Identification of dispersion effects in replicated two-level fractional factorial experiments

Cheryl Dingus[1], Bruce Ankenman[2], Angela Dean[3,4] and Fangfang Sun[4]

[1]Battelle Memorial Institute

[2]Department of Industrial Engineering, Northwestern University

[3]Department of Mathematics, University of Southampton

[4]Department of Statistics, The Ohio State University

Abstract

Tests for dispersion effects in replicated two-level factorial experiments assuming a location-dispersion model are presented. The tests use individual measures of dispersion which remove the location effects and also provide an estimate of pure error. Empirical critical values for two such tests are given for two-level full or regular fractional factorial designs with 8, 16, 32 and 64 runs. The powers of the tests are examined under normal, exponential, and Cauchy distributed errors. Our recommended test uses dispersion measures calculated as deviations of the data values from their cell medians, and this test is illustrated via an example.

Keywords: Dispersion measures; location-dispersion model; fractional factorial experiment

1 Introduction

An important aspect of designing quality into a product or process is the identification of those factors that contribute most to the mean and the variability of a measured response. If the average response differs substantially as the level of a factor is changed from one level to another, then this factor is said to have a *location main effect*. Similarly, if the variability of the response differs substantially as the factor is changed from one level to another, then the factor is said to have a *dispersion main effect*. If the size of a location or dispersion main effect of one factor differs as the level of a second factor changes then the two factors are said to have a two-factor *location* or *dispersion interaction*. Higher order interactions can be defined in a similar way.

A factor has a *location (dispersion) effect* if it has either a location (dispersion) main effect or is involved in a location (dispersion) interaction or both. Factors can be classified as having (i) a location effect only, (ii) a dispersion effect only, (iii) both a location and a dispersion effect, (iv) neither. If factors can be classified correctly, then the quality of a product or process can be improved by setting the levels of factors in group (ii) and possibly some of those in group (iii) to the level combination that gives rise to minimum response variability, and then setting the remaining factors at the level combination that results in a mean response close to a required target value or optimum value (see, for example, Box 1988; Pignatiello and Ramberg 1987; Ankenman and Dean 2003).

Location effects have been discussed for many years, and tests for their detection in both replicated and unreplicated experiments can be found in text books on experimental designs (for example, Box, Hunter, and Hunter 1978; Dean and Voss 1999; Montgomery 2009). Detection of dispersion effects in unreplicated experiments is difficult in the presence of location effects. A number of methods have been proposed (see, for example, Box and Meyer 1986; Bergman and Hynén 1997; Wang 1989; Brenneman and Nair 2001), but incorrect specification of the location model can lead to incorrect conclusions about dispersion effects and detailed discussions of such problems are given by Brenneman and Nair (2001), Pan (1999), McGrath and Lin (2001a,b) and Bursztyn and Steinberg (2006).

In this paper, we discuss detection of dispersion effects in *replicated* experiments. One standard methodology combines the observations in each cell (factor combination) into a single value such as sample variance or standard deviation (see, for example, Nair and Pregibon 1988), in which case the methodology is similar to that for unreplicated experiments and there is no opportunity to obtain a pure estimate of error variance. In such a case, a non-saturated model is often fitted and degrees of freedom from the assumed negligible interactions are used to obtain an estimate of error. A test based on this strategy requires an accurate specification of the dispersion model as well as effect sparsity. Alternatively, the method of Lenth (1989) (see, also, Haaland and O'Connell 1995) and the subsequent variations (see Miller 2005) that allow for estimation of error in unreplicated fractional factorials can be used, but again effect sparsity must be assumed.

For replicated experiments, Nair and Pregibon (1988) and Pan (1999) studied extensions of the methods of Bartlett and Kendall (1946), Box and Meyer (1986) and Bergman and Hynén (1997), but these are all still effectively based on a single dispersion measure in each cell. Mackertich, Benneyan, and Kraus (2003) suggested an alternative approach which transforms each observation in such a way as to provide an *individual* measure of dispersion which removes the location effects. Similar to test statistics for testing homogeneity of population variances (such as those of Levene 1960; Brown and Forsythe 1974), the measures of Mackertich et al. (2003) were based on deviations of data values from the corresponding cell means and then raised to a power to obtain approximate normality of the measure. Analysis of variance test statistics were used together with critical values obtained either from the F distribution or from the empirical distribution of the test statistic under assumed location and dispersion models. Based on a few selected location and dispersion models, these authors found that their proposed tests, based on individual dispersion measures, have increased power over tests that use a single combined measure.

In this paper, we investigate modified versions of the measures suggested by Mackertich et al. (2003) and, like these authors, we work with test statistics which are similar to those arising from the analysis of variance. However, our model is the location-dispersion model (see Section 2.1) rather than the usual analysis of variance model of Mackertich et al. (2003). Also, unlike the simulation study of Mackertich et al. (2003), we do not make assumptions about which factorial effects are in the true location or dispersion model but instead standardize our test statistic using an estimate of pure error obtained from the within cell replicates. Rather than attempting to transform our measures to approximate normality, we obtain critical values from the empirical distributions of our test statistics under a null model. We show, in Section 3.2, that the average empirical Type I errors for our recommended test hold reasonably close to the nominal levels when simulated under randomly selected location-dispersion models.

Our selection of dispersion measures, our model and dispersion test statistics are described in Sections 2.2 and 2.3. The critical values are given in Section 2.4 for experiments with $r \geq 3$ observations on each of v factor combinations in a full or regular fractional factorial design. In Sections 3.1 and 3.2, the levels and powers of the tests under normal, exponential, and Cauchy error distributions are examined in simulations in which the true location and dispersion models are randomly generated. These results are presented for a 2^{5-1} fraction with $r = 4$ observations per cell for both first-order and second-order models. Our tests based on the natural logarithm of the absolute deviation of the data values from either the mean or the median are shown to have high power for detecting a single dispersion effect under a normal error distribution and across a wide range of true model and effect sizes. Under non-normal error distributions, our test based on the deviation from the median maintains a Type I error rate close to the nominal level with high power (see Section 3.2) and this is our recommended test for dispersion effects. In Section 4, we illustrate our test via data from Pignatiello and Ramberg (1985) as used by Nair and Pregibon (1988).

2 Tests for dispersion

2.1 Location-Dispersion Model

For a two-level factorial experiment with f factors, we assume the following location-dispersion linear model.

$$Y_{ij} \;=\; \mu_i + \sigma_i \epsilon_{ij} \qquad \text{with} \quad \mu_i = \boldsymbol{x}'_{\mu,i}\boldsymbol{\beta}, \quad \sigma_i \;=\; exp(\boldsymbol{x}'_{\sigma,i}\boldsymbol{\gamma})\,, \tag{2.1}$$

where Y_{ij} is the response for the jth observation at factor combination $i = i_1 i_2 \ldots i_f$, where i_h is the level of factor h ($i_h \in \{1, 2\}$; $h = 1, \ldots, f$; $j = 1, \ldots, r$) and where $\boldsymbol{x}'_{\mu,i}$ and $\boldsymbol{x}'_{\sigma,i}$ are the rows of the model matrices $\boldsymbol{X}_\mu$ and $\boldsymbol{X}_\sigma$ corresponding to factor level combination $i = i_1 i_2 \ldots i_f$ in the design ($i = 1, \ldots, v$). The model matrices $\boldsymbol{X}_\mu$ and $\boldsymbol{X}_\sigma$ each contain columns for the mean, main effects and interactions to be included in the location and dispersion models, respectively, and may or may not be identical. The error variables, ϵ_{ij}, are assumed to be independent and identically distributed. The vectors $\boldsymbol{\beta} = [\beta_0, \beta_1, \ldots, \beta_{v-1}]$ and $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_{v-1}]$ are, respectively, the parameter vectors for the location and dispersion effects (means, main effects, and interactions) that can be supported by the design.

One significant advantage of the multiplicative variance model (2.1) over an additive model is that the standard deviations σ_i are necessarily positive. The bounded nature of the likelihood function using (2.1), the simpler form of the likelihood ratio test for the multiplicative model, and the consistency of the dispersion effect estimators were cited by Harvey (1976) as reasons why this multiplicative variance model is attractive. The multiplicative variance model has also been supported by Cook and Weisberg (1983), Aitkin (1987), Nair and Pregibon (1988), Verbyla (1993), and Wolfinger and Tobias (1998), among others.

2.2 Choice of dispersion measures

Our goal is to find a dispersion measure that gives rise to powerful tests for identifying dispersion main effects and interactions for a wide variety of location and dispersion models and also that are robust to non-normal errors. Following Mackertich et al. (2003), Levene (1960), and Brown and Forsythe (1974), we transform every observation y_{ij} in cell $i = (i_i i_2 \ldots i_f)$ to an individual dispersion measure m_{ij}. For example, Mackertich et al. (2003) used $m_{ij} = |y_{ij} - \overline{y}_{i.}|^p$, for various values of p. Extending this idea, Dingus (2005) presented an initial study of thirty-seven different dispersion measures. She obtained empirical critical values for tests for detecting dispersion effects based on each of these measures and examined the powers of the tests under a large range of randomly selected models. The 37 dispersion measures of Dingus (2005) included the traditional measures,

s_i, s_i^2, and $ln(s_i+1)$ as well as the measures $|y_{ij}-\bar{y}_i|$, $ln(|y_{ij}-\bar{y}_i|)$, $|y_{ij}-\bar{y}_i|^{0.42}$ and $|y_{ij}-\bar{y}_i|^{1.5}$, examined by Mackertich et al. (2003). The absolute deviations of the data from the within-cell means, $|y_{ij}-\overline{y}_i|^p$, are functions of the least squares residuals arising from a saturated location model with additive error. The studies of Mackertich et al. (2003) and Dingus (2005) showed that, for an additive-error model, the tests based on these measures with exponent $p=1$ have the most stable Type I errors and are the most powerful.

The means $\overline{y}_i$ in the above measures can be replaced by different estimates of central tendency, and Dingus (2005) recommended the use of the median. Test statistics and critical values using the natural logarithm of the absolute deviations of the data values from the cell medians, $ln(|y_{ij}-\widetilde{y}_i|)$, are not available when r is odd due to the fact that one value of $|y_{ij}-\widetilde{y}_i|$, $j=1,\ldots,r$, is always zero. In this paper, we avoid this problem by adding 1.0 to the absolute deviation, and we also exclude the minimum absolute deviation value, which is zero for r odd, and a duplicate value for r even.

As pointed out in Section 2.1, there are significant advantages to using the multiplicative-variance model (2.1). Consequently, in preliminary research for this paper, we investigated the performances of the dispersion measures mentioned above, using data generated from the multiplicative model as described in Section 3.1. The two measures that lead to the most powerful tests are discussed below and compared with a traditional measure. Specifically, the three measures of dispersion discussed in this paper are:

$$m_{ij}^{(1)}=\ln(|y_{ij}-\widetilde{y}_i|_{-1}+1), \quad m_{ij}^{(2)}=ln(|y_{ij}-\bar{y}_i|+1)\ , \quad m_i^{(3)}=ln(s_i+1), \quad (2.2)$$

where $\widetilde{y}_i$ and $\overline{y}_i$, denote, respectively, the median and the mean of the r observations on factor combination $i=i_1,\ldots,i_f$, and $|\cdot|_{-1}$ denotes that (one of) the smallest value(s) is not included in the calculation of the test statistic. In Section 2.3, we give the form of the test statistics that we use in conjunction with these three measures to test for a single dispersion effect, and in Section 2.4 we obtain the critical values for the tests.

2.3 Test statistics

For a full or regular fractional factorial design with v distinct factor combinations, it is possible to estimate $v-1$ dispersion main effects and interactions independently. We write

$$\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_f, \gamma_{f+1}, \ldots, \gamma_{v-1}]$$

where γ_0 is the dispersion mean, $\gamma_1, \ldots, \gamma_f$ are the dispersion main effect parameters for factors $1, \ldots, f$, and $\gamma_{f+1}, \ldots, \gamma_{v-1}$ are a set of parameters for the $v-f-1$ dispersion interaction effects that can be independently estimated in the design (i.e. a saturated model).

When all factors have two levels, each main effect and interaction effect can be measured by a single contrast. This allows our discussion, without loss of generality, to be in terms of testing a non-specific null hypothesis

$$\begin{aligned} &H_0^{\gamma_t}: \ \gamma_t \ = \ 0; \text{ all } \gamma_q \ (q \neq t) \text{ unrestricted, all } \beta_0, \ldots, \beta_{v-1} \text{ unrestricted,} \\ &\text{versus} \\ &H_1^{\gamma_t}: \ \text{ all } \gamma_q \text{ and } \beta_q \text{ unrestricted,} \quad q = 0, \ldots, v-1\,, \end{aligned} \tag{2.3}$$

for any $t \in \{1, \ldots, v-1\}$.

Our test statistic M_t for testing $H_0^{\gamma_t}$ against $H_1^{\gamma_t}$ using dispersion measures $m_{ij}^{(1)}$ or $m_{ij}^{(2)}$ in (2.2) is of the same form as that used for a standard partial F test and can be calculated by statistical software packages if a saturated dispersion model is fitted. For two-level factors, this can be written as

$$M_t \ = \ \frac{(\overline{m}_+ - \overline{m}_-)^2 \, vr^* \ / \ 4}{\Sigma_{i=1}^{v} \Sigma_{j=1}^{r} \left(m_{ij} - \overline{m}_{i.}\right)^2 \ / \ v(r^*-1)}\,, \tag{2.4}$$

where m_{ij} is the dispersion measure for the jth observation on factor combination $i = (i_1 i_2 \ldots i_f)$, $\overline{m}_{i.}$ is the average of the r^* dispersion measures m_{ij} in cell i, where $r^* = r$ for $m_{ij}^{(2)}$ and $m_i^{(3)}$, and $r^* = r-1$ for $m_{ij}^{(1)}$ (omitting the smallest $m_{ij}^{(1)}$), and r is the number of observations per cell. Also, $\overline{m}_+$ and $\overline{m}_-$ are the averages of the $v/2$ values $\overline{m}_{i.}$ whose factor combination i enters into contrast t with contrast coefficient $+1$ and

-1, respectively. Critical values for testing $H_0^{\gamma_t}$ (2.3) are discussed in Section 2.4. A traditional test for dispersion is based on the natural log of the variance or standard deviation s_i of the observations for factor combination $i = (i_1 \ldots i_f)$, thereby reducing all observations within cell i to a single value (see Bursztyn and Steinberg 2006, for an excellent review). If a saturated model is fitted, no degrees for freedom are then available for estimating σ^2 but, under effect sparsity, the method of Lenth (1989) can be used for testing $H_0^{\gamma_t}$ in (2.3) as follows. Let $\hat{\gamma}_t$ be the difference between the average values of $m_i^{(3)} = ln(s_i + 1)$ for all cells $i = 1, \ldots, v$ corresponding to coefficient $+1$ and coefficient -1 in contrast t $(t = 1, \ldots, v-1)$. Then, following Lenth (1989), we use the following test statistic for testing $H_0^{\gamma_t}$ in (2.3):

$$M_{PSE,t} = \frac{|\hat{\gamma}_t|}{PSE} \tag{2.5}$$

where

$$PSE = 1.5 \times \underset{|\hat{\gamma}_j|<2.5s_0}{median}|\hat{\gamma}_j| \quad \text{and} \quad s_0 = 1.5 \times \underset{j=1,\ldots v-1}{median}|\hat{\gamma}_j|\,.$$

Critical values for testing $H_0^{\gamma_t}$ using this test statistic are obtained in the next section.

2.4 Critical values

We now obtain the empirical distributions of the test statistics (2.4) using $m_{ij}^{(1)}$ and $m_{ij}^{(2)}$, and of (2.5) using $m_i^{(3)} = ln(s_i+1)$. Although the test statistics M_t have similar forms to analysis of variance test statistics, they do not have F distributions even when the error variables ϵ_{ij} follow a normal distribution, nor do test statistics $M_{PSE,t}$ have t distributions. Thus, it is necessary to obtain critical values for testing $H_0^{\gamma_t}$ from their empirical distributions. There are many possible error distributions and an infinite number of values for each unrestricted parameter under $H_0^{\gamma_t}$ (2.3) and, therefore, we obtain the critical values under the null model assuming normally distributed errors; that is, under the more restrictive null hypothesis

$$H_0^*: \quad \text{all } \gamma_q = 0 \text{ and } \beta_q = 0, \quad q = 1, \ldots, v-1\,, \tag{2.6}$$

(c.f. Nair and Pregibon 1988; Wolfe, Dean, Wiers, and Hartlaub 1992). Using the critical values so obtained, in Section 3.2 we simulate the true average levels of the tests of the hypothesis $H_0^{\gamma_t}$ (2.3) of interest under randomly generated location-dispersion models and normal, exponential and Cauchy error distributions.

The empirical critical values for testing H_0^* (2.6) against the general alternative hypothesis were obtained by simulation as follows. For each $i = 1, \ldots, v$ and each $r = 3, 4, \ldots, 10$, we simulated r data values from a $N(0, 1)$ distribution and calculated the dispersion measures $m_{ij}^{(k)}$, $j = 1, \ldots, r$, $k = 1, 2, 3$. Without loss of generality, a contrast with first $v/2$ coefficients equal to -1 and the remainder equal to $+1$ was taken for γ_t. Then the test statistic M_t (2.4) or $M_{PSE,t}$ (2.5) was calculated depending on the measure used. This simulation was done 2,500,000 times and the critical values were obtained as percentiles of the empirical distributions of M_t and $M_{PSE,t}$ for 8–, 16–, 32–, and 64–run designs. These values are shown in Tables 4–6 in Appendix A. It can be seen that the empirical critical values for $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ and $m_{ij}^{(2)} = ln(|y_{ij} - \bar{y}_i| + 1)$ using (2.4) are generally larger than the corresponding percentiles of an F-distribution, and those for $m_i^{(3)} = ln(s_i + 1)$ using (2.5) are greater than the corresponding percentiles of a t distribution.

The empirical critical values for all three tests decrease as v increases, as one would expect. Similarly, the critical values for the test based on $m_{ij}^{(2)}$ decrease as r increases, and those based on $m_i^{(3)}$ remain relatively constant. The behavior of the critical values for $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ sometimes behave differently for r even and r odd due to the difference in the calculation of the cell median. There also appears to be some interaction between v and r as to the size of the critical values. In the next section, it shown that these empirical critical values lead to tests which hold their Type I error levels and have good power characteristics in the numerical studies.

3 Type I errors and powers

We now check, via simulation, whether the Type I errors remain at their nominal values when the tests are conducted under the null hypothesis $H_0^{\gamma_t}$ (2.3) of interest rather than H_0^* (2.6). In addition, we examine the power of our proposed tests and the effect of non-normal error distributions.

3.1 Simulation studies

For the simulations, we took a 2_V^{5-1} fractional factorial design with 16 observations and defining relation $I = ABCDE$. This allows a second-order model to be fitted with all main effects and two-factor interactions. Separate sets of simulations were conducted for first-order dispersion models (i.e. where $\boldsymbol{X}_\sigma$ in (2.1) contains only the six columns for the mean and the dispersion main effects, while all interaction parameters are set to zero) and second-order dispersion models (i.e. where $\boldsymbol{X}_\sigma$ contains 16 columns one for each main effect, two-factor interaction, and the mean). The error variables ϵ_{ij} and the location means μ_i were generated from independent Normal distributions with mean 0 and standard deviations 1 and 3, respectively. (Thus the location means were generated directly rather than through the location main effect and interaction parameters $\boldsymbol{X}_\mu\boldsymbol{\beta}$). The value of dispersion parameter γ_1 in vector $\boldsymbol{\gamma}$ was incremented from $\gamma_1 = 0$ to $\gamma_1 = (1/2)ln5$ in fifty steps. The value $(1/2)ln5$ translates to a ratio of 5 in the standard deviations at two different levels of a factor in the model (2.1). The values of the other parameters γ_q in vector $\boldsymbol{\gamma}$ were generated from a Normal distribution with mean 0 and standard deviation $(1/2)ln2$ (corresponding to a standard deviation ratio equal to 2) with probability 0.4, and γ_q was set to zero with probability 0.6. This mimics a situation of effect sparsity, so that the test using (2.5) can be done.

The simulation was repeated 100,000 times for each value of γ_1 and a new location-dispersion model was generated for each data set. For each data set, r independent observations were generated per factor level combination (cell), according to model (2.1). The test statistic was calculated and compared against the appropriate critical value in Tables 4–6 corresponding to a nominal error rate of $\alpha = 0.05$. The proportion of times that the null hypothesis $H_0^{\gamma_t}$ (2.3) was rejected for each value of γ_1 gives the empirical

Table 1: Average empirical Type I errors for testing $H_0^{\gamma_1}$ (2.3) for the tests based on (2.4) and (2.5) at nominal level $\alpha = 0.05$ and using critical values in Tables 4–6 and a 2_V^{5-1} fraction. Values of $\gamma_q,\ q \neq 1$, and $\mu_i,\ i = 1, \dots, v-1$, generated as in Section 3.1 with $100,000$ simulations

Normally distributed errors		First order models			Second order models		
		Number of observations per cell (r)					
Measure	Test	3	5	10	3	5	10
$\ln(\|y_{ij} - \widetilde{y}_i\|_{-1} + 1)$	M_t, (2.4)	0.0525	0.0503	0.0514	0.0713	0.0790	0.1083
$ln(\|y_{ij} - \bar{y}_i\| + 1)$	M_t, (2.4)	0.0514	0.0512	0.0501	0.0677	0.0802	0.1103
$ln(s_i + 1)$	$M_{PSE,t}$, (2.5)	0.0347	0.0311	0.0292	0.0215	0.0213	0.0258
Exponentially distributed errors		**First order models**			**Second order models**		
		Number of observations per cell (r)					
Measure	Test	3	5	10	3	5	10
$\ln(\|y_{ij} - \widetilde{y}_i\|_{-1} + 1)$	M_t, (2.4)	0.0589	0.0693	0.0773	0.0756	0.0944	0.1250
$ln(\|y_{ij} - \bar{y}_i\| + 1)$	M_t, (2.4)	0.1835	0.1720	0.1756	0.1981	0.2017	0.2261
$ln(s_i + 1)$	$M_{PSE,t}$, (2.5)	0.0365	0.0338	0.0312	0.0228	0.0213	0.0198
Cauchy distributed errors		**First order models**			**Second order models**		
		Number of observations per cell (r)					
Measure	Test	3	5	10	3	5	10
$\ln(\|y_{ij} - \widetilde{y}_i\|_{-1} + 1)$	M_t, (2.4)	0.1093	0.0822	0.0649	0.1137	0.0877	0.0764
$ln(\|y_{ij} - \bar{y}_i\| + 1)$	M_t, (2.4)	0.4748	0.4952	0.5635	0.4642	0.4954	0.5614
$ln(s_i + 1)$	$M_{PSE,t}$, (2.5)	0.0381	0.0371	0.0368	0.0233	0.0234	0.0239

Type I error (for $\gamma_1 = 0$) and the power of the test (for $\gamma_1 > 0$).

3.2 Type I errors and powers for testing $H_0^{\gamma_t}$ (2.3)

Table 1 shows that, for first order models and normally distributed errors, the two measures $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ and $m_{ij}^{(2)} = ln(|y_{ij} - \bar{y}_i| + 1)$, with test statistic (2.4) for testing $H_0^{\gamma_t}$ (2.3), hold the nominal 0.05 level reasonably well. For second order models, the empirical Type I error is a little inflated and reaches 0.1 for $r = 10$ observations per cell. On the other hand, the Type I error for test (2.5) using measure $m_i^{(3)} = ln(s_i + 1)$ is reduced to 0.02–0.03, and it will be seen below that the power of the test is also depressed.

For both first-order and second-order models under normally distributed error variables, the power of all three tests increases as the number of replicates increases, but at different rates (see Figures 1–4). It is clear from these figures that the test (2.4) based on

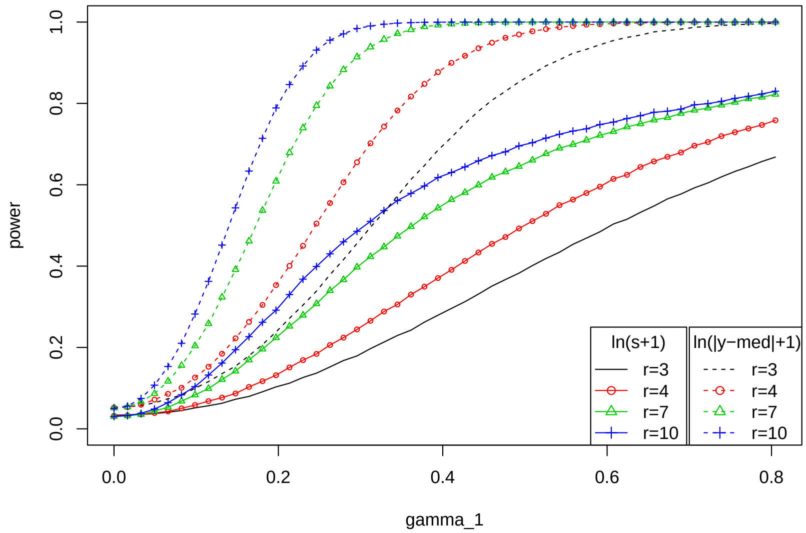

Figure 1: Power curves for tests based on $ln(s+1)$ and $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ using empirical critical values from Tables 4 and 6, with data from randomly generated location models and *first-order* dispersion models, with $r = 3, 4, 7, 10$ replicates per cell, and normal error distribution

either $m_{ij}^{(1)}$ or $m_{ij}^{(2)}$ shows greater power than the test (2.5) based on $m_i^{(3)}$. As mentioned above, this is partly due to the fact that, although the test based on $ln(s+1)$ is run at nominal level $\alpha = 0.05$, its actual level is 0.02–0.03 (see Table 1). It also highlights the difficulty for this measure of detecting a dispersion effect in the presence of location effects (see Section 1). The tests using both $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ and $ln(|y_{ij} - \bar{y}_i| + 1)$ have reasonable power for detecting a dispersion effect of 0.3 or more for $r \geq 7$ and of 0.6 or more for smaller r under model (2.1) while controlling the Type I error rate. Figures 2 and 4 indicate that the power for the test based on $ln(|y_{ij} - \bar{y}_i| + 1)$ is slightly greater for smaller r, but for larger r, there is little difference in the power of these two tests.

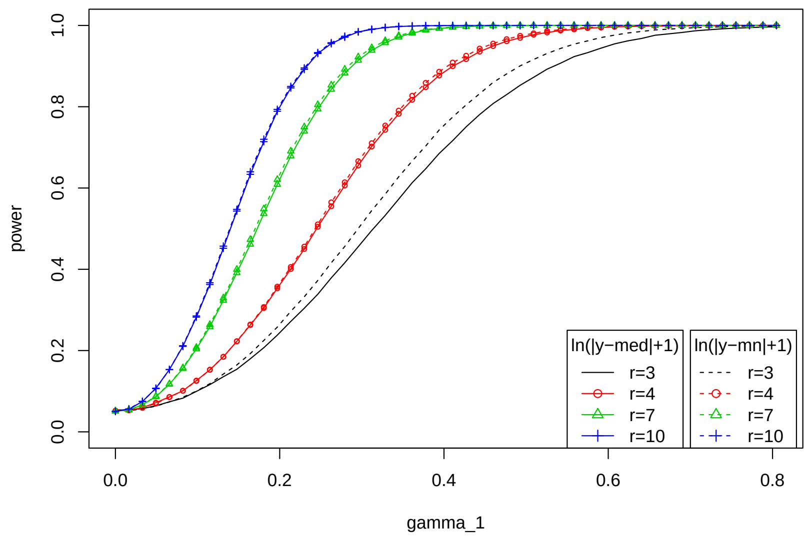

Figure 2: Power curves for tests based on $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ and $ln(|y_{ij} - \bar{y}_i| + 1)$ using empirical critical values from Tables 4 and 5, with data from randomly generated location models and *first-order* dispersion models, with $r = 3, 4, 7, 10$ replicates per cell, and normal error distribution

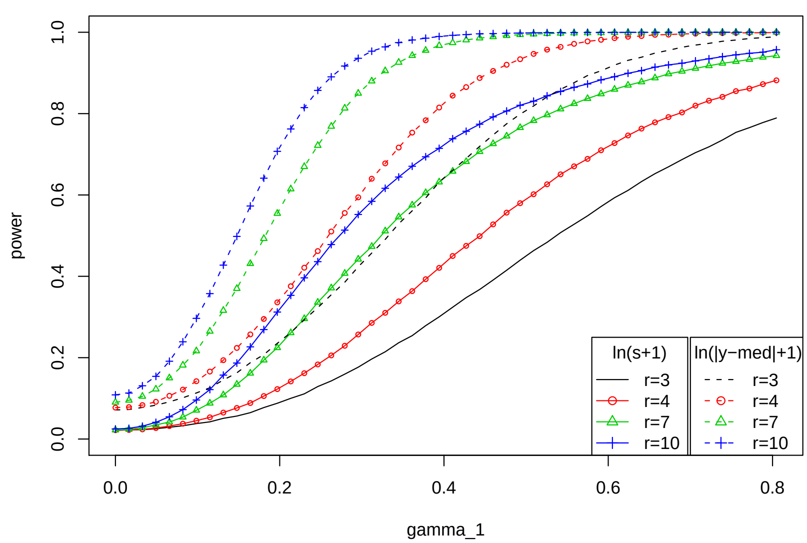

Figure 3: Power curves for tests based on $ln(s + 1)$ and $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ using empirical critical values from Tables 4 and 6, with data from randomly generated location models and *second-order* dispersion models, with $r = 3, 4, 7, 10$ replicates per cell, and normal error distribution

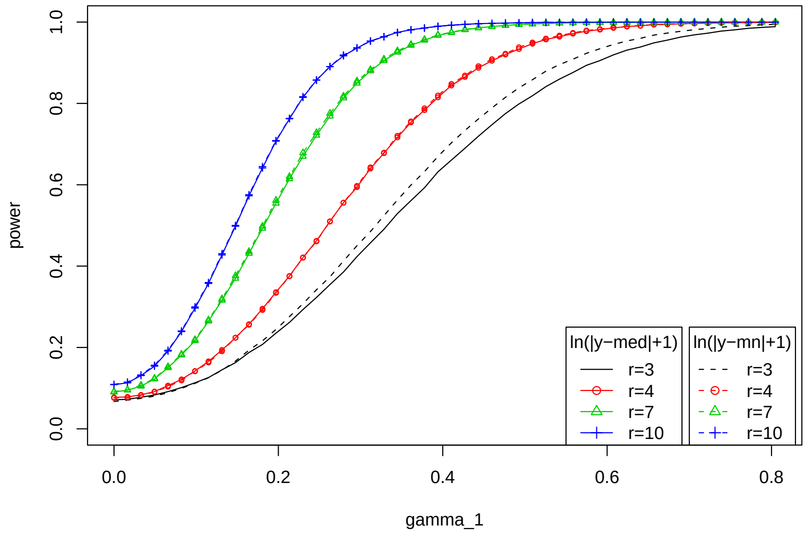

Figure 4: Power curves for tests based on $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ and $ln(|y_{ij} - \bar{y}_i| + 1)$ using empirical critical values from Tables 4 and 5, with data from randomly generated location models and *second-order* dispersion models, with $r = 3, 4, 7, 10$ replicates per cell, and normal error distribution

To study the effect of non-normal error distributions, additional simulations were run for both a Cauchy and an exponential error distribution, for $r = 4$ replicates per treatment combination. The power curves based on the results from these simulations are shown in Figures 5 and 6. These figures, together with Table 1, show that for exponentially distributed errors, with first or second order dispersion models, the Type I error rates for the test of $H_0^{\gamma_t}$ (2.3) based on $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ are raised slightly above the nominal $\alpha = 0.05$ level, but those for the test based on $ln(|y_{ij} - \bar{y}_i| + 1)$ are considerably higher.

The situation is even more exaggerated for Cauchy distributed errors and here, clearly, the test based on $ln(|y_{ij} - \bar{y}_i| + 1)$ is not usable. Consequently, we recommend the test based on $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ rather than $ln(|y_{ij} - \bar{y}_i| + 1)$ unless r is small and the errors are "known" to be identically and independently normally distributed.

4 Example

Pignatiello and Ramberg (1985) discussed an experiment which studied the effect of five

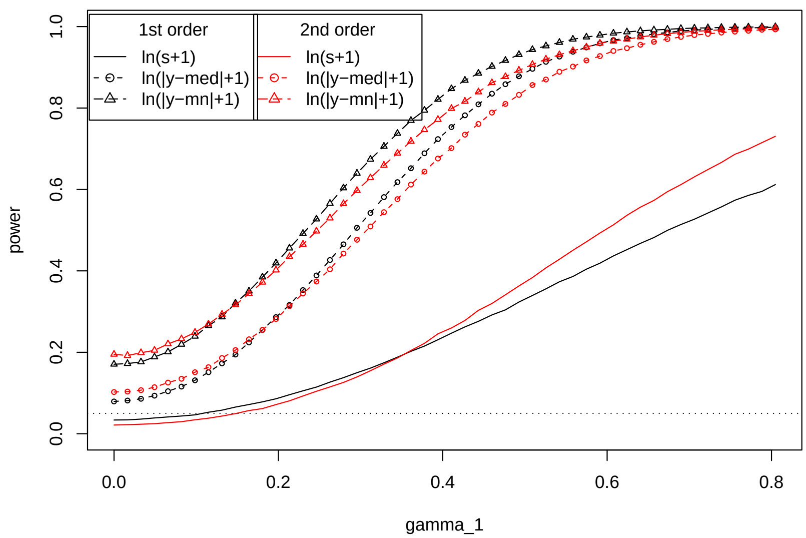

Figure 5: Power curves for tests based on $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$, $ln(|y_{ij} - \bar{y}_i| + 1)$ and $ln(s + 1)$, using empirical critical values from Tables 4–6, with data from randomly generated location models and *first-order and second-order* dispersion models, with $r = 4$ replicates per cell, and exponential error distribution

factors on the robust design of leaf springs in trucks; the experiment has further been analysed for dispersion effects by Nair and Pregibon (1988) and Wu and Hamada (2000). The experiment examined five factors, each at two levels: furnace temperature (B), heating time (C), transfer time (D), hold down time (E), and quench-oil temperature (O). The response of interest was the free height (Y) of a spring in an unloaded condition. Pignatiello and Ramberg (1985) first used factor O as a noise factor that could not be controlled and was folded into the experimental error. Then, in a separate analysis, factor O was used as a control factor. These two analyses result in different main effects being classified as significant, owing to sizeable interactions involving factors B and O.

Here, in order to illustrate our dispersion tests, we use the first setting with O contributing to the experimental error. The design is then a 2^{4-1} fractional factorial, with four factors B, C, D and E and defining contrast $I = BCDE$. There are $r = 6$ replicates at each of the $v = 8$ treatment combinations. The contrasts of interest and data were presented by both Pignatiello and Ramberg (1985) and Nair and Pregibon (1988). In our Table 2, we show the design, where -1 and $+1$ represent the two levels of each factor,

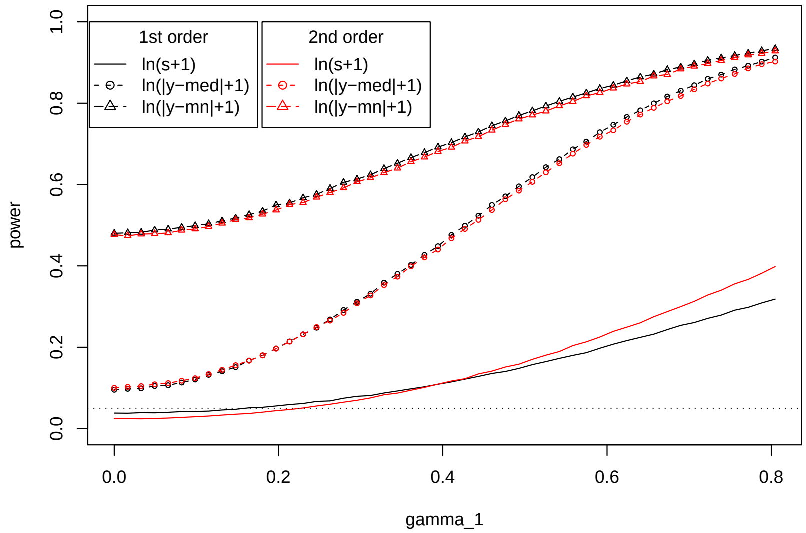

Figure 6: Power curves for tests based on $ln(s+1)$, $\ln(|y_{ij}-\widetilde{y}_i|_{-1}+1)$ and $ln(|y_{ij}-\bar{y}_i|+1)$ using empirical critical values from Tables 4–6, with data from randomly generated location models and *first-order and second-order* dispersion models, with $r=4$ replicates per cell, and Cauchy error distribution

and the calculated dispersion measures $m_{ij}^{(1)} = \ln(|y_{ij}-\widetilde{y}_i|_{-1}+1)$.

Consider the test for the null hypothesis $H_0 : \gamma_C = 0$ against the alternative hypothesis $H_1 : \gamma_C \neq 0$, where γ_C is the dispersion main effect of factor C. The test statistic M_C (2.4) is

$$M_C = \frac{(0.09503 - 0.205207)^2 \times 8 \times 5/4}{0.315530/(8\times 4)} = 12.31.$$

Comparing M_C with the empirical critical value of 6.58 given in Table 4 at level $\alpha = 0.01$ for $v = 8$ treatment combinations and $r = 6$ observations per cell, we see that $M_C = 12.31 > 6.58$, and we conclude that heating time (C) has statistically significant dispersion main effect at level 0.01.

The values of M_t for each of the contrasts in a second order saturated model are shown in Table 3. If $H_0^{\gamma_t}$ (2.3) is tested for each γ_t at level 0.01, the overall level is at most 0.07 for the 7 tests using a Bonferroni correction, and only a significant main effect of factor

Table 2: Design and dispersion measures $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ for the leaf spring 2^{4-1} experiment with factor O as an uncontrolled noise factor. The values $(m_{ij}^{(1)})$ are the smallest dispersion measures to be omitted from the calculation of M_t

B	C	D	E	$m_{i1}^{(1)}$	$m_{i2}^{(1)}$	$m_{i3}^{(1)}$	$m_{i4}^{(1)}$	$m_{i5}^{(1)}$	$m_{i6}^{(1)}$	$\overline{m}_{i.}^{(1)}$
-1	-1	-1	-1	0.131	0.131	0.157	(0.131)	0.329	0.418	0.2332
1	-1	-1	1	0.239	0.262	0.000	0.000	(0.000)	0.364	0.1730
-1	1	-1	1	0.000	0.058	0.000	0.000	0.058	(0.000)	0.0232
1	1	-1	-1	0.019	0.048	0.131	(0.019)	0.131	0.048	0.0754
-1	-1	1	1	0.246	0.292	0.198	0.292	0.198	(0.198)	0.2452
1	-1	1	-1	0.000	0.336	0.314	0.121	(0.000)	0.067	0.1676
-1	1	1	-1	0.194	0.242	0.090	0.152	0.152	(0.090)	0.1660
1	1	1	1	0.076	0.157	0.048	0.157	0.131	(0.048)	0.1138

Table 3: Test statistics based on dispersion measures $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$ for the leaf spring experiment data; critical value from Table 4 for $\alpha = 0.01$ is 6.583

Source	B	C	D	E	$BC(DE)$	$BD(CE)$	$CD(BE)$
M_t	1.21	12.31**	2.27	0.49	1.21	0.96	1.79

C is detected. This conclusion is consistent with those of previous analysis of these data using half-normal probability plots; see the results of Pignatiello and Ramberg (1985) who used signal to noise ratios, and also the results of Nair and Pregibon (1988) who used γ_t estimates based on the work of Bartlett and Kendall (1946) and Box and Meyer (1986). If we were to analyse the experiment with factor O as a control factor, so that we have a 2^{5-1} experiment with $r = 3$ observations per cell and $v = 16$ treatment combinations, then we find that factor B has the only significant dispersion effect (the change being due to the masking interactions between the factors O, B and C). This is in line with the results of Nair and Pregibon (1988) and also the work of Wu and Hamada (2000), Section 4.3.

5 Summary

In this paper, we have proposed tests for the detection of dispersion effects in a location-dispersion model (2.1) with $r \geq 3$ observations within each cell. Measures which combine observations into a single dispersion measure (such as $ln(s^2)$ or $ln(s+1)$) have been shown previously to be affected by location effects (see Section 1) and our study confirms that Lenth's test using $ln(s+1)$ does not perform well even when using empirical critical values. Instead, we follow the recommendation of Mackertich et al. (2003) and obtain individual measures of dispersion for all observations in a cell, rather than combining these into a single measure. These measures remove the location effects and also allow for an estimate of pure error.

We show that tests using a test statistic of the form (2.4) with dispersion measure $\ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$, the natural logarithm of the absolute deviation from the median omitting (one of) the minimum observation(s), and critical values from Table 4 control the Type I error rate close to the nominal significance level for data following a normal distribution. Figures 5 and 6 and Table 1 show that the test has a slightly elevated Type I error when the errors follow an exponential or Cauchy distribution. This test has good power whether data are generated from a first- or second-order dispersion model under effect sparsity and randomly generated location models. When more than one dispersion effect is to be tested, a Bonferroni correction can be made as in Section 4.

When errors are "known" to be normally distributed and r is small, slightly higher power can be achieved by basing the test on the dispersion measure $m_{ij}^{(2)} = ln(|y_{ij} - \bar{y}_i| + 1)$ rather than $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$, together with critical values from Table 5. If $m_{ij}^{(2)} = ln(|y_{ij} - \bar{y}_i| + 1)$ is used for the example of Section 4, the same conclusions are reached.

In summary, for experimental designs, such as regular fractional factorial designs, in which effects can be estimated independently, the dispersion effect test methodology proposed in the current work provides a more powerful alternative to traditional dispersion test methodologies for data from replicated two-level experiments.

Acknowledgements

The authors thank a referee for helpful and thoughtful comments. The work of Dean was supported by grants SES-0437251 and DMS-0806134 from the National Science Foundation, and the work of Ankenman by EPSRC grant GR/S91673 while visiting Southampton Statistical Sciences Research Institute.

A Empirical Critical Values

Table 4: Critical values for test statistic (2.4) using $m_{ij}^{(1)} = \ln(|y_{ij} - \widetilde{y}_i|_{-1} + 1)$

v	α	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
8	0.1	2.60	2.41	2.59	2.51	2.63	2.58	2.65	2.61
	0.05	4.03	3.57	3.81	3.65	3.79	3.71	3.79	3.76
	0.01	8.76	6.81	7.06	6.58	6.79	6.65	6.80	6.63
	0.005	11.54	8.45	8.70	8.00	8.20	7.97	8.19	8.02
16	0.1	2.31	2.27	2.50	2.45	2.56	2.54	2.59	2.58
	0.05	3.41	3.28	3.59	3.51	3.66	3.63	3.70	3.68
	0.01	6.51	5.96	6.42	6.21	6.45	6.36	6.48	6.43
	0.005	8.11	7.22	7.75	7.48	7.77	7.64	7.74	7.68
32	0.1	2.18	2.21	2.45	2.42	2.53	2.51	2.57	2.56
	0.05	3.15	3.16	3.49	3.45	3.61	3.57	3.66	3.64
	0.01	5.72	5.59	6.14	6.04	6.29	6.21	6.37	6.34
	0.005	6.94	6.70	7.37	7.23	7.47	7.39	7.59	7.55
64	0.1	2.12	2.18	2.43	2.40	2.52	2.49	2.56	2.55
	0.05	3.03	3.10	3.45	3.42	3.58	3.55	3.64	3.63
	0.01	5.37	5.42	6.01	5.94	6.22	6.15	6.31	6.27
	0.005	6.44	6.47	7.16	7.08	7.39	7.33	7.53	7.48

Table 5: Critical values for test statistic (2.4) using $m_{ij}^{(2)} = ln(|y_{ij} - \bar{y}_i| + 1)$

v	α	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
8	0.1	5.19	4.10	3.61	3.36	3.24	3.16	3.08	3.03
	0.05	7.48	6.00	5.26	4.88	4.69	4.51	4.41	4.36
	0.01	13.57	11.28	9.60	8.81	8.35	8.11	7.93	7.65
	0.005	16.58	14.05	11.75	10.72	10.04	9.75	9.51	9.14
16	0.1	4.93	3.87	3.49	3.29	3.17	3.09	3.04	3.00
	0.05	7.08	5.60	5.00	4.72	4.54	4.43	4.33	4.28
	0.01	12.53	10.08	8.91	8.35	7.99	7.77	7.60	7.46
	0.005	15.09	12.22	10.71	10.02	9.59	9.29	9.08	8.91
32	0.1	4.82	3.80	3.43	3.25	3.14	3.07	3.01	3.00
	0.05	6.88	5.43	4.90	4.63	4.48	4.37	4.29	4.24
	0.01	12.07	9.57	8.58	8.10	7.79	7.58	7.46	7.37
	0.005	14.42	11.44	10.28	9.68	9.27	9.04	8.88	8.75
64	0.1	4.76	3.74	3.41	3.23	3.12	3.05	3.00	2.97
	0.05	6.77	5.34	4.85	4.59	4.43	4.37	4.27	4.22
	0.01	11.76	9.30	8.43	7.98	7.69	7.53	7.39	7.31
	0.005	14.03	11.11	10.03	9.49	9.18	8.94	8.80	8.69

Table 6: Critical values for Lenth's test statistic (2.5) using $m_i^{(3)} = ln(s_i + 1)$

v	α	$r=3$	$r=4$	$r=5$	$r=6$	$r=7$	$r=8$	$r=9$	$r=10$
8	0.1	1.73	1.73	1.72	1.72	1.72	1.72	1.72	1.71
	0.05	2.34	2.32	2.32	2.31	2.31	2.31	2.30	2.30
	0.01	5.20	5.17	5.12	5.10	5.10	5.10	5.10	5.10
	0.005	7.00	6.98	6.90	6.87	6.87	6.87	6.87	6.87
16	0.1	1.71	1.71	1.71	1.70	1.70	1.70	1.70	1.70
	0.05	2.18	2.17	2.17	2.16	2.16	2.16	2.16	2.16
	0.01	3.69	3.66	3.65	3.64	3.63	3.63	3.63	3.63
	0.005	4.44	4.41	4.41	4.39	4.37	4.37	4.37	4.37
32	0.1	1.68	1.68	1.68	1.68	1.68	1.68	1.68	1.68
	0.05	2.07	2.07	2.07	2.07	2.07	2.07	2.07	2.07
	0.01	3.07	3.06	3.06	3.05	3.05	3.05	3.05	3.05
	0.005	3.50	3.49	3.48	3.48	3.48	3.48	3.47	3.47
64	0.1	1.67	1.67	1.67	1.67	1.67	1.66	1.66	1.66
	0.05	2.02	2.02	2.02	2.02	2.01	2.01	2.01	2.01
	0.01	2.80	2.80	2.80	2.80	2.80	2.80	2.80	2.80
	0.005	3.12	3.12	3.12	3.12	3.12	3.12	3.12	3.12

References

Aitkin, M. (1987), "Modelling variance heterogeneity in normal regression using GLIM," *Applied Statistics*, 36, 332–339.

Ankenman, B. E. and Dean, A. M. (2003), "Quality improvement and robustness via design of experiments," in *Handbook of Statistics*, eds. Rao, C. R. and Khattree, R., Elsevier, vol. 22, pp. 263–317.

Bartlett, M. S. and Kendall, D. G. (1946), "The statistical analysis of variance-heterogeneity and the logarithmic transformation," *Supplement to the Journal of the Royal Statistical Society*, 8, 128–138.

Bergman, B. and Hynén, A. (1997), "Dispersion effects from unreplicated designs in the 2^{k-p} series," *Technometrics*, 39, 191–198.

Box, G. E. and Meyer, R. D. (1986), "Dispersion effects from fractional designs," *Technometrics*, 28, 19–27.

Box, G. E. P. (1988), "Signal-to-noise ratios, performance criteria, and transformations," *Technometrics*, 30, 1–17.

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons.

Brenneman, W. A. and Nair, V. N. (2001), "Methods for identifying dispersion effects in unreplicated factorial experiments," *Technometrics*, 43, 388–405.

Brown, M. and Forsythe, A. (1974), "Robust tests for the equality of variances," *Journal of the American Statistical Association*, 69, 364–367.

Bursztyn, D. and Steinberg, D. M. (2006), "Screening Experiments for Dispersion Effects," in *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*, eds. Dean, A. M. and Lewis, S. M., New York: Springer, pp. 21–47.

Cook, R. D. and Weisberg, S. (1983), "Diagnostics for heteroscedasticity in regression," *Biometrika*, 70, 1–10.

Dean, A. and Voss, D. (1999), *Design and Analysis of Experiments*, Springer.

Dingus, C. A. V. (2005), "Designs and methods for the identification of active location and dispersion effects," Ph.D. thesis, The Ohio State University.

Haaland, P. and O'Connell, M. (1995), "Inference for effect-saturated fractional fatorials," *Technometrics*, 37, 82–93.

Harvey, A. C. (1976), "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, 44, 461–465.

Lenth, R. V. (1989), "Quick and easy analysis of unreplicated factorials," *Technometrics*, 31, 469–473.

Levene, H. (1960), "Robust tests for equality of variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, ed. et al., I. O., Stanford University Press.

Mackertich, N. A., Benneyan, J. C., and Kraus, P. D. (2003), "Alternate dispersion measures in replicated factorial Experiments," Unpublished Manuscript.

McGrath, R. N. and Lin, D. K. J. (2001a), "The confounding relationships of location and dispersion effects in unreplicated fractional factorials," *Journal of Quality Technology*.

— (2001b), "Testing multiple dispersion effects in unreplicated two-level fractional factorial designs," *Technometrics*, 43, 406–414.

Miller, A. (2005), "The analysis of unreplicated factorial experiments using all possible comparisons," *Technometrics*, 47, 51–63.

Montgomery, D. C. (2009), *Design and Analysis of Experiments*, John Wiley & Sons, 7th ed.

Nair, V. J. and Pregibon, D. (1988), "Analyzing dispersion effects from replicated factorial experiments," *Technometrics*, 30, 247–257.

Pan, G. (1999), "The impact of unidentified location effects on dispersion- effects identification from unreplicated designs," *Technometrics*, 41, 313–326.

Pignatiello, Joseph J., J. and Ramberg, J. S. (1985), "Discussion of "Off-line Quality Control, Parameter Design, and the Taguschi Method" by R.N. Kackar," *Journal of Quality Technology*, 17, 198–206.

— (1987), "Comments on "Performance Measures Independent of Adjustment: An Explanation and Extension of Taguchi's Signal-to-noise Ratios"," *Technometrics*, 29, 274–277.

Verbyla, A. P. (1993), "Modelling variance heterogeneity: residual maximum likelihood and diagnostics," *Journal of the Royal Statistical Society, Series B*, 55, 493–508.

Wang, P. C. (1989), "Tests for dispersion effects from orthogonal arrays," *Computational Statistics and Data Analysis*, 8, 109–117.

Wolfe, D. A., Dean, A. M., Wiers, M. D., and Hartlaub, B. A. (1992), "Nonparametric rank based main effects test procedures for the two-way layout in the presence of interaction," *Journal of Nonparametric Statistics*, 1, 241–252.

Wolfinger, R. D. and Tobias, R. D. (1998), "Joint estimation of location, dispersion, and random effects in robust design," *Technometrics*, 40, 62–71.

Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, John Wiley & Sons.