Portable Power

Robert E. Wheeler

E. I. du Pont de Nemours and Co., Inc.
Engineering Department
Wilmington, Delaware

In applications one must relate the probability constants α and β to differences of practical importance in the field of application. This paper states and proves a fundamental theorem which establishes this relationship and then derives elementary formulae for making power calculations in the most frequently encountered situations such as for main effects, interactions, regression coefficients and response surfaces. After an examination of the sensitivity of power calculations to ranges of α, β, and error degrees of freedom, modifications are made in the formulae which allow them to be used without reference to tables and charts: i.e., they are made portable.

KEY WORDS

Power
Experimental Design
Efficiency
Sample Size

1. INTRODUCTION

Probably the single most important question an applied statistician is called upon to answer is "how many observations do I need" It is sometimes possible to take this question under advisement and return to one's office and consult whatever tables and charts one customarily uses, but in most cases a practicing consultant really needs a portable method which will enable him to juggle the following two large unknowns and two major problems in an on-the-spot conference with a client: Unknown 1. His estimate of σ is seldom within a factor of 2 of the true value. Unknown 2. He is usually unable to specify precisely the experimental design that he will finally produce. Problem 1. He must advise on the return in terms of detectable differences for the investment in terms of numbers of observations, as a function of the number of variables, the number of levels of the variables, etc. Problem 2. He must not grossly underestimate the numbers of observations that will be required for the experimental design that he will produce: increasing the number of observations in a proposed experiment after the conference is embarrassing and difficult, if not impossible on occasion.

Such portable methods are possible, and the most useful one of them is expressed in the following formula:

$$n = (4r\sigma/\Delta)^2, \tag{1.1}$$

where $r \geq 1$ is the number of levels of a factor, σ^2 is the variance of the observations, n is the total number of observations in the experiment, and $\Delta > 0$ is the minimum absolute pairwise difference between the expected values of the means of the r − level factor that one desires to detect with an $\alpha = 0.05$ level test and a power of $\beta = 0.90$.

This equation and others related to it are what one really needs in the applications of statistics, for one must be able to translate the probability constants α and β into constants like Δ which are interpretable in the field of application. Without such a translation in the planning stages, it is all too easy to over- or underpower an experiment by choosing the number of observations according to the availability of a design. It will be our purpose in this paper to describe a method for this translation which is applicable to a wide variety of situations and which, at the same time, will be portable in the sense that one can use it in the midst of a consultation with no tools other than perhaps a pocket calculator.

We provide formulas by which the most precise calculations desired may be made. Indeed, as it should, changes of symbols and manipulations with one of our formulas, formula (3.3), may be used to obtain a formula that has appeared in the literature for the main effect situation: see Pearson and Hartley (1951), or Scheffé (1959, Section 3.3). Our formulas for other situations (regression coefficients, interactions, response surfaces) are, of course, quite new. We do not, however, think it always desirable to use the precise formulas because 1) their precision is in part illusionary due to poorly known parameter values, and 2) their richness produces client-consultant dialogues which are counter-productive: e.g., should the power be 0.90, 0.95, or 0.99?

In line with this, we have set certain parameters

Received Dec. 1971; revised Nov. 1973.

in our precise formulas to nominal values and called them portable. In particular (1.1) results from fixing α at 0.05, β at 0.90, and ν_2, the number of degrees of freedom for error, at 5. This works quite well for most situations and in particular for the common sort of multifactor experiment. An alternative formula (3.3)″ is provided in Section 6 for large ν_2.

Specialized experimental situations are of course possible, and for such, suitable portable formulas may be worked out from the precise formulas given here. For instance, Kastenbaum et al (1970) and Bratcher et al (1970) provide tables for a class of designs which appear more in textbooks than in practice: the single factor design. For such a class, an examination of the behavior of the function φ in (3.3) when ν_2 is equal to $n - r$ and $\alpha = 0.05, \beta = 0.90$ results in the following portable formula:

$$n = (3.6r\sigma/\Delta)^2/(r - 1)^{\frac{1}{2}} \qquad (1.2)$$

We will not pursue this line further, since each applied statistician will have a class of experimental designs that he favors, and if the portable formulas offered here do not satisfy his needs he may build others to suit from the precise formulas given. We strongly urge that each use of a portable formula in a consultation be followed by a calculation with the appropriate precise formula when the consultation is over. In this way each applied statistician can learn of the range of applicability of the portable formulas for his class of problems.

In Section 2 we state and prove a fundamental theorem relating minimum power and values of linear functionals. In Section 3 we apply this theorem in a variety of contexts, showing how to make exact power calculations for main effects, interactions, regression coefficients, and response surfaces. In Section 4 we discuss the effect of the efficiency of the experimental design on power calculations. In Section 5 we explore the sensitivity of power calculations to the values of α, β, and ν_2, and conclude that when one takes the usual indeterminacy of σ into account, that α and β can vary to a considerable degree without substantially changing the total number of observations required in an experiment, and that any ν_2 much greater than 5 is wasteful. In Section 6 this lack of sensitivity guides us in our selection of fixed values to make our formulas portable. Finally, in Section 7 we give an example which embodies our ideas of good practice and the proper use of the formulas.

2. The Theorem

In preparation for the theorem, consider a factorial experiment and a factor with r levels. For some specified contrast among the means of the r levels, one would like the usual α level F-test to be significant with at least probability β whenever the expected value of this contrast exceeds a quantity $\Delta > 0$ in absolute value. This desideratum cannot, however, be achieved because β is not uniquely determined by α, Δ, and the specified contrast; β also depends on the expected values of the other contrasts among the means of the r levels. One thing that can be achieved is a relation ketween the specified contrast, α, Δ, and the *minimum* value of β over all values of the other contrasts; and this is what the following theorem gives in a slightly more general setting. The theorem's use is obvious: let i index some finite set of contrasts of interest among the means of the r levels, then find $\{\beta_i\}$, the minimum values of β for each i over the other contrasts and adjust the experiment so that the minimum β_i in $\{\beta_i\}$ is adequately large.

We will consider the usual fixed effects model

$$\eta = EY = \mathbf{X}\theta$$

where Y is a vector of n normally and independently distributed observations, each with variance σ^2, $\mathbf{X}$ is an n by k matrix of known values, and θ is a k-element vector of unknown parameters. Let S be a space of estimable parametric functionals of dimension $\nu_1 \leq k$. If we choose as a basis for S, ν_1 functionals $\mu_1, \cdots, \mu_{\nu_1}$ which have statistically independent least squares estimates with unit variances, then the noncentrality parameter δ^2 in the distribution of the F-statistic under the usual alternative hypothesis, "not all of the elements of S are zero", is given by $\delta^2 = \Sigma\mu_i^2$: see Scheffé (1959), pp. 13–14 and 38 for this terminology. The above notation and assumptions will be used without specific mention in later sections: in addition, "functional" is used throughout this paper in preference to the perhaps more familiar "function" to emphasize that the values of the linear functions of interest inhabit vector spaces.

Theorem: If $\hat{\psi}$ is the least squares estimate of $\psi \in S$, then the minimum power when $\psi^2 \geq \Delta^2$ may be determined from

$$\delta^2 = \Delta^2/\sigma_{\hat{\psi}}^2, \qquad (2.1)$$

where $\sigma_{\hat{\psi}}^2$ is the variance of $\hat{\psi}$.

Proof: Minimizing δ^2 is equivalent to minimizing power because power is a monotonically increasing function of δ^2. Let $\{\psi_i\}$, $i = 1, \cdots, \nu_1$ be a basis for S such that $\psi \equiv \psi_1$ and such that the corresponding set of least squares estimates $\{\hat{\psi}_i\}$ are statistically independent. Then

$$(\hat{\psi}_i/\sigma_{\hat{\psi}_i}) \sim N(\psi_i/\sigma_{\hat{\psi}_i}, 1)$$

and

$$\delta^2 = \Sigma(\psi_i^2/\sigma_{\hat{\psi}_i}^2).$$

Clearly δ^2 is minimized for $\psi^2 \geq \Delta^2$ when $\psi_1^2 = \Delta^2$, and $\psi_i^2 = 0$ for $i \neq 1$.

3. Applications

We shall call $\Delta > 0$ the "minimum detectable value", because it is the smallest value that will produce a significant result with at least the probability β: i.e., Δ is the smallest value that can be "detected" with power β. By (2.1) this minimum detectable value may be translated into δ^2 which may then be referred to any of several published tables and charts to obtain β. We shall assume that the reader has Scheffé (1959) at hand, and phrase our discussion in terms of the Fox (1956) and of the Pearson and Hartley (1951) charts reproduced at the rear of this book. These charts tabulate $\varphi = [\delta^2/(\nu_1 + 1)]^{\frac{1}{2}}$ as a function of ν_1, the level α of the test, the probability of rejection β, and the number of degrees of freedom for error ν_2. We will rewrite (2.1) in terms of φ, as

$$\varphi(\nu_1 + 1)^{\frac{1}{2}} = \Delta/\sigma_{\hat{\psi}} \tag{3.1}$$

3.1 *Regression Coefficients*

Let $\{\theta_i\}$, $i = 1, \cdots, \nu_1$, be a set of estimable regression coefficients with least squares estimates $\{\hat{\theta}_i\}$ and covariance matrix $(\sigma^2/n)B = (\sigma^2/n)||b_{ij}||$, $i, j = 1, \cdots, \nu_1$, so $\sigma_{\hat{\theta}_i} = \sigma(b_{ii}/n)^{\frac{1}{2}}$. Substituting in (3.1) and rearranging we have

$$n = \varphi^2\sigma^2 b_{ii}(\nu_1 + 1)/\Delta^2.$$

This can be used as is, but in general it will be difficult to decide on Δ. To assist in this, consider the column of $\mathbf{X}$ associated with, say, θ_i. Let ρ be the range of values in this column; then $\rho\,|\theta_i|$ is the range of the contribution to η, and it is usually quite easy to decide on the smallest value for this, i.e., a value for $\rho\Delta$. Let $\gamma > 0$ be the smallest difference of interest among the elements of η, then $\Delta = \gamma/\rho$, and by substituting into the above formula we have

$$n = \varphi^2\sigma^2 b_{ii}(\nu_1 + 1)\rho^2/\gamma^2. \tag{3.2}$$

Note that φ depends on n, and b_{ii} and ρ may also depend on n.

Later we shall again have occasion to make this distinction between γ and Δ: Δ is the minimum detectable value for a linear functional ψ, but γ is the smallest difference of interest among the elements of η. In general, γ is a function of Δ: in the above case $\gamma = \rho\Delta$. Sometimes the function is the identity, but the distinction between γ and Δ must be kept clearly in mind.

To give an example, suppose one has a univariate quadratic polynomial model, $\theta_1 + \theta_2 x + \theta_3 x^2$, on the interval $[-1, 1]$, then taking data only at the points -1, 0, and 1 has much to recommend it [cf. Guest (1958)], and $\mathbf{X}$ for these three points is

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}.$$

Let us agree to use only experimental designs which are uniform replicates of these 3 points, then n is a multiple of 3, and B is constant with

$$B = \begin{bmatrix} 3 & 0 & -3 \\ 0 & \frac{3}{2} & 0 \\ -3 & 0 & \frac{9}{2} \end{bmatrix}, \text{ and } B^{-1} = \begin{bmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{3} \end{bmatrix};$$

here $nB^{-1} = \mathbf{X}'\mathbf{X}$ is the cross-product matrix of the experimental design. In this case $\nu_1 = 2$ because there are two regression coefficients of interest: linear and quadratic. Suppose changes of less than γ units in the elements of η are of negligible value. From $\mathbf{X}$ we see that $\rho = 2$ for the linear coefficient, while $\rho = 1$ for the quadratic coefficient. Articulating γ and ρ with (3.2), we find

$$n_l = 18\varphi^2\sigma^2/\gamma^2$$
$$n_q = 13.5\varphi^2\sigma^2/\gamma^2,$$

where n_l and n_q are sample sizes implied by the linear and quadratic minimum detectable values, respectively. No matter what values we choose for φ and σ, n_l will always be the larger; hence, if a sample size for the linear coefficient is adequate, so too will it be for the quadratic.

The value of φ must be found from the charts with $\nu_1 = 2$, $\nu_2 = n_l - 3$. Some iteration may be necessary. Usually σ is not known, but this is seldom critical since some guess can always be made and the value of power calculations lies not so much in a nice adjustment of n as it does in a confrontation with reality in terms of contemplated n's which may prove grossly out of line with detectable values.

3.2 *Main Effects*

Consider a replicated $2^2 \cdot 3$ factorial involving 24 observations with variance σ^2. There are 12 degrees of freedom for error; hence $\nu_2 = 12$. The estimable functionals of interest are contrasts, and so for the 3-level factor $\nu_1 = 2$. The contrasts of most interest are usually pairwise differences, and the 3 possible such differences for the 3-level factor all have the same variance, namely $\sigma^2\,(1/8 + 1/8) = \sigma^2/4$. From (3.1)

$$3^{\frac{1}{2}}\varphi = 2\Delta/\sigma,$$

or

$$\Delta = (3/4)^{\frac{1}{2}}\varphi\sigma,$$

and for $\alpha = 0.05$ and $\beta = 0.90$, we find from the Pearson and Hartley charts that φ is approximately 2.3, hence Δ is approximately equal to 2σ. Note that Δ and γ are identical here. This means that if pairwise differences of practical importance are less in absolute value than 2σ, they will not be detectable by this experiment; if the smallest absolute pairwise difference of practical importance is equal to the "true" difference and if this is somewhat less than 2σ, then because power changes rapidly, most realizations of this experiment will fail to produce a statistically significant result.

When all pairwise differences between the means of a $\nu_1 + 1$ level factor have variance

$$2(\nu_1 + 1)\sigma^2/n,$$

then the following general formula may be obtained by simple substitution and manipulation of (3.1):

$$n = 2\varphi^2\sigma^2(\nu_1 + 1)^2/\Delta^2. \tag{3.3}$$

This will occur, for instance, in an m-way layout with equal numbers of observations in the cells. When these assumptions do not hold, (4.4) may be used.

For a 2-level factor, in the above $2^2 \cdot 3$ example, the minimum detectable value from (3.3) is approximately 1.44σ. It is only reasonable that the minimum detectable value should be smaller for a 2-level factor since the means for the 2 levels are each based on 12 observations, while the means for the 3 levels of the 3-level factor are each based on only 8: there are additional more complicated reasons, which we will not go into here.

It is worthwhile noting that for the 3-level factor the 3 pairwise contrasts of interest span the 2 dimensional space, S, of contrasts among these levels but do not form a basis, as did the regression coefficients in example 3.1. It is entirely possible to adjust the power for a set of functionals which span S and to have it inadequate for other functionals in S.

3.3 *Interactions*

Interactions may be treated in a fashion somewhat similar to regression coefficients. The idea is, as always, to relate differences in the elements of η to Δ. Consider the interaction between two factors, and let t be a vector of expected values for the two-way table (i.e., for an unreplicated m-way layout the elements of t would form a subset of the elements of η). Let $\psi = c't$ be an estimable contrast in the ν_1 dimensional interaction space, then $c(c't/c'c)$ is that part of t ascribable to this contrast. If ρ is the range of the elements of c, then $\rho(|c't|/c'c)$ is the range of the contribution to t made by ψ, and it follows that if the smallest interesting range for the elements of t is γ units, then the minimum detectable value desired should be found from

$$\gamma = \rho\Delta/c'c;$$

that is

$$\Delta = \gamma c'c/\rho.$$

Usually, γ will also be the smallest interesting range for the elements of η.

Analogous to pairwise contrasts for main effects, there is a class of "two-level" contrasts which are usually of interest for interactions. For such contrasts, a c will have ± 1 as its nonzero elements in the positions of an m-way table which correspond to a 2^m factorial obtained by ignoring all except two levels for each factor.

For (3.1), after substitution and manipulation, we then have, in general, for m-factor interactions and two-level contrasts

$$n = \varphi^2\sigma^2(\nu_1 + 1)\pi/(\gamma^2 2^{m-2}), \tag{3.4}$$

where π is the number of cells in the m-way table. Formula (4.4) may be used when the assumptions required for (3.4) do not hold.

For the interaction between two 3-level factors, this gives

$$n = 45\varphi^2\sigma^2/\gamma^2.$$

For the interaction between two 2-level factors in a replicated $2^2 \cdot 3$, as in the example of section 3.2, formula (3.4) gives

$$\gamma = \varphi\sigma/3^{\frac{1}{2}},$$

which is 1.44σ for $\alpha = 0.05$, $\beta = 0.90$—the same as for differences in the means of the levels of a 2-level factor. For the interaction between a 2-level and the 3-level factor one finds

$$\gamma = \varphi\sigma(3/4)^{\frac{1}{2}},$$

or for the above α and β

$$\gamma = 2.04\sigma.$$

3.4 *Response Surfaces*

There are many things of interest about a response surface, but surely a minimal requirement is that the difference between the maximum and minimum values on the surface shall be detectable with adequate power. Therefore, consider the functional ψ which is the difference between the responses at the points $u, v \in R$, where R is the experimental region of points x. Assume $\mathbf{X}$ is of full rank, and as in section 3.1 define $nB^{-1} = \mathbf{X}'\mathbf{X}$, and let $\mathbf{x}'$ be the "model expanded" version of $x \in R$ that would be used to form a row of $\mathbf{X}$ if x were a point in the experimental design; e.g., if R is the real line, and the model is an mth degree polynomial, then for $x \varepsilon R$

$$\mathbf{x}' = (1, x, x^2, \cdots, x^m).$$

The variance of $\hat{\psi}$ is then

$$\sigma_{\hat{\psi}}^2 = (\sigma^2/n)(\mathbf{u} - \mathbf{v})'B(\mathbf{u} - \mathbf{v}).$$

Assume that the model has a constant term, then $(\mathbf{u} - \mathbf{v})$ depends only on $k - 1$ parameters, and ψ, therefore, lies in a space of dimension $\nu_1 = k - 1$. The functional ψ is clearly estimable since $(\mathbf{u} - \mathbf{v})'$ lies in the row space of $\mathbf{X}$, as do all k dimensional row vectors when $\mathbf{X}$ is of full rank. From (3.1) we have

$$n = \varphi^2\sigma^2 kw/\Delta^2, \tag{3.5}$$

where $w = (\mathbf{u} - \mathbf{v})'B(\mathbf{u} - \mathbf{v})$, and by choosing w a maximum for $u, v \varepsilon R$ one is assured that the difference between the minimum and maximum responses will be detectable.

An easier-to-evaluate formula may be obtained in terms of the maximum deviation of the response surface from its mean value. Subtract the means of the last $k - 1$ columns of $\mathbf{X}$ from all model expanded points $\mathbf{x}$. This puts B in "correlation" form where

$$B = \begin{bmatrix} 1 & 0 \\ 0 & C \end{bmatrix},$$

and C is $k - 1$ by $k - 1$. The value of w is not changed by such a transformation which is the only thing of concern here. Now if $\mathbf{v}$ is chosen with leading element unity and all others zero, then

$$w = \mathbf{u}'B\mathbf{u} - 1 = d(u) - 1, \text{ say,}$$

and (3.5) becomes

$$n = \varphi^2\sigma^2 k(d(u) - 1)/\Delta^2.$$

Hence, one need only find $d = \max_{u \varepsilon R} d(u)$ to obtain

$$n = \varphi^2\sigma^2 k(d - 1)/\Delta^2, \tag{3.6}$$

which assures that the maximum deviation of the response surface from the mean response in the design will be detectable, since the mean response over the n points in the design is given by the functional with coefficient vector $\mathbf{v}$.

Even more may be done. Kiefer and Wolfowitz (1960) have shown that for D-optimal (approximate theory) designs $d = k$, hence

$$n = \varphi^2\sigma^2 k(k - 1)/\Delta^2. \tag{3.7}$$

Consider the example in section 3.1. If the smallest difference of interest on the response surface is γ, then $\Delta = \gamma/2$ by arguing that γ should be twice Δ, and

$$n = 24\varphi^2\sigma^2/\gamma^2.$$

Compare this with a calculation using (3.6) on the same example. Here B is not in "correlation" form, so $\mathbf{v}$ must be taken as the average of the rows of $\mathbf{X}$, and this is

$$\mathbf{v}' = (1, 0, 2/3).$$

Taking

$$\mathbf{u}' = (1, 1, 1)$$

gives $d(u)$ its maximum value $d = 3$ in R, and hence $w = d - 1 = 2$ and

$$n = 24\varphi^2\sigma^2/\gamma^2 \text{ as above.}$$

These two calculations will agree only when the design is D-optimal, as it is here.

On the other hand, the maximum value of w in R can be shown to be approximately 6.31, and when (3.5) is used, one finds

$$n = 18.9\varphi^2\sigma^2/\gamma^2,$$

which indicates the magnitude of the error in the argument that γ should be twice the Δ in (3.6) and (3.7).

4. Efficiency

The influence of the efficiency of an experimental design on power is only through $\sigma_{\hat{\psi}}^2$ as may be seen from (3.1). In general, $\sigma_{\hat{\psi}}^2$ has the form

$$\sigma_{\hat{\psi}}^2 = (\sigma^2/n)c'Bc, \tag{4.1}$$

where c is the coefficient vector of $\psi = c'\theta$. For any given functional ψ, the design is optimum when $c'Bc$ is a minimum, but there are always several functionals of interest, and so some compromise must be made. Two compromise criteria of great interest are labeled D and G in the literature. The D (for determinant) criterion minimizes the determinant of B. This results in the simultaneous minimization of $c'Bc$ for all c such that $c'c$ is a constant, as may be seen by transforming B to diagonal form. The G (for global) criterion minimizes the maximum value of $\mathbf{x}'B\mathbf{x}$ for all x in the experimental region R. Both criteria are invariant under linear transformation of the column space of $\mathbf{X}$, and moreover, Kiefer and Wolfowitz (1960) have shown that they are equivalent when all convex combinations of matrices B^{-1} are allowed. In addition, they have shown that the minimax value of $\mathbf{x}'B\mathbf{x}$ is k, which means that one has an ever fixed mark against which any candidate design may be judged.

Atwood (1969) has shown how to define efficiency for these criteria and has shown the D-efficiency is always greater than G-efficiency. As before, let $d = \max_{x \varepsilon R} \mathbf{x}'B\mathbf{x}$, then if the maximum variance

$\sigma^2 d/m$ for some design on m observations is equal to the maximum variance $\sigma^2 k/n$ for the optimal design based on n observations,

$$n/m = k/d,$$

and hence the G-efficiency of a design is given by

$$e = k/d;$$

although we prefer to use the inefficiency

$$i = d/k$$

in equations such as (3.6):

$$n = \varphi^2\sigma^2 k(ik - 1)/\Delta^2. \tag{4.2}$$

Main effects and interactions may be treated more completely than was possible in Sections 3.2 and 3.3, which treated pairwise contrasts for main effects and two-level contrasts for interactions, respectively, only in the case of "orthogonal" designs. Defining $\eta = \mathbf{X}\theta$ as in Section 2, suppose that, without loss of generality, the first r columns of $\mathbf{X}$ span the space of a factor (interaction), and partition $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$, $\theta' = (\theta_1', \theta_2')$ and B accordingly so that $\mathbf{X}_1$ is $n \times r$, θ_1 is $r \times 1$, and B_1 is the leading $r \times r$ submatrix of B. This means, when $\mathbf{X}_1'\mathbf{X}_2 = 0$, that $\mathbf{X}_1\theta_1$ will be a column of the expected values of the main effects (interaction effects) for the n observations. In general, $\mathbf{X}_1'\mathbf{X}_2 \neq 0$. The argument leading to (3.6) made use of a linear transformation which caused the first column of $\mathbf{X}$ to become orthogonal to the remaining columns, while leaving the first column of $\mathbf{X}$ unchanged. A similar transformation will be used here, except that $\mathbf{X}_2$ will remain unchanged. The orthogonalizing transformation

$$\mathbf{X}_{(1)} = \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1$$

gives $\mathbf{X}_{(1)}$ as the first r columns of the new $\mathbf{X}$. This leaves B_1 unchanged while the inverse transformation on the θ leaves θ_1 unchanged. It may now be seen that $X_{(1)}\theta_1$ is a column of expected values of the effects for the factor (interaction) of interest. For $x \in R$, the expected value of the effect due to the factor (interaction) of interest is given by $\psi = \mathbf{x}_{(1)}'\theta_1$, where

$$\mathbf{x}_{(1)} = \mathbf{x}_1 - \mathbf{x}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1$$

and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2')$ is the model expanded version of x.

Let $g(x) = \mathbf{x}_{(1)}'B\mathbf{x}_{(1)}$, then from (3.1)

$$n = \varphi^2\sigma^2(r + 1)g(x)/\Delta^2,$$

or by defining $g = \max_{x \in R} g(x)$,

$$n = \varphi^2\sigma^2(r + 1)g/\Delta^2, \tag{4.3}$$

which assures that the maximum, absolute, expected effect due to the factor (interaction) of interest will be detectable. In addition, Kiefer (1961) has shown that when all k parameters are estimable, $g = r$ for a design which is D-optimal for the r parameters, and so by taking $i_0 = g/r$, one has from (4.3)

$$n = \varphi^2\sigma^2(r + 1)i_0 r/\Delta^2 \tag{4.4}$$

For the entire response surface when the model has a constant term, $r = k - 1$, hence

$$n = \varphi^2\sigma^2 k i_0(k - 1)/\Delta^2$$

which may be compared with (4.2) as an indication of the distinction between i and i_0.

Equation (4.4) may be used instead of (3.3) when the factorial design does not possess the orthogonality properties required by (3.3); Δ is then to be interpreted as the minimum desired difference for the maximum, absolute, expected main effect. Similarly, (4.4) may be used instead of (3.4) with the same type of interpretation for Δ. Most common designs are D-optimal and have $i_0 = 1$.

5. Sensitivity

Equation (3.1) indicates something quite important about experimental design. Consider the following table which gives the ratio of φ for $\nu_2 = 6$ to that of φ for $\nu_2 = \infty$ as a function of α, β, and ν_1.

	$\alpha = 0.05$		$\alpha = 0.01$	
	$\beta = 0.50$	$\beta = 0.90$	$\beta = 0.50$	$\beta = 0.90$
$\nu_1 = 2$	1.3	1.3	1.5	1.6
$\nu_1 = 8$	1.7	1.7	2.0	2.1

The change in the minimum detectable value, ascribable to φ, is only 30% between ∞ and 6 degrees of freedom for error for $\beta = 0.90$, $\alpha = 0.05$, and $\nu_1 = 2$. Since the effect of replication on the estimate of σ^2 appears only through φ, it is clear that the function of replication is not to "improve" the estimate of σ^2.

The practical effect of replication is to increase n, thereby making parameter estimates (and cell averages) more precise. This may be seen clearly by rewriting (3.1) using (4.1) with $q = c'Bc$:

$$\Delta = \sigma\varphi[q(\nu_1 + 1)/n]^{\frac{1}{2}}. \tag{5.1}$$

In the case of replication, q will be a constant and Δ inversely proportional to $n^{\frac{1}{2}}$, except for the small changes introduced by φ as ν_2 increases.

From (5.1) we may also observe the effect of ν_1 on Δ: it is directly proportional to $(\nu_1 + 1)^{\frac{1}{2}}$. For a given φ and n, it is harder to detect a given Δ with large ν_1 than with small. This is reasonable since the n observations are parceled out among the ν_1 degrees of freedom; for a $\nu_1 + 1$ level factor, the mean for each level is based on only $n/(\nu_1 + 1)$ observations, and obviously the fewer the levels the better; or, in other words, two-level factors are best.

A similar lack of sensitivity is revealed by the following two tables. The first indicates sensitivity to β by giving the ratio of φ for $\beta = 0.90$ to that of φ for $\beta = 0.50$.

	$\nu_1 = 3$		$\nu_1 = 10$	
	$\nu_2 = 6$	$\nu_2 = \infty$	$\nu_2 = 6$	$\nu_2 = \infty$
$\alpha = 0.05$	1.6	1.6	1.6	1.5
$\alpha = 0.01$	1.5	1.4	1.5	1.3

The second table indicates the sensitivity to α and gives the ratio of φ for $\alpha = 0.01$ to φ for $\alpha = 0.05$.

	$\nu_1 = 3$		$\nu_1 = 10$	
	$\nu_2 = 6$	$\nu_2 = \infty$	$\nu_2 = 6$	$\nu_2 = \infty$
$\beta = 0.90$	1.4	1.2	1.4	1.1
$\beta = 0.50$	1.5	1.3	1.5	1.3

These results are quite impressive. Of course it does matter what α, β, and ν_2 one chooses, but the distinction between, say, $\beta = 0.50$ and $\beta = 0.90$ will seldom be critical, since under almost any circumstance φ and thus Δ will only change by about 60%, and the available guesses at σ will differ by much more than this.

6. Portability

If one has a precise estimate of σ, then clearly a precise power calculation is feasible, and (3.1) or one of its derivative formulas such as (3.2), (3.3), (3.4), or (3.7) may be used. Usually σ is not known with any precision, and it is almost always unknown in a consultation. What is needed in a consultation is a portable procedure that is reasonably accurate and that gives insight into the relations between Δ, ν_1 , and n, since perhaps the most important contribution that an applied statistician will make is to bring his client to a realistic understanding of the "possible".

In view of the results of the preceding section, it is only reasonable to pick some value of φ to use in most situations, and this will, therefore, be the purpose of the present section. In addition, mnemonic definitions of the symbols are given in Table 1 so that this section forms a substantially self-contained souvenir.

6.1 *General*

For $\alpha = 0.05$, $\beta = 0.95$, $\nu_2 = 5$, and $\nu_1 > 0$, φ is approximately 3; therefore, the portable version of (3.1) is

$$\Delta = 3(\nu_1 + 1)^{\frac{1}{2}}\sigma_{\hat{\psi}} . \qquad (3.1)'$$

Multiplying the right-hand side by, say, 0.7 will correct to large ν_2 .

Table 1—*Notation*

Symbol	Mnemonic Definition
σ^2	Variance of the observations
n	Total number of observations
α	Level of the test
β	Power
ψ	Estimable parametric functional
$\hat{\psi}$	Least squares estimate of ψ
ν_1	Numerator degrees of freedom for F-test
ν_2	Denominator degrees of freedom for F-test
Δ	Minimum detectable value
γ	Smallest difference of interest
φ	Noncentrality function
$\sigma_{\hat{\psi}}^2$	Variance of $\hat{\psi}$
π	Number of cells in an "interaction table"
m	Number of factors
i	Inefficiency of the design
k	Number of terms in the "model"

6.2 *Main Effects*

For $\alpha = 0.05$, $\beta = 0.90$, $\nu_2 = 5$, and $\nu_1 > 0$, φ is approximately 2.8, and if the factor has r levels, one obtains from (3.3)

$$n = (4r\sigma/\Delta)^2. \qquad (3.3)'$$

Here $\gamma = \Delta$. For large ν_2 one may use

$$n = (3r\sigma/\Delta)^2 \qquad (3.3)''$$

These are by far the most useful portable expressions, and by a direct derivation from (3.1) it may be seen that they hold, with $r = 1$, for a two-sided test of the hypothesis that a single given ψ is equal to a constant—such as when testing a mean for equality to a specified constant.

6.3 *Interactions*

For (3.4) the substitution $\varphi = 3$, as in (3.1)′, may be used to give

$$n = 9\sigma^2(\nu_1 + 1)\pi/(\gamma^2 2^{m-2}), \qquad (3.4)'$$

although for an m-factor interaction of 2-level factors (3.3)′ and (3.3)″ apply with r replaced by 2. Multiply the right-hand side of (3.4)′ by $\frac{1}{2}$ to correct to large ν_2 .

6.4 *Response Surfaces*

When designing for response surfaces, inefficiency should be taken into account since constraints on the design space and uncertainty about the model will usually lead one to a compromise design. A good value of i to use is 2, hence using $\varphi = 3$, one obtains from (4.2)

$$n = 9\sigma^2 k(2k - 1)/\Delta^2, \qquad (4.2)'$$

with a multiplier of $\frac{1}{2}$ to correct to large ν_2 .

For an r term model which is part of a larger k term model, (4.4) becomes

$$n = 9\sigma^2(r + 1)2r/\Delta^2, \qquad (4.4)'$$

by assuming $i_0 = 2$, with a multiplier of $\frac{1}{2}$ to correct to large ν_2. Here one usually takes $\Delta = \frac{1}{2}\gamma$, although an experimenter may sometimes be able to specify directly the smallest deviation of the response surface from the mean response in the design that will be of interest.

7. Example

Sheets of plastic film are coated and run through two chambers of a heated coating tower to drive off solvent. The problem as posed was to investigate the influence of the temperature, M, in the moist heated first chamber, the temperature, D, in the dry heated second chamber, and the film travel speed, S, on the residual solvent. High throughput was desired, and so learning of speed-temperature interactions was considered as important as learning about main effects.

The experimenter decided that any true difference in responses exceeding 150 was of practical importance ($\gamma = 150$), and he was able to supply data from 7 trial runs which, by regression analysis, provided an estimated σ of 80, based on 3 degrees of freedom. His maximum limit on the number of runs was 20.

Because the process was to be scaled up and because a fundamental understanding of the effects of the variables and their interactions was needed, response surface experiments were not considered appropriate. Consequently, to get a feel for the "possible", a 3^3 factorial design was considered with the thought that the mean square for the 3-way interaction could be used as an 8 degrees of freedom estimate of σ^2. From (3.3)′ for main effects,

$$\Delta = 4\cdot 3\cdot 80/(27)^{1/2} = 185,$$

which was large but not too large; however, from (3.4)′ for interactions,

$$\gamma = 3\cdot 80\cdot (5\cdot 9/27)^{1/2} = 310,$$

which was clearly too large. Not only did the 3^3 have too many runs, but it clearly failed to provide power adequate for the interactions.

Next, a 2×3^2 factorial was considered with the thought that the highest order interaction with 4 degrees of freedom could be used to estimate σ^2. From (3.3)′ for a 3-level factor,

$$\Delta = 4\cdot 3\cdot 80/(18)^{1/2} = 226,$$

while for the 2-level factor,

$$\Delta = 4\cdot 2\cdot 80/(18)^{1/2} = 151.$$

From (3.4)′ for the 3×3 interaction,

$$\gamma = 3\cdot 80\cdot (5\cdot 9/18)^{1/2} = 379,$$

and for a 2×3 interaction,

$$\gamma = 3\cdot 80\cdot (2\cdot 6/18)^{1/2} = 196.$$

By taking S as the 2-level factor, the power for the important interactions would not have been too bad.

Although the 2×3^2 was a possible design, it was felt that a preferable design would be one which allowed freedom for blocking and one which did not confound interaction effects with error effects. A variety of replicated 2^3 in 16 runs seemed to meet these requirements, since it had 6 degrees of freedom from replicate runs to estimate error and would permit blocks of 4 runs each. From (3.3)′ for either main effects or interactions,

$$\Delta = 4\cdot 2\cdot 80/(16)^{1/2} = 160.$$

The 2^3 was used, and an analysis of variance on the resulting data indicated significance only for the main effects of D and S. The estimate of σ was 144 and the M, D, and S means were

	M	D	S
Low level	234	332	136
High level	238	140	336

Using 140 as σ in (3.3)′ gives

$$\Delta = 4\cdot 2\cdot 140/(16)^{1/2} = 280,$$

and so statements of the following form were made about the nonsignificant main effect of M and about the not-significant interactions: "If this source has an effect, it is less than about 280". The omission of such statements (crude though the numbers in them may be) is a major shortcoming of many statistical analyses.

The maximum number of observations tends to be a fixed quantity, but γ is mutable to a degree. Using the initial estimate 80 of σ, a 60% confidence interval for σ would have been [64, 139]. Had the above designs been compared using 139, it is more likely that the 2^3 design would still have been chosen, thus compromising $\gamma = 150$ as specified, than that the experimenter would have decided to run no experiment at all.

8. Acknowledgment

We are extremely grateful for the assistance and advice of a most unusual referee.

REFERENCES

[1] ATWOOD, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Statist. 40*, 1570–1602.

[2] BRATCHER, T. L., MORAN, M. A., and ZIMMER, W. J. (1970). Tables of Sample Sizes in the Analysis of Variance. *J. Qual. Tech. 2*, 156–164.

[3] FOX, M. (1956). Charts of the power of the F-test. *Ann. Math. Statist. 27*, 484–497.

[4] GUEST, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist. 29*, 294–299.

[5] KASTENBAUM, M. A., HOEL, D. G., and BOWMAN, K. O. (1970). Sample size requirements: One-way analysis of variance. *Biometrika. 57*, 421–430.

[6] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist. 32*, 298–325.

[7] KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math. 12*, 363–366.

[8] PEARSON, E. S., and HARTLEY, H. O. (1951). Charts of the power function of the analysis of variance tests, derived from the non-central F-distribution. *Biometrika. 38*, 112–130.

[9] SCHEFFÉ, H. (1959). *The Analysis of Variance*, New York: John Wiley and Sons, Inc.