

## Chapter 6. Supplemental Text Material

### S6-1. Factor Effect Estimates are Least Squares Estimates

We have given heuristic or intuitive explanations of how the estimates of the factor effects are obtained in the textbook. Also, it has been pointed out that in the regression model representation of the  $2^k$  factorial, the regression coefficients are exactly one-half the effect estimates. It is straightforward to show that the model coefficients (and hence the effect estimates) are least squares estimates.

Consider a  $2^2$  factorial. The regression model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \varepsilon_i$$

The data for the  $2^2$  experiment is shown in the following table:

Run, $i$	$X_{i1}$	$X_{i2}$	$X_{i1}X_{i2}$	Response total
1	-1	-1	1	(1)
2	1	-1	-1	$a$
3	-1	1	-1	$b$
4	1	1	1	$ab$

The least squares estimates of the model parameters  $\beta$  are chosen to minimize the sum of the squares of the model errors:

$$L = \sum_{i=1}^4 (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_{12} x_{i1} x_{i2})^2$$

It is straightforward to show that the least squares normal equations are

$$\begin{aligned} 4\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^4 x_{i1} + \hat{\beta}_2 \sum_{i=1}^4 x_{i2} + \hat{\beta}_{12} \sum_{i=1}^4 x_{i1} x_{i2} &= (1) + a + b + ab \\ \hat{\beta}_0 \sum_{i=1}^4 x_{i1} + \hat{\beta}_1 \sum_{i=1}^4 x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^4 x_{i1} x_{i2} + \hat{\beta}_{12} \sum_{i=1}^4 x_{i1}^2 x_{i2} &= -(1) + a - b + ab \\ \hat{\beta}_0 \sum_{i=1}^4 x_{i2} + \hat{\beta}_1 \sum_{i=1}^4 x_{i1} x_{i2} + \hat{\beta}_2 \sum_{i=1}^4 x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^4 x_{i1} x_{i2}^2 &= -(1) - a + b + ab \\ \hat{\beta}_0 \sum_{i=1}^4 x_{i1} x_{i2} + \hat{\beta}_1 \sum_{i=1}^4 x_{i1}^2 x_{i2} + \hat{\beta}_2 \sum_{i=1}^4 x_{i1} x_{i2}^2 + \hat{\beta}_{12} \sum_{i=1}^4 x_{i1}^2 x_{i2}^2 &= (1) - a - b + ab \end{aligned}$$

Now since  $\sum_{i=1}^4 x_{i1} = \sum_{i=1}^4 x_{i2} = \sum_{i=1}^4 x_{i1} x_{i2} = \sum_{i=1}^4 x_{i1}^2 x_{i2} = \sum_{i=1}^4 x_{i1} x_{i2}^2 = 0$  because the design is orthogonal, the normal equations reduce to a very simple form:

$$\begin{aligned}
4\hat{\beta}_0 &= (1) + a + b + ab \\
4\hat{\beta}_1 &= -(1) + a - b + ab \\
4\hat{\beta}_2 &= -(1) - a + b + ab \\
4\hat{\beta}_{12} &= (1) - a - b + ab
\end{aligned}$$

The solution is

$$\begin{aligned}
\hat{\beta}_0 &= \frac{[(1) + a + b + ab]}{4} \\
\hat{\beta}_1 &= \frac{[-(1) + a - b + ab]}{4} \\
\hat{\beta}_2 &= \frac{[-(1) - a + b + ab]}{4} \\
\hat{\beta}_{12} &= \frac{[(1) - a - b + ab]}{4}
\end{aligned}$$

These regression model coefficients are exactly one-half the factor effect estimates. Therefore, the effect estimates are least squares estimates. We will show this in a more general manner in Chapter 10.

### S6-2. Yates's Method for Calculating Effect Estimates

While we typically use a computer program for the statistical analysis of a  $2^k$  design, there is a very simple technique devised by Yates (1937) for estimating the effects and determining the sums of squares in a  $2^k$  factorial design. The procedure is occasionally useful for manual calculations, and is best learned through the study of a numerical example.

Consider the data for the  $2^3$  design in Example 6-1. These data have been entered in Table 1 below. The treatment combinations are always written down in standard order, and the column labeled "Response" contains the corresponding observation (or total of all observations) at that treatment combination. The first half of column (1) is obtained by adding the responses in adjacent pairs. The second half of column (1) is obtained by changing the sign of the first entry in each of the pairs in the Response column and adding the adjacent pairs. For example, in column (1) we obtain for the fifth entry  $5 = -(-4) + 1$ , for the sixth entry  $6 = -(-1) + 5$ , and so on.

Column (2) is obtained from column (1) just as column (1) is obtained from the Response column. Column (3) is obtained from column (2) similarly. In general, for a  $2^k$  design we would construct  $k$  columns of this type. Column (3) [in general, column ( $k$ )] is the contrast for the effect designated at the beginning of the row. To obtain the estimate of the effect, we divide the entries in column (3) by  $n2^{k-1}$  (in our example,  $n2^{k-1} = 8$ ). Finally, the sums of squares for the effects are obtained by squaring the entries in column (3) and dividing by  $n2^k$  (in our example,  $n2^k = (2)2^3 = 16$ ).

**Table 1.** Yates's Algorithm for the Data in Example 6-1

Treatment	Response	(1)	(2)	(3)	Effect	Estimate of Effect	Sum of Squares
Combination						$(3) \div n 2^{k-1}$	$(3)^2 \div n 2^{k-1}$
(1)	-4	-3	1	16	<i>I</i>	---	---
<i>a</i>	1	4	15	24	<i>A</i>	3.00	36.00
<i>b</i>	-1	2	11	18	<i>B</i>	2.25	20.25
<i>ab</i>	5	13	13	6	<i>AB</i>	0.75	2.25
<i>c</i>	-1	5	7	14	<i>C</i>	1.75	12.25
<i>ac</i>	3	6	11	2	<i>AC</i>	0.25	0.25
<i>bc</i>	2	4	1	4	<i>BC</i>	0.50	1.00
<i>abc</i>	11	9	5	4	<i>ABC</i>	0.50	1.00

The estimates of the effects and sums of squares obtained by Yates' algorithm for the data in Example 6-1 are in agreement with the results found there by the usual methods. Note that the entry in column (3) [in general, column (*k*)] for the row corresponding to (1) is always equal to the grand total of the observations.

In spite of its apparent simplicity, it is notoriously easy to make numerical errors in Yates's algorithm, and we should be extremely careful in executing the procedure. As a partial check on the computations, we may use the fact that the sum of the squares of the elements in the *j*th column is  $2^j$  times the sum of the squares of the elements in the response column. Note, however, that this check is subject to errors in sign in column *j*. See Davies (1956), Good (1955, 1958), Kempthorne (1952), and Rayner (1967) for other error-checking techniques.

### S6-3. A Note on the Variance of a Contrast

In analyzing  $2^k$  factorial designs, we frequently construct a normal probability plot of the factor effect estimates and visually select a tentative model by identifying the effects that appear large. These effect estimates are typically relatively far from the straight line passing through the remaining plotted effects.

This method works nicely when (1) there are not many significant effects, and (2) when all effect estimates have the same variance. It turns out that all contrasts computed from a  $2^k$  design (and hence all effect estimates) have the same variance even if the individual observations have different variances. This statement can be easily demonstrated.

Suppose that we have conducted a  $2^k$  design and have responses  $y_1, y_2, \dots, y_{2^k}$  and let the variance of each observation be  $\sigma_1^2, \sigma_2^2, \dots, \sigma_{2^k}^2$  respectively. Now each effect estimate is a linear combination of the observations, say

$$Effect = \frac{\sum_{i=1}^{2^k} c_i y_i}{2^k}$$

where the contrast constants  $c_i$  are all either  $-1$  or  $+1$ . Therefore, the variance of an effect estimate is

$$\begin{aligned} V(\text{Effect}) &= \frac{1}{(2^k)^2} \sum_{i=1}^{2^k} c_i^2 V(y_i) \\ &= \frac{1}{(2^k)^2} \sum_{i=1}^{2^k} c_i^2 \sigma_i^2 \\ &= \frac{1}{(2^k)^2} \sum_{i=1}^{2^k} \sigma_i^2 \end{aligned}$$

because  $c_i^2 = 1$ . Therefore, all contrasts have the same variance. If each observation  $y_i$  in the above equations is the total of  $n$  replicates at each design point, the result still holds.

#### S6-4. The Variance of the Predicted Response

Suppose that we have conducted an experiment using a  $2^k$  factorial design. We have fit a regression model to the resulting data and are going to use the model to predict the response at locations of interest inside the design space  $-1 \leq x_i \leq +1$ ,  $i = 1, 2, \dots, k$ . What is the variance of the predicted response at the point of interest, say  $\mathbf{x}' = [x_1, x_2, \dots, x_k]$ ?

Problem 6-32 asks the reader to answer this question, and while the answer is given in the Instructors Resource CD, we also give the answer here because it is useful information. Assume that the design is balanced and every treatment combination is replicated  $n$  times. Since the design is orthogonal, it is easy to find the variance of the predicted response.

We consider the case where the experimenters have fit a “main effects only” model, say

$$\hat{y}(\mathbf{x}) \equiv \hat{y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i$$

Now recall that the variance of a model regression coefficient is  $V(\hat{\beta}) = \frac{\sigma^2}{n2^k} = \frac{\sigma^2}{N}$ ,

where  $N$  is the total number of runs in the design. The variance of the predicted response is

$$\begin{aligned} V[\hat{y}(\mathbf{x})] &= V\left(\hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i\right) \\ &= V(\hat{\beta}_0) + \sum_{i=1}^k V(\hat{\beta}_i x_i) \\ &= V(\hat{\beta}_0) + \sum_{i=1}^k x_i^2 V(\hat{\beta}_i) \\ &= \frac{\sigma^2}{N} + \frac{\sigma^2}{N} \sum_{i=1}^k x_i^2 \\ &= \frac{\sigma^2}{N} \left(1 + \sum_{i=1}^k x_i^2\right) \end{aligned}$$

In the above development we have used the fact that the design is orthogonal, so there are no nonzero covariance terms when the variance operator is applied

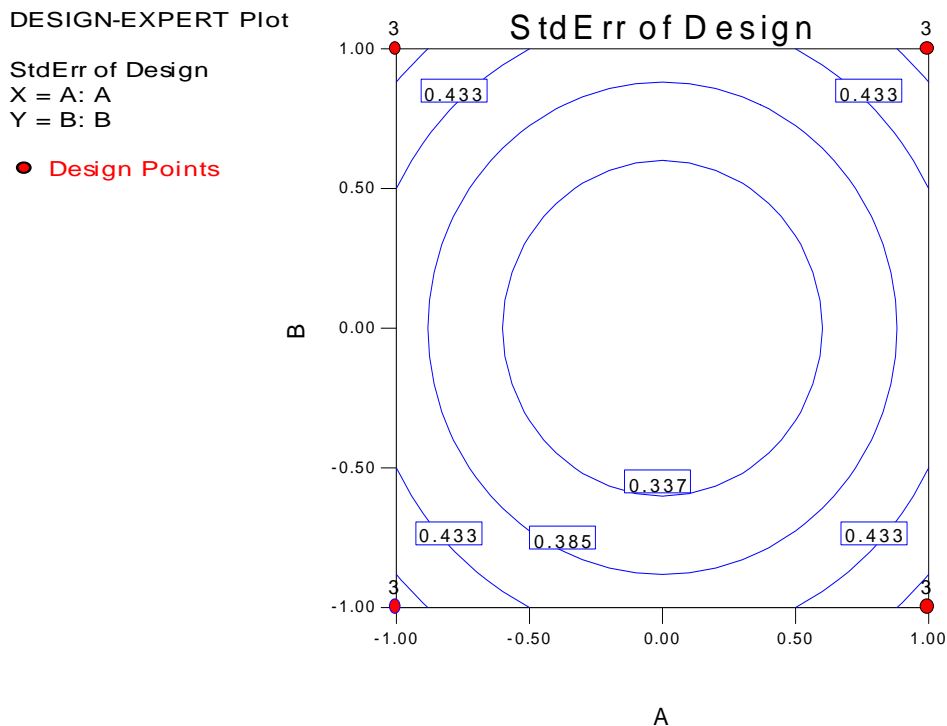
The Design-Expert software program plots contours of the standard deviation of the predicted response; that is the square root of the above expression. If the design has already been conducted and analyzed, the program replaces  $\sigma^2$  with the error mean square, so that the plotted quantity becomes

$$\sqrt{\widehat{V}[\hat{y}(\mathbf{x})]} = \sqrt{\frac{MS_E}{N} \left( 1 + \sum_{i=1}^k x_i^2 \right)}$$

If the design has been constructed but the experiment has not been performed, then the software plots (on the design evaluation menu) the quantity

$$\sqrt{\frac{V[\hat{y}(\mathbf{x})]}{\sigma^2}} = \sqrt{\frac{1}{N} \left( 1 + \sum_{i=1}^k x_i^2 \right)}$$

which can be thought of as a **standardized** standard deviation of prediction. To illustrate, consider a  $2^2$  with  $n = 3$  replicates, the first example in Section 6-2. The plot of the standardized standard deviation of the predicted response is shown below.



The contours of constant standardized standard deviation of predicted response should be exactly circular, and they should be a maximum within the design region at the point  $x_1 = \pm 1$  and  $x_2 = \pm 1$ . The maximum value is

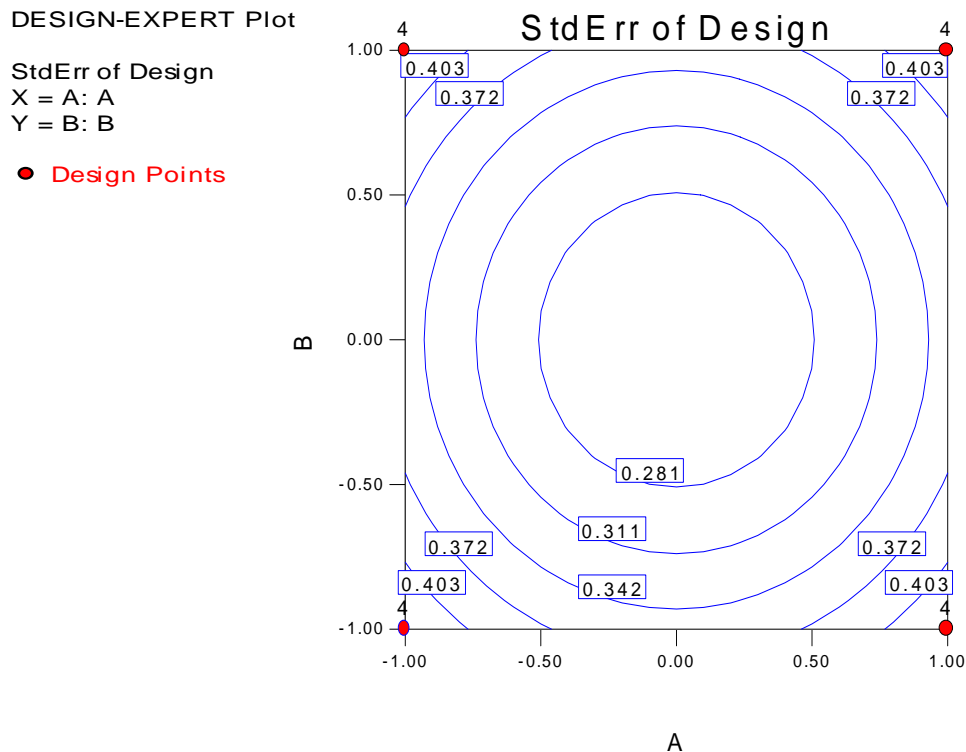
$$\begin{aligned}\sqrt{\frac{V[\hat{y}(\mathbf{x} = 1)]}{\sigma^2}} &= \sqrt{\frac{1}{12}(1 + (1)^2 + (1)^2)} \\ &= \sqrt{\frac{3}{12}} \\ &= 0.5\end{aligned}$$

This is also shown on the graph at the corners of the square.

Plots of the standardized standard deviation of the predicted response can be useful in comparing designs. For example, suppose the experimenter in the above situation is considering adding a fourth replicate to the design. The maximum standardized prediction standard deviation in the region now becomes

$$\begin{aligned}\sqrt{\frac{V[\hat{y}(\mathbf{x} = 1)]}{\sigma^2}} &= \sqrt{\frac{1}{12}(1 + (1)^2 + (1)^2)} \\ &= \sqrt{\frac{3}{16}} \\ &= 0.433\end{aligned}$$

The plot of the standardized prediction standard deviation is shown below.



Notice that adding another replicate has reduced the maximum prediction variance from  $(0.5)^2 = 0.25$  to  $(0.433)^2 = 0.1875$ . Comparing the two plots shown above reveals that the standardized prediction standard deviation is uniformly lower throughout the design region when an additional replicate is run.

Sometimes we like to compare designs in terms of **scaled prediction variance**, defined as

$$\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2}$$

This allows us to evaluate designs that have different numbers of runs. Since adding replicates (or runs) to a design will generally always make the prediction variance get smaller, the scaled prediction variance allows us to examine the prediction variance on a *per-observation basis*. Note that for a  $2^k$  factorial and the “main effects only” model we have been considering, the scaled prediction variance is

$$\begin{aligned} \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= \left( 1 + \sum_{i=1}^k x_i^2 \right) \\ &= (1 + \rho^2) \end{aligned}$$

where  $\rho^2$  is the distance of the design point where prediction is required from the center of the design space ( $\mathbf{x} = \mathbf{0}$ ). Notice that the  $2^k$  design achieves this scaled prediction variance *regardless* of the number of replicates. The maximum value that the scaled prediction variance can have over the design region is

$$\text{Max} \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} = (1 + k)$$

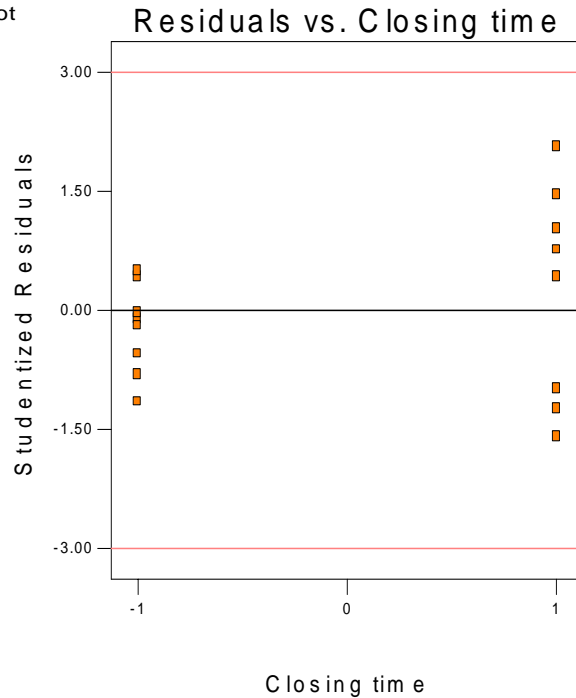
It can be shown that no other design over this region can achieve a smaller maximum scaled prediction variance, so the  $2^k$  design is in some sense an **optimal design**. We will discuss optimal designs more in Chapter 11.

### S6-5. Using Residuals to Identify Dispersion Effects

We illustrated in Example 6-4 (Section 6-5 on unreplicated designs) that plotting the residuals from the regression model versus each of the design factors was a useful way to check for the possibility of dispersion effects. These are factors that influence the variability of the response, but which have little effect on the mean. A method for computing a measure of the dispersion effect for each design factor and interaction that can be evaluated on a normal probability plot was also given. However, we noted that these residual analyses are fairly sensitive to correct specification of the location model. That is, if we leave important factors out of the regression model that describes the mean response, then the residual plots may be unreliable.

To illustrate, reconsider Example 6-4, and suppose that we leave out one of the important factors,  $C = \text{Resin flow}$ . If we use this incorrect model, then the plots of the residuals versus the design factors look rather different than they did with the original, correct model. In particular, the plot of residuals versus factor  $D = \text{Closing time}$  is shown below.

DESIGN-EXPERT Plot  
Defects



This plot indicates that factor  $D$  has a potential dispersion effect. The normal probability plot of the dispersion statistic  $F_i^*$  in Figure 6-28 clearly reveals that factor  $B$  is the only factor that has an effect on dispersion. Therefore, if you are going to use model residuals to search for dispersion effects, it is really important to select the *right* model for the location effects.

### S6-6. Center Points versus Replication of Factorial Points

In some design problems an experimenter may have a choice of replicating the corner or “cube” points in a  $2^k$  factorial, or placing replicate runs at the design center. For example, suppose our choice is between a  $2^2$  with  $n = 2$  replicates at each corner of the square, or a single replicate of the  $2^2$  with  $n_c = 4$  center points.

We can compare these designs in terms of prediction variance. Suppose that we plan to fit the first-order or “main effects only” model

$$\hat{y}(\mathbf{x}) \equiv \hat{y} = \hat{\beta}_0 + \sum_{i=1}^2 \hat{\beta}_i x_i$$

If we use the replicated design the scaled prediction variance is (see Section 6-4 above):

$$\begin{aligned} \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= \left( 1 + \sum_{i=1}^2 x_i^2 \right) \\ &= (1 + \rho^2) \end{aligned}$$



Now consider the prediction variance when the design with center points is used. We have

$$\begin{aligned}
 V[\hat{y}(\mathbf{x})] &= V\left(\hat{\beta}_0 + \sum_{i=1}^2 \hat{\beta}_i x_i\right) \\
 &= V(\hat{\beta}_0) + \sum_{i=1}^2 V(\hat{\beta}_i x_i) \\
 &= V(\hat{\beta}_0) + \sum_{i=1}^k x_i^2 V(\hat{\beta}_i) \\
 &= \frac{\sigma^2}{8} + \frac{\sigma^2}{4} \sum_{i=1}^2 x_i^2 \\
 &= \frac{\sigma^2}{8} \left(1 + 2 \sum_{i=1}^k x_i^2\right) \\
 &= \frac{\sigma^2}{8} (1 + 2\rho^2)
 \end{aligned}$$

Therefore, the scaled prediction variance for the design with center points is

$$\frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} = (1 + 2\rho^2)$$

Clearly, replicating the corners in this example outperforms the strategy of replicating center points, at least in terms of scaled prediction variance. At the corners of the square, the scaled prediction variance for the replicated factorial is

$$\begin{aligned}
 \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= (1 + \rho^2) \\
 &= (1 + 2) \\
 &= 3
 \end{aligned}$$

while for the factorial design with center points it is

$$\begin{aligned}
 \frac{NV[\hat{y}(\mathbf{x})]}{\sigma^2} &= (1 + 2\rho^2) \\
 &= (1 + 2(2)) \\
 &= 5
 \end{aligned}$$

However, prediction variance might not tell the complete story. If we only replicate the corners of the square, we have no way to judge the lack of fit of the model. If the design has center points, we can check for the presence of pure quadratic (second-order) terms, so the design with center points is likely to be preferred if the experimenter is at all uncertain about the order of the model he or she should be using.

### S6-7. Testing for “Pure Quadratic” Curvature using a $t$ -Test

In Section 6-6 of the textbook we discuss the addition of center points to a  $2^k$  factorial design. This is a very useful idea as it allows an estimate of “pure error” to be obtained even though the factorial design points are not replicated and it permits the experimenter to obtain an assessment of model adequacy with respect to certain second-order terms. Specifically, we present an  $F$ -test for the hypotheses

$$\begin{aligned}H_0 &: \beta_{11} + \beta_{22} + \cdots + \beta_{kk} = 0 \\H_1 &: \beta_{11} + \beta_{22} + \cdots + \beta_{kk} \neq 0\end{aligned}$$

An equivalent  $t$ -statistic can also be employed to test these hypotheses. Some computer software programs report the  $t$ -test instead of (or in addition to) the  $F$ -test. It is not difficult to develop the  $t$ -test and to show that it is equivalent to the  $F$ -test.

Suppose that the appropriate model for the response is a complete quadratic polynomial and that the experimenter has conducted an unreplicated full  $2^k$  factorial design with  $n_F$  design points plus  $n_C$  center points. Let  $\bar{y}_F$  and  $\bar{y}_C$  represent the averages of the responses at the factorial and center points, respectively. Also let  $\hat{\sigma}^2$  be the estimate of the variance obtained using the center points. It is easy to show that

$$\begin{aligned}E(\bar{y}_F) &= \frac{1}{n_F}(n_F\beta_0 + n_F\beta_{11} + n_F\beta_{22} + \cdots + n_F\beta_{kk}) \\&= \beta_0 + \beta_{11} + \beta_{22} + \cdots + \beta_{kk}\end{aligned}$$

and

$$\begin{aligned}E(\bar{y}_C) &= \frac{1}{n_C}(n_C\beta_0) \\&= \beta_0\end{aligned}$$

Therefore,

$$E(\bar{y}_F - \bar{y}_C) = \beta_{11} + \beta_{22} + \cdots + \beta_{kk}$$

and so we see that the difference in averages  $\bar{y}_F - \bar{y}_C$  is an unbiased estimator of the **sum** of the pure quadratic model parameters. Now the variance of  $\bar{y}_F - \bar{y}_C$  is

$$V(\bar{y}_F - \bar{y}_C) = \sigma^2 \left( \frac{1}{n_F} + \frac{1}{n_C} \right)$$

Consequently, a test of the above hypotheses can be conducted using the statistic

$$t_0 = \frac{\bar{y}_F - \bar{y}_C}{\sqrt{\hat{\sigma}^2 \left( \frac{1}{n_F} + \frac{1}{n_C} \right)}}$$

which under the null hypothesis follows a  $t$  distribution with  $n_C - 1$  degrees of freedom. We would reject the null hypothesis (that is, no pure quadratic curvature) if  $|t_0| > t_{\alpha/2, n_C - 1}$ .

This  $t$ -test is equivalent to the  $F$ -test given in the book. To see this, square the  $t$ -statistic above:

$$t_0^2 = \frac{(\bar{y}_F - \bar{y}_C)^2}{\hat{\sigma}^2 \left( \frac{1}{n_F} + \frac{1}{n_C} \right)}$$

$$= \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{(n_F + n_C) \hat{\sigma}^2}$$

This ratio is computationally identical to the  $F$ -test presented in the textbook. Furthermore, we know that the square of a  $t$  random variable with (say)  $\nu$  degrees of freedom is an  $F$  random variable with 1 numerator and  $\nu$  denominator degrees of freedom, so the  $t$ -test for “pure quadratic” effects is indeed equivalent to the  $F$ -test.

### Supplemental References

- Good, I. J. (1955). “The Interaction Algorithm and Practical Fourier Analysis”. *Journal of the Royal Statistical Society, Series B*, Vol. 20, pp. 361-372.
- Good, I. J. (1958). Addendum to “The Interaction Algorithm and Practical Fourier Analysis”. *Journal of the Royal Statistical Society, Series B*, Vol. 22, pp. 372-375.
- Rayner, A. A. (1967). “The Square Summing Check on the Main Effects and Interactions in a  $2^n$  Experiment as Calculated by Yates’ Algorithm”. *Biometrics*, Vol. 23, pp. 571-573.