ELSEVIER Computational Statistics & Data Analysis 33 (2000) 291–298

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Identification of dispersion effects from unreplicated 2^{n-k} fractional factorial designs

C.T. Liao

Division of Biometry, Institute of Agronomy, National Taiwan University, No.1, Sec. 4 Roosevelt Road, Taipei 10764, Taiwan

Received 1 July 1998; received in revised form 1 June 1999

Abstract

In this article, we present a test for dispersion effects from the unreplicated 2^{n-k} regular fractional factorial designs. The proposed procedure for the identification of dispersion effects uses the log-likelihood ratio based on normal errors. Some practical examples are given to illustrate the applicability of the test. It is shown that the proposed method is a useful and economical means for the identification of dispersion effects at the screening stage of experiments. Comparing the power of our method with the two methods published in the literature, we suggest that our test might be more sensitive for identifying the dispersion effects.

Keywords: Taguchi; Location effects; Experimental designs; Quality improvement; Robust designs

1. Introduction

Traditionally, factorial screening experiments have been used in product design, process design and other industrial applications for the estimation of location effects. One of the important contributions of Taguchi was to point out that product quality is highly related to the variability of the quality characteristics. What is required of a process is to produce products whose quality characteristics are tightly distributed around a specified target value. Therefore, Taguchi applied experimental designs to study dispersion effects in addition to the location effects. A useful discussion of Taguchi's method can be seen in the review edited by Nair (1992).

PII: S 0167-9473(99)00058-4

Methods for identifying dispersion effects usually involve replicated factorial experiments. See Taguchi (1987), Nair and Pregibon (1988), and Ghosh and Lagergren (1990). Although such designs have proved to be effective in practice, they can lead to very costly experiments in certain situations as the required number of runs can be large. Therefore, the study of the identification of dispersion effects from the unreplicated factorial designs has been drawing more attention due to its economical cost of experimentation.

Most approaches for the identification of dispersion effects from the unreplicated factorial designs involve two separate steps. The first step is to apply the normal probability plot or the half-normal probability plot to identify unusually large location effects. The second step is to compute a statistic that is relevant to the dispersion effect, mostly based on the residuals of the linear model fitting those identified for larger location effects. Then one popular method is to apply the normal plot again on the computed statistic to identify unusually large dispersion effects. See Glejser (1969), Daniel (1976), Hoaglin et al. (1983), Box and Meyer (1986), and Davidian and Carroll (1987). However, as this is a subjective model-discrimination method, one would have to trust "eye-balling". Another method is to develop a significance test. Wang (1989) and Bergman and Hynen (1997) provided two different tests based on the χ^2 and F distributions, respectively. Bergman and Hynen (1997) also gave a good review of these methods.

In the next section, we review the methods of Wang (1989) and Bergman and Hynen (1997), and present our method. In Section 3, we illustrate these three methods by some real data sets. In Section 4, we report a comparison of the power among the methods.

2. Dispersion effects

Let $\boldsymbol{y}$ be the $N \times 1$ response vector of a regular $N = 2^{n-k}$ fractional factorial design with an $N \times N$ orthogonal design matrix $\boldsymbol{X}$ of -1s and $+1$s in column vectors $\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$. Column $\boldsymbol{x}_0$ corresponds to the grand mean with all elements equal to $+1$; the remaining columns correspond to the all possibly estimated main effects and interactions from this design. The observations $y_1, y_2, \ldots, y_N$ are assumed to be realizations of uncorrelated normally distributed random variables.

The *sparsity principle* frequently used in practice suggests that in most cases only a few factorial effects are non-negligible. Therefore, one can use the normal plotting techniques provided by Daniel (1959, 1976) to identify these unusually large location effects. As pointed out by Bergman and Hynen (1997), the use of the normal plot for the identification of location effects is still legitimate even though the dispersion of the response may vary with the levels of some factors in the experiment.

We will use the term "*active*" location effects to denote these unusually large location effects identified by the normal plot. Bergman and Hynen (1997) fitted separate regression models to two sets of data associated with the high and low levels of column i in the $\boldsymbol{X}$ matrix, respectively. Let $\boldsymbol{e}_{i+}$ and $\boldsymbol{e}_{i-}$ denote the corresponding residuals vectors. They suggested the following statistic for the dispersion effect of column i in the $\boldsymbol{X}$ matrix.

$$D_i^{\mathrm{BH}} = \sum_{u=1}^{N/2} e_{ui+}^2 \Big/ \sum_{u=1}^{N/2} e_{ui-}^2. \tag{1}$$

Bergman and Hynen (1977) claimed that D_i^{BH} has an $F(m,m)$ distribution assuming that the active location effects are responsible for all of the true location effects. The degree of freedom m is equal to $N/2-q-1$ if column i is not associated with the active location effects; otherwise it is $N/2-q$, where q is the number of the active location effects.

Most methods for the identification of dispersion effects from unreplicated designs are not based on the residuals of e_{ui}, but on the residuals, say r_u, $u=1,2,\ldots,N$, which are computed from the model fitting the active location effects for all observations. Wang (1989) presented the following statistics representing the dispersion effect of column i in the $\boldsymbol{X}$ matrix:

$$D_i^{\mathrm{W}} = \frac{1}{2N}\left(\frac{S_{i+}-S_{i-}}{S_{i+}+S_{i-}}\right)^2, \tag{2}$$

where $S_{i+}=\sum_{u:x_{ui}=+1} r_u^2$ and $S_{i-}=\sum_{u:x_{ui}=-1} r_u^2$. Wang (1989) claimed that D_i^{W} is approximately $\chi^2(1)$ for a large sample size N.

It is assumed that r_u are normal errors with expectation $E(r_u)=0$. Then we consider the null hypothesis as

$$\mathrm{H}_0\colon \mathrm{Var}(r_u)=\sigma^2$$

and alternative hypothesis as

$$\mathrm{H}_1\colon \mathrm{Var}(r_u)=\begin{cases}\sigma_{i+}^2 & \text{if} x_{ui}=+1,\\ \sigma_{i-}^2 & \text{if} x_{ui}=-1\end{cases}$$

for $u=1,2,\ldots,N$. Let $\widehat{L}_0$ and $\widehat{L}_1$ denote the likelihood evaluated at MLE of H_0 and H_1, respectively. It is well known that $-2\log(\widehat{L}_0/\widehat{L}_1)$ approximates to $\chi^2(1)$ for large N. This approximation is often quite accurate for small values of N (see, McCullagh and Nelder, 1989). We thus use this statistics to represent the dispersion effect of column i. Let the notation D_i^{L} denote this $-2\log(\widehat{L}_0/\widehat{L}_1)$; then it is easy to verify that

$$D_i^{\mathrm{L}} = \frac{N}{2}\log\left(\frac{1}{4S_{i+}}+\frac{1}{4S_{i-}}\right). \tag{3}$$

3. Examples

Some practical examples will be given to illustrate the three tests described in the previous section. For each example, the p-value will be calculated for the dispersion effect of each column i in the $\boldsymbol{X}$ matrix.

Example 1. Montgomery (1997 p. 391) gave an example to illustrate the method of Box and Meyer (1986) for identifying dispersion effects. A quality-improvement team used a 2^{6-2} fractional factorial design to study the injection molding process so that the excessive shrinkage can be reduced. There are six factors, *A*–*F*, involved

Table 1
Design matrix, response, and confounding structure up to two-factor interactions for the injection molding experiment.

u	A	B	AB CE	C	AC BE	AE BC	E	D	AD EF	BD CE	ABD	BF CD	ACD	F	AF DE	y_u
1	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+	6
2	+	−	−	−	−	+	+	−	−	+	+	+	+	−	−	10
3	−	+	−	−	+	−	+	−	+	−	+	+	−	+	−	32
4	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	60
5	−	−	+	+	−	−	+	−	+	+	−	−	+	+	−	4
6	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	15
7	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	26
8	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	60
9	−	−	+	−	+	+	−	+	−	−	+	−	+	+	−	8
10	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	12
11	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+	34
12	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	60
13	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	16
14	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	5
15	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	37
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	52

Table 2
Statistics for identifying dispersion effects from the injection molding experiment.

i	D_i^{BH}	df	p-value (D_i^{BH})	D_i^{W}	p-value (D_i^{W})	D_i^{L}	p-value (D_i^{L})
1	0.68	(5,5)	0.69	0.29	0.60	0.28	0.59
2	0.83	(5,5)	0.84	0.07	0.79	0.07	0.79
3	1.11	(5,5)	0.91	0.02	0.88	0.02	0.88
4	35.75	(4,4)	0.004	5.62	0.02	9.70	0.002
5	0.64	(4,4)	0.68	0.30	0.58	0.31	0.57
6	0.78	(4,4)	0.81	0.10	0.75	0.10	0.75
7	0.96	(4,4)	0.97	0.002	0.96	0.002	0.96
8	2.86	(4,4)	0.33	0.47	0.49	0.48	0.48
9	1.56	(4,4)	0.68	0.98	0.75	0.98	0.75
10	0.68	(4,4)	0.72	0.07	0.79	0.07	0.79
11	3.05	(4,4)	0.31	0.52	0.47	0.54	0.46
12	2.40	(4,4)	0.41	0.51	0.48	0.52	0.47
13	1.26	(4,4)	0.83	0.04	0.84	0.04	0.84
14	0.60	(4,4)	0.64	0.18	0.67	0.18	0.67
15	3.59	(4,4)	0.24	0.94	0.33	1.01	0.31

in this experiment. Table 1 shows the design matrix, the confounding structure and in the last column, the data.

It can be shown, by normal plot (see Montgomery, 1997, p. 393), that factorial effects associated with A, B and AB are considered as active location effects. The computed D_i^{BH}, D_i^{W} and D_i^{L} values and the respective p-values are displayed in Table 2.

Table 3
Design matrix, response, and confounding structure up to two-factor interactions for the welding strength experiment.

u	$-BC$ $-FG$ $-AE$ $-D$	$-EI$ $-CI$ H	$-FI$ $-DF$ G	AI $-DE$ A	GI BI DH	CE DG $-F$	CF AD $-E$	DI AB GH	EF $-BE$ BG AH	$-BF$ BH AG	$-AF$ $-EG$ $-FH$ $-AC$	HI $-EH$ $-CG$	$-CH$ I	$-CD$ B	BD $-C$	y_u
1	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	43.7
2	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	40.2
3	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	42.4
4	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	44.7
5	−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	42.4
6	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	45.9
7	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	42.2
8	+	+	+	−	+	+	−	+	−	−	+	−	−	−	−	40.6
9	−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	42.4
10	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	45.5
11	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	43.6
12	+	+	−	+	+	−	+	−	+	−	−	+	−	−	−	40.6
13	−	−	+	+	+	−	−	−	−	+	+	+	−	−	+	44.0
14	+	−	+	+	−	+	+	−	−	+	−	−	+	−	−	40.2
15	−	+	+	+	−	−	−	+	+	+	−	−	−	+	−	42.5
16	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	46.5

Table 4
Statistics for identifying the dispersion effects from the welding strength experiment

i	D_i^{BH}	df	p-value (D_i^{BH})	D_i^{W}	p-value (D_i^{W})	D_i^{L}	p-value (D_i^{L})
1	0.97	(6,6)	0.97	0.002	0.96	0.002	0.96
2	15.93	(5,5)	0.009	4.13	0.04	5.82	0.01
3	4.38	(5,5)	0.13	1.89	0.17	2.16	0.14
4	0.34	(5,5)	0.26	0.92	0.34	0.97	0.32
5	1.37	(5,5)	0.74	0.11	0.74	0.11	0.74
6	0.21	(5,5)	0.11	1.39	0.24	1.52	0.22
7	2.18	(5,5)	0.41	0.80	0.37	0.84	0.36
8	2.20	(5,5)	0.41	0.78	0.38	0.82	0.37
9	0.20	(5,5)	0.10	1.41	0.23	1.55	0.21
10	1.15	(5,5)	0.88	0.03	0.86	0.03	0.86
11	0.34	(5,5)	0.27	0.90	0.34	0.95	0.33
12	4.21	(5,5)	0.14	1.92	0.17	2.20	0.14
13	20.96	(5,5)	0.005	3.60	0.06	4.79	0.03
14	0.82	(6,6)	0.82	0.07	0.79	0.07	0.79
15	21.72	(6,6)	0.002	6.44	0.01	13.07	0.0003

Even though it shows that D_i^{BH} and D_i^{L} are more distinct, these three tests give fairly similar results that point out C as the potential dispersion effect.

Example 2. A welding strength experiment was carried out by the National Railway Corporation in Japan (see Taguchi and Wu, 1980). There are nine factors, A–I, considered in a 2^{9-5} fractional factorial experiment. Table 3 shows the design matrix, the confounding structure and in the last column, the data.

It can be shown, by normal plotting technique (see Box and Meyer, 1986), that the location factorial effects associated with B and C are the active location effects. The computed D_i^{BH}, D_i^{W} and D_i^{L} values and the respective p-values are given below (Table 4).

These three tests give fairly similar results that point out C, H, and I as the likely dispersion effects. It can be seen that D_i^{L} and D_i^{BH} might be more sensitive.

4. Comparison of power

The examples discussed in the previous section show that the three methods have a similar performance. It is of interest to compare the power of the tests. Under the assumption that the active locations are responsible for all of the true location effects, these three statistics are all functions of an F-random variable. Let $\rho = (\sigma_{i+}^2/\sigma_{i-}^2)$. Then it is easy to verify the following:

$$\begin{aligned}\mathrm{Power}^{\mathrm{BH}}(\rho) &= \Pr(D_i^{\mathrm{BH}} \geq F_{\alpha/2,m,m}) + \Pr(D_i^{\mathrm{BH}} \leq F_{1-\alpha/2,m,m}) \\ &= \Pr(z \geq F_{\alpha/2,m,m}) + \Pr(z \leq F_{1-\alpha/2,m,m}). \qquad (4)\end{aligned}$$

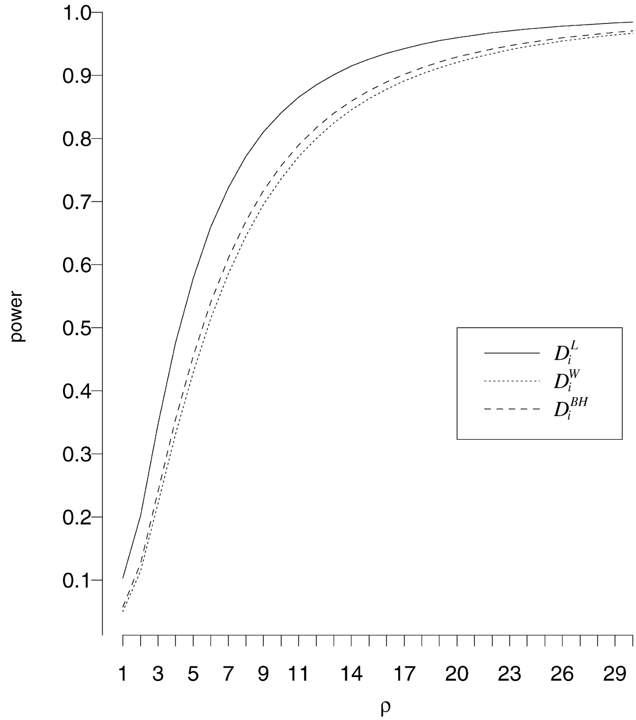

Fig. 1. The power of the tests for the case where $N = 16$, $m = 6$, $\alpha = 0.05$.

$$\begin{aligned}\text{Power}^{\text{W}}(\rho) &= \Pr(D_i^{\text{W}} \geq \chi^2_{\alpha,1}) \\ &= \Pr\left(\frac{N}{2}\left(\frac{z-1}{z+1}\right)^2 \geq \chi^2_{\alpha,1}\right)\end{aligned} \tag{5}$$

and

$$\begin{aligned}\text{Power}^{\text{L}}(\rho) &= \Pr(D_i^{\text{L}} \geq \chi^2_{\alpha,1}) \\ &= \Pr\left(\frac{N}{2}\log\left[\frac{1}{4}\left(z+\frac{1}{z}\right)\right] \geq \chi^2_{\alpha,1}\right),\end{aligned} \tag{6}$$

where the random variable z is a scaled F distribution, i.e., z multiplied by the constant ρ is $F(m,m)$. The degrees of freedom m must be adjusted to the number of active location effects. For detailed discussions on parameter m one can refer to Bergman and Hynen (1997). For the case where $N = 16$, $m = 6$, ρ ranges from 1 to 30 in step of 1, and the significance level $\alpha = 0.05$; the power for these tests is graphed in Fig 1.

It is shown that D_i^{L} is more liberal and is capable of being more sensitive for identifying the possibly non-negligible dispersion effects at the screening stage of experimentation.

5. Concluding remarks

At the initial stage of the experimentation, the investigators addressed the issue that all possibly non-negligible location effects and dispersion effects can be identified within the fixed budget. We are convinced that the proposed method is a good alternative for identifying the dispersion effects, and this method might be more sensitive for the screening experiment. Once the influential factors for the location and dispersion of the response have been identified, one can plan more elaborate follow-up experiments to characterize the relationship between the responsible variable and experimental factors by means of suitable mathematical functions. The mixed linear model discussed by Wolfinger and Tobias (1998) can be a good choice.

References

Bergman, B., Hynen, A., 1997. Dispersion effects from unreplicated designs in the 2^{k-p} series. Technometrics 39, 191–198.

Box, G.E.P., Meyer, R.D., 1986. Dispersion effects from fractional designs. Technometrics 28, 19–27.

Daniel, C., 1959. Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1, 311–341.

Daniel, C., 1976. Application of Statistics to Industrial Experiments. Wiley, New York.

Davidian, M., Carroll, R.J., 1987. Variance Function Estimation. J. Amer. Statist. Assoc. 82, 1079–1091.

Ghosh, S., Lagergren, E.S., 1990. Dispersion models and estimation of dispersion effects in replicated factorial experiments. J. Statist. Plann. Inference 26, 253–262.

Glejser, H., 1969. A New Test for Heteroscedasticity. J. Amer. Statist. Assoc. 64, 494–499.

Hoaglin, D.C., Mosteller, F., Tukey, J.W., 1983. Understanding robust and exploratory data analysis. New York: Wiley.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. New York: Chapman & Hall.

Montgomery, D.C., 1997. Design and Analysis of Experiments. Wiley, New York.

Nair, V. N. (Ed.), 1992. Taguchi's parameter design: a panel discussion. Technometrics, 34, 127–159.

Nair, V.N., Pregibon, D., 1988. Analyzing dispersion effects from replicated factorial experiments. Technometrics 30 (8), 247–257.

Taguchi, G., 1987. System of Experimental Designs, 2 Vols., UNIPUB, New York.

Taguchi, G., Wu, Y., 1980. Introduction to Off-Line Quality Control, Central Japan Quality Control Association, Nagoya, Japan.

Wang, P.C., 1989. Tests for dispersion effects from orthogonal arrays. Comput. Statist. Data Anal. 8, 109–117.

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