Behavior Research Methods
2009, 41 (4), 1149-1160
doi:10.3758/BRM.41.4.1149

Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses

Franz Faul
Christian-Albrechts-Universität, Kiel, Germany

Edgar Erdfelder
Universität Mannheim, Mannheim, Germany

and

Axel Buchner and Albert-Georg Lang
Heinrich-Heine-Universität, D sseldorf, Germany

G*Power is a free power analysis program for a variety of statistical tests. We present extensions and improvements of the version introduced by Faul, Erdfelder, Lang, and Buchner (2007) in the domain of correlation and regression analyses. In the new version, we have added procedures to analyze the power of tests based on (1) single-sample tetrachoric correlations, (2) comparisons of dependent correlations, (3) bivariate linear regression, (4) multiple linear regression based on the random predictor model, (5) logistic regression, and (6) Poisson regression. We describe these new features and provide a brief introduction to their scope and handling.

G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) is a stand-alone power analysis program for many statistical tests commonly used in the social, behavioral, and biomedical sciences. It is available free of charge via the Internet for both Windows and Mac OS X platforms (see the Concluding Remarks section for details). In this article, we present extensions and improvements of G*Power 3 in the domain of correlation and regression analyses. G*Power now covers (1) one-sample correlation tests based on the tetrachoric correlation model, in addition to the bivariate normal and point biserial models already available in G*Power 3, (2) statistical tests comparing both dependent and independent Pearson correlations, and statistical tests for (3) simple linear regression coefficients, (4) multiple linear regression coefficients for both the fixed- and random-predictors models, (5) logistic regression coefficients, and (6) Poisson regression coefficients. Thus, in addition to the generic power analysis procedures for the z, t, F, χ^2, and binomial tests, and those for tests of means, mean vectors, variances, and proportions that have already been available in G*Power 3 (Faul et al., 2007), the new version, G*Power 3.1, now includes statistical power analyses for six correlation and nine regression test problems, as summarized in Table 1.

As usual in G*Power 3, five types of power analysis are available for each of the newly available tests (for more thorough discussions of these types of power analyses, see Erdfelder, Faul, Buchner, & Cüpper, in press; Faul et al., 2007):

1. *A priori analysis* (see Bredenkamp, 1969; Cohen, 1988). The necessary sample size is computed as a function of user-specified values for the required significance level α, the desired statistical power $1-\beta$, and the to-be-detected population effect size.
2. *Compromise analysis* (see Erdfelder, 1984). The statistical decision criterion ("critical value") and the associated α and β values are computed as a function of the desired error probability ratio β/α, the sample size, and the population effect size.
3. *Criterion analysis* (see Cohen, 1988; Faul et al., 2007). The required significance level α is computed as a function of power, sample size, and population effect size.
4. *Post hoc analysis* (see Cohen, 1988). Statistical power $1-\beta$ is computed as a function of significance level α, sample size, and population effect size.
5. *Sensitivity analysis* (see Cohen, 1988; Erdfelder, Faul, & Buchner, 2005). The required population effect size is computed as a function of significance level α, statistical power $1-\beta$, and sample size.

As already detailed and illustrated by Faul et al. (2007), G*Power provides for both numerical and graphical output options. In addition, any of four parameters—α, $1-\beta$, sample size, and effect size—can be plotted as a function of each of the other three, controlling for the values of the remaining two parameters.

Below, we briefly describe the scope, the statistical background, and the handling of the correlation and regression

E. Erdfelder, erdfelder@psychologie.uni-mannheim.de

Table 1
Summary of the Correlation and Regression Test Problems Covered by G*Power 3.1

Correlation Problems Referring to One Correlation
Comparison of a correlation ρ with a constant ρ_0 (bivariate normal model)
Comparison of a correlation ρ with 0 (point biserial model)
Comparison of a correlation ρ with a constant ρ_0 (tetrachoric correlation model)
Correlation Problems Referring to Two Correlations
Comparison of two dependent correlations ρ_{jk} and ρ_{jh} (common index)
Comparison of two dependent correlations ρ_{jk} and ρ_{hm} (no common index)
Comparison of two independent correlations ρ_1 and ρ_2 (two samples)
Linear Regression Problems, One Predictor (Simple Linear Regression)
Comparison of a slope b with a constant b_0
Comparison of two independent intercepts a_1 and a_2 (two samples)
Comparison of two independent slopes b_1 and b_2 (two samples)
Linear Regression Problems, Several Predictors (Multiple Linear Regression)
Deviation of a squared multiple correlation ρ^2 from zero (F test, fixed model)
Deviation of a subset of linear regression coefficients from zero (F test, fixed model)
Deviation of a single linear regression coefficient b_j from zero (t test, fixed model)
Deviation of a squared multiple correlation ρ^2 from constant (random model)
Generalized Linear Regression Problems
Logistic regression
Poisson regression

power analysis procedures that are new in G*Power 3.1. Further technical details about the tests described here, as well as information on those tests in Table 1 that were already available in the previous version of G*Power, can be found on the G*Power Web site (see the Concluding Remarks section). We describe the new tests in the order shown in Table 1 (omitting the procedures previously described by Faul et al., 2007), which corresponds to their order in the "Tests → Correlation and regression" dropdown menu of G*Power 3.1 (see Figure 1).

1. The Tetrachoric Correlation Model

The "Correlation: Tetrachoric model" procedure refers to samples of two dichotomous random variables X and Y as typically represented by 2×2 contingency tables. The tetrachoric correlation model is based on the assumption that these variables arise from dichotomizing each of two standardized continuous random variables following a bivariate normal distribution with correlation ρ in the underlying population. This latent correlation ρ is called the tetrachoric correlation. G*Power 3.1 provides power analysis procedures for tests of H_0: $\rho = \rho_0$ against H_1: $\rho \neq \rho_0$ (or the corresponding one-tailed hypotheses) based on (1) a precise method developed by Brown and Benedetti (1977) and (2) an approximation suggested by Bonett and Price (2005). The procedure refers to the Wald z statistic $W = (r - \rho_0)/\mathrm{se}_0(r)$, where $\mathrm{se}_0(r)$ is the standard error of the sample tetrachoric correlation r under H_0: $\rho = \rho_0$. W follows a standard normal distribution under H_0.

Effect size measure. The tetrachoric correlation under H_1, ρ_1, serves as an effect size measure. Using the effect size drawer (i.e., a subordinate window that slides out from the main window after clicking on the "Determine" button), it can be calculated from the four probabilities of the 2×2 contingency tables that define the joint distribution of X and Y. To our knowledge, effect size conventions have not been defined for tetrachoric correlations. However, Cohen's (1988) conventions for correlations in the framework of the bivariate normal model may serve as rough reference points.

Options. Clicking on the "Options" button opens a window in which users may choose between the exact approach of Brown and Benedetti (1977) (default option) or an approximation suggested by Bonett and Price (2005).

Input and output parameters. Irrespective of the method chosen in the options window, the power of the tetrachoric correlation z test depends not only on the values of ρ under H_0 and H_1 but also on the marginal distributions of X and Y. For post hoc power analyses, one therefore needs to provide the following input in the lower left field of the main window: The number of tails of the test ("Tail(s)": one vs. two), the tetrachoric correlation under H_1 ("H1 corr ρ"), the α error probability, the "Total sample size" N, the tetrachoric correlation under H_0 ("H0 corr ρ"), and the marginal probabilities of $X = 1$ ("Marginal prob x") and $Y = 1$ ("Marginal prob y")—that is, the proportions of values exceeding the two criteria used for dichotomization. The output parameters include the "Critical z" required for deciding between H_0 and H_1 and the "Power ($1-\beta$ err prob)." In addition, critical values for the sample tetrachoric correlation r ("Critical r upr" and "Critical r lwr") and the standard error $\mathrm{se}(r)$ of r ("Std err r") under H_0 are also provided. Hence, if the Wald z statistic $W = (r - \rho_0)/\mathrm{se}(r)$ is unavailable, G*Power users can base their statistical decision on the sample tetrachoric r directly. For a two-tailed test, H_0 is retained whenever r is not less than "Critical r lwr" and not larger than "Critical r upr"; otherwise H_0 is rejected. For one-tailed tests, in contrast, "Critical r lwr" and "Critical r upr" are identical; H_0 is rejected if and only if r exceeds this critical value.

Illustrative example. Bonett and Price (2005, Example 1) reported the following "yes" (= 1) and "no" (= 2)

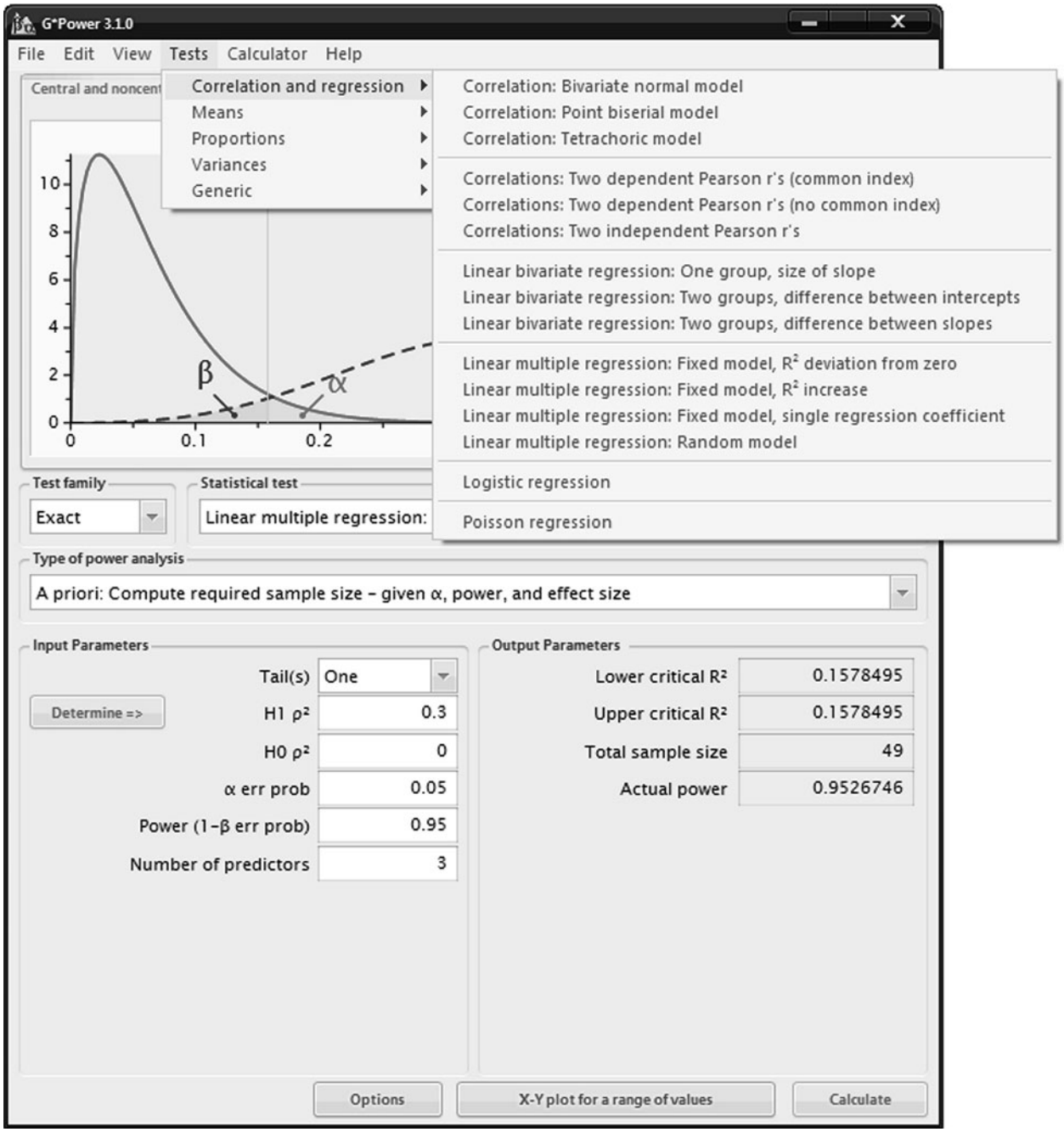

Figure 1. The main window of G*Power, showing the contents of the "Tests → Correlation and regression" drop-down menu.

answer frequencies of 930 respondents to two questions in a personality inventory: $f_{11} = 203$, $f_{12} = 186$, $f_{21} = 167$, $f_{22} = 374$. The option "From C.I. calculated from observed freq" in the effect size drawer offers the possibility to use the $(1-\alpha)$ confidence interval for the sample tetrachoric correlation r as a guideline for the choice of the correlation ρ under H_1. To use this option, we insert the observed frequencies in the corresponding fields. If we assume, for example, that the population tetrachoric correlation under H_1 matches the sample tetrachoric correlation r, we should choose the center of the C.I. (i.e., the correlation coefficient estimated from the data) as "H1 corr ρ" and press "Calculate." Using the exact calculation method, this results in a sample tetrachoric correlation $r = .334$ and marginal proportions $p_x = .602$ and $p_y = .582$. We then click on "Calculate and transfer to main window" in the effect size drawer. This copies the calculated parameters to the corresponding input fields in the main window.

For a replication study, say we want to know the sample size required for detecting deviations from H_0: $\rho = 0$ consistent with the above H_1 scenario using a one-tailed test

and a power $(1-\beta) = .95$, given $\alpha = .05$. If we choose the a priori type of power analysis and insert the corresponding input parameters, clicking on "Calculate" provides us with the result "Total sample size" = 229, along with "Critical z" = 1.644854 for the z test based on the exact method of Brown and Benedetti (1977).

2. Correlation Problems Referring to Two Dependent Correlations

This section refers to z tests comparing two dependent Pearson correlations that either share (Section 2.1) or do not share (Section 2.2) a common index.

2.1. Comparison of Two Dependent Correlations ρ_{ab} and ρ_{ac} (Common Index)

The "Correlations: Two dependent Pearson r's (common index)" procedure provides power analyses for tests of the null hypothesis that two dependent Pearson correlations ρ_{ab} and ρ_{ac} are identical (H_0: $\rho_{ab} = \rho_{ac}$). Two correlations are dependent if they are defined for the same population. Correspondingly, their sample estimates, r_{ab} and r_{ac}, are observed for the same sample of N observations of three continuous random variables X_a, X_b, and X_c. The two correlations share a common index because one of the three random variables, X_a, is involved in both correlations. Assuming that X_a, X_b, and X_c are multivariate normally distributed, Steiger's (1980, Equation 11) $Z1^*$ statistic follows a standard normal distribution under H_0 (see also Dunn & Clark, 1969). G*Power's power calculations for dependent correlations sharing a common index refer to this test.

Effect size measure. To specify the effect size, both correlations ρ_{ab} and ρ_{ac} are required as input parameters. Alternatively, clicking on "Determine" opens the effect size drawer, which can be used to compute ρ_{ac} from ρ_{ab} and Cohen's (1988, p. 109) effect size measure q, the difference between the Fisher r-to-z transforms of ρ_{ab} and ρ_{ac}. Cohen suggested calling effects of sizes $q = .1$, .3, and .5 "small," "medium," and "large," respectively. Note, however, that Cohen developed his q effect size conventions for comparisons between *independent* correlations in different populations. Depending on the size of the third correlation involved, ρ_{bc}, a specific value of q can have very different meanings, resulting in huge effects on statistical power (see the example below). As a consequence, ρ_{bc} is required as a further input parameter.

Input and output parameters. Apart from the number of "Tail(s)" of the z test, the post hoc power analysis procedure requires ρ_{ac} (i.e., "H1 Corr ρ_ac"), the significance level "α err prob," the "Sample Size" N, and the two remaining correlations "H0 Corr ρ_ab" and "Corr ρ_bc" as input parameters in the lower left field of the main window. To the right, the "Critical z" criterion value for the z test and the "Power (1−β err prob)" are displayed as output parameters.

Illustrative example. Tsujimoto, Kuwajima, and Sawaguchi (2007, p. 34, Table 2) studied correlations between age and several continuous measures of working memory and executive functioning in children. In 8- to 9-year-olds, they found age correlations of .27 and .17 with visuospatial working memory (VWM) and auditory working memory (AWM), respectively. The correlation of the two working memory measures was r(VWM, AWM) = .17. Assume we would like to know whether VWM is more strongly correlated with age than AWM in the underlying population. In other words, H_0: ρ(Age, VWM) ≤ ρ(Age, AWM) is tested against the one-tailed H_1: ρ(Age, VWM) > ρ(Age, AWM). Assuming that the true population correlations correspond to the results reported by Tsujimoto et al., what is the sample size required to detect such a correlation difference with a power of $1-\beta = .95$ and $\alpha = .05$? To find the answer, we choose the a priori type of power analysis along with "Correlations: Two dependent Pearson r's (common index)" and insert the above parameters ($\rho_{ac} = .27$, $\rho_{ab} = .17$, $\rho_{bc} = .17$) in the corresponding input fields. Clicking on "Calculate" provides us with N = 1,663 as the required sample size. Note that this N would drop to N = 408 if we assumed ρ(VWM, AWM) = .80 rather than ρ(VWM, AWM) = .17, other parameters being equal. Obviously, the third correlation ρ_{bc} has a strong impact on statistical power, although it does not affect whether H_0 or H_1 holds in the underlying population.

2.2. Comparison of Two Dependent Correlations ρ_{ab} and ρ_{cd} (No Common Index)

The "Correlations: Two dependent Pearson r's (no common index)" procedure is very similar to the procedure described in the preceding section. The single difference is that H_0: $\rho_{ab} = \rho_{cd}$ is now contrasted against H_1: $\rho_{ab} \neq \rho_{cd}$; that is, the two dependent correlations here do *not* share a common index. G*Power's power analysis procedures for this scenario refer to Steiger's (1980, Equation 12) $Z2^*$ test statistic. As with $Z1^*$, which is actually a special case of $Z2^*$, the $Z2^*$ statistic is asymptotically z distributed under H_0 given a multivariate normal distribution of the four random variables X_a, X_b, X_c, and X_d involved in ρ_{ab} and ρ_{cd} (see also Dunn & Clark, 1969).

Effect size measure. The effect size specification is identical to that for "Correlations: Two dependent Pearson r's (common index)," except that all six pairwise correlations of the random variables X_a, X_b, X_c, and X_d under H_1 need to be specified here. As a consequence, ρ_{ac}, ρ_{ad}, ρ_{bc}, and ρ_{bd} are required as input parameters in addition to ρ_{ab} and ρ_{cd}.

Input and output parameters. By implication, the input parameters include "Corr ρ_ac," "Corr ρ_ad," "Corr ρ_bc," and "Corr ρ_bd" in addition to the correlations "H1 corr ρ_cd" and "H0 corr ρ_ab" to which the hypotheses refer. There are no other differences from the procedure described in the previous section.

Illustrative example. Nosek and Smyth (2007) reported a multitrait–multimethod validation using two attitude measurement methods, the Implicit Association Test (IAT) and self-report (SR). IAT and SR measures of attitudes toward Democrats versus Republicans (DR) were correlated at r(IAT-DR, SR-DR) = .51 = r_{cd}. In contrast, when measuring attitudes toward whites versus blacks (WB), the correlation between both methods was only r(IAT-WB, SR-WB) = .12 = r_{ab}, probably because

SR measures of attitudes toward blacks are more strongly biased by social desirability influences. Assuming that (1) these correlations correspond to the true population correlations under H_1 and (2) the other four between-attitude correlations ρ(IAT-WB, IAT-DR), ρ(IAT-WB, SR-DR), ρ(SR-WB, IAT-DR), and ρ(SR-WB, SR-DR) are zero, how large must the sample be to make sure that this deviation from H_0: ρ(IAT-DR, SR-DR) = ρ(IAT-WB, SR-WB) is detected with a power of $1-\beta = .95$ using a one-tailed test and $\alpha = .05$? An a priori power analysis for "Correlations: Two dependent Pearson r's (no common index)" computes $N = 114$ as the required sample size.

The assumption that the four additional correlations are zero is tantamount to assuming that the two correlations under test are statistically independent (thus, the procedure in G*Power for independent correlations could alternatively have been used). If we instead assume that ρ(IAT-WB, IAT-DR) = $\rho_{ac} = .6$ and ρ(SR-WB, SR-DR) = $\rho_{bd} = .7$, we arrive at a considerably lower sample size of $N = 56$. If our resources were sufficient for recruiting not more than $N = 40$ participants and we wanted to make sure that the "β/α ratio" equals 1 (i.e., balanced error risks with $\alpha = \beta$), a compromise power analysis for the latter case computes "Critical z" = 1.385675 as the optimal statistical decision criterion, corresponding to $\alpha = \beta = .082923$.

3. Linear Regression Problems, One Predictor (Simple Linear Regression)

This section summarizes power analysis procedures addressing regression coefficients in the bivariate linear standard model $Y_i = a + b \cdot X_i + E_i$, where Y_i and X_i represent the criterion and the predictor variable, respectively, a and b the regression coefficients of interest, and E_i an error term that is independently and identically distributed and follows a normal distribution with expectation 0 and homogeneous variance σ^2 for each observation unit i. Section 3.1 describes a one-sample procedure for tests addressing b, whereas Sections 3.2 and 3.3 refer to hypotheses on differences in a and b between two different underlying populations. Formally, the tests considered here are special cases of the multiple linear regression procedures described in Section 4. However, the procedures for the special case provide a more convenient interface that may be easier to use and interpret if users are interested in bivariate regression problems only (see also Dupont & Plummer, 1998).

3.1. Comparison of a Slope b With a Constant b_0

The "Linear bivariate regression: One group, size of slope" procedure computes the power of the t test of H_0: $b = b_0$ against H_1: $b \neq b_0$, where b_0 is any real-valued constant. Note that this test is equivalent to the standard bivariate regression t test of H_0: $b^* = 0$ against H_1: $b^* \neq 0$ if we refer to the modified model $Y_i^* = a + b^* \cdot X_i + E_i$, with $Y_i^* := Y_i - b_0 \cdot X_i$ (Rindskopf, 1984). Hence, power could also be assessed by referring to the standard regression t test (or global F test) using Y^* rather than Y as a criterion variable. The main advantage of the "Linear bivariate regression: One group, size of slope" procedure is that power can be computed as a function of the slope values under H_0 and under H_1 directly (Dupont & Plummer, 1998).

Effect size measure. The slope b assumed under H_1, labeled "Slope H1," is used as the effect size measure. Note that the power does not depend only on the difference between "Slope H1" and "Slope H0," the latter of which is the value $b = b_0$ specified by H_0. The population standard deviations of the predictor and criterion values, "Std dev σ_x" and "Std dev σ_y," are also required. The effect size drawer can be used to calculate "Slope H1" from other basic parameters such as the correlation ρ assumed under H_1.

Input and output parameters. The number of "Tail(s)" of the t test, "Slope H1," "α err prob," "Total sample size," "Slope H0," and the standard deviations ("Std dev σ_x" and "Std dev σ_y") need to be specified as input parameters for a post hoc power analysis. "Power (1−β err prob)" is displayed as an output parameter, in addition to the "Critical t" decision criterion and the parameters defining the noncentral t distribution implied by H_1 (the noncentrality parameter δ and the df of the test).

Illustrative example. Assume that we would like to assess whether the standardized regression coefficient β of a bivariate linear regression of Y on X is consistent with H_0: $\beta \geq .40$ or H_1: $\beta < .40$. Assuming that $\beta = .20$ actually holds in the underlying population, how large must the sample size N of X–Y pairs be to obtain a power of $1-\beta = .95$ given $\alpha = .05$? After choosing "Linear bivariate regression: One group, size of slope" and the a priori type of power analysis, we specify the above input parameters, making sure that "Tail(s)" = one, "Slope H1" = .20, "Slope H0" = .40, and "Std dev σ_x" = "Std dev σ_y" = 1, because we want to refer to regression coefficients for *standardized* variables. Clicking on "Calculate" provides us with the result "Total sample size" = 262.

3.2. Comparison of Two Independent Intercepts a_1 and a_2 (Two Samples)

The "Linear bivariate regression: Two groups, difference between intercepts" procedure is based on the assumption that the standard bivariate linear model described above holds within each of two different populations with the same slope b and possibly different intercepts a_1 and a_2. It computes the power of the two-tailed t test of H_0: $a_1 = a_2$ against H_1: $a_1 \neq a_2$ and for the corresponding one-tailed t test, as described in Armitage, Berry, and Matthews (2002, ch. 11).

Effect size measure. The absolute value of the difference between the intercepts, $|\Delta \text{ intercept}| = |a_1 - a_2|$, is used as an effect size measure. In addition to |Δ intercept|, the significance level α, and the sizes n_1 and n_2 of the two samples, the power depends on the means and standard deviations of the criterion and the predictor variable.

Input and output parameters. The number of "Tail(s)" of the t test, the effect size "|Δ intercept|," the "α err prob," the sample sizes in both groups, the standard deviation of the error variable E_{ij} ("Std dev residual σ"), the means ("Mean μ_x1," "Mean μ_x2"), and the standard deviations ("Std dev σ_x1," "Std dev σ_x2") are required as input parameters for a "Post hoc" power analysis. The

"Power (1−β err prob)" is displayed as an output parameter in addition to the "Critical t" decision criterion and the parameters defining the noncentral *t* distribution implied by H_1 (the noncentrality parameter δ and the *df* of the test).

3.3. Comparison of Two Independent Slopes b_1 and b_2 (Two Samples)

The linear model used in the previous two sections also underlies the "Linear bivariate regression: Two groups, differences between slopes" procedure. Two independent samples are assumed to be drawn from two different populations, each consistent with the model $Y_{ij} = a_j + b_j \cdot X_i + E_i$, where Y_{ij}, X_{ij}, and E_{ij} respectively represent the criterion, the predictor, and the normally distributed error variable for observation unit i in population j. Parameters a_j and b_j denote the regression coefficients in population j, $j = 1, 2$. The procedure provides power analyses for two-tailed *t* tests of the null hypothesis that the slopes in the two populations are equal (H_0: $b_1 = b_2$) versus different (H_1: $b_1 \neq b_2$) and for the corresponding one-tailed *t* test (Armitage et al., 2002, ch. 11, Equations 11.18–11.20).

Effect size measure. The absolute difference between slopes, $|\Delta \text{ slope}| = |b_1 - b_2|$, is used as an effect size measure. Statistical power depends not only on $|\Delta \text{ slope}|$, α, and the two sample sizes n_1 and n_2. Specifically, the standard deviations of the error variable E_{ij} ("Std dev residual σ"), the predictor variable ("Std dev σ_X"), and the criterion variable ("Std dev σ_Y") in both groups are required to fully specify the effect size.

Input and output parameters. The input and output parameters are similar to those for the two procedures described in Sections 3.1 and 3.2. In addition to |Δ slope| and the standard deviations, the number of "Tail(s)" of the test, the "α err prob," and the two sample sizes are required in the Input Parameters fields for the post hoc type of power analysis. In the Output Parameters fields, the "Noncentrality parameter δ" of the *t* distribution under H_1, the decision criterion ("Critical t"), the degrees of freedom of the *t* test ("Df"), and the "Power (1−β err prob)" implied by the input parameters are provided.

Illustrative application. Perugini, O'Gorman, and Prestwich (2007, Study 1) hypothesized that the criterion validity of the IAT depends on the degree of self-activation in the test situation. To test this hypothesis, they asked 60 participants to circle certain words while reading a short story printed on a sheet of paper. Half of the participants were asked to circle the words "the" and "a" (control condition), whereas the remaining 30 participants were asked to circle "I," "me," "my," and "myself" (self-activation condition). Subsequently, attitudes toward alcohol versus soft drinks were measured using the IAT (predictor *X*). In addition, actual alcohol consumption rate was assessed using the self-report (criterion *Y*). Consistent with their hypothesis, they found standardized *Y*–*X* regression coefficients of $\beta = .48$ and $\beta = -.09$ in the self-activation and control conditions, respectively. Assuming that (1) these coefficients correspond to the actual coefficients under H_1 in the underlying populations and (2) the error standard deviation is .80, how large is the power of the one-tailed *t* test of H_0: $b_1 \leq b_2$ against H_1: $b_1 > b_2$, given $\alpha = .05$? To answer this question, we select "Linear bivariate regression: Two groups, difference between slopes" along with the post hoc type of power analysis. We then provide the appropriate input parameters ["Tail(s)" = one, "|Δ slope|" = .57, "α error prob" = .05, "Sample size group 1" = "Sample size group 2" = 30, "Std dev. residual σ" = .80, and "Std dev σ_X1" = "Std dev σ_X2" = 1] and click on "Calculate." We obtain "Power (1−β err prob)" = .860165.

4. Linear Regression Problems, Several Predictors (Multiple Linear Regression)

In multiple linear regression, the linear relation between a criterion variable *Y* and *m* predictors $X = (X_1, \ldots, X_m)$ is studied. G*Power 3.1 now provides power analysis procedures for both the conditional (or fixed-predictors) and the unconditional (or random-predictors) models of multiple regression (Gatsonis & Sampson, 1989; Sampson, 1974). In the fixed-predictors model underlying previous versions of G*Power, the predictors *X* are assumed to be fixed and known. In the random-predictors model, by contrast, they are assumed to be random variables with values sampled from an underlying multivariate normal distribution. Whereas the fixed-predictors model is often more appropriate in experimental research (where known predictor values are typically *assigned* to participants by an experimenter), the random-predictors model more closely resembles the design of observational studies (where participants and their associated predictor values are *sampled* from an underlying population). The test procedures and the maximum likelihood estimates of the regression weights are identical for both models. However, the models differ with respect to statistical power.

Sections 4.1, 4.2, and 4.3 describe procedures related to *F* and *t* tests in the fixed-predictors model of multiple linear regression (cf. Cohen, 1988, ch. 9), whereas Section 4.4 describes the procedure for the random-predictors model. The procedures for the fixed-predictors model are based on the general linear model (GLM), which includes the bivariate linear model described in Sections 3.1–3.3 as a special case (Cohen, Cohen, West, & Aiken, 2003). In other words, the following procedures, unlike those in the previous section, have the advantage that they are not limited to a single predictor variable. However, their disadvantage is that effect size specifications cannot be made in terms of regression coefficients under H_0 and H_1 directly. Rather, variance proportions, or ratios of variance proportions, are used to define H_0 and H_1 (cf. Cohen, 1988, ch. 9). For this reason, we decided to include both bivariate and multiple linear regression procedures in G*Power 3.1. The latter set of procedures is recommended whenever statistical hypotheses are defined in terms of proportions of explained variance or whenever they can easily be transformed into hypotheses referring to such proportions.

4.1. Deviation of a Squared Multiple Correlation ρ^2 From Zero (Fixed Model)

The "Linear multiple regression: Fixed model, R^2 deviation from zero" procedure provides power analyses for

omnibus (or "global") F tests of the null hypothesis that the squared multiple correlation between a criterion variable Y and a set of m predictor variables $X_1, X_2, \ldots, X_m$ is zero in the underlying population (H_0: $\rho^2_{Y.X1,\ldots,Xm} = 0$) versus larger than zero (H_1: $\rho^2_{Y.X1,\ldots,Xm} > 0$). Note that the former hypothesis is equivalent to the hypothesis that all m regression coefficients of the predictors are zero (H_0: $b_1 = b_2 = \ldots = b_m = 0$). By implication, the omnibus F test can also be used to test fully specified linear models of the type $Y_i = b_0 + c_1 \cdot X_{1i} + \ldots + c_m \cdot X_{mi} + E_i$, where $c_1, \ldots, c_m$ are user-specified real-valued constants defining H_0. To test this fully specified model, simply define a new criterion variable $Y_i^* := Y_i - c_1 \cdot X_{1i} - \ldots - c_m \cdot X_{mi}$ and perform a multiple regression of Y^* on the m predictors X_1 to X_m. H_0 holds if and only if $\rho^2_{Y^*.X1,\ldots,Xm} = 0$—that is, if all regression coefficients are zero in the transformed regression equation pertaining to Y^* (see Rindskopf, 1984).

Effect size measure. Cohen's f^2, the ratio of explained variance and error variance, serves as the effect size measure (Cohen, 1988, ch. 9). Using the effect size drawer, f^2 can be calculated directly from the squared multiple correlation $\rho^2_{Y.X1,\ldots,Xm}$ in the underlying population. For omnibus F tests, the relation between f^2 and $\rho^2_{Y.X1,\ldots,Xm}$ is simply given by $f^2 = \rho^2_{Y.X1,\ldots,Xm} / (1 - \rho^2_{Y.X1,\ldots,Xm})$. According to Cohen (ch. 9), f^2 values of .02, .15, and .35 can be called "small," "medium," and "large" effects, respectively. Alternatively, one may compute f^2 by specifying a vector u of correlations between Y and the predictors X_i along with the $(m \times m)$ matrix B of intercorrelations among the predictors. Given u and B, it is easy to derive $\rho^2_{Y.X1,\ldots,Xm} = u^T B^{-1} u$ and, via the relation between f^2 and $\rho^2_{Y.X1,\ldots,Xm}$ described above, also f^2.

Input and output parameters. The post hoc power analysis procedure requires the population "Effect size f²," the "α err prob," the "Total sample size" N, and the "Total number of predictors" m in the regression model as input parameters. It provides as output parameters the "Noncentrality parameter λ" of the F distribution under H_1, the decision criterion ("Critical F"), the degrees of freedom ("Numerator df," "Denominator df"), and the power of the omnibus F test ["Power (1−β err prob)"].

Illustrative example. Given three predictor variables X_1, X_2, and X_3, presumably correlated to Y with $\rho_1 = .3$, $\rho_2 = -.1$, and $\rho_3 = .7$, and with pairwise correlations of $\rho_{13} = .4$ and $\rho_{12} = \rho_{23} = 0$, we first determine the effect size f^2. Inserting $b = (.3, -.1, .7)$ and the intercorrelation matrix of the predictors in the corresponding input dialog in G*Power's effect size drawer, we find that $\rho^2_{Y.X1,\ldots,Xm} = .5$ and $f^2 = 1$. The results of an a priori analysis with this effect size reveals that we need a sample size of $N = 22$ to achieve a power of .95 in a test based on $\alpha = .05$.

4.2. Deviation of a Subset of Linear Regression Coefficients From Zero (Fixed Model)

Like the previous procedure, this procedure is based on the GLM, but this one considers the case of two Predictor Sets A (including $X_1, \ldots, X_k$) and B (including $X_{k+1}, \ldots, X_m$) that define the full model with m predictors. The "Linear multiple regression: Fixed model, R² increase" procedure provides power analyses for the special F test of H_0: $\rho^2_{Y.X1,\ldots,Xm} = \rho^2_{Y.X1,\ldots,Xk}$ (i.e., Set B *does not* increase the proportion of explained variance) versus H_1: $\rho^2_{Y.X1,\ldots,Xm} > \rho^2_{Y.X1,\ldots,Xk}$ (i.e., Set B increases the proportion of explained variance). Note that H_0 is tantamount to claiming that all $m-k$ regression coefficients of Set B are zero.

As shown by Rindskopf (1984), special F tests can also be used to assess various constraints on regression coefficients in linear models. For example, if $Y_i = b_0 + b_1 \cdot X_{1i} + \ldots + b_m \cdot X_{mi} + E_i$ is the full linear model and $Y_i = b_0 + b \cdot X_{1i} + \ldots + b \cdot X_{mi} + E_i$ is a restricted H_0 model claiming that all m regression coefficients are equal (H_0: $b_1 = b_2 = \ldots = b_m = b$), we can define a new predictor variable $X_i := X_{1i} + X_{2i} + \ldots + X_{mi}$ and consider the restricted model $Y_i = b_0 + b \cdot X_i + E_i$. Because this model is equivalent to the H_0 model, we can compare $\rho^2_{Y.X1,\ldots,Xm}$ and $\rho^2_{Y.X}$ with a special F test to test H_0.

Effect size measure. Cohen's (1988) f^2 is again used as an effect size measure. However, here we are interested in the proportion of variance explained by predictors from Set B only, so that $f^2 = (\rho^2_{Y.X1,\ldots,Xm} - \rho^2_{Y.X1,\ldots,Xk}) / (1 - \rho^2_{Y.X1,\ldots,Xm})$ serves as an effect size measure. The effect size drawer can be used to calculate f^2 from the variance explained by Set B (i.e., $\rho^2_{Y.X1,\ldots,Xm} - \rho^2_{Y.X1,\ldots,Xk}$) and the error variance (i.e., $1 - \rho^2_{Y.X1,\ldots,Xm}$). Alternatively, f^2 can also be computed as a function of the partial correlation squared of Set B predictors (Cohen, 1988, ch. 9).

Cohen (1988) suggested the same effect size conventions as in the case of global F tests (see Section 4.1). However, we believe that researchers should reflect the fact that a certain effect size, say $f^2 = .15$, may have very different substantive meanings depending on the proportion of variance explained by Set A (i.e., $\rho^2_{Y.X1,\ldots,Xk}$).

Input and output parameters. The inputs and outputs match those of the "Linear multiple regression: Fixed model, R² deviation from zero" procedure (see Section 4.1), with the exception that the "Number of tested predictors" $:= m - k$—that is, the number of predictors in Set B—is required as an additional input parameter.

Illustrative example. In Section 3.3, we presented a power analysis for differences in regression slopes as analyzed by Perugini et al. (2007, Study 1). Multiple linear regression provides an alternative method to address the same problem. In addition to the self-report of alcohol use (criterion Y) and the IAT attitude measure (predictor X), two additional predictors are required for this purpose: a binary dummy variable G representing the experimental condition ($G = 0$, control group; $G = 1$, self-activation group) and, most importantly, a product variable $G \cdot X$ representing the interaction of the IAT measure and the experimental conditional. Differences in Y–X regression slopes in the two groups will show up as a significant effect of the $G \cdot X$ interaction in a regression model using Y as the criterion and X, G, and $G \cdot X$ as $m = 3$ predictors.

Given a total of $N = 60$ participants (30 in each group), $\alpha = .05$, and a medium size $f^2 = .15$ of the interaction effect in the underlying population, how large is the power of the special F test assessing the increase in explained variance due to the interaction? To answer this question, we choose the post hoc power analysis in the "Linear mul-

tiple regression: Fixed model, R^2 increase" procedure. In addition to "Effect size f^2" = .15, "α err prob" = .05, and "Total sample size" = 60, we specify "Number of tested predictors" = 1 and "Total number of predictors" = 3 as input parameters, because we have $m = 3$ predictors in the full model and only one of these predictors captures the effect of interest—namely, the interaction effect. Clicking on "Calculate" provides us with the result "Power (1−β err prob)" = .838477.

4.3. Deviation of a Single Linear Regression Coefficient b_j From Zero (t Test, Fixed Model)

Special F tests assessing effects of a single predictor X_j in multiple regression models (hence, numerator $df = 1$) are equivalent to two-tailed t tests of H_0: $b_j = 0$. The "Linear multiple regression: Fixed model, single regression coefficient" procedure has been designed for this situation. The main reason for including this procedure in G*Power 3.1 is that t tests for single regression coefficients can take the form of one-tailed tests of, for example, H_0: $b_j \leq 0$ versus H_1: $b_j > 0$. Power analyses for one-tailed tests can be done most conveniently with the "Linear multiple regression: Fixed model, single regression coefficient" t test procedure.

Effect size measures, input and output parameters. Because two-tailed regression t tests are special cases of the special F tests described in Section 4.2, the effect size measure and the input and output parameters are largely the same for both procedures. One exception is that "Numerator df" is not required as an input parameter for t tests. Also, the number of "Tail(s)" of the test (one vs. two) is required as an additional input parameter in the t test procedure.

Illustrative application. See the example described in Section 4.2. For the reasons outlined above, we would obtain the same power results if we were to analyze the same input parameters using the "Linear multiple regression: Fixed model, single regression coefficient" procedure with "Tail(s)" = two. In contrast, if we were to choose "Tail(s)" = one and keep the other parameters unchanged, the power would increase to .906347.

4.4. Deviation of Multiple Correlation Squared ρ^2 From Constant (Random Model)

The random-predictors model of multiple linear regression is based on the assumption that $(Y, X_1, \ldots, X_m)$ are random variables with a joint multivariate normal distribution. Sampson (1974) showed that choice of the fixed or the random model has no bearing on the test of significance or on the estimation of the regression weights. However, the choice of model affects the power of the test. Several programs exist that can be used to assess the power for random-model tests (e.g., Dunlap, Xin, & Myers, 2004; Mendoza & Stafford, 2001; Shieh & Kung, 2007; Steiger & Fouladi, 1992). However, these programs either rely on other software packages (Mathematica, Excel) or provide only a rather inflexible user interface. The procedures implemented in G*Power use the exact sampling distribution of the squared multiple correlation coefficient (Benton & Krishnamoorthy, 2003). Alternatively, the three-moment F approximation suggested by Lee (1972) can also be used.

In addition to the five types of power analysis and the graphic options that G*Power 3.1 supports for any test, the program provides several other useful procedures for the random-predictors model: (1) It is possible to calculate exact confidence intervals and confidence bounds for the squared multiple correlation; (2) the critical values of R^2 given in the output may be used in hypothesis testing; and (3) the probability density function, the cumulative distribution function, and the quantiles of the sampling distribution of the squared multiple correlation coefficients are available via the G*Power calculator.

The implemented procedures provide power analyses for tests of the null hypothesis that the population squared multiple correlation coefficient ρ^2 equals ρ_0^2 (H_0: $\rho^2 = \rho_0^2$) versus a one- or a two-tailed alternative.

Effect size measure. The squared population correlation coefficient ρ^2 under the alternative hypothesis (H_1 ρ^2) serves as an effect size measure. To fully specify the effect size, the squared multiple correlation under the null hypothesis (H_0 ρ^2) is also needed.

The effect size drawer offers two ways to calculate the effect size ρ^2. First, it is possible to choose a certain percentile of the $(1-\alpha)$ confidence interval calculated for an observed R^2 as H_1 ρ^2. Alternatively, H_1 ρ^2 can be obtained by specifying a vector u of correlations between criterion Y and predictors X_i and the $(m \times m)$ matrix B of correlations among the predictors. By definition, $\rho^2 = u^T B^{-1} u$.

Input and output parameters. The post hoc power analysis procedure requires the type of test ("Tail(s)": one vs. two), the population ρ^2 under H_1, the population ρ^2 under H_0, the "α error prob," the "Total sample size" N, and the "Number of predictors" m in the regression model as input parameters. It provides the "Lower critical R^2," the "Upper critical R^2," and the power of the test "Power (1−β err prob)" as output. For a two-tailed test, H_0 is retained whenever the sample R^2 lies in the interval defined by "Lower critical R^2" and "Upper critical R^2"; otherwise, H_0 is rejected. For one-tailed tests, in contrast, "Lower critical R^2" and "Upper critical R^2" are identical; H_0 is rejected if and only if R^2 exceeds this critical value.

Illustrative examples. In Section 4.1, we presented a power analysis for a three-predictor example using the fixed-predictors model. Using the same input procedure in the effect size drawer (opened by pressing "Insert/edit matrix" in the "From predictor correlations" section) described in Section 4.1, we again find "H1 ρ^2" = .5. However, the result of the same a priori analysis (power = .95, α = .05, 3 predictors) previously computed for the fixed model shows that we now need a sample size of at least $N = 26$ for the random model, instead of only the $N = 22$ previously found for the fixed model. This difference illustrates the fact that sample sizes required for the random-predictors model are always slightly larger than those required for the corresponding fixed-predictors model.

As a further example, we use the "From confidence interval" procedure in the effect size drawer to determine the effect size based on the results of a pilot study. We insert the values from a previous study ("Total sample

size" = 50, "Number of predictors" = 5, and "Observed R^2" = .3), choose "Confidence level (1−α)" = .95, and choose the center of the confidence interval as "H1 ρ^2" ["Rel. C.I. pos to use (0=left,1=right)" = .5]. Pressing "Calculate" produces the two-sided interval (.0337, .4603) and the two one-sided intervals (0, .4245) and (.0589, 1), thus confirming the values computed by Shieh and Kung (2007, p. 733). In addition, "H1 ρ^2" is set to .2470214, the center of the interval. Given this value, an a priori analysis of the one-sided test of H_0: $\rho^2 = 0$ using $\alpha = .05$ shows that we need a sample size of at least 71 to achieve a power of .95. The output also provides "Upper critical R^2" = .153427. Thus, if in our new study we found R^2 to be larger than this value, the result here implies that the test would be significant at the .05 α level.

Suppose we were to find a value of $R^2 = .19$. Then we could use the G*Power calculator to compute the p value of the test statistic: The syntax for the cumulative distribution function of the sampling distribution of R^2 is "mr2cdf(R^2, ρ^2, m+1, N)". Thus, in our case, we need to insert "1−mr2cdf(0.19, 0, 5+1, 71)" in the calculator. Pressing "Calculate" shows that $p = .0156$.

5. Generalized Linear Regression Problems

G*Power 3.1 includes power procedures for logistic and Poisson regression models in which the predictor variable under test may have one of six different predefined distributions (binary, exponential, log-normal, normal, Poisson, and uniform). In each case, users can choose between the enumeration approach proposed by Lyles, Lin, and Williamson (2007), which allows power calculations for Wald and likelihood ratio tests, and a large-sample approximation of the Wald test based on the work of Demidenko (2007, 2008) and Whittemore (1981). To allow for comparisons with published results, we also implemented simple but less accurate power routines that are in widespread use (i.e., the procedures of Hsieh, Bloch, & Larsen, 1998, and of Signorini, 1991, for logistic and Poisson regressions, respectively). Problems with Whittemore's and Signorini's procedures have already been discussed by Shieh (2001) and need not be reiterated here.

The enumeration procedure of Lyles et al. (2007) is conceptually simple and provides a rather direct, simulation-like approach to power analysis. Its main disadvantage is that it is rather slow and requires much computer memory for analyses with large sample sizes. The recommended practice is therefore to begin with the large-sample approximation and to use the enumeration approach mainly to validate the results (if the sample size is not too large).

Demidenko (2008, pp. 37f) discussed the relative merits of power calculations based on Wald and likelihood ratio tests. A comparison of both approaches using the Lyles et al. (2007) enumeration procedure indicated that in most cases the difference in calculated power is small. In cases in which the results deviated, the simulated power was slightly higher for the likelihood ratio test than for the Wald test procedure, which tended to underestimate the true power, and the likelihood ratio procedure also appeared to be more accurate.

5.1. Logistic Regression

Logistic regression models address the relationship between a binary dependent variable (or criterion) Y and one or more independent variables (or predictors) X_j, with discrete or continuous probability distributions. In contrast to linear regression models, the logit transform of Y, rather than Y itself, serves as the criterion to be predicted by a linear combination of the independent variables. More precisely, if $y = 1$ and $y = 0$ denote the two possible values of Y, with probabilities $p(y = 1)$ and $p(y = 0)$, respectively, so that $p(y = 1) + p(y = 0) = 1$, then $\text{logit}(Y) := \ln[p(y = 1) / p(y = 0)]$ is modeled as $\text{logit}(Y) = \beta_0 + \beta_1 \cdot X_1 + \ldots + \beta_m \cdot X_m$. A logistic regression model is called *simple* if $m = 1$. If $m > 1$, we have a *multiple* logistic regression model. The implemented procedures provide power analyses for the Wald test $z = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$ assessing the effect of a specific predictor X_j (e.g., H_0: $\beta_j = 0$ vs. H_1: $\beta_j \neq 0$, or H_0: $\beta_j \leq 0$ vs. H_1: $\beta_j > 0$) in both simple and multiple logistic regression models. In addition, the procedure of Lyles et al. (2007) also supports power analyses for likelihood ratio tests. In the case of multiple logistic regression models, a simple approximation proposed by Hsieh et al. (1998) is used: The sample size N is multiplied by $(1-R^2)$, where R^2 is the squared multiple correlation coefficient when the predictor of interest is regressed on the other predictors. The following paragraphs refer to the simple model $\text{logit}(Y) = \beta_0 + \beta_1 \cdot X$.

Effect size measure. Given the conditional probability $p_1 := p(Y = 1 \mid X = 1)$ under H_0, we may define the effect size either by specifying $p_2 := p(Y = 1 \mid X = 1)$ under H_1 or by specifying the odds ratio $OR := [p_2/(1 - p_2)] / [p_1/(1 - p_1)]$. The parameters β_0 and β_1 are related to p_1 and p_2 as follows: $\beta_0 = \ln[p_1/(1 - p_1)]$, $\beta_1 = \ln[OR]$. Under H_0, $p_1 = p_2$ or $OR = 1$.

Input and output parameters. The post hoc type of power analysis for the "Logistic regression" procedure requires the following input parameters: (1) the number of "Tail(s)" of the test (one vs. two); (2) "Pr(Y = 1 | X = 1)" under H_0, corresponding to p_1; (3) the effect size [either "Pr(Y = 1 | X = 1)" under H_1 or, optionally, the "Odds ratio" specifying p_2]; (4) the "α err prob"; (5) the "Total sample size" N; and (6) the proportion of variance of X_j explained by additional predictors in the model ("R^2 other X"). Finally, because the power of the test also depends on the distribution of the predictor X, the "X distribution" and its parameters need to be specified. Users may choose between six predefined distributions (binomial, exponential, log-normal, normal, Poisson, or uniform) or select a manual input mode. Depending on this selection, additional parameters (corresponding to the distribution parameters or, in the manual mode, the variances v_0 and v_1 of β_1 under H_0 and H_1, respectively) must be specified. In the manual mode, sensitivity analyses are not possible. After clicking "Calculate," the statistical decision criterion ("Critical z" in the large-sample procedures, "Noncentrality parameter λ," "Critical χ^2," and "df" in the enumeration approach) and the power of the test ["Power (1−β err prob)"] are displayed in the Output Parameters fields.

Illustrative examples. Using multiple logistic regression, Friese, Bluemke, and Wänke (2007) assessed the effects of (1) the intention to vote for a specific political party x (= predictor X_1) and (2) the attitude toward party x as measured with the IAT (= predictor X_2) on actual voting behavior in a political election (= criterion Y). Both X_1 and Y are binary variables taking on the value 1 if a participant intends to vote for x or has actually voted for x, respectively, and the value 0 otherwise. In contrast, X_2 is a continuous implicit attitude measure derived from IAT response time data.

Given the fact that Friese et al. (2007) were able to analyze data for $N = 1{,}386$ participants, what would be a reasonable statistical decision criterion (critical z value) to detect effects of size $p_1 = .30$ and $p_2 = .70$ (so that $OR = .7/.3 \cdot .7/.3 = 5.44$) for predictor X_1, given a base rate $B = .10$ for the intention to vote for the Green party, a two-tailed z test, and balanced α and β error risks so that $q = \beta/\alpha = 1$? A compromise power analysis helps find the answer to this question. In the effect size drawer, we enter "Pr(Y = 1 | X = 1) H1" = .70, "Pr(Y = 1 | X = 1) H0" = .30. Clicking "Calculate and transfer to main window" yields an odds ratio of 5.44444 and transfers this value and the value of "Pr(Y = 1 | X = 1) H0" to the main window. Here we specify "Tail(s)" = two, "β/α ratio" = 1, "Total sample size" = 1,386, "X distribution" = binomial, and "x parm π" = .1 in the Input Parameters fields. If we assume that the attitude toward the Green party (X_2) explains 40% of the variance of X_1, we need "R² other X" = .40 as an additional input parameter. In the Options dialog box, we choose the Demidenko (2007) procedure (without variance correction). Clicking on "Calculate" provides us with "Critical z" = 3.454879, corresponding to $\alpha = \beta = .00055$. Thus, very small α values may be reasonable if the sample size, the effect size, or both are very large, as is the case in the Friese et al. study.

We now refer to the effect of the attitude toward the Green party (i.e., predictor X_2, assumed to be standard normally distributed). Assuming that a proportion of $p_1 = .10$ of the participants with an average attitude toward the Greens actually vote for them, whereas a proportion of .15 of the participants with an attitude one standard deviation above the mean would vote for them [thus, $OR = (.15/.85) \cdot (.90/.10) = 1.588$ and $b_1 = \ln(1.588) = .4625$], what is a reasonable critical z value to detect effects of this size for predictor X_2 with a two-tailed z test and balanced α and β error risks (i.e., $q = \beta/\alpha = 1$)? We set "X distribution" = normal, "X parm μ" = 0, and "X parm σ" = 1. If all other input parameters remain unchanged, a compromise power analysis results in "Critical z" = 2.146971, corresponding to $\alpha = \beta = .031796$. Thus, different decision criteria and error probabilities may be reasonable for different predictors in the same regression model, provided we have reason to expect differences in effect sizes under H_1.

5.2. Poisson Regression

A Poisson regression model describes the relationship between a Poisson-distributed dependent variable (i.e., the criterion Y) and one or more independent variables (i.e., the predictors X_j, $j = 1, \ldots, m$). For a count variable Y indexing the number $Y = y$ of events of a certain type in a fixed amount of time, a Poisson distribution model is often reasonable (Hays, 1972). Poisson distributions are defined by a single parameter λ, the so-called *intensity* of the Poisson process. The larger the λ, the larger the number of critical events per time, as indexed by Y. Poisson regression models are based on the assumption that the logarithm of λ is a linear combination of m predictors $X_j, j = 1, \ldots, m$. In other words, $\ln(\lambda) = \beta_0 + \beta_1 \cdot X_1 + \ldots + \beta_m \cdot X_m$, with β_j measuring the "effect" of predictor X_j on Y.

The procedures currently implemented in G*Power provide power analyses for the Wald test $z = \hat{\beta}_j / \text{se}(\hat{\beta}_j)$, which assesses the effect of a specific predictor X_j (e.g., H_0: $\beta_j = 0$ vs. H_1: $\beta_j \neq 0$ or H_0: $\beta_j \leq 0$ vs. H_1: $\beta_j > 0$) in both simple and multiple Poisson regression models. In addition, the procedure of Lyles et al. (2007) provides power analyses for likelihood ratio tests. In the case of multiple Poisson regression models, a simple approximation proposed by Hsieh et al. (1998) is used: The sample size N is multiplied by $(1-R^2)$, where R^2 is the squared multiple correlation coefficient when the predictor of interest is regressed on the other predictors. In the following discussion, we assume the simple model $\ln(\lambda) = \beta_0 + \beta_1 X$.

Effect size measure. The ratio $R = \lambda(H_1)/\lambda(H_0)$ of the intensities of the Poisson processes under H_1 and H_0, given $X_1 = 1$, is used as an effect size measure. Under H_0, $R = 1$. Under H_1, in contrast, $R = \exp(\beta_1)$.

Input and output parameters. In addition to "Exp(β1)," the number of "Tail(s)" of the test, the "α err prob," and the "Total sample size," the following input parameters are required for post hoc power analysis tests in Poisson regression: (1) The intensity $\lambda = \exp(\beta_0)$ assumed under H_0 ["Base rate exp(β0)"], (2) the mean exposure time during which the Y events are counted ("Mean exposure"), and (3) "R² other X," a factor intended to approximate the influence of additional predictors X_k (if any) on the predictor of interest. The meaning of the latter factor is identical to the correction factor proposed by Hsieh et al. (1998, Equation 2) for multiple logistic regression (see Section 5.1), so that "R² other X" is the proportion of the variance of X explained by the other predictors. Finally, because the power of the test also depends on the distribution of the predictor X, the "X distribution" and its parameters need to be specified. Users may choose from six predefined distributions (binomial, exponential, lognormal, normal, Poisson, or uniform) or select a manual input mode. Depending on this choice, additional parameters (corresponding to distribution parameters or, in the manual mode, the variances v_0 and v_1 of β_1 under H_0 and H_1, respectively) must be specified. In the manual mode, sensitivity analyses are not possible.

In post hoc power analyses, the statistical decision criterion ("Critical z" in the large-sample procedures, "Noncentrality parameter λ," "Critical χ²," and "df" in the enumeration approach) and the power of the test ["Power (1−β err prob)"] are displayed in the Output Parameters fields.

Illustrative example. We refer to the example given in Signorini (1991, p. 449). The number of infections Y (modeled as a Poisson-distributed random variable) of swimmers ($X = 1$) versus nonswimmers ($X = 0$) during a swimming season (defined as "Mean exposure" = 1) is analyzed using a Poisson regression model. How many participants do we need to assess the effect of swimming versus not swimming (X) on the infection counts (Y)? We consider this to be a one-tailed test problem ("Tail(s)" = one). The predictor X is assumed to be a binomially distributed random variable with $\pi = .5$ (i.e., equal numbers of swimmers and nonswimmers are sampled), so that we set "X distribution" to Binomial and "X param π" to .5. The "Base rate exp(β0)"—that is, the infection rate in nonswimmers—is estimated to be .85. We want to know the sample size required to detect a 30% increase in infection rate in swimmers with a power of .95 at $\alpha = .05$. Using the a priori analysis in the "Poisson regression" procedure, we insert the values given above in the corresponding input fields and choose "Exp(β1)" = 1.3 = (100% + 30%)/100% as the to-be-detected effect size. In the single-predictor case considered here, we set "R^2 other X" = 0. The sample size calculated for these values with the Signorini procedure is $N = 697$. The procedure based on Demidenko (2007; without variance correction) and the Wald test procedure proposed by Lyles et al. (2007) both result in $N = 655$. A computer simulation with 150,000 cases revealed mean power values of .952 and .962 for sample sizes 655 and 697, respectively, confirming that Signorini's procedure is less accurate than the other two.

6. Concluding Remarks

G*Power 3.1 is a considerable extension of G*Power 3.0, introduced by Faul et al. (2007). The new version provides a priori, compromise, criterion, post hoc, and sensitivity power analysis procedures for six correlation and nine regression test problems as described above. Readers interested in obtaining a free copy of G*Power 3.1 may download it from the G*Power Web site at www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/. We recommend registering as a G*Power user at the same Web site. Registered G*Power users will be informed about future updates and new G*Power versions.

The current version of G*Power 3.1 runs on Windows platforms only. However, a version for Mac OS X 10.4 (or later) is currently in preparation. Registered users will be informed when the Mac version of G*Power 3.1 is made available for download.

The G*Power site also provides a Web-based manual with detailed information about the statistical background of the tests, the program handling, the implementation, and the validation methods used to make sure that each of the G*Power procedures works correctly and with the appropriate precision. However, although considerable effort has been put into program evaluation, there is no warranty whatsoever. Users are kindly asked to report possible bugs and difficulties in program handling by writing an e-mail to gpower-feedback@uni-duesseldorf.de.

AUTHOR NOTE

Manuscript preparation was supported by Grant SFB 504 (Project A12) from the Deutsche Forschungsgemeinschaft. We thank two anonymous reviewers for valuable comments on a previous version of the manuscript. Correspondence concerning this article should be addressed to E. Erdfelder, Lehrstuhl für Psychologie III, Universität Mannheim, Schloss Ehrenhof Ost 255, D-68131 Mannheim, Germany (e-mail: erdfelder@psychologie.uni-mannheim.de) or F. Faul, Institut für Psychologie, Christian-Albrechts-Universität, Olshausenstr. 40, D-24098 Kiel, Germany (e-mail: ffaul@psychologie.uni-kiel.de).

Note—This article is based on information presented at the 2008 meeting of the Society for Computers in Psychology, Chicago.

REFERENCES

Armitage, P., Berry, G., & Matthews, J. N. S. (2002). *Statistical methods in medical research* (4th ed.). Oxford: Blackwell.

Benton, D., & Krishnamoorthy, K. (2003). Computing discrete mixtures of continuous distributions: Noncentral chisquare, noncentral *t* and the distribution of the square of the sample multiple correlation coefficient. *Computational Statistics & Data Analysis*, **43**, 249-267.

Bonett, D. G., & Price, R. M. (2005). Inferential methods for the tetrachoric correlation coefficient. *Journal of Educational & Behavioral Statistics*, **30**, 213-225.

Bredenkamp, J. (1969). Über die Anwendung von Signifikanztests bei theorie-testenden Experimenten [On the use of significance tests in theory-testing experiments]. *Psychologische Beiträge*, **11**, 275-285.

Brown, M. B., & Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika*, **42**, 347-355.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Demidenko, E. (2007). Sample size determination for logistic regression revisited. *Statistics in Medicine*, **26**, 3385-3397.

Demidenko, E. (2008). Sample size and optimal design for logistic regression with binary interaction. *Statistics in Medicine*, **27**, 36-46.

Dunlap, W. P., Xin, X., & Myers, L. (2004). Computing aspects of power for multiple regression. *Behavior Research Methods, Instruments, & Computers*, **36**, 695-701.

Dunn, O. J., & Clark, V. A. (1969). Correlation coefficients measured on the same individuals. *Journal of the American Statistical Association*, **64**, 366-377.

Dupont, W. D., & Plummer, W. D. (1998). Power and sample size calculations for studies involving linear regression. *Controlled Clinical Trials*, **19**, 589-601.

Erdfelder, E. (1984). Zur Bedeutung und Kontrolle des beta-Fehlers bei der inferenzstatistischen Prüfung log-linearer Modelle [On significance and control of the beta error in statistical tests of log-linear models]. *Zeitschrift f r Sozialpsychologie*, **15**, 18-32.

Erdfelder, E., Faul, F., & Buchner, A. (2005). Power analysis for categorical methods. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1565-1570). Chichester, U.K.: Wiley.

Erdfelder, E., Faul, F., Buchner, A., & Cüpper, L. (in press). Effektgröße und Teststärke [Effect size and power]. In H. Holling & B. Schmitz (Eds.), *Handbuch der Psychologischen Methoden und Evaluation*. Göttingen: Hogrefe.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, **39**, 175-191.

Friese, M., Bluemke, M., & Wänke, M. (2007). Predicting voting behavior with implicit attitude measures: The 2002 German parliamentary election. *Experimental Psychology*, **54**, 247-255.

Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, **106**, 516-524.

Hays, W. L. (1972). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston.

HSIEH, F. Y., BLOCH, D. A., & LARSEN, M. D. (1998). A simple method of sample size calculation for linear and logistic regression. *Statistics in Medicine*, **17**, 1623-1634.

LEE, Y.-S. (1972). Tables of upper percentage points of the multiple correlation coefficient. *Biometrika*, **59**, 175-189.

LYLES, R. H., LIN, H.-M., & WILLIAMSON, J. M. (2007). A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. *Statistics in Medicine*, **26**, 1632-1648.

MENDOZA, J. L., & STAFFORD, K. L. (2001). Confidence intervals, power calculation, and sample size estimation for the squared multiple correlation coefficient under the fixed and random regression models: A computer program and useful standard tables. *Educational & Psychological Measurement*, **61**, 650-667.

NOSEK, B. A., & SMYTH, F. L. (2007). A multitrait–multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, **54**, 14-29.

PERUGINI, M., O'GORMAN, R., & PRESTWICH, A. (2007). An ontological test of the IAT: Self-activation can increase predictive validity. *Experimental Psychology*, **54**, 134-147.

RINDSKOPF, D. (1984). Linear equality restrictions in regression and loglinear models. *Psychological Bulletin*, **96**, 597-603.

SAMPSON, A. R. (1974). A tale of two regressions. *Journal of the American Statistical Association*, **69**, 682-689.

SHIEH, G. (2001). Sample size calculations for logistic and Poisson regression models. *Biometrika*, **88**, 1193-1199.

SHIEH, G., & KUNG, C.-F. (2007). Methodological and computational considerations for multiple correlation analysis. *Behavior Research Methods*, **39**, 731-734.

SIGNORINI, D. F. (1991). Sample size for Poisson regression. *Biometrika*, **78**, 446-450.

STEIGER, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, **87**, 245-251.

STEIGER, J. H., & FOULADI, R. T. (1992). R2: A computer program for interval estimation, power calculations, sample size estimation, and hypothesis testing in multiple regression. *Behavior Research Methods, Instruments, & Computers*, **24**, 581-582.

TSUJIMOTO, S., KUWAJIMA, M., & SAWAGUCHI, T. (2007). Developmental fractionation of working memory and response inhibition during childhood. *Experimental Psychology*, **54**, 30-37.

WHITTEMORE, A. S. (1981). Sample size for logistic regression with small response probabilities. *Journal of the American Statistical Association*, **76**, 27-32.

(Manuscript received December 22, 2008;
revision accepted for publication June 18, 2009.)