Power Analysis and Sample Size Estimation in Multivariate Analysis
James H. Steiger
Psychology 312

Textbooks never treat it in detail, and often do not treat it at all. (Morrison and Anderson, the two classic references, do not even have an index item for "power.")

In practice, power and sample size calculation is linked to an approach to confidence interval estimation that I call "noncentrality interval estimation." However, most references on power and sample size calculation do not discuss the accompanying interval estimation procedures. An exception (besides my own work) is the work of Ken Kelley and his colleagues at Notre Dame.

In the pages that follow, I'll present basic notes on techniques covered in the course, with the exception of canonical correlation. We begin with a detailed analysis of the simplest special case, ANOVA and *t*-tests, because many of the concepts developed there are employed in the other procedures.

Basic Concepts via t-test and 1-Way ANOVA

Standardized Effect Size

One Sample *t*

$$E_s = \frac{\mu - \mu_0}{\sigma}$$

2-Sample *t*

$$E_s = \frac{\mu_1 - \mu_2}{\sigma}$$

General Distribution

One Sample *t*

$$t_{n-1,\lambda}, \; \lambda = \sqrt{n} E_s$$

2-Sample *t*

$$t_{n_1+n_2-2,\lambda}, \; \lambda = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} E_s$$

Interval Estimation

Estimate E_s rather than simply test a null hypothesis. See Steiger and Fouladi (1997) for details. See the MBESS package in R and its documentation for routines.

Equivalence Testing

A simple approach is to see if the entire $1-2\alpha$ confidence interval for E_s fits within a "zone of triviality."

Precision Planning versus Power Planning

Choose a sample size so that the standard error of E_s is sufficiently small. Alternatively, the AIPE (Accuracy in Parameter Estimation) approach of Ken Kelley is to plan sample size so that the expected width of the confidence interval is sufficiently small.

Factorial ANOVA

1-Way

Standardized Effect Size. We seek a multi-sample analogue of E_s. There are several alternative measures. With *a* levels of the *A* factor, Cohen's *f* is

$$f = \sqrt{\frac{\sum_{j=1}^{a}\left(\frac{\alpha_j}{\sigma}\right)^2}{a}} = \sqrt{\frac{\sum_{j=1}^{a}\left(\frac{\mu_j - \mu}{\sigma}\right)^2}{a}}$$

The distribution of the *F* statistic is noncentral $F_{a-1,a(n-1),\lambda}$, with noncentrality parameter λ given by

$$\lambda = n\sum_{j=1}^{a}\left(\frac{\alpha_j}{\sigma}\right)^2 = naf^2 = N_{tot} f^2$$

One may calculate power and sample size for a given *f* directly from the noncentral *F* distribution in R. This is automated very nicely in the program *Gpower 3*.

This relationship may also be "turned around" to generate a confidence interval for λ, and ultimately, for *f.* This calculation may be performed using the routines in the R package MBESS.

2-Way and Beyond

Except for degrees of freedom, little changes. For main effect or interaction θ, the noncentrality parameter may be calculated directly as

$$\lambda_\theta = N_{tot} f_\theta^2$$

For a detailed treatment of confidence interval estimation in the context of ANOVA and regression, consult the book chapter by Steiger and Fouladi (1997), and the journal article by Steiger (2004).

Multiple Regression with Fixed Regressors

Test that $\rho^2 = 0$

With a sample size of *n*, number of predictors *k*, the *F* statistic is

$$F_{k,n-k,\lambda} = \frac{R^2 / k}{(1-R^2)/(n-k-1)}$$

This is distributed as a noncentral *F* with noncentrality parameter

$$\lambda = n\frac{\rho^2}{1-\rho^2}$$

Hence, it is rather straightforward to calculate power for a given ρ^2, k, n. In any given situation, one may plot power as a function of *n* for a given ρ^2 and k, and then compute required sample size by inverting the function plot. This is nicely automated in numerous programs, including the freeware program *Gpower 3.*

This same approach can also be used to perform power and sample size analysis for tests of an additional predictor or group of predictors.

For details, see Faul, et al (2007), p. 181.

Multiple Regression with Random Regressors

The non-null distribution in the case of random regressors is much more complicated than for the fixed regressors case. As a consequence, most people relied on fixed regressor calculations as an approximation. In 1992, Steiger and Fouladi produced R2, the first program that could calculate the exact distribution of R^2. The program performed a full range of power and sample size calculations for tests that $\rho^2 = c$, where c need not be zero. In addition, the program produced an exact confidence interval on ρ^2. The power and sample size calculations are available in Gpower 3 and MBESS, and the confidence interval calculations can be performed by MBESS.

Structural Equation Modeling, Factor Analysis, Confirmatory Factor Analysis

Fitting a model by maximum likelihood involves minimizing the maximum likelihood criterion $F(\mathbf{S},\mathbf{M}(\theta))$. The standard test statistic is the chi-square statistic $(N-1)F(\mathbf{S},\mathbf{M}(\theta))$. Steiger, Shapiro, and Browne (1985) showed that this statistic has approximately a noncentral chi-square distribution with noncentrality parameter

$$\lambda = (N-1)F^*$$

where F^* is the population discrepancy function, i.e., the value of the discrepancy function that would be obtained if the population covariance matrix $\mathbf{\Sigma}$ were available and analyzed by maximum likelihood.

Steiger and Lind (1980) proposed the RMSEA as an index of population badness of fit. (See handout on Fit Indices in SEM at the course website.) This index is

$$RMSEA = \sqrt{\frac{F^*}{\nu}}$$

where ν is the degrees of freedom for the model. Work by Browne (1977) had shown that F^* is closely approximated by a sum of squared orthogonalized model errors, much the same as the squared Mahalanobis distance in form, i.e.,

$$F^* \approx \mathbf{e}'\mathbf{\Gamma}^{-1}\mathbf{e}$$

where **e** is a vector of discrepancies between the elements of $\mathbf{\Sigma}$ and the model's approximation of them, and $\mathbf{\Gamma}$ is the covariance matrix of the elements of S. So the *RMSEA* is essentially a root-mean-square-error of approximation of the model to the data. This means that the noncentrality parameter may be calculated as

$$\lambda = (n = 1)\nu \times RMSEA^2$$

MacCallum, Browne, and Sugawara (1996) suggested a formal hypothesis test of target values of the *RMSEA.* The traditional test is, of course, a test that the population *RMSEA* = 0.

Steiger (1990) favored a confidence interval based approach, centering on what MacCallum, Browne, and Sugawara termed a test of *not-close-fit,* corresponding to the standard approach in biostatistics "bioequivalence testing."

MacCallum, Browne, and Sugawara produced sample size tables. Steiger (1999) included a power and sample size calculator in the program *Statistica Power Analysis.* Ken Kelley includes routines to perform calculations on these procedures in the R package MBESS.

Hotelling's T^2

The Squared Mahalanobis Distance $\Delta^2 = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)$

The Ψ index: $\Psi = \sqrt{\dfrac{\Delta^2}{k}}$

One Sample

$$T^2 = n(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0) = n\hat{\Delta}^2$$

$$F_{k-1,n-k-1,\lambda} = \frac{n-k+1}{(n-1)(k-1)}T^2$$

$$\lambda = n\Delta^2 = nk\Psi^2$$

Two Sample

$$T^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2)'\boldsymbol{\Sigma}^{-1}(\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2) = \frac{n_1 n_2}{n_1 + n_2}\hat{\Delta}^2$$

$$\lambda = \frac{n_1 n_2}{n_1 + n_2}\Delta^2 = \frac{n_1 n_2}{n_1 + n_2}k\Psi^2$$

Binary Logistic Regression

GPower 3 gives power calculation for a single predictor, in terms of the null hypothesized probability of the response = 1 given $X = 1$. One simply specifies the odds ratio and sample size, and the program calculates power. Or, alternatively, one specifies power and the odds ratio, and the program computes sample size.

GPower 3 also computes power and sample size for testing an additional predictor. One must specify the ρ^2 for the other predictors, which is of course largely based on guesswork.

Theory underlying the tests in *GPower3* is given in Hsieh (1989) and discussed in detail in Hosmer and Lemeshow, *Applied Logistic Regression*, p. 339–347.

MANOVA

Power and sample size analysis in MANOVA requires specification of numerous parameters that you are unlikely to know.   It is not for the faint of heart, but an approach is implemented in *Gpower 3* and discussed in the tutorial, available for download online.

Faul, et al. (2007) give a thorough discussion of the theory behind the methods employed in *GPower3.* An alternate, somewhat simpler approach that Faul et al. claim is slightly less accurate is given by Muller et al (1992). Their approach is very general.