or power. Specifically, when H_0 is true (Case 1), the nominal $\alpha = 0.05$ should be contained within the confidence interval if the Weibull distribution had no significant effect on the significance level α of the test. When H_0 is not true (Cases 2-5), the Weibull confidence interval should overlap with the baseline confidence interval if the Weibull data had no significant effect on the power of the test. The second objective of this research is to isolate the effects of the Weibull shape and scale parameters to determine which, if any, of these parameters have a significant effect on ANOVA power and significance. Accordingly, a set of three designed experiments based on the two-way ANOVA method will be conducted corresponding to three of the five major cases from Table 2. Case 1 will be investigated to derive results for test significance. Cases 2 and 4 will be investigated to derive results for test power in the presence of small and large treatment effects, respectively. Each of the three two-way ANOVA experiments will consist of two factors (shape and scale) and the applicable response variable (power or significance). The levels used for the shape and scale values represent common values used in the preceding ANOVA experiments based on the Barringer database (Barringer and Associates, Inc., 2001). Specifically, the 4×3 design shown in Table 4 will be used. Common random numbers will be used across these design combinations. Referencing Table 4, ten independent estimates y_1, y_2, \dots, y_{10} of the mean response value will be generated for each of the 12 design combinations using Monte Carlo simulation.

Table 4
Experimental design for shape and scale parameter analysis

| Scale Parameter | Shape Parameter | | | |
|-----------------|-----------------|----------|----------|----------|
| | 1 | 1.5 | 2 | 2.5 |
| 50 | y_1 | y_1 | y_1 | y_1 |
| | . | . | . | . |
| | . | . | . | . |
| | y_{10} | y_{10} | y_{10} | y_{10} |
| 75 | y_1 | y_1 | y_1 | y_1 |
| | . | . | . | . |
| | . | . | . | . |
| | y_{10} | y_{10} | y_{10} | y_{10} |
| 100 | y_1 | y_1 | y_1 | y_1 |
| | . | . | . | . |
| | . | . | . | . |
| | y_{10} | y_{10} | y_{10} | y_{10} |

Table 5**Experimental Results, Significance and Power**

| Sub - case | Shape Value (γ) | Scale Value (δ) | Significance (α) | | Power (1- β) CASE 2: (-0.5 σ , 0, 0.5 σ) | | Power (1- β) CASE 3: (-0.75 σ , 0, 0.75 σ) | | Power (1- β) CASE 4: (- σ , 0, σ) | | Power (1- β) CASE 5: (-0.5 σ , -0.5 σ , σ) | |
|------------------|--------------------------------|--------------------------------|------------------------------|-------------------------------|--|----------------|---|----------------|---|----------------|--|----------------|
| | | | Lower Limit | Upper Limit | Lower Limit | Upper Limit | Lower Limit | Upper Limit | Lower Limit | Upper Limit | Lower Limit | Upper Limit |
| | | | 1 | Baseline (normal dist.) | | 0.047 | 0.054 | 0.452 | 0.462 | 0.814 | 0.822 | 0.971 |
| 2 | 0.7 | 100 | 0.035 | 0.038 | 0.567 | 0.577 | 0.823 | 0.834 | 0.936 | 0.943 | 0.899 | 0.907 |
| 3 | 1 | 100 | 0.040 | 0.044 | 0.505 | 0.516 | 0.815 | 0.823 | 0.949 | 0.954 | 0.903 | 0.913 |
| 4 | 1.1 | 15 | 0.041 | 0.049 | 0.498 | 0.509 | 0.816 | 0.824 | 0.954 | 0.958 | 0.906 | 0.918 |
| 5 | 1.2 | 50 | 0.044 | 0.049 | 0.482 | 0.496 | 0.813 | 0.826 | 0.958 | 0.963 | 0.911 | 0.920 |
| 6 | 1.3 | 50 | 0.044 | 0.051 | 0.479 | 0.489 | 0.817 | 0.826 | 0.958 | 0.963 | 0.908 | 0.917 |
| 7 | 1.4 | 40 | 0.046 | 0.051 | 0.474 | 0.485 | 0.819 | 0.826 | 0.965 | 0.969 | 0.912 | 0.920 |
| 8 | 1.5 | 100 | 0.045 | 0.051 | 0.463 | 0.477 | 0.813 | 0.822 | 0.967 | 0.970 | 0.912 | 0.920 |
| 9 | 1.7 | 65 | 0.044 | 0.052 | 0.464 | 0.478 | 0.813 | 0.824 | 0.969 | 0.971 | 0.913 | 0.919 |
| 10 | 1.8 | 50 | 0.043 | 0.051 | 0.453 | 0.469 | 0.816 | 0.821 | 0.968 | 0.973 | 0.917 | 0.926 |
| 11 | 2 | 300 | 0.049 | 0.054 | 0.462 | 0.472 | 0.809 | 0.823 | 0.971 | 0.975 | 0.918 | 0.924 |
| 12 | 2.5 | 75 | 0.049 | 0.051 | 0.453 | 0.464 | 0.816 | 0.823 | 0.974 | 0.977 | 0.914 | 0.921 |
| 13 | 3 | 65 | 0.049 | 0.052 | 0.446 | 0.453 | 0.812 | 0.823 | 0.973 | 0.976 | 0.915 | 0.923 |

Each estimate y_i represents the mean power or significance averaged across 3000 independent simulation replications. These values will be used to render judgments upon the significance of the shape and scale parameters as will be discussed in the following subsection. The next subsection will present and discuss the results of the experimentation.

4.2 Results and Discussion

This subsection will present and analyze the results associated with the two research objectives described in Subsection 4.1. Regarding the first objective to investigate the robustness of ANOVA when underlying data follows the Weibull distribution, Table 5 displays the results for each of the five major cases in terms of 95% confidence intervals on the test significance level (α) when the null hypothesis is true (Case 1) and on the test power (1- β) when the null hypothesis is not true (Cases 2-5). Prior to discussing each case individually, we note that, in general, power increases as the treatment effects increase under Weibull data. Larger effects are easier to detect. Specifically, the highest power values are achieved in Case 4 where the effect coefficients are the largest.

In Case 1, all treatment effects are zero and, thus, H_0 should not be rejected. As expected based upon the nominal significance level of $\alpha = 0.05$ utilized, the confidence interval on α for the baseline

(normal) distribution is well centered around the value 0.05, thus serving as a check on the simulation model validity itself. For the Weibull distributions, it is noted that as the shape parameter increases, the corresponding confidence intervals become centered about the nominal $\alpha = 0.05$ value. Confidence intervals for eight of the twelve Weibull distributions contain the nominal significance level (0.05) and, thus, are not significantly affected by non-normality. The remaining four Weibull confidence intervals (bold-faced) did not contain 0.05. These Weibull distributions resulted in a smaller (more conservative) Type I error. Donaldson (1968) made the following surprising remark on similar findings in his prior research: "...if a test is designed with α level of protection against Type I error under the assumption of a normal distribution, even more protection against a Type I error exists if the distribution is of the non-normal type."

In Case 2, confidence intervals for eight Weibull distributions (bold-faced) do not overlap with the baseline confidence interval and, in fact, result in higher power values than the baseline. These eight distributions correspond to the smaller shape parameter values. Confidence intervals for the remaining four Weibull distributions overlap with the baseline interval and, thus, the power differences are insignificant. Moreover, as the shape parameter γ increases to 2.5-3.0, the power decreases and converges to near the baseline values. The effect of the scale parameter is low relative to the effect of the shape parameter. However, these effects will be formally discussed later in this section using designed experiments.

In Case 3, only the Weibull distribution with the smallest shape parameter results in a significantly larger power than the baseline interval. No significant differences in power are observed for the remaining eleven Weibull distributions. Once again, a clear relationship is observed between the shape parameter value and the resultant power.

In Case 4, the case with the highest treatment effect coefficients, the power of seven of the twelve Weibull distributions is significantly less than the baseline distribution. These distributions correspond to the smaller shape parameter settings. As the shape parameter increases, the power increases and approaches that of the baseline value.

In Case 5, the two Weibull distributions with the smallest shape values have much lower power value than the baseline. Again, as the shape parameter increases, the power tends to increase and to level off to near the baseline distribution value.

In summary, the Case 1 results indicate that the Type I error incurred by using the ANOVA method under Weibull data is generally less than or equal to the baseline normal value. In Cases 2 and 3, the power obtained under Weibull data is generally greater than or equal to that of the respective baseline. The most extreme *increase* in power is 0.115, or 25.2%, based on the midpoint of the confidence interval in Case 2 Sub-case 2. In Cases 4 and 5, the power obtained under Weibull data is generally less than or equal to that of the baseline. The most extreme *decrease* in power is just 0.034, or 3.5% less than the baseline, in Case 4 Sub-case 2. Both of these extreme values occur at the smallest shape parameter setting (0.7). The practical implication of these results is that, under the experimental conditions tested, the ANOVA method is reasonably robust to violations of the normality assumption in the presence of Weibull data. An increase in α error was not observed for any of the 12 Weibull distributions studied. Although several Weibull distributions have significantly lower power than the baseline, the maximum power degradation in any case is 3.5%. Most cases had much less than 3.5% power degradation and, thus, although statistically significant, may not be practically significant relative to real-world applications.

Table 6Results of designed experiment # 1 ($\tau_1=0$, $\tau_2=0$, $\tau_3=0$)

| Shape | Scale | α |
|-------|-------|----------|-------|-------|----------|-------|-------|----------|-------|-------|----------|
| 1 | 50 | 0.045 | 1.5 | 50 | 0.046 | 2 | 50 | 0.046 | 2.5 | 50 | 0.055 |
| 1 | 50 | 0.048 | 1.5 | 50 | 0.054 | 2 | 50 | 0.050 | 2.5 | 50 | 0.054 |
| 1 | 50 | 0.039 | 1.5 | 50 | 0.042 | 2 | 50 | 0.050 | 2.5 | 50 | 0.057 |
| 1 | 50 | 0.040 | 1.5 | 50 | 0.043 | 2 | 50 | 0.056 | 2.5 | 50 | 0.048 |
| 1 | 50 | 0.043 | 1.5 | 50 | 0.054 | 2 | 50 | 0.043 | 2.5 | 50 | 0.053 |
| 1 | 50 | 0.043 | 1.5 | 50 | 0.046 | 2 | 50 | 0.050 | 2.5 | 50 | 0.045 |
| 1 | 50 | 0.039 | 1.5 | 50 | 0.047 | 2 | 50 | 0.063 | 2.5 | 50 | 0.048 |
| 1 | 50 | 0.039 | 1.5 | 50 | 0.043 | 2 | 50 | 0.051 | 2.5 | 50 | 0.047 |
| 1 | 50 | 0.045 | 1.5 | 50 | 0.053 | 2 | 50 | 0.048 | 2.5 | 50 | 0.050 |
| 1 | 50 | 0.048 | 1.5 | 50 | 0.045 | 2 | 50 | 0.050 | 2.5 | 50 | 0.052 |
| 1 | 75 | 0.045 | 1.5 | 75 | 0.052 | 2 | 75 | 0.048 | 2.5 | 75 | 0.052 |
| 1 | 75 | 0.045 | 1.5 | 75 | 0.044 | 2 | 75 | 0.048 | 2.5 | 75 | 0.050 |
| 1 | 75 | 0.052 | 1.5 | 75 | 0.048 | 2 | 75 | 0.048 | 2.5 | 75 | 0.050 |
| 1 | 75 | 0.043 | 1.5 | 75 | 0.044 | 2 | 75 | 0.045 | 2.5 | 75 | 0.048 |
| 1 | 75 | 0.042 | 1.5 | 75 | 0.040 | 2 | 75 | 0.050 | 2.5 | 75 | 0.054 |
| 1 | 75 | 0.045 | 1.5 | 75 | 0.039 | 2 | 75 | 0.047 | 2.5 | 75 | 0.052 |
| 1 | 75 | 0.038 | 1.5 | 75 | 0.050 | 2 | 75 | 0.052 | 2.5 | 75 | 0.047 |
| 1 | 75 | 0.054 | 1.5 | 75 | 0.051 | 2 | 75 | 0.045 | 2.5 | 75 | 0.054 |
| 1 | 75 | 0.047 | 1.5 | 75 | 0.053 | 2 | 75 | 0.051 | 2.5 | 75 | 0.052 |
| 1 | 75 | 0.039 | 1.5 | 75 | 0.045 | 2 | 75 | 0.055 | 2.5 | 75 | 0.050 |
| 1 | 100 | 0.044 | 1.5 | 100 | 0.048 | 2 | 100 | 0.050 | 2.5 | 100 | 0.045 |
| 1 | 100 | 0.048 | 1.5 | 100 | 0.049 | 2 | 100 | 0.042 | 2.5 | 100 | 0.052 |
| 1 | 100 | 0.050 | 1.5 | 100 | 0.046 | 2 | 100 | 0.043 | 2.5 | 100 | 0.049 |
| 1 | 100 | 0.037 | 1.5 | 100 | 0.051 | 2 | 100 | 0.052 | 2.5 | 100 | 0.053 |
| 1 | 100 | 0.043 | 1.5 | 100 | 0.045 | 2 | 100 | 0.045 | 2.5 | 100 | 0.049 |
| 1 | 100 | 0.037 | 1.5 | 100 | 0.051 | 2 | 100 | 0.043 | 2.5 | 100 | 0.046 |
| 1 | 100 | 0.046 | 1.5 | 100 | 0.042 | 2 | 100 | 0.045 | 2.5 | 100 | 0.050 |
| 1 | 100 | 0.044 | 1.5 | 100 | 0.052 | 2 | 100 | 0.050 | 2.5 | 100 | 0.052 |
| 1 | 100 | 0.040 | 1.5 | 100 | 0.050 | 2 | 100 | 0.048 | 2.5 | 100 | 0.045 |
| 1 | 100 | 0.038 | 1.5 | 100 | 0.053 | 2 | 100 | 0.050 | 2.5 | 100 | 0.041 |

Table 7Results of Designed Experiment # 2 ($\tau_1=-0.5\sigma$, $\tau_2=0$, $\tau_3=0.5\sigma$)

| shape | scale | power |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 50 | 0.51 | 1.5 | 50 | 0.47 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.50 | 1.5 | 50 | 0.48 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.51 | 1.5 | 50 | 0.48 | 2 | 50 | 0.46 | 2.5 | 50 | 0.45 |
| 1 | 50 | 0.51 | 1.5 | 50 | 0.48 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.52 | 1.5 | 50 | 0.47 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.53 | 1.5 | 50 | 0.48 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.52 | 1.5 | 50 | 0.47 | 2 | 50 | 0.44 | 2.5 | 50 | 0.47 |
| 1 | 50 | 0.51 | 1.5 | 50 | 0.49 | 2 | 50 | 0.46 | 2.5 | 50 | 0.46 |
| 1 | 50 | 0.52 | 1.5 | 50 | 0.46 | 2 | 50 | 0.46 | 2.5 | 50 | 0.48 |
| 1 | 50 | 0.50 | 1.5 | 50 | 0.47 | 2 | 50 | 0.47 | 2.5 | 50 | 0.45 |
| 1 | 75 | 0.52 | 1.5 | 75 | 0.47 | 2 | 75 | 0.46 | 2.5 | 75 | 0.46 |
| 1 | 75 | 0.52 | 1.5 | 75 | 0.47 | 2 | 75 | 0.47 | 2.5 | 75 | 0.44 |
| 1 | 75 | 0.51 | 1.5 | 75 | 0.49 | 2 | 75 | 0.47 | 2.5 | 75 | 0.45 |
| 1 | 75 | 0.51 | 1.5 | 75 | 0.46 | 2 | 75 | 0.44 | 2.5 | 75 | 0.45 |
| 1 | 75 | 0.52 | 1.5 | 75 | 0.47 | 2 | 75 | 0.44 | 2.5 | 75 | 0.45 |
| 1 | 75 | 0.51 | 1.5 | 75 | 0.46 | 2 | 75 | 0.45 | 2.5 | 75 | 0.44 |
| 1 | 75 | 0.50 | 1.5 | 75 | 0.48 | 2 | 75 | 0.45 | 2.5 | 75 | 0.45 |
| 1 | 75 | 0.51 | 1.5 | 75 | 0.47 | 2 | 75 | 0.47 | 2.5 | 75 | 0.46 |
| 1 | 75 | 0.50 | 1.5 | 75 | 0.47 | 2 | 75 | 0.46 | 2.5 | 75 | 0.45 |
| 1 | 75 | 0.51 | 1.5 | 75 | 0.47 | 2 | 75 | 0.48 | 2.5 | 75 | 0.45 |
| 1 | 100 | 0.50 | 1.5 | 100 | 0.47 | 2 | 100 | 0.45 | 2.5 | 100 | 0.45 |
| 1 | 100 | 0.50 | 1.5 | 100 | 0.47 | 2 | 100 | 0.47 | 2.5 | 100 | 0.45 |
| 1 | 100 | 0.52 | 1.5 | 100 | 0.48 | 2 | 100 | 0.46 | 2.5 | 100 | 0.45 |
| 1 | 100 | 0.51 | 1.5 | 100 | 0.48 | 2 | 100 | 0.46 | 2.5 | 100 | 0.47 |
| 1 | 100 | 0.51 | 1.5 | 100 | 0.48 | 2 | 100 | 0.45 | 2.5 | 100 | 0.46 |
| 1 | 100 | 0.53 | 1.5 | 100 | 0.46 | 2 | 100 | 0.46 | 2.5 | 100 | 0.46 |
| 1 | 100 | 0.50 | 1.5 | 100 | 0.47 | 2 | 100 | 0.44 | 2.5 | 100 | 0.44 |
| 1 | 100 | 0.51 | 1.5 | 100 | 0.48 | 2 | 100 | 0.47 | 2.5 | 100 | 0.47 |
| 1 | 100 | 0.53 | 1.5 | 100 | 0.48 | 2 | 100 | 0.45 | 2.5 | 100 | 0.46 |
| 1 | 100 | 0.52 | 1.5 | 100 | 0.48 | 2 | 100 | 0.47 | 2.5 | 100 | 0.44 |

Table 8Results of Designed Experiment # 3 ($\tau_1=-\sigma$, $\tau_2=0$, $\tau_3=\sigma$)

| shape | scale | power |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 50 | 0.95 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.97 |
| 1 | 50 | 0.95 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.98 |
| 1 | 50 | 0.96 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.97 |
| 1 | 50 | 0.95 | 1.5 | 50 | 0.96 | 2 | 50 | 0.98 | 2.5 | 50 | 0.97 |
| 1 | 50 | 0.94 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.97 |
| 1 | 50 | 0.96 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.98 |
| 1 | 50 | 0.95 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.98 |
| 1 | 50 | 0.96 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.98 |
| 1 | 50 | 0.95 | 1.5 | 50 | 0.97 | 2 | 50 | 0.97 | 2.5 | 50 | 0.98 |
| 1 | 50 | 0.96 | 1.5 | 50 | 0.97 | 2 | 50 | 0.98 | 2.5 | 50 | 0.97 |
| 1 | 75 | 0.95 | 1.5 | 75 | 0.96 | 2 | 75 | 0.98 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.96 | 2 | 75 | 0.98 | 2.5 | 75 | 0.97 |
| 1 | 75 | 0.95 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.95 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.97 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.96 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.97 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.97 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 75 | 0.96 | 1.5 | 75 | 0.96 | 2 | 75 | 0.97 | 2.5 | 75 | 0.98 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.96 | 2 | 100 | 0.98 | 2.5 | 100 | 0.98 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.98 |
| 1 | 100 | 0.96 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.97 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.97 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.96 | 2 | 100 | 0.97 | 2.5 | 100 | 0.98 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.98 |
| 1 | 100 | 0.96 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.97 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.97 |
| 1 | 100 | 0.95 | 1.5 | 100 | 0.96 | 2 | 100 | 0.97 | 2.5 | 100 | 0.98 |
| 1 | 100 | 0.96 | 1.5 | 100 | 0.97 | 2 | 100 | 0.97 | 2.5 | 100 | 0.97 |

Table 9
Summary results for shape and scale parameter significance

| Shape (γ) | Scale (δ) | Experiment # 1 Significance (α) Results (0, 0, 0) | Experiment # 2 Power (1- β) Results (-0.5 σ , 0, 0.5 σ) | Experiment # 3 Power (1- β) Results (- σ , 0, σ) |
|--------------------------------------|-----------------------|--|---|--|
| 1 | 50 | 0.043 | 0.513 | 0.953 |
| 1 | 75 | 0.045 | 0.511 | 0.957 |
| 1 | 100 | 0.043 | 0.513 | 0.953 |
| 1.5 | 50 | 0.047 | 0.475 | 0.969 |
| 1.5 | 75 | 0.047 | 0.471 | 0.966 |
| 1.5 | 100 | 0.049 | 0.475 | 0.967 |
| 2 | 50 | 0.051 | 0.459 | 0.972 |
| 2 | 75 | 0.049 | 0.459 | 0.972 |
| 2 | 100 | 0.047 | 0.458 | 0.971 |
| 2.5 | 50 | 0.051 | 0.461 | 0.975 |
| 2.5 | 75 | 0.051 | 0.450 | 0.977 |
| 2.5 | 100 | 0.048 | 0.455 | 0.975 |
| p-value for shape parameter | | 0.000 | 0.000 | 0.000 |
| p-value for scale parameter | | 0.271 | 0.097 | 0.065 |
| p-value for shape, scale interaction | | 0.254 | 0.688 | 0.086 |

With regards to the second research objective to isolate the effects of the Weibull shape and scale parameters, Tables 6 through 8 show detailed results for each of the three two-way ANOVA experiments. The response variable in Table 6 is test significance (α), and the response in Tables 7 and 8 is power (1- β). Table 9 summarizes the results and associated p-values. The inner cells of Table 9 display response values averaged across the ten sets of 3000 independent replications (i.e., average of the ten y_i values). The bottom rows show p-values for the shape, scale and shape/scale interaction effects in each of the three experiments. The shape parameter effect is significant in Experiment 1 involving test significance and in Experiments 2 and 3 involving test power under low and high treatment effects, respectively. The scale parameter effect and shape/scale interaction effect are not significant with respect to significance or power at the target $\alpha = 0.05$ level in any of the three experiments. However, as indicated by decreasing p-values from Case 1 to Case 2 to Case 3, the scale parameter becomes significant with respect to power (Experiments 2 and 3) when the target α is raised to 0.10. The realized significance level (p-value) of the shape/scale interactive effect with respect to power (Experiments 2 and 3) appears to be very sensitive to the magnitude of the effect as evidenced by the change in p-value from 0.688 to 0.086. Thus, a more detailed investigation of the scale parameter and interactive effect at more extreme settings than those found in the Barringer database (Barringer & Associates, Inc., 2001) may be warranted in future research.

5. Conclusions and Future Directions

The results of this research indicate that the one-way fixed effects ANOVA method generally performs robustly under the conditions examined whenever the normality assumption is violated in light of Weibull data. The Weibull shape parameter is most often significant; the scale parameter and shape/scale interactive effect can become significant at higher target significance levels (e.g., $\alpha \geq 0.10$). However, this study considers only a finite number of Weibull distributions and, thus, the

conclusions can only be applied to those specific parameter settings. Additional research may be conducted to study more extreme parameter values than those found in the Barringer database of real-world systems and components (Barringer and Associates, Inc., 2001). Another interesting research direction would be to study the case in which *individual* treatment data follow different Weibull distributions. This case would be somewhat comparable to the situation where the ANOVA assumption of constant variance has been violated, since the variance of Weibull populations is a function of the respective Weibull parameters.

References

- Barringer and Associates, Inc. (2001). Weibull Reliability Database for failure data for various components. *Computer database* (<http://www.barringer1.com/wdbase.htm>).
- Brown, M. B., & Forsythe, A. B. (1974). The small sample behavior of some statistics which test the equality of several means. *Technometrics*, 16, 385-389.
- David, F.N., & N.L. Johnson. (1951). The effect of Non-normality on the Power Function of the F-Test the Analysis of Variance. *Biometrika*, 38, 43-57.
- Donaldson, T. S. (1968). Robustness of the F-test to Errors of both kinds and the Correlation Between the Numerator and the Denominator of the F-Ratio. *Journal of American Statistical Association* (June), 660-676.
- Driscoll, W. C. (1990). Bootstrapping: An Alternative to ANOVA. *Computers & Industrial Engineering* 19, 562-566.
- Driscoll, W. C. (1996). Robustness of the ANOVA and Tukey-Kramer statistical tests. *Computers & Industrial Engineering*, 31, 265-268.
- Games, P. A., & Lucas, P.A. (1966). Power of the Analysis of Variance of Independent Groups and Non-normal and Normally Transformed Data. *Educational and Psychological Measurement*, 26(2), 311-327.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42 (3), 237-288.
- Harwell, M., Rubinstein, E., Hayes, W., & Olds, C. (1992). Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics*, 17(4), 315-339.
- James, G.S. (1951). The Comparison of Several Groups of Observations When the Ratios of the Population Variances are Unknown. *Biometrika*, 38, 324-329.
- Kruskal, W. H. & Wallis, W. A. (1952) Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583-621.
- Li Li (2007). *Monte Carlo Methods for Modified ANOVA*. Thesis. Kaiserslautern University of Technology, Germany.
- Lix, Lisa M., Keselman, J. C. & Keselman, H.J. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F-Test. *Review of Educational Research* (Winter), 579-620.
- Mendes, M. (2007). The Effects of Non-Normality on Type III Error for Comparing Independent Means. *Journal of Applied Quantitative Methods*, 2(4), 444-454.
- Montgomery, D. C. (2001). *Design and Analysis of Experiments* (5th Edition). John Wiley and Sons, Inc, New York, NY. 684pp.
- Montgomery, D. & George Runger (2010). *Applied Statistics and Probability for Engineers* (5th Edition), John Wiley and Sons, Inc, New York, NY. 784pp.
- Srivastava, A. B. L. (1959). Effect of Non-normality on the Power Function of the Analysis of Variance. *Biometrika* 46, 114-122.
- Smith, J.R. (1966). *The Robustness to Non-Normality of the Size and Power of the F-Test in the One-way Fixed Effects Analysis of Variance*. M.S. Thesis. Virginia Polytechnic Institute.
- Welch, B. L. (1951). On the comparison of several mean values: An alternative approach. *Biometrika*, 38, 330-336.