Psychological Bulletin
1969, Vol. 72, No. 4, 273–276

ESTIMATING THE MAGNITUDE OF EXPERIMENTAL EFFECTS[1]

JOSEPH L. FLEISS[2]

Biometrics Research, New York State Department of Mental Hygiene

Researchers are becoming increasingly aware that the mere statistical significance of an experimental effect is insufficient to warrant the conclusion that the effect is large and practically important. A number of related measures of the magnitude of experimental effects, $\hat{\omega}^2$, $\hat{\epsilon}^2$, and $\hat{\eta}^2$, are available. Any one of these can confidently be applied to the results of a one-way analysis of variance but not so to the results of a more complicated design. The proper measure for a complex design depends on whether other factors are fixed or random, and the uncritical following of advice given in the literature can result in serious over- or underestimation of the magnitude of experimental effects.

The assessment of the practical significance of experimental effects, given that they have been found to be statistically significant, has received renewed attention recently (Cohen, 1966; Friedman, 1968; Hays, 1963). The advice is given to express the magnitude of experimental effects as a correlationlike measure of association. The greater the proportion of the total variance attributable to experimental effects, the more confident the investigator can be that the effects are sizable and important.

Such a measure of the magnitude of experimental effects is indeed a useful one. It is the purpose of the present note to point out, however, that just as the statistics used in tests of significance must depend on the experimental design, so must the statistics used in measuring association. Hays (1963, Section 12.34) seems to be aware of this fact, but the routine application of the formulas he gives may grossly underestimate the magnitude of experimental effects. Neither Cohen (1966) nor Friedman (1968), on the other hand, seems to be aware of the fact, and the routine application of the formulas they give may grossly overestimate the magnitude of experimental effects.

Little complication exists in the simple one-way layout, although an investigator has at least three formulas to choose from. Let there be J levels of the experimental factor and I independent observations on each, and let F denote the usual ratio of mean squares. Hays (1963, Formula 12.18.5) recommends calculating

$$\hat{\omega}^2 = \frac{(J-1)(F-1)}{(J-1)(F-1)+IJ}; \qquad [1]$$

Cohen (1966, Formula 2) recommends calculating

$$\hat{\epsilon}^2 = \frac{(J-1)(F-1)}{(J-1)(F-1)+IJ-1}; \qquad [2]$$

and Friedman (1968, Formula 2) recommends calculating

$$\hat{\eta}^2 = \frac{(J-1)F}{(J-1)F+J(I-1)}. \qquad [3]$$

The statistics differ in the extent to which the numerators and denominators are unbiased. In practice, their values will differ only slightly provided F and I are large. Hays' $\hat{\omega}^2$ is nevertheless preferable, for the reason now to be given.

Let σ^2 denote the within-levels variance of an observation, and α_j the effect of the *jth* level of the experimental factor. The within-levels mean square MS_w has expectation

$$\mathrm{E}(MS_w) = \sigma^2 \qquad [4]$$

and the between-levels mean square MS_b has expectation

$$\mathrm{E}(MS_b) = \sigma^2 + \frac{I}{J-1}\sum_{j=1}^{J}\alpha_j^2 = \sigma^2 + \frac{IJ}{J-1}\theta^2, \qquad [5]$$

[1] This work was supported in part by Grants MH08534 and MH09191 from the National Institute of Mental Health.

[2] Also at Columbia University.

Requests for reprints should be addressed to Joseph L. Fleiss, Biometrics Research, 722 West 168th Street, New York, New York 10032

where

$$\theta^2 = \frac{1}{J}\sum_{j=1}^{J}\alpha_j^2, \qquad [6]$$

a variancelike measure of the variability among the levels of the experimental factor. The total variance of an observation may be taken to be $\sigma^2 + \theta^2$, and the proportion of this total variance attributable to experimental effects is

$$\omega^2 = \frac{\theta^2}{\sigma^2 + \theta^2}. \qquad [7]$$

An unbiased estimate of the numerator is $(J-1)(MS_b - MS_w)/IJ$ and one of the denominator is $[(J-1)(MS_b - MS_w) + IJMS_w]/IJ$. The most reasonable (though not an unbiased) estimate of ω^2 is thus the ratio of these two estimates,

$$\hat{\omega}^2 = \frac{(J-1)(MS_b - MS_w)/IJ}{[(J-1)(MS_b - MS_w) + IJMS_w]/IJ} = \frac{(J-1)(F-1)}{(J-1)(F-1) + IJ}. \qquad [8]$$

Serious complications begin to emerge in designs more complicated than the one-way layout. Consider, for example, a two-way layout with J levels of the experimental factor, I levels of the second factor (assumed for simplicity to be assigned to the rows of the layout), and no replications within cells. It is assumed that no interactions between row and column effects exist. Friedman (1968) implies, and Cohen (1966) states explicitly, that no modification is needed in the estimates (Expressions 2 and 3) of the magnitude of experimental effects, no matter what the nature of the row classification. Hays (1963, Section 12.34), on the other hand, gives the formula for a modified version of Expression 1 and implies that his modification is appropriate no matter what the nature of the row classification. Not only are these implications contradictory, but no one of them is uniformly correct.

Suppose for specificity that the effects of J drugs on the behavior of mice are to be compared, and that the row classification consists of I methods of administering the drugs (intravenously, orally through drinking water, orally through food, etc.). Interest is assumed to be only in these I methods. A total of IJ mice are available for study, and the IJ mice are randomly paired with the IJ treatment combinations. Let θ_r^2 denote the variance of the effects of the different methods of administration. The analysis of variance table is given in Table 1 (no interactions are assumed to exist), and the column labeled Rows fixed gives the appropriate expected mean squares. Applying Hays' general reasoning (1963, Section 12.34) to this study, one would estimate the magnitude of drug effects as

$$\hat{\omega}_d^2 = \frac{(J-1)(F_d-1)}{(J-1)(F_d-1)+(I-1)(F_r-1)+IJ}. \qquad [9]$$

One advantage of using the statistic given by Expression 9 is that, with $\hat{\omega}_r^2$ defined similarly, and added to $\hat{\omega}_d^2$, one can estimate the proportion of the total variance attributable to both

TABLE 1

ANALYSIS OF VARIANCE TABLE AND EXPECTED MEAN SQUARES WHEN THE LEVELS OF THE ROW CLASSIFICATION ARE ASSUMED TO BE FIXED OR TO BE RANDOM

Source	df	Mean square	Expected mean squares		F ratio
			Rows fixed	Rows random	
Drugs	$J-1$	MS_d	$\sigma^2 + \frac{IJ}{J-1}\theta_d^2$	$\sigma_e^2 + \frac{IJ}{J-1}\theta_d^2$	F_d
Rows	$I-1$	MS_r	$\sigma^2 + \frac{IJ}{I-1}\theta_r^2$	$\sigma_e^2 + J\sigma_r^2$	F_r
Residual	$(I-1)(J-1)$	MS_e	σ^2	σ_e^2	

drugs and the row classification. For, $\hat{\omega}_d^2$ estimates (though not unbiasedly)

$$\omega_d^2 = \frac{\theta_d^2}{\theta_d^2 + \theta_r^2 + \sigma^2} \qquad [10]$$

and $\hat{\omega}_r^2$ estimates

$$\omega_r^2 = \frac{\theta_r^2}{\theta_d^2 + \theta_r^2 + \sigma^2}. \qquad [11]$$

The statistic given by Expression 9 is thus useful when one wishes to compare the relative magnitudes of the two experimental effects.

There is a decided disadvantage to the use of Expression 9 however. Suppose θ_d^2 and θ_r^2 are approximately equal and very large relative to σ^2, with no two drugs and no two methods of administration having similar effects. Then, if one could imagine different populations of mice, each given a different drug with the same method of administration, or each given the same drug but with a different method of administration, one would expect little overlap among these populations, and he would hope that his measure of the magnitude of the experimental effects would reflect such separation among the populations. Nevertheless, both ω_d^2 and ω_r^2 would be close to $\frac{1}{2}$, and so would their estimates probably be.

By using Expression 9 to measure the magnitude of one kind of experimental effect, the investigator is penalizing himself whenever other kinds of experimental effects are large. The use of any one of the Expressions 1, 2 or 3, without introducing any other F ratios, is therefore indicated whenever all mean squares in the analysis of variance except that for error are associated with a fixed experimental factor.

Suppose, now, that the same J drugs are to be compared and that the investigator, in order to increase the precision of his experiment, decides to perform his study as a randomized-blocks design. He obtains I litters of J mice each and, independently for each litter, randomly pairs mice with drugs. Litters now form the row classification instead of methods of administration. Letting σ_r^2 denote the variance due to litter differences and σ_e^2 the residual error variance, the appropriate expected mean squares are given in the indicated column of Table 1. The variance of an observation σ^2 has been split into two components σ_e^2 and σ_r^2, with $\sigma^2 = \sigma_e^2 + \sigma_r^2$.

Applying any one of Formulas 1 to 3, as recommended in effect by Cohen (1966) and Friedman (1968), an investigator would essentially be estimating

$$\lambda^2 = \frac{\theta_d^2}{\theta_d^2 + \sigma_e^2}. \qquad [12]$$

If θ_d^2 is large relative to σ_e^2, then the estimated measure of the magnitude of experimental effects will tend to be close to unity, implying, presumably, a marked separation among at least some of the hypothetical populations of mice given the various drugs. This implication is not necessarily true, however, because none of these estimates incorporates that component of the variability among mice due to differences among litters, namely σ_r^2. If σ_r^2 is positive, then all these measures assume less variability than actually exists, and thus overestimate the proportion of the total variance due to drugs.

The proper parameter is

$$\omega_d^2 = \frac{\theta_d^2}{\sigma_e^2 + \sigma_r^2 + \theta_d^2}, \qquad [13]$$

for the variance of an observation is $\sigma_e^2 + \sigma_r^2$ rather than σ_e^2, and a ratio of unbiased estimates may be shown from the next to last column of Table 1 to be

$$\hat{\omega}_d^2 = \frac{\dfrac{(J-1)(MS_d - MS_e)}{IJ}}{\dfrac{(J-1)(MS_d - MS_e)}{IJ} + \dfrac{(MS_r - MS_e)}{J} + MS_e}$$

$$= \frac{(J-1)(F_d - 1)}{(J-1)(F_d - 1) + I(F_r - 1) + IJ}. \qquad [14]$$

The larger is F_r (i.e., the more successful has the investigator been in removing from experimental error an identifiable component of variance), the greater will be the overestimation ot the magnitude of experimental effects using Formulas 1, 2, or 3. Formula 14 differs from Formula 9 only in that the multiplier of $(F_r - 1)$ is I rather than $I - 1$. The reason for the difference resides in the different expected mean squares under fixed and mixed analysis

TABLE 2

MEASURES OF THE MAGNITUDE OF THE MAIN EXPERIMENTAL EFFECTS APPROPRIATE IN A LATIN SQUARE UNDER DIFFERENT ASSUMPTIONS ABOUT THE ROWS AND COLUMNS

Nature of row and column classification	$\hat{\omega}_d^2$
Rows and columns fixed	$\frac{(J-1)(F_d-1)}{(J-1)(F_d-1)+J^2}$
Rows fixed, columns random	$\frac{(J-1)(F_d-1)}{(J-1)(F_d-1)+J(F_c-1)+J^2}$
Rows random, columns fixed	$\frac{(J-1)(F_d-1)}{(J-1)(F_d-1)+J(F_r-1)+J^2}$
Rows and columns random	$\frac{(J-1)(F_d-1)}{(J-1)(F_d-1)+J(F_r-1)+J(F_c-1)+J^2}$

of variance models, as seen in the next to last row of Table 1.

For a Latin square of dimension J, with the levels of the main experimental factor distributed over the cells of the square, cases where the rows, the columns, or both represent sources of variation removed from the total variance of an observation must be distinguished. In the study of drug effects on mice, for example, let the rows of the square be the J litters. If the columns are an ordering of birth weights, from lightest to heaviest in each litter, then both rows and columns represent actual components of variance. If the columns are different methods of administering the drugs, then only the row classification represents an actual component of variance.

Let F_d denote the F ratio for the main experimental factor, F_r the F ratio for row effects, and F_c the F ratio for column effects. Table 2 gives the appropriate measure of the magnitude of experimental effects for each possible nature of the row and column classifications.

Yet more complicated designs are frequently employed in psychological research: The treatments may assume a factorial structure; some random factors may be nested within the levels of other factors; etc. No simple formulas for estimating the magnitude of experimental effects in all such situations seem possible. Rather, the investigator must examine the column of expected mean squares in his analysis of variance table, identify the bona fide components of the variance of an observation, and use as an estimate of total variance that linear combination of mean squares appropriate to his design.

REFERENCES

COHEN, J. Some statistical issues in psychological research. In B. B. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill, 1966.

FRIEDMAN, H. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 1968, **70**, 245–251.

HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

(Received November 11, 1968)