PSYCHOMETRIKA—VOL. 23, NO. 3
SEPTEMBER, 1958

# POWER FUNCTION CHARTS FOR SPECIFICATION OF SAMPLE SIZE IN ANALYSIS OF VARIANCE

LEONARD S. FELDT
STATE UNIVERSITY OF IOWA
AND
MOHARRAM W. MAHMOUD
EGYPTIAN MINISTRY OF EDUCATION

The specification of sample size is an important aspect of the planning of every experiment. When the investigator intends to use the techniques of analysis of variance in the study of treatments effects, he should, in specifying sample size, take into consideration the power of the $F$ tests which will be made. The charts presented in this paper make possible a simple and direct estimate of the sample size required for $F$ tests of specified power.

A primary consideration in the design of any experiment is the specification of the number of subjects to be selected from the various treatment populations. This number should be such that the important statistical tests will be reasonably sensitive in detecting false null hypotheses. Statistical theory provides the basis for designing such tests; in many psychological and educational experiments sufficient preliminary information is available to permit an application of this theory. The purpose of this paper is to provide power function charts which will simplify the application of the theory and thus facilitate the specification of sample size in experiments employing the techniques of analysis of variance.

The power of the statistical test in any experimental setup—that is, the probability of rejecting the null hypothesis when it is false—depends on the level of significance $\alpha$ at which the test is made, the number of observations or subjects $n$ on which data are available, and the degree of falsity $\phi'$ of the hypothesis under test. The latter factor is defined as the square root of the ratio of the variance of the treatment population means to the variance for error within the treatment populations. Symbolically,

$$\phi' = \sqrt{\frac{\sum_i^k (\mu_i - \mu)^2/k}{\sigma^2}}.$$

For every $F$ test at a given level of significance in any given design, the power $P$ against any specified alternative to the null hypothesis is uniquely determined by the value of $n$. Conversely, for every test there exists a value of $n$ which will result in a test of specified power against a specified alternative.

In those experiments in which the power requirements of the $F$ test can be rationally fixed against a specific alternative, it is possible to determine the appropriate sample size. It is for such situations that the present charts are intended.

*Nature of the Charts*

The charts presented in this paper are for use with tests of the main effects of treatments in experiments involving two to five levels of the treatment variable. The charts are strictly valid only for the completely randomized design; however, they may be applied with relatively little error to tests of treatments effects in randomized block designs and factorial designs employing a within-cells estimate of error variance. A chart presents two families of three curves each. The families pertain to the .05 and .01 levels of significance; the curves within families correspond to power values of .5, .7 and .9.

A separate chart is provided for each value of $k$, the number of levels of the treatment variable, $f_1 = k - 1$, from 2 through 5. The chart and family appropriate for a given experimental test is entered with the parameter $\phi'$ along the abscissa. The value of $n$, the number of observations required per treatment for a test of specified power, is read directly from the ordinate of the chart.

*Historical Development*

The distribution of the $F$ statistic under hypotheses alternative to the null hypothesis was first considered by Fisher [1] and Wishart [9], who derived expressions for the noncentral $F$ distribution in the form of the correlation ratio. Later Tang [8] derived the same result from the distribution of the noncentral $\chi^2$. Tang also presented extensive tables of the power function. These tables are entered with the parameter $\phi$, defined as

$$\phi = \sqrt{\frac{\sum_i^k n(\mu_i - \mu)^2/k}{\sigma^2}}.$$

For fixed values of $\alpha$, $\phi$, $f_1$, and $f_2$ the probability of retaining a false null hypothesis may be determined. Unfortunately, the interval of tabulation for $\phi$ is .50, an interval which is not sufficiently fine for satisfactory interpolation.

Following Tang's procedure, Lehmer [4] tabulated the values of $\phi$ for $\alpha = .05$ and .01, $P = .7$ and .8 over a wide range of $f_1$ and $f_2$. These tables are quite complete within the power range considered; however they cannot be conveniently used in the planning of experiments. From the tables the experimenter can tell only that a projected test will have a power less than .7, between .7 and .8, or greater than .8 against a specified alternative. A greater range of power values would make such tables considerably more useful.

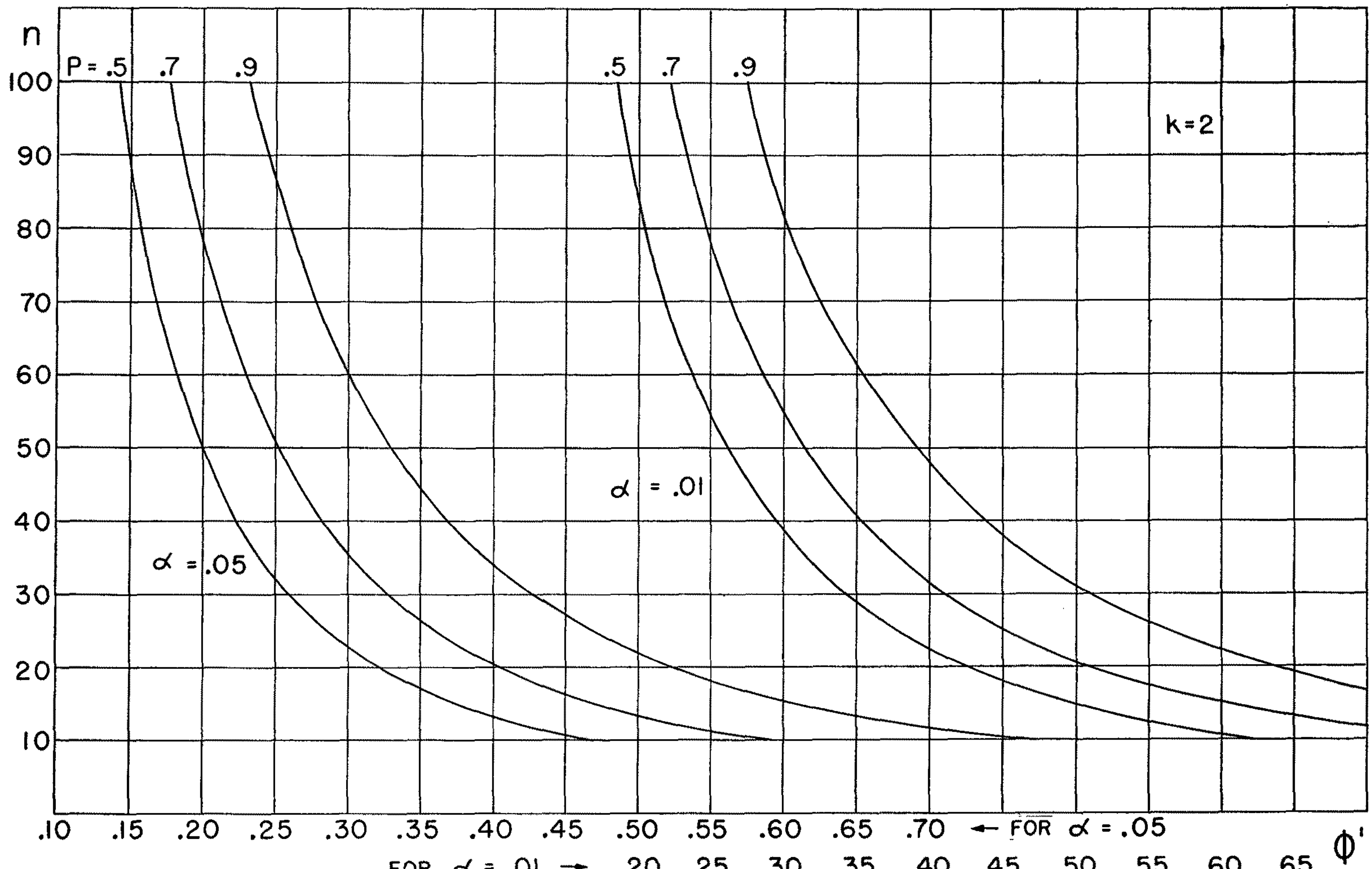


FIGURE 1

Curves of Constant Power for the Test of Main Effects with $k = 2$

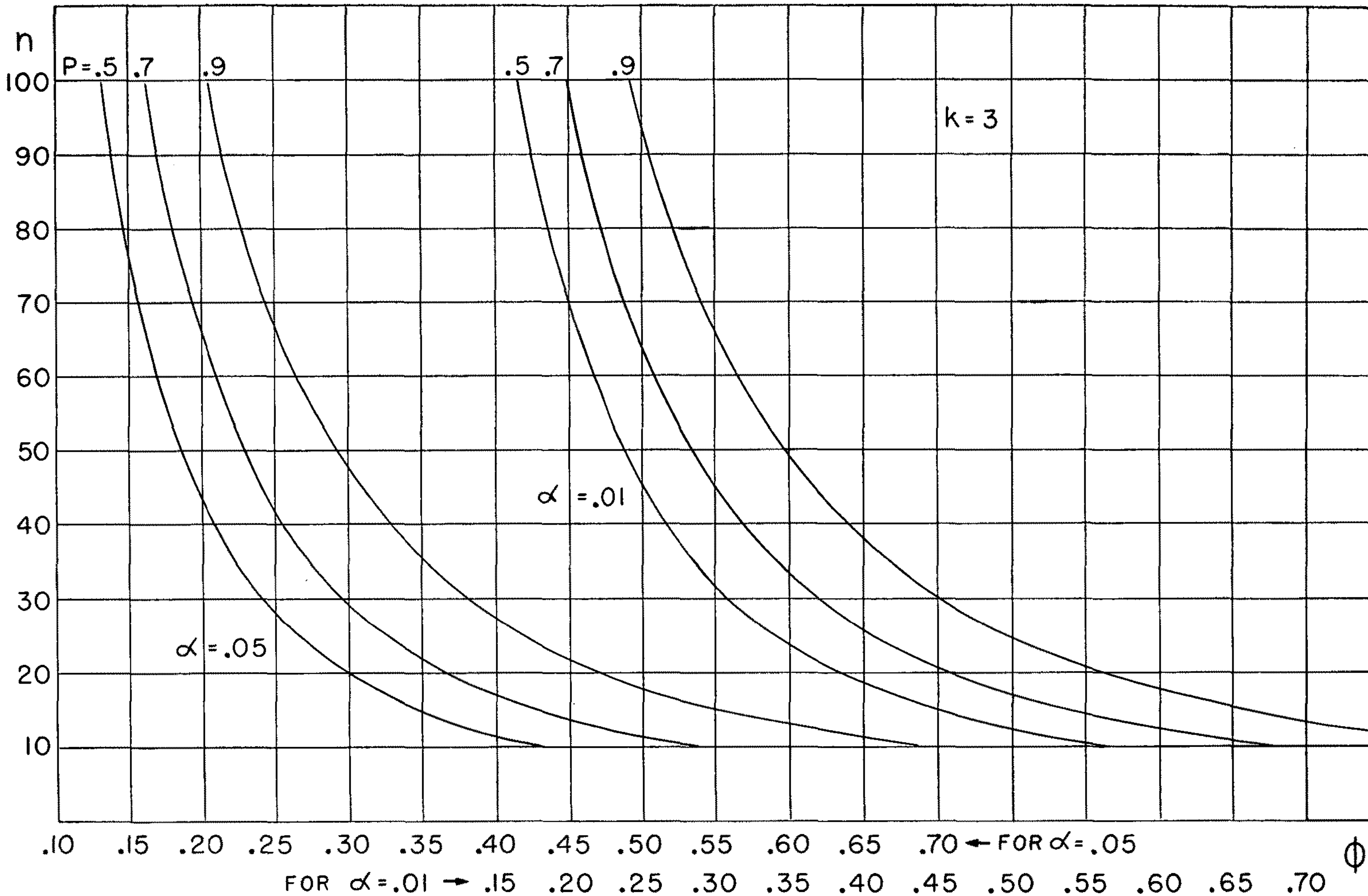


FIGURE 2
Curves of Constant Power for the Test of Main Effects with $k = 3$

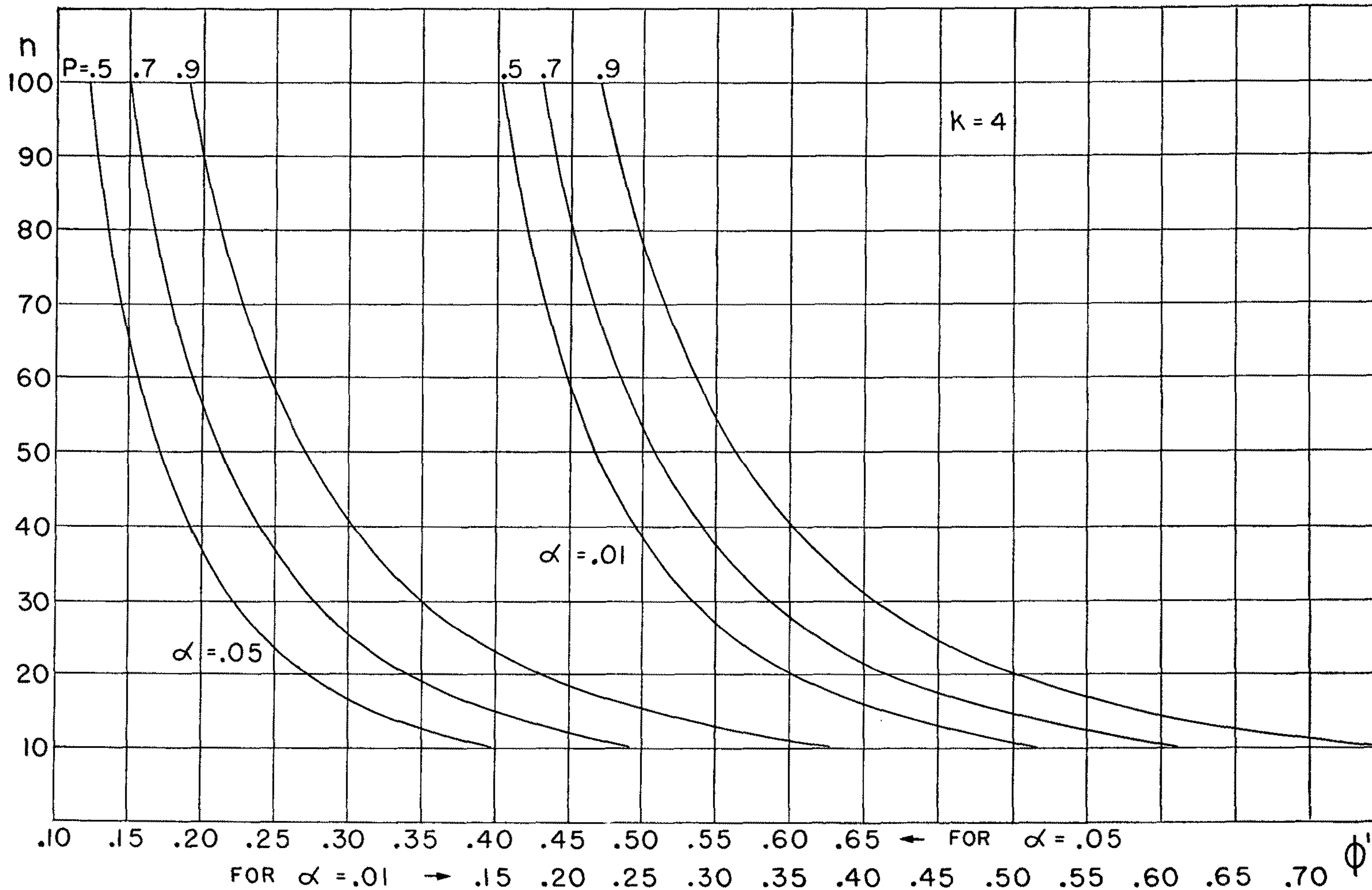


FIGURE 3
Curves of Constant Power for the Test of Main Effects with $k = 4$

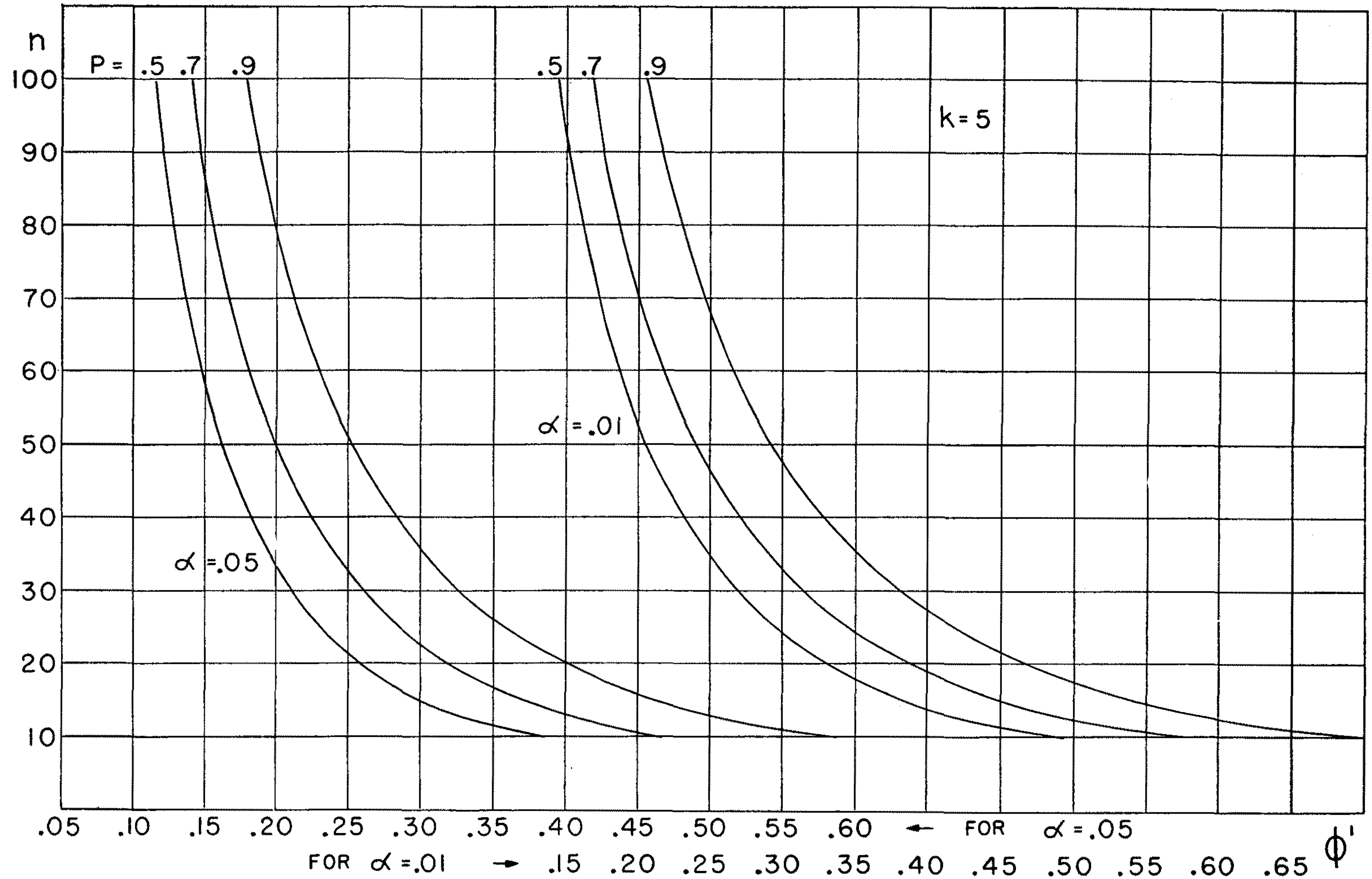


FIGURE 4
Curves of Constant Power for the Test of Main Effects with $k = 5$

Patnaik [6] made an extensive study of the power function of analysis of variance tests. By the method of moments, he derived an approximation to the noncentral $F$ distribution based on the central $F$ distribution. This approximation is computationally feasible but somewhat tedious, especially if a number of power estimates are required. Its primary limitation to psychological experimenters is the labor involved in utilizing Pearson's *Tables of the Incomplete Beta Function* in obtaining power values. This limitation is especially marked in the many instances which demand interpolation within these tables.

Pearson and Hartley [7] presented families of power curves for various combinations of $\alpha$, $f_1$, and $f_2$, which make possible a direct estimate of the power of analysis of variance tests. These curves, like the tables of Tang, are entered with the parameter $\phi$. For any given experimental setup, the power of the test may be read directly from the ordinate of the curve. These curves are well suited to the evaluation of the power of any given test. They cannot be easily employed in the inverse manner, however, to indicate the value of $n$ which should be adopted in order to secure a test of specified power. For this purpose, the experimenter must adopt the relatively inefficient approach of making repeated approximations until the value of $n$ has been estimated with sufficient accuracy.

Nicholson [5] and Hodges [3] have derived general formulas for the computation of the power of the analysis of variance test when $f_2$ is an even number. The formulas involve the evaluation of terms in a certain series, the number of terms being dependent on the number of degrees of freedom for error. The latter feature is a serious practical limitation, for when $f_2$ is greater than 20, as it often is in psychological experiments, the evaluation becomes too laborious to be of practical utility.

Fox [2] developed charts which overcome some of the objections to earlier works and facilitate the determination of sample size. These charts were constructed from the tables of Tang and Lehmer and are essentially graphs of constant $\phi$ for varying values of $f_1$ and $f_2$. By a method of successive approximations, the value of $n$ may be determined for a fixed value of $\alpha$ and a fixed value of $P$ against a specified alternative hypothesis. These charts are somewhat laborious to apply, however, because of the iterative nature of the approximation for $n$. Also, the charts do not extend below $f_1 = 3$. For psychological experimenters, who typically deal with fixed treatments effects, this limitation considerably restricts their usefulness.

In theory, the problem of evaluating the power of the test of treatments effects in the simpler analyses of variance has been completely solved. Exact, approximate, and graphical solutions have been derived. However, neither the computational formulas nor the graphical solutions make possible a simple, direct, noniterative approximation of the sample size required for a test of specified power. The charts presented in this paper permit this direct approximation for $n$, and hence should be of considerable value.

### *Construction of the Charts*

Each chart in this series presents, for $\alpha = .05$ and .01, a family of three curves which correspond to $P = .5$, .7, and .9. The numerical calculations for the coordinates of the points on the curve $P = .7$ were carried out from the tables of Lehmer; the calculations for the remaining curves were based on data read from the charts of Pearson and Hartley. The three basic steps in the calculations were as follows.

(1) Determine (from chart or table) pairs of values for $\phi$ and $f_2$ for a specified value of $P$, $f_1$, and $\alpha$.
(2) Solve $f_2$ for $n$ from the relationship $n = 1 + f_2/k$, where $k$ is the number of treatments and $n$ the number of observations per treatment.
(3) Divide $\phi$ by $\sqrt{n}$ to obtain $\phi'$.

The pairs of coordinates for $n$ and $\phi'$ were then plotted and smooth curves fitted through the points.

### *Example*

An experimenter wishes to compare the level of mastery reached by three groups of college subjects who memorize a list of paired adjectives under three levels of motivation. From a tentative theoretical formulation of the learning task the experimenter predicts the following array of mean differences for the treatment populations at the three levels:

$$M_I - M_{II} = 5.0;$$
$$M_I - M_{III} = 8.0;$$
$$M_{II} - M_{III} = 3.0.$$

Against this alternative to the null hypothesis the experimenter wishes a power of .90 for a test made at the 5 per cent level. Previous experimentation with this list has given rise to an error variance of 100.0, a value which may be taken as a population parameter for this purpose.

From the array of differences the variance of the treatment population means can be computed equal to 10.89. The value of $\phi'$ is therefore equal to .33. Entering Figure 2 with this value, the required number of subjects is equal to 40.

### *Note on the Generality of the Charts*

The charts presented in this paper are strictly valid only for the test of main effects in the completely randomized design. However, values of $n$ read from these charts underestimate only slightly the values of $n$ which would be required in the randomized block or factorial designs.

The specificity of the charts stems from the unique relationship which holds between $n$ and $f_2$ in each experimental design. For example, for the completely randomized design (and the one used in the construction of these charts) the relationship is

$$n = 1 + f_2/k.$$

In the randomized block design

$$n = 1 + f_2/(k - 1).$$

For the test of the factor with $k$ levels in the $k \times h$ factorial design (mean square within cells being used as the error term) the relationship is

$$n = h + f_2/k.$$

Because of the differences among these relationships, the numerical relationship of $\phi$ to $\phi'$ varies from one design to another, and charts based on that which holds for the completely randomized design will be only approximately correct for the other setups. However, for values of $f_2 \geq 20$, the relationship of $\phi$ to $\phi'$ is almost identical for all three designs. We may demonstrate the relatively small error involved in using the present charts for planning randomized block and factorial designs by applying the charts to two examples which tend to maximize the extent of the inaccuracy. This occurs in the randomized block design when $k = 2$; it occurs in the factorial design employing a within-cells error term and proportional frequencies when the number of measures per cell approaches 2. According to Figure 1, a completely randomized design involving two levels of the treatment variable and $\alpha = .05$ will require $n = 11$ ($f_2 = 20$) for $P = .90$ against $\phi' = .725$. The value of $n$ which is actually needed in a randomized block design for $P = .90$ is approximately 12.0. The comparable value for a $2 \times 6$ factorial design is 11.9.

This discrepancy, which for even this extreme case is probably of little consequence in the planning of most experiments, is considerably smaller for larger values of $f_1$ and $f_2$. Therefore, for practical purposes of approximating the necessary sample size in randomized block and factorial experiments, the tables presented in this paper would seem sufficiently precise.

## REFERENCES

[1] Fisher, R. A. The general sampling distribution of the multiple correlation coefficient. *Proc. roy. Soc. London*, 1928, *A*, **121**, 654–673.

[2] Fox, M. Charts for the power of the *F*-test. *Ann. math. Statist.*, 1956, **27**, 484–497.

[3] Hodges, J. L. On the noncentral Beta distribution. *Ann. math. Statist.*, 1955, **26**, 648–653.

[4] Lehmer, E. Inverse tables of probabilities of errors of the second kind. *Ann. math. Statist.*, 1944, **15**, 388–398.

[5] Nicholson, W. L. A computing formula for the power of the analysis of variance test. *Ann. math. Statist.*, 1954, **25**, 607–610.

[6] Patnaik, P. B. The noncentral $\chi^2$ and $F$-distribution and their applications. *Biometrika*, 1949, **36**, 202–232.
[7] Pearson, E. S. and Hartley, H. O. Charts of the power function for analysis of variance tests, derived from the noncentral $F$-distribution. *Biometrika*, 1951, **38**, 112–130.
[8] Tang, P. C. The power function of the analysis of variance test with tables and illustrations of their use. *Statist. res. Memoirs*. 1938, **2**, 126–149.
[9] Wishart, J. A note on the distribution of the correlation ratio. *Biometrika*, 1932, **24**, 441–456.

*Manuscript received 9/25/57*