POWER FUNCTION CHARTS FOR SPECIFYING NUMBERS OF OBSERVATIONS IN ANALYSES OF VARIANCE OF FIXED EFFECTS

By Leonard S. Feldt and Moharram W. Mahmoud

State University of Iowa and Egyptian Ministry of Education

1. Summary. The charts presented in this paper are designed to facilitate the estimation of the number of observations per treatment required for analysis of variance tests of specified power. They are intended for use by experimenters dealing with fixed treatments effects. With these charts the experimenter may answer the following question: How many observations must I use per treatment to obtain a power of P_1 against alternative hypothesis H_1? Charts are presented for use with tests of treatments effects 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 the randomized block and factorial designs, the latter employing a within-cells estimate of error variance.

2. Nature of the charts. Charts are presented for α equal to .05 and .01 and k, the number of levels of the treatment variable, equal to 2, 3, 4 and 5. The charts are entered with the parameter λ, which is defined as follows:

$$\lambda = \sqrt{\frac{\sum_{j}^{k} (\mu_j - \mu)^2}{k\sigma^2}}$$

where μ_j is the mean of treatment population j, μ the mean of the combined treatment populations, σ^2 the population error variance, and k the number of treatments. The value of n, the number of observations which will be required per treatment for a test of specified power, is read directly from the ordinate of the appropriate chart. It is assumed that the same number of observations will be used in every treatment. The relation of λ to ϕ, the parameter customarily employed in the definition of the power function of the F-test, is simply

$$\lambda = \phi\sqrt{n}.$$

3. Historical development. The first extensive tables of the power function of analysis of variance tests were published by Tang [5]. The tables given by Tang are designed in such a way that for fixed values of α, ϕ, f_1 (degrees of freedom for treatments) and f_2 (degrees of freedom for error) the probability of a Type II error may be determined. The interval of tabulation of Tang's tables is .50, however, which is not sufficiently fine for accurate interpolation.

Following Tang's procedure, Lehmer [3] tabulated the values of ϕ for α =

Received December 17, 1957; revised January 31, 1958.

.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 can not 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.

Pearson and Hartley [4] presented families of power curves for various combinations of α, 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 ϕ. For any given experimental setup, the power of the test may be read directly from the ordinate of the curve. These charts are well suited to the evaluation of the power of any given test. They can not be easily employed, however, to indicate the value of n which should be adopted in order to secure a 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.

Fox [2] contributed charts which facilitate the determination of sample size. These charts were constructed from the tables of Tang and Lehmer and are essentially graphs of constant ϕ 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 α and a fixed value of P against a specified alternative. These charts are somewhat more convenient than the curves of Pearson and Hartley for this purpose, but they are somewhat laborious to use because of the iterative nature of the method of approximating n. Also, the charts do not extend below $f_1 = 3$. For experimenters dealing with fixed treatments effects, this limitation considerably restricts their usefulness.

Duncan [1] published a special condensation of the Pearson and Hartley charts. He plotted on a single set of axes the values of ϕ corresponding to $P =$ .50 and .90 for various values of f_1 and f_2. Separate charts are presented for $\alpha =$ .05 and .01. Having f_1 and f_2 on the same chart facilitates computations which involve both of these elements. For use in planning experiments, however, these charts are subject to the same weaknesses as those of Pearson and Hartley

Though several types of charts and tables of the power function of F-tests have been published, none permits a direct, non-iterative approximation for the number of observations required for a test of specified power. The charts presented in this paper make possible such an approximation for experiments which include 2 to 5 levels of the treatment variable.

4. Construction of the charts. Each chart presents, for $\alpha = .05$ and .01, a family of five curves which correspond to the following values of P: .5, .7, .8, .9 and .95. The number of observations per treatment (n) is plotted on the ordinate, the value of λ is plotted on the abscissa.

The numerical calculations for the coordinates of the points on the curves .7 and .8 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 table or chart) pairs of values for ϕ and f_2 for specified value of P, f_1 and α.

(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 ϕ by $\sqrt{n}$ to obtain λ.

The pairs of coordinates, n and λ, were then plotted and smooth curves fitted through these points.

5. Example. An experimenter wishes to investigate the legibility of two common styles of handwriting: manuscript and cursive. These styles, which constitute the two "levels" of the treatment variable, are to be compared for a population of fourth grade children. The measure of legibility to be employed is based on the number of regressions in the eye movements of adult readers as they read a standard passage written in one or the other of these styles. Previous research with this measure has given rise to an error variance of 10.00, an estimate which may be taken as a population value for this purpose. The completely randomized design is to be used. For a difference of 3.0 between the population means, the experimenter wishes the power of the F-test to equal .90. The .05 level of significance has been adopted.

For this situation

$$\lambda = \sqrt{\frac{\Sigma (\mu_j - \mu)^2}{k\sigma^2}} = \sqrt{\frac{(1.5)^2 + (1.5)^2}{2(10)}} = .47 .$$

Entering Figure 1 with this value, and using the curve for $P = .90$, $\alpha = .05$, we read the required number of observations per treatment to equal 24+ or 25.

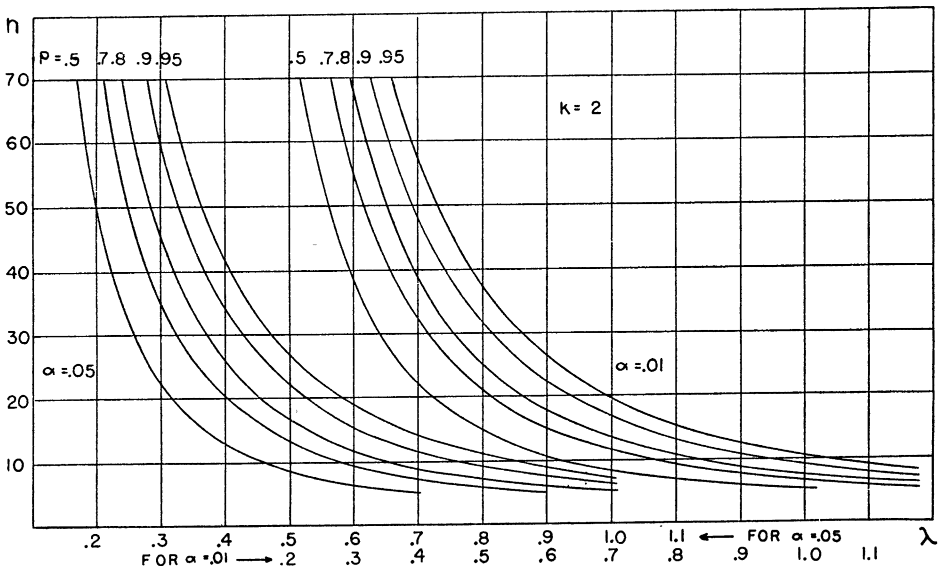

FIG. 1. Curves of constant power (P) for the test of main effects w th k=2.

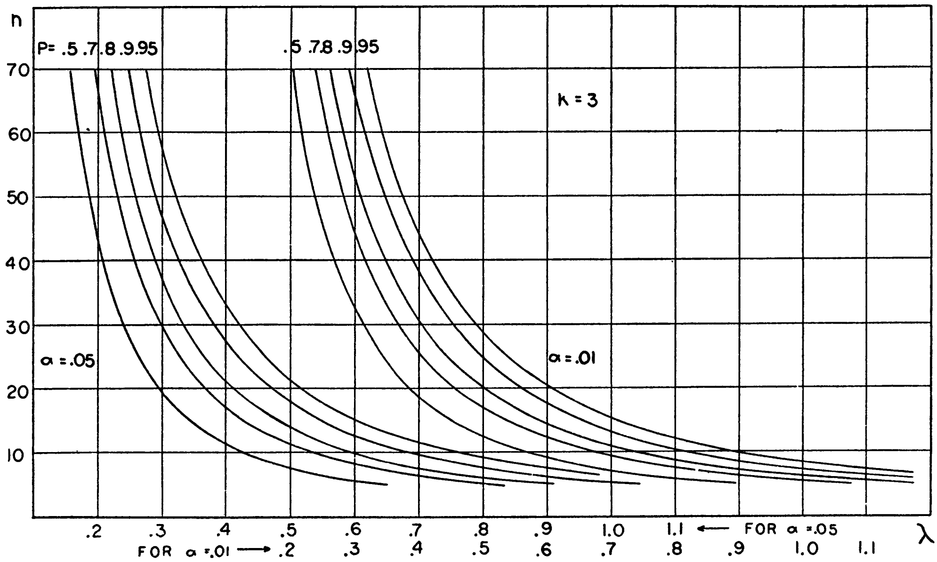

FIG. 2. Curves of constant power (P) for the test of main effects with k=3.

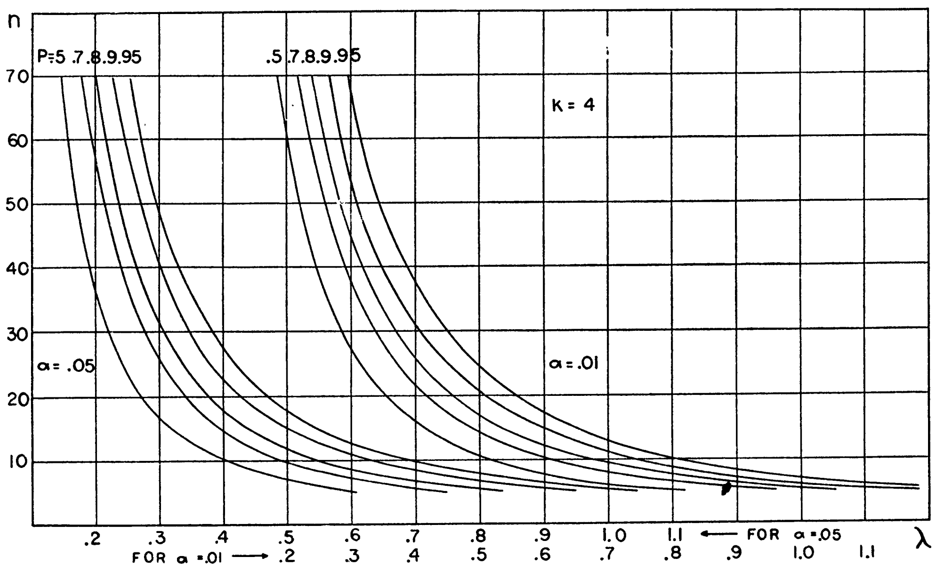

FIG. 3. Curves of constant power (P) for the test of main effects with k=4.

In this example the difference between the population means and the error variance were separately specified. It is often the case, however, that the alternative hypothesis can be defined as a proportion of the error variance. For example, the experimenter might desire a certain power against the alternative

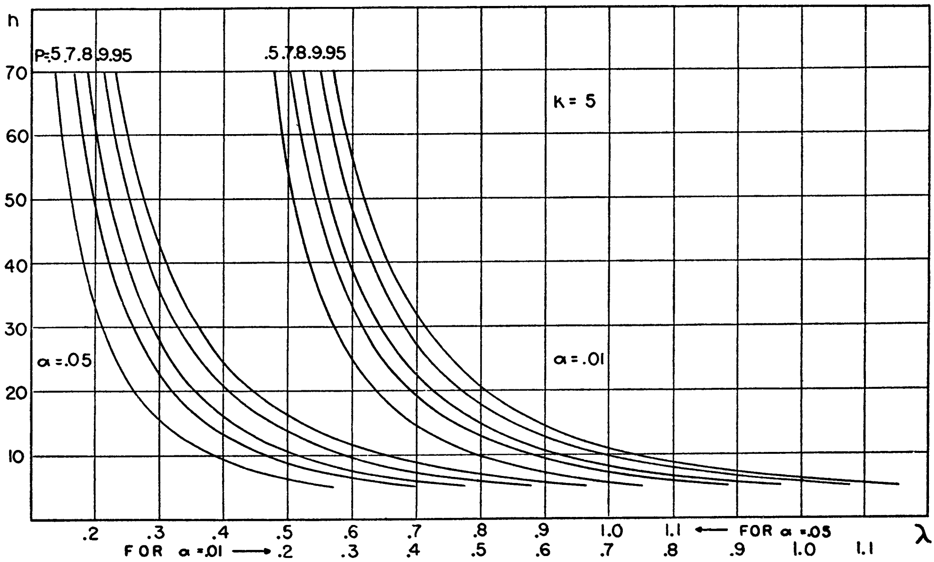

FIG. 4. Curves of constant power (P) for the test of main effects with k=5.

$$\frac{\sum (\mu_j - \mu)^2}{k} = .10\sigma^2.$$

In this case

$$\lambda = \sqrt{\frac{\sum (\mu_j - \mu)^2}{k\sigma^2}} = \sqrt{.10} = .32.$$

The value of λ is thus specified without an explicit statement of the absolute differences between treatment population means.

6. Note. Steps 2 and 3 in the derivation of these charts are based on the relationship which holds between f_2 and n for the completely randomized design. Since this relationship varies from one experimental design to another, these charts are strictly valid only for the completely randomized setup. For precisely accurate determination of the value of n in any other design, a unique set of charts for that design would be required. Charts for the randomized block design, for example, would be based on the relationship

$$f_2 = (k - 1)(n - 1)$$

or

$$n = 1 + \frac{f_2}{(k - 1)}$$

Charts for the test of the factor with k levels in the $k \times h$ factorial design would be based on the relationship

$$f_2 = k(n - h)$$

or

$$n = h + \frac{f_2}{k}.$$

However, from charts specifically constructed for each of these designs it was found that when $k(n - 1) \geqq 20$ the relationship between λ and n is almost identical for all three designs. Little inaccuracy results from the application of charts based upon the relationship which holds for the completely randomized design.

The relatively small error involved in using the present charts for planning randomized block and factorial experiments is demonstrated by the values in Table 1. In this table the appropriate numbers of observations are indicated for selected experimental conditions involving the three types of designs. The values of n for the randomized block and factorial designs were derived from the charts specially constructed for these designs. It may be seen that in every instance the value of n read from charts constructed for the completely randomized design is only slightly smaller than that read from charts specific to the other designs. The underestimate is less than one observation in almost

TABLE 1

Comparative Values of n for Completely Randomized, Randomized Block, and Factorial Experiments (α = .05)

k	P	λ	n		
			Completely Randomized	Randomized Block	$k \times k$ Factorial
2	.5	.525	8.0	8.9	8.2
		.358	16.0	16.9	16.1
		.257	30.0	30.9	30.1
		.181	60.0	60.7	60.0+
	.95	.967	8.0	9.1	8.2
		.657	16.0	17.1	16.1
		.473	30.0	30.9	30.0+
		.333	60.0	60.5	60.0+
4	.5	.450	8.0	8.6	8.8
		.308	16.0	16.3	16.2
		.220	30.0	30.1	30.0+
		.156	60.0	60.0	60.0+
	.95	.770	8.0	8.6	8.8
		.532	16.0	16.5	16.3
		.382	30.0	30.3	30.0+
		.270	60.0	60.0+	60.0

all cases. Therefore, for the practical purpose of approximating the necessary number of observations per treatment in randomized block and factorial experiments, it would seem sufficiently precise to use values read from the present charts.

REFERENCES

[1] A. J. Duncan, "Charts of the 10% and 50% points of the operating characteristic curves for fixed effects analysis of variance F tests, $\alpha = .01$ and .05," *J. Amer. Stat. Assoc.*, Vol. 52 (1957), pp. 345–349.

[2] M. Fox, "Charts for the power of the F-test," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 484–497.

[3] E. Lehmer, "Inverse tables of probabilities of errors of the second kind," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 388–398.

[4] E. S. Pearson and H. O. Hartley, "Charts for the power function for analysis of variance tests, derived from the noncentral F-distribution," *Biometrika*, Vol. 38 (1951), pp. 112–130.

[5] P. C. Tang, "The power function of the analysis of variance test with tables and illustrations of their use," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 126–149 + tables.