

# Determining the power of multiple regression analyses both with and without repeated measures

ROGER BAKEMAN and DUNCAN McARTHUR  
Georgia State University, Atlanta, Georgia

Power analysis guides researchers in planning how much data to collect. This article describes BWPower, a computer program for the Windows 95 environment that performs power analyses for research designs that may or may not include both between- and within-subjects factors. We discuss how BWPower easily accommodates both between- and within-subjects factors and provide examples of BWPower's use in performing power analyses on designs with only between-subjects factors, designs with only repeated measures, and with mixed between- and within-subjects designs. We highlight the major features of BWPower's user interface, such as the ability to iteratively increment or decrement the number of subjects and the automatic recalculation of power when the number of subjects or effect sizes is changed.

When planning a study, common sense and good practice lead investigators to ask how many scores are required in order to have a reasonable chance of finding reasonably sized effects statistically significant (Cohen, 1992). After all, collecting so few data that effects are unlikely to be detected or far more data than are needed squander research resources. Thus, it is not surprising that both investigators and institutions that fund research have embraced Cohen's methods for determining statistical power (Cohen, 1969, 1978, 1988), where power is understood as the probability that effects of a specified size will be detected as significant, given a specified number of scores and alpha level.

Usually, power analysis is discussed in terms of the number of cases or participants needed for a study; however, in the preceding sentences, we have referred to number of scores instead to make a point. Investigators whose research design includes one or more factors with repeated measures are unlikely to find much explicit help in the power-analysis literature. For example, the second edition of Cohen's (1988) *Statistical Power Analysis for the Behavioral Sciences* includes hundreds of worked examples but not one that includes repeated measures. Cohen and Cohen's (1983) *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* does describe how to perform a power analysis for a design that includes one within-subjects factor and no between-subjects factors (pp. 435-437) and one within-subjects factor and one between-subjects factor (pp. 449-450), but these procedures are based on a section of the first, revised edition of *Statistical Power Analysis for the Behavioral Sciences* (Cohen,

1978, chap. 9) that is modified in the second edition (Cohen, 1988).

This article has two purposes. First, primarily on the basis of Cohen and Cohen's (1983) discussion of the multiple regression approach to analysis of repeated measures data (chap. 11, pp. 428-451; see also Bakeman, 1992), we describe specific procedures for performing a power analysis given a design that includes repeated measures on one or more factors. Second, we describe and make available at no cost a Windows 95 computer program (BWPower, for *between-within power*) that performs these computations. However, the multiple-regression-based approach is quite general. As a result, this easy-to-use and flexible program can be used to compute power, whether or not the research design includes repeated measures.

## Power for Multiple Regression Analyses

Beginning with his seminal article, *Multiple Regression as a General Data-Analytic System*, Cohen (1968) has repeatedly emphasized the generality of a multiple regression approach whenever investigators want to explain or account for a single interval-scaled criterion variable. Indeed, in different chapters of *Statistical Power Analysis for the Behavioral Sciences*, Cohen (1988) provides power analytic methods for different statistical approaches, such as *t* tests, product moment correlations, and analyses of variance and covariance, and then notes how identical results are achieved when diverse numeric examples are recast in a multiple regression framework. Thus, the simplest way to embody power analysis computations in a general-purpose computer program is to take a multiple regression approach. This is the approach taken by Borenstein and Cohen (1988), by Borenstein, Cohen, Rothstein, Pollack, and Kane (1990), and by Rothstein, Borenstein, Cohen, and Pollack (1990) in their DOS-based power

---

The authors wish to thank Patrick Onghena for helpful comments on the initial draft of this paper. Correspondence should be addressed to R. Bakeman, Department of Psychology, University Plaza, Georgia State University, Atlanta, GA 30303 (e-mail: bakeman@gsu.edu).

analysis programs, and it is the approach we take in BWPower.

A somewhat different approach is taken by Erdfelder, Faul, and Buchner (1996) in GPOWER. This MS-DOS and Macintosh program performs high-precision statistical power analyses for *t*, *F*, and  $\chi^2$  tests. As both Erdfelder et al. and Bradley, Russell, and Reeve (1996) note, Cohen's more approximate approach applied to complex factorial designs results in a conservative bias, and, in particular, users with such designs should consult these two references.

**Effect sizes.** In the next section, we describe how a multiple regression approach to power analysis generalizes to repeated measures designs, but first we briefly describe the overarching approach. Cohen (1988) recommends that investigators estimate the proportions of variability in their criterion variable as additional sources of variance are taken into account (understanding that these estimates refer to population values). In the simplest case, only a single predictor or set of predictors is considered, symbolized *B*. Cohen (1988) terms this Case 0; his effect-size index for this case is

$$f^2 = \frac{R_{Y.B}^2}{1 - R_{Y.B}^2}, \quad (1)$$

where the numerator is the variance accounted for by *B* and the denominator is the residual variance. Thus, the effect-size index can be regarded as a signal-to-noise ratio. Case 1 considers the effect of *B*, after a set of other variables, symbolized *A* (e.g., background variables or covariates), has been taken into account. The effect-size index for Case 1 is

$$f^2 = \frac{R_{Y.A.B}^2 - R_{Y.A}^2}{1 - R_{Y.A,B}^2}, \quad (2)$$

where the numerator is the additional variance accounted for by *B*, above that accounted for by *A* alone. Finally, Case 2 considers the effect of *B*, but additional variables, symbolized *C*, are taken into account, thereby reducing the noise in the signal-to-noise ratio (Cohen, 1988, terms this Model II error). For example, with a  $2 \times 2$  analysis of variance (ANOVA), *A* would be the *A* main effect, *B* would be the *B* main effect, and *C* would be the *AB* interaction, when the investigator is interested in the power of the *B* main effect. The effect-size index for Case 2 is

$$f^2 = \frac{R_{Y.A,B}^2 - R_{Y.A}^2}{1 - R_{Y.A,B,C}^2}. \quad (3)$$

Thus, Case 2 is the general case; when *C* is 0, it reduces to Case 1, and when *A* is also 0, it reduces to Case 0. Consistent with his definition of .01, .09, and .25 as small, medium, and large values for  $R^2$ , Cohen defines values of .02, .15, and .35 as small, medium, and large values for  $f^2$  (Cohen, 1988, pp. 412–414).

**Using BWPower when studies do not include repeated measures.** To perform a power analysis using

BWPower, and, for the moment, assuming no factors with repeated measures, the user first selects an alpha level (either .05 or .01) and then identifies the number of sources of variance. BWPower permits up to 99 sources (whereas the program described by Rothstein et al., 1990, permits 3 sources), but values even approaching 99 seem unlikely. The program labels these as *A*, *B*, *C*, and so forth, but the user can change the labels to something more descriptive. Next, the user enters the additional proportions of criterion variance accounted for by each source (labeled *incR2* in BWPower) along with its degrees of freedom. Each source reflects an individual variable or a set of variables. The sources are arranged as for the steps of a hierarchical multiple regression, and the degrees of freedom for each source reflect the number of variables entered at that step. For example, two sources might be specified. The first could represent a set of 3 covariates estimated to account for 20% of the variability, and the second could represent a single variable that codes whether or not participants received an intervention that is estimated to account for an additional 15%. The program computes the error or residual variance; thus, the increments in  $R^2$  the user enters must sum to less than 1 (or else error variance would be negative).

Pressing the "compute" button displays effect sizes and power for all sources. The "compute" button is a toggle; thus, it remains depressed until clicked again. As long as the "compute" button remains depressed, any changes in degrees of freedom, increases in  $R^2$ , alpha level, or number of cases cause power to be recomputed. In particular, the number of cases can be changed continuously (by pressing the "up" or "down" button associated with number of cases) and associated changes in power observed (see Figure 1). In this way, the user can readily determine the number of cases required for any given value of power (.80 and .90 are often used). For the example given in the previous paragraph, and assuming an alpha level of .05, a sample size of 52 is required to have a 90% chance of detecting a treatment effect that accounts for 15% additional variance above the 20% accounted for by 3 covariates; for an alpha level of .01, a sample size of 71 is required.

### Power for Repeated Measures Analyses

Power analysis for designs that include repeated measures is a straightforward extension of the procedure described in the previous section. We say straightforward, but the solution becomes so only when one takes a view of repeated measures analyses that is not widely utilized (Cohen & Cohen, 1983, chap. 11; Bakeman, 1992, chaps. 14 and 15). Required is the view that repeated measures analyses can be recast as a hierarchical multiple regression in which between-subjects variability serves as a covariate.

Consider the analysis of a 1 between-subjects 1 within-subjects gender  $\times$  age (3 levels) ANOVA summarized in Table 1 (this table is from Bakeman & McArthur, 1996). The first two sources, gender and subject, represent between-subjects variability and serve as covariates for the within-subjects effect of age and the gender  $\times$  age interaction that follow. For this example, approximately 50%

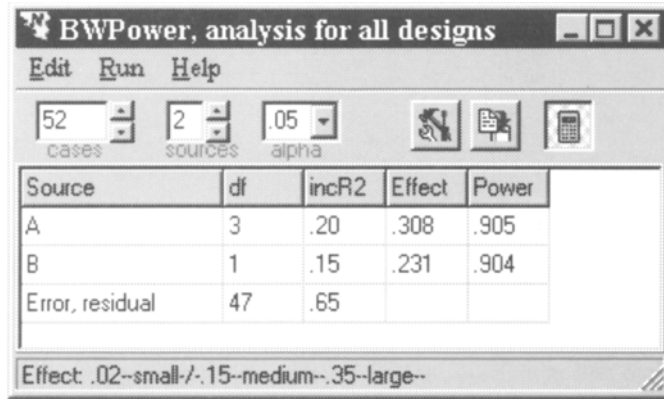


Figure 1. The BWPow display showing the “compute” button depressed.

of the variance in the scores is accounted for by between-subjects variability: About 20% is associated with gender, and the remaining 30% is associated with between-subjects error. Thus, the gender effect accounts for 40% of between-subjects variability. The remaining 50% of the variance in the scores is accounted for by within-subjects variability: About 20% is associated with age, and an additional 1% is associated with the gender  $\times$  age interaction. Thus, the interaction, which could be of primary theoretical interest here, accounts for 2% of the within-subjects variability.

**Using BWPow when studies include repeated measures.** To perform a power analysis using BWPow, the user first clicks the “Design Assist” button or selects “Design Assist” on the Edit menu (see Figure 2). For the example presented in the previous paragraph, the user indicates one between-subjects variable with two levels and one within-subjects variable with three levels. The program then generates the appropriate sources for this analysis along with their degrees of freedom. As many as five factors may be specified in all. Between-subjects variables are labeled *A*, *B*, and so forth, and within-subjects variables are labeled *P*, *Q*, and so forth; but, again, the user can change the labels. After increases in  $R^2$  are specified, pressing the “compute” button displays effect sizes and power for all sources. Changing the number of cases causes degrees of freedom for both between- and within-subjects error terms to be recomputed. Assuming

the values suggested in the previous paragraph (20% gender, 30% residual between-subjects error, 20% age, 1% interaction), 188 cases would be required to have a 90% chance of detecting the interaction as significant at the .05 level, and 142 would be required for an 80% chance (see Figure 3; power = .995 for the other effects is the highest value BWPow displays). However, if the user were only interested in detecting interaction effects that accounted for 10% of the variance instead of 1%, and assuming power of .80 and alpha of .05, then 12 cases would be sufficient (actually power = .81 for this example, because BWPow will not compute power if the final error term is associated with fewer than 20 degrees of freedom, which is the minimum value given in Cohen’s, 1988, tables).

One final example, adopted from Cohen and Cohen (1983), is as follows: Assume a single within-subjects variable with four conditions and no between-subjects variables. Assume further that between-subjects variability accounts for 50% of criterion score variance and that the condition variable accounts for an additional 10%. BWPow computes that 16 cases are required for power = .80 and alpha = .05, which agrees with Cohen and Cohen (1983, p. 436), although BWPow uses the refined methods of Cohen (1988).

**Effect sizes with repeated measures.** Power analysis, no matter the design, requires that users provide values for the strength of the various effects (BWPow re-

Table 1  
Source Table for the Analysis of the 36 Play Level Scores  $\times$   
Gender (Between-Subjects) and Age (Within-Subjects)

Source	$R^2$	$\Delta R^2$	SS	df	MS	F
Gender (G)	.189	.189	12.25	1	12.25	5.73
Subject (S)	.518	.329	21.39	10	2.14	
Age	.729	.211	13.72	2	6.86	8.02
G $\times$ Age	.737	.008	0.50	2	0.25	0.29
S $\times$ Age	1	.263	17.11	20	0.86	
Total			64.97	35		

Note—This table is reproduced from Bakeman and McArthur (1996). Copyright 1996 by Psychonomic Society, Inc.

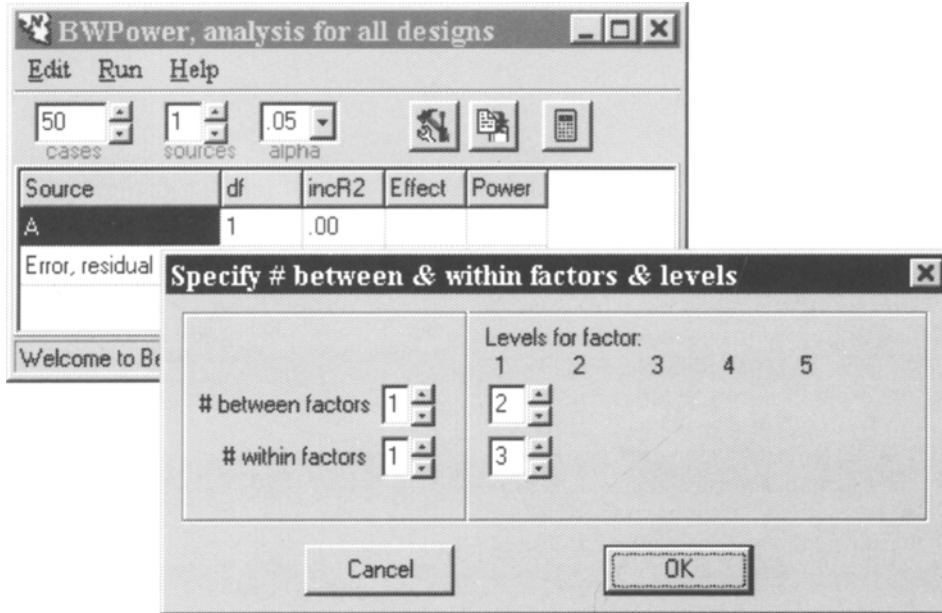


Figure 2. The BWPower display showing the “Design Assist” window.

quires  $R^2$ s and then computes  $f^2$ s). The best guide is previous experience (as reflected in the published literature) or an investigator’s sense of what size effects are worth detecting, although one can always fall back on Cohen’s (1988) definitions of small, medium, and large. The published literature is not always a certain guide because authors often fail to provide effect-size statistics; this is especially so when studies include repeated measures. However, investigators can at least compute effect sizes for their own previous studies. The general linear model repeated measures procedure in SPSS for Windows, for example, provides the appropriate sums of squares from which  $R^2$ s, like those in Table 1, can be computed. The user should keep in mind, however, that these sample  $R^2$ s

are biased estimates and that population values will be somewhat lower.

**Program Availability**

BWPower was written using Borland’s Dephi 2. It requires Windows 95 and is available to interested researchers at no cost. It consists of a single executable file and can be downloaded from [www.gsu.edu/psychology/bakeman](http://www.gsu.edu/psychology/bakeman).

**Conclusion**

Whenever investigators want to account for variability in a single, interval-scaled criterion measure with one or more predictor variables (i.e., research factors), a multiple-

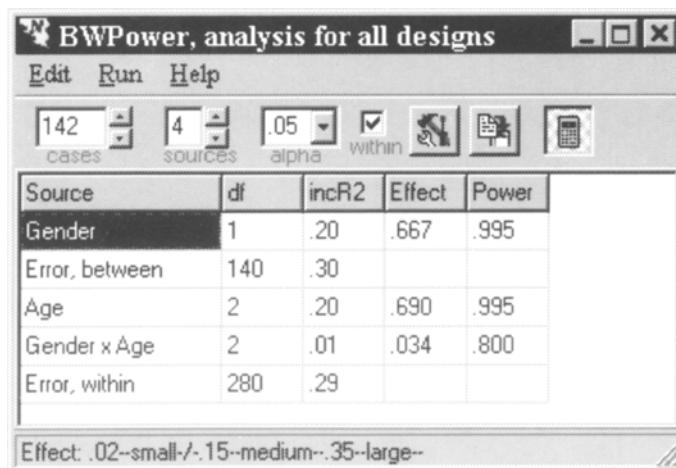


Figure 3. The BWPower display showing results for a design that includes a repeated age factor.

regression-based version of power analysis (Cohen, 1988) can be used to determine the number of cases needed to have a reasonable chance (usually 80% or 90%) of detecting effects of a specified size as significant at a given alpha (usually .05 or .01). Using quite general procedures, power can be computed for multiple regression analyses, for analyses based on between-subjects ANOVAs, and for analyses that include repeated measures if between-subjects variability is conceptualized as a covariate. BWPower is a Windows 95 program that computes power for all such cases. Users specify the alpha level, the sources of variance and their degrees of freedom, and the number of cases. For ANOVAs, users specify the number of between-subjects and within-subjects factors and levels for each, in which case the program automatically generates the sources and their degrees of freedom. BWPower then computes effects sizes and the power for each. The program is flexible and easy to use. In particular, users can continuously change the number of cases until the desired power is achieved.

#### REFERENCES

- BAKEMAN, R. (1992). *Understanding social science statistics: A spreadsheet approach*. Hillsdale, NJ: Erlbaum.
- BAKEMAN, R., & McARTHUR, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers*, **28**, 584-589.
- BORENSTEIN, M., & COHEN, J. (1988). *Statistical power analysis: A computer program*. Hillsdale, NJ: Erlbaum.
- BORENSTEIN, M., COHEN, J., ROTHSTEIN, H. R., POLLACK, S., & KANE, J. M. (1990). Statistical power analysis for one-way analysis of variance: A computer program. *Behavior Research Methods, Instruments, & Computers*, **22**, 271-282.
- BRADLEY, D. R., RUSSELL, R. L., & REEVE, C. P. (1996). Statistical power in complex experimental designs. *Behavior Research Methods, Instruments, & Computers*, **28**, 319-326.
- COHEN, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, **70**, 426-443.
- COHEN, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- COHEN, J. (1978). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155-159.
- COHEN, J., & COHEN, P. (1983). *Applied multiple regression correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- ERDFELDER, E., FAUL, F., & BUCHNER, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1-11.
- ROTHSTEIN, H. R., BORENSTEIN, M., COHEN, J., & POLLACK, S. (1990). Statistical power analysis for multiple regression/correlation: A computer program. *Educational & Psychological Measurement*, **50**, 819-830.

(Manuscript received December 10, 1997;  
revision accepted for publication March 31, 1998.)