STATISTICAL POWER ANALYSIS for the BEHAVIORAL SCIENCES

Second Edition

Jacob Cohen

Psychology Press
Taylor & Francis Group

and fairly readily understood measure of strength of relationship or effect size when the dependent variable is an interval, ratio, or dichotomous scale, the need to think in terms of $\mathbf{f}^2$ is reduced, and with it, the need to rely on conventional operational definitions of "small," "medium," and "large" values for $\mathbf{f}^2$. We nevertheless offer such conventions for the frame of reference that they provide, and for use in power surveys and other methodological investigations. We reiterate the caveat that they can represent only a crude guide in as diverse a colleciton of areas as fall under the rubric of behavioral science.

The values for $\mathbf{f}^2$ that follow are somewhat larger than strict equivalence with the operational definitions for the other tests in this book would dictate. For example, when there is only 1 (= $\mathbf{u}$) independent variable, the $\mathbf{F}$ test for $\mathbf{R}^2$ specializes to $\mathbf{t}^2$ of the $\mathbf{t}$ test for $\mathbf{r}$, whose ES operational definitions are respectively .10, .30, and .50 (Section 3.2.1), hence, for $\mathbf{r}^2$, .01, .09, and .25. These in turn yield $\mathbf{f}^2$ values (for Case 0), respectively, of .01, .10, and .33 (from formula 9.2.2), each smaller than the respective $\mathbf{f}^2$ value given below. The reason for somewhat higher standards for $\mathbf{f}^2$ for the operational definitions in MRC is the expectation that the number of IVs in typical applications will be several (if not many). It seems intuitively evident that, for example, if $\mathbf{f}^2 = .10$ defines a "medium" $\mathbf{r}^2$ (= .09), it is reasonable for $\mathbf{f}^2 = .15$ to define a "medium" $\mathbf{R}^2$ (or partial $\mathbf{R}^2$) of .15 when several IVs are involved.

SMALL EFFECT SIZE: $\mathbf{f}^2 = .02$. Translated into $\mathbf{R}^2$ (9.2.5) or partial $\mathbf{R}^2$ for Case 1 (9.1.8), this gives $.02/(1 + .02) = .0196$. We thus define a small effect as one that accounts for 2% of the $\mathbf{Y}$ variance (in contrast with 1% for $\mathbf{r}$), and translate to an $\mathbf{R} = \sqrt{.0196} = .14$ (compared to .10 for $\mathbf{r}$). This is a modest enough amount, just barely escaping triviality and (alas!) all too frequently in practice represents the true order of magnitude of the effect being tested. The discussion under "Small Effect Size" in Section 3.2.1 is relevant here: what may be a moderate theoretical ES may easily, in a "noisy" research, be no larger than what is defined here as small.

MEDIUM EFFECT SIZE: $\mathbf{f}^2 = .15$. In PV terms, this amounts to an $\mathbf{R}^2$ or partial $\mathbf{R}^2$ of $.15/(1 + .15) = .13$, hence $\mathbf{R}$ or partial $\mathbf{R} = .36$ (compared to $\mathbf{r} = .30$ for a medium ES). It may seem that 13% is a paltry amount of variance to define as "medium" when a set made up of several variables is used, but keep in mind that we are defining population values—these are not subject to the inflation (least squares overfitting) which requires correction for shrinkage of a sample $\mathbf{R}^2$ (Cohen & Cohen, 1983, pp. 105–107). In any case, if an investigator finds this criterion too small (or, for that matter, too large) for an area in which he is experienced, he clearly has no need for conventions—he should specify the $\mathbf{R}^2$ (or partial $\mathbf{R}^2$) appropriate to his substantive content and type of $\mathbf{F}$ test, and determine the $\mathbf{f}^2$ from the relevant formula in the preceding material.

LARGE EFFECT SIZE: $\mathbf{f}^2 = .35$. This translates into PV = .26 for $\mathbf{R}^2$ and partial $\mathbf{R}^2$, which in terms of correlation, gives .51 (slightly larger than the $\mathbf{r} = .50$ defining a "large mount" of correlation). This value seems about right for defining a large effect in the middle of the range of fields we cover. It will undoubtedly be often found to be small in sociology, economics, and psychophysics on the one hand, and too large in personality, clinical, and social psychology on the other. As always, this criterion is a compromise that should be rejected when it seems unsuited to the substantive content of any given investigation.

9.3 POWER TABLES

The determination of power as a function of the other parameters proceeds differently in this chapter than in those preceding. Whereas for the other tests, the power tables were entered with the ES index and sample size, here the noncentrality parameter of the noncentral **F** distribution, λ, is used. λ is a simple function of the ES index and the numerator and denominator **df**, respectively **u** and **v**:

$$\lambda = \mathbf{f}^2\,(\mathbf{u} + \mathbf{v} + 1). \tag{9.3.1}$$

We have seen that $\mathbf{f}^2$ and the error model differ in the three cases, so each case has its own function of population $\mathbf{R}^2$ values for $\mathbf{f}^2$ and its own function of **N** and number of IVs (**u**) for **v**. These will be made explicit as each case is discussed.

The three tables in this section yield power values for the **F** tests on the proportion of **Y** variance accounted for by a set of **u** variables **B** (or a partialled set, **B•A**). To read out power, the tables are entered with **a**, λ, **u**, and **v**.

1. *Significance Criterion,* **a**. Tables 9.3.1 and 9.3.2 are for **a** = .01 and .05, respectively.

2. λ, *the Noncentrality Parameter*. λ is tabled over the most useful range for typical MRC applications. Power values are provided at the following 15 λ values: 2 (2) 20 (4) 40. Since λ is a continuous function, interpolation will generally be necessary. Linear interpolation is quite adequate for virtually all purposes, and, because of the intervals tabled, can frequently be done by mental arithmetic. (For interpolation when $\lambda < 2$, note that at $\lambda = 0$, for all values of **u**, power = **a**.)

3. *Degrees of Freedom of the Numerator of the* **F** *Ratio,* **u**. This is also the number of variables in the set **B** which represents the source of variance under study. Each table provides entries for the following 23 values of **u**: 1 (1) 15, 18, 20, 24, 30, 40, 48, 60, 120. (The larger values are rarely used in