

# G \* POWER 3.1 manual

## January 21, 2021

This manual is not yet complete. We will be adding help on more tests in the future. If you cannot find help for your test in this version of the manual, then please check the G\*Power website to see if a more up-to-date version of the manual has been made available.

### Contents

<b>1 Introduction</b>	<b>2</b>	<b>21 t test: Means - difference between two independent means (two groups)</b>	<b>52</b>
<b>2 The G * Power calculator</b>	<b>7</b>	<b>22 Wilcoxon signed-rank test: Means - difference from constant (one sample case)</b>	<b>53</b>
<b>3 Exact: Correlation - Difference from constant (one sample case)</b>	<b>9</b>	<b>23 Wilcoxon signed-rank test: (matched pairs)</b>	<b>55</b>
<b>4 Exact: Proportion - difference from constant (one sample case)</b>	<b>11</b>	<b>24 Wilcoxon-Mann-Whitney test of a difference between two independent means</b>	<b>59</b>
<b>5 Exact: Proportion - inequality, two dependent groups (McNemar)</b>	<b>14</b>	<b>25 t test: Generic case</b>	<b>63</b>
<b>6 Exact: Proportions - inequality of two independent groups (Fisher's exact-test)</b>	<b>17</b>	<b>26 <math>\chi^2</math> test: Variance - difference from constant (one sample case)</b>	<b>64</b>
<b>7 Exact test: Multiple Regression - random model</b>	<b>18</b>	<b>27 z test: Correlation - inequality of two independent Pearson r's</b>	<b>65</b>
<b>8 Exact: Proportion - sign test</b>	<b>22</b>	<b>28 z test: Correlation - inequality of two dependent Pearson r's</b>	<b>66</b>
<b>9 Exact: Generic binomial test</b>	<b>23</b>	<b>29 Z test: Multiple Logistic Regression</b>	<b>70</b>
<b>10 F test: Fixed effects ANOVA - one way</b>	<b>24</b>	<b>30 Z test: Poisson Regression</b>	<b>75</b>
<b>11 F test: Fixed effects ANOVA - special, main effects and interactions</b>	<b>26</b>	<b>31 Z test: Tetrachoric Correlation</b>	<b>80</b>
<b>12 t test: Linear Regression (size of slope, one group)</b>	<b>31</b>	<b>References</b>	<b>84</b>
<b>13 F test: Multiple Regression - omnibus (deviation of <math>R^2</math> from zero), fixed model</b>	<b>33</b>		
<b>14 F test: Multiple Regression - special (increase of <math>R^2</math>), fixed model</b>	<b>36</b>		
<b>15 F test: Inequality of two Variances</b>	<b>39</b>		
<b>16 t test: Correlation - point biserial model</b>	<b>40</b>		
<b>17 t test: Linear Regression (two groups)</b>	<b>42</b>		
<b>18 t test: Linear Regression (two groups)</b>	<b>45</b>		
<b>19 t test: Means - difference between two dependent means (matched pairs)</b>	<b>48</b>		
<b>20 t test: Means - difference from constant (one sample case)</b>	<b>50</b>		

# 1 Introduction

G\*POWER (Fig. 1 shows the main window of the program) covers statistical power analyses for many different statistical tests of the

- $F$  test,
- $t$  test,
- $\chi^2$ -test and
- $z$  test families and some
- exact tests.

G\*POWER provides effect size calculators and graphics options. G\*POWER supports both a distribution-based and a design-based input mode. It contains also a calculator that supports many central and noncentral probability distributions.

G\*POWER is free software and available for Mac OS X and Windows XP/Vista/7/8.

## 1.1 Types of analysis

G\*POWER offers five different types of statistical power analysis:

1. A priori (sample size  $N$  is computed as a function of power level  $1 - \beta$ , significance level  $\alpha$ , and the to-be-detected population effect size)
2. Compromise (both  $\alpha$  and  $1 - \beta$  are computed as functions of effect size,  $N$ , and an error probability ratio  $q = \beta/\alpha$ )
3. Criterion ( $\alpha$  and the associated decision criterion are computed as a function of  $1 - \beta$ , the effect size, and  $N$ )
4. Post-hoc ( $1 - \beta$  is computed as a function of  $\alpha$ , the population effect size, and  $N$ )
5. Sensitivity (population effect size is computed as a function of  $\alpha$ ,  $1 - \beta$ , and  $N$ )

## 1.2 Program handling

**Perform a Power Analysis** Using G\*POWER typically involves the following three steps:

1. Select the statistical test appropriate for your problem.
2. Choose one of the five types of power analysis available
3. Provide the input parameters required for the analysis and click "Calculate".

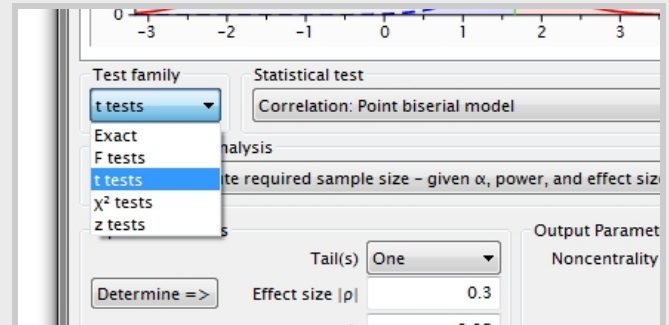
**Plot parameters** In order to help you explore the parameter space relevant to your power analysis, one parameter ( $\alpha$ , power ( $1 - \beta$ ), effect size, or sample size) can be plotted as a function of another parameter.

### 1.2.1 Select the statistical test appropriate for your problem

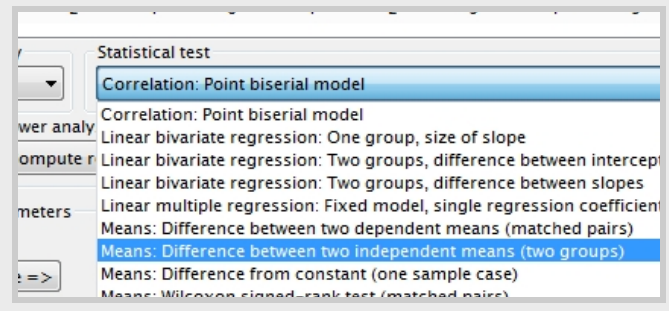
In Step 1, the statistical test is chosen using the distribution-based or the design-based approach.

**Distribution-based approach to test selection** First select the family of the test statistic (i.e., exact,  $F$ -,  $t$ -,  $\chi^2$ , or  $z$ -test) using the *Test family* menu in the main window. The *Statistical test* menu adapts accordingly, showing a list of all tests available for the test family.

**Example:** For the two groups  $t$ -test, first select the test family based on the  $t$  distribution.



Then select *Means: Difference between two independent means (two groups)* option in the *Statistical test* menu.

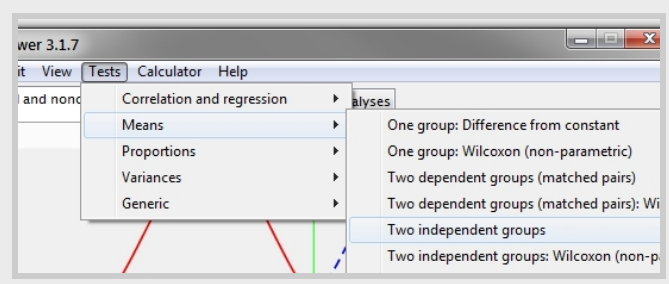


**Design-based approach to the test selection** Alternatively, one might use the design-based approach. With the *Tests* pull-down menu in the top row it is possible to select

- the parameter class the statistical test refers to (i.e., correlations and regression coefficients, means, proportions, or variances), and
- the design of the study (e.g., number of groups, independent vs. dependent samples, etc.).

The design-based approach has the advantage that test options referring to the same parameter class (e.g., means) are located in close proximity, whereas they may be scattered across different distribution families in the distribution-based approach.

**Example:** In the *Tests* menu, select *Means*, then select *Two independent groups* to specify the two-groups  $t$  test.



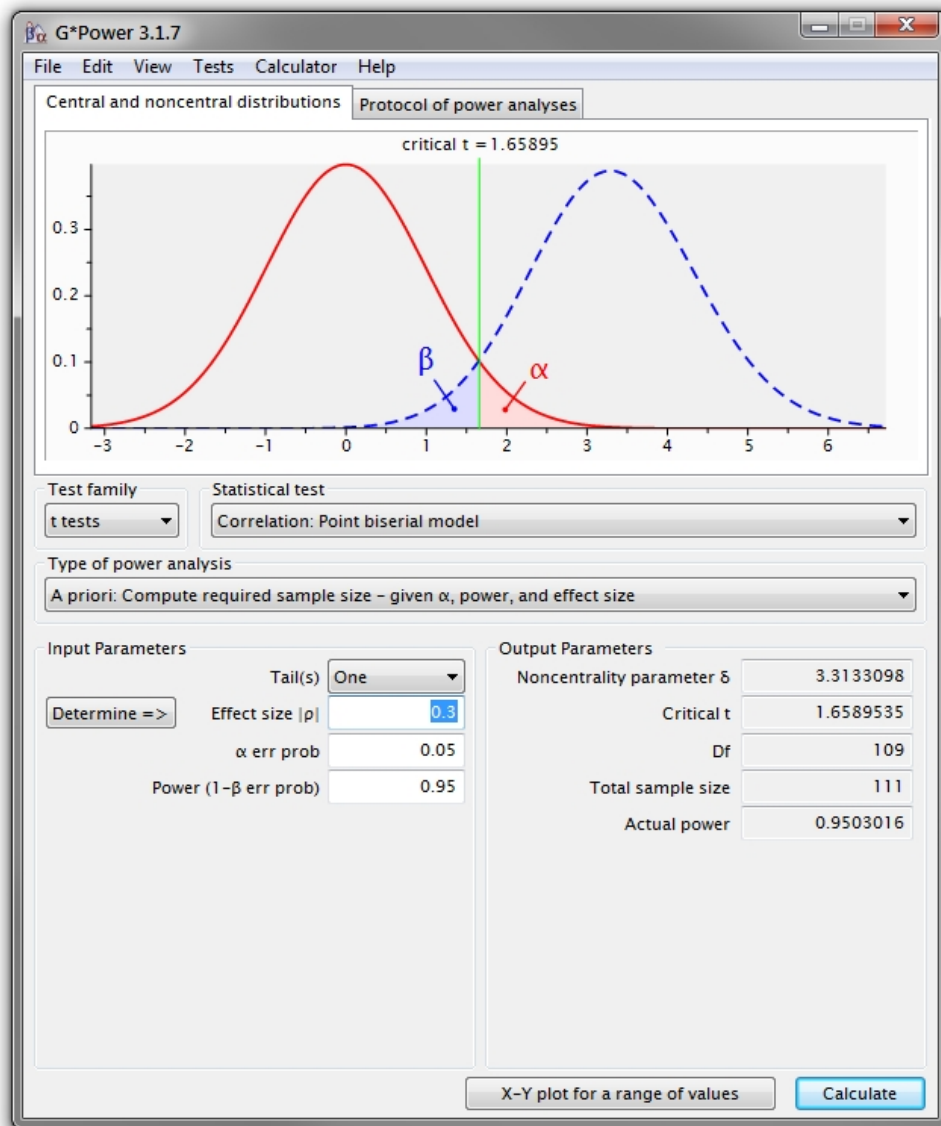
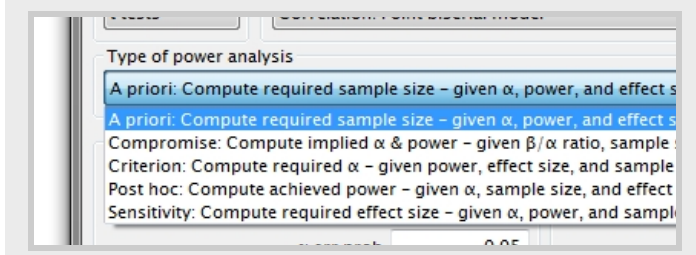


Figure 1: The main window of G\*POWER

### 1.2.2 Choose one of the five types of power analysis available

In Step 2, the *Type of power analysis* menu in the center of the main window is used to choose the appropriate analysis type and the input and output parameters in the window change accordingly.

**Example:** If you choose the first item from the *Type of power analysis* menu the main window will display input and output parameters appropriate for an a priori power analysis (for t tests for independent groups if you followed the example provided in Step 1).



In an **a priori power analysis**, sample size  $N$  is computed as a function of

- the required power level  $(1 - \beta)$ ,
- the pre-specified significance level  $\alpha$ , and
- the population effect size to be detected with probability  $(1 - \beta)$ .

In a **criterion power analysis**,  $\alpha$  (and the associated decision criterion) is computed as a function of

- $1 - \beta$ ,
- the effect size, and
- a given sample size.

In a **compromise power analysis** both  $\alpha$  and  $1 - \beta$  are computed as functions of

- the effect size,
- $N$ , and
- an error probability ratio  $q = \beta/\alpha$ .

In a **post-hoc power analysis** the power  $(1 - \beta)$  is computed as a function of

- $\alpha$ ,
- the population effect size parameter, and
- the sample size(s) used in a study.

In a **sensitivity power analysis** the critical population effect size is computed as a function of

- $\alpha$ ,
- $1 - \beta$ , and
- $N$ .

### 1.2.3 Provide the input parameters required for the analysis

In Step 3, you specify the power analysis input parameters in the lower left of the main window.

**Example:** An a priori power analysis for a two groups  $t$  test would require a decision between a one-tailed and a two-tailed test, a specification of Cohen's (1988) effect size measure  $d$  under  $H_1$ , the significance level  $\alpha$ , the required power  $(1 - \beta)$  of the test, and the preferred group size allocation ratio  $n_2/n_1$ .

Let us specify input parameters for

- a one-tailed  $t$  test,
- a medium effect size of  $d = .5$ ,
- $\alpha = .05$ ,
- $(1 - \beta) = .95$ , and
- an allocation ratio of  $n_2/n_1 = 1$

This would result in a total sample size of  $N = 176$  (i.e., 88 observation units in each group). The noncentrality parameter  $\delta$  defining the  $t$  distribution under  $H_1$ , the decision criterion to be used (i.e., the critical value of the  $t$  statistic), the degrees of freedom of the  $t$  test and the actual power value are also displayed.

Note that the actual power will often be slightly larger than the pre-specified power in a priori power analyses. The reason is that non-integer sample sizes are always rounded up by G\*POWER to obtain integer values consistent with a power level not less than the pre-specified one.

Because Cohen's book on power analysis Cohen (1988) appears to be well known in the social and behavioral sciences, we made use of his effect size measures whenever possible. In addition, wherever available G\*POWER provides his definitions of "small", "medium", and "large" effects as "Tool tips". The tool tips may be obtained by moving the cursor over the "effect size" input parameter field (see below). However, note that these conventions may have different meanings for different tests.

**Example:** The tooltip showing Cohen's measures for the effect size  $d$  used in the two groups  $t$  test

If you are not familiar with Cohen's measures, if you think they are inadequate for your test problem, or if you have more detailed information about the size of the to-be-expected effect (e.g., the results of similar prior studies), then you may want to compute Cohen's measures from more basic parameters. In this case, click on the Determine button to the left the effect size input field. A drawer will open next to the main window and provide access to an effect size calculator tailored to the selected test.

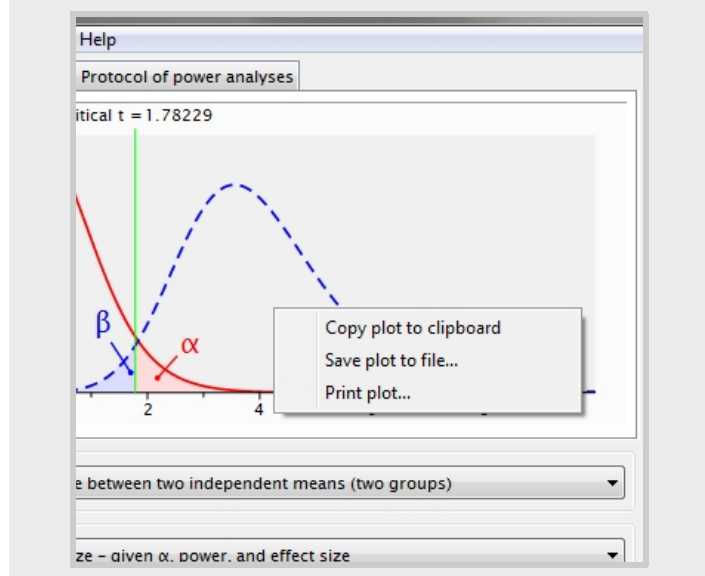
**Example:** For the two-group  $t$ -test users can, for instance, specify the means  $\mu_1, \mu_2$  and the common standard deviation ( $\sigma = \sigma_1 = \sigma_2$ ) in the populations underlying the groups to calculate Cohen's  $d = |\mu_1 - \mu_2|/\sigma$ . Clicking the Calculate and transfer to main window button copies the computed effect size to the appropriate field in the main window

In addition to the numerical output, G\*POWER displays the central ( $H_0$ ) and the noncentral ( $H_1$ ) test statistic distributions along with the decision criterion and the associated error probabilities in the upper part of the main window. This supports understanding the effects of the input parameters and is likely to be a useful visualization tool in

the teaching of, or the learning about, inferential statistics. The distributions plot may be copied, saved, or printed by clicking the right mouse button inside the plot area.

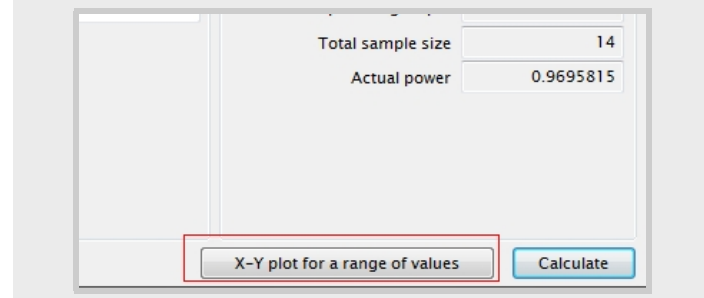
pendent variable  $y$ . In an a prior analysis, for instance, this is the sample size.

**Example:** The menu appearing in the distribution plot for the  $t$ -test after right clicking into the plot.



The input and output of each power calculation in a G\*Power session are automatically written to a protocol that can be displayed by selecting the "Protocol of power analyses" tab in the main window. You can clear the protocol, or to save, print, and copy the protocol in the same way as the distributions plot.

The button X-Y plot for a range of values at to bottom of the main window opens the plot window.



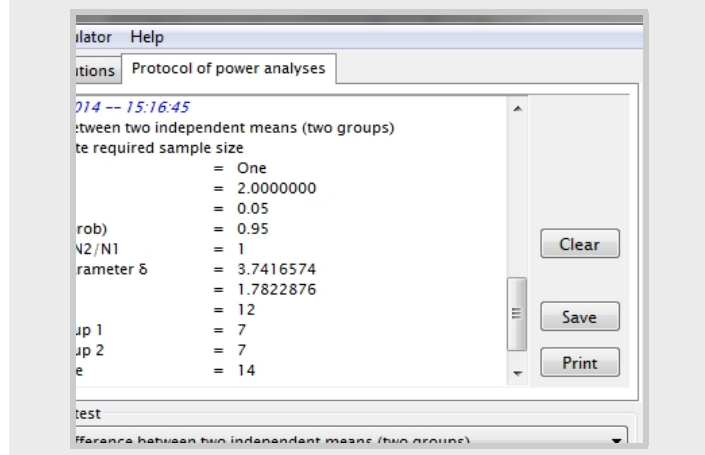
By selecting the appropriate parameters for the  $y$  and the  $x$  axis, one parameter ( $\alpha$ , power ( $1 - \beta$ ), effect size, or sample size) can be plotted as a function of another parameter. Of the remaining two parameters, one can be chosen to draw a family of graphs, while the fourth parameter is kept constant. For instance, power ( $1 - \beta$ ) can be drawn as a function of the sample size for several different population effects sizes, keeping  $\alpha$  at a particular value.

The plot may be printed, saved, or copied by clicking the right mouse button inside the plot area.

Selecting the Table tab reveals the data underlying the plot (see Fig. 3); they may be copied to other applications by *selecting, cut* and *paste*.

**Note:** The Power Plot window inherits all input parameters of the analysis that is active when the X-Y plot for a range of values button is pressed. Only some of these parameters can be directly manipulated in the Power Plot window. For instance, switching from a plot of a two-tailed test to that of a one-tailed test requires choosing the Tail(s): one option in the main window, followed by pressing the X-Y plot for range of values button.

(Part of) the protocol window.



## 1.2.4 Plotting of parameters

G\*POWER provides to possibility to generate plots of one of the parameters  $\alpha$ , effectsize, power and sample size, depending on a range of values of the remaining parameters.

The Power Plot window (see Fig. 2) is opened by clicking the X-Y plot for a range of values button located in the lower right corner of the main window. To ensure that all relevant parameters have valid values, this button is only enabled if an analysis has successfully been computed (by clicking on calculate).

The main output parameter of the type of analysis selected in the main window is by default selected as the de-

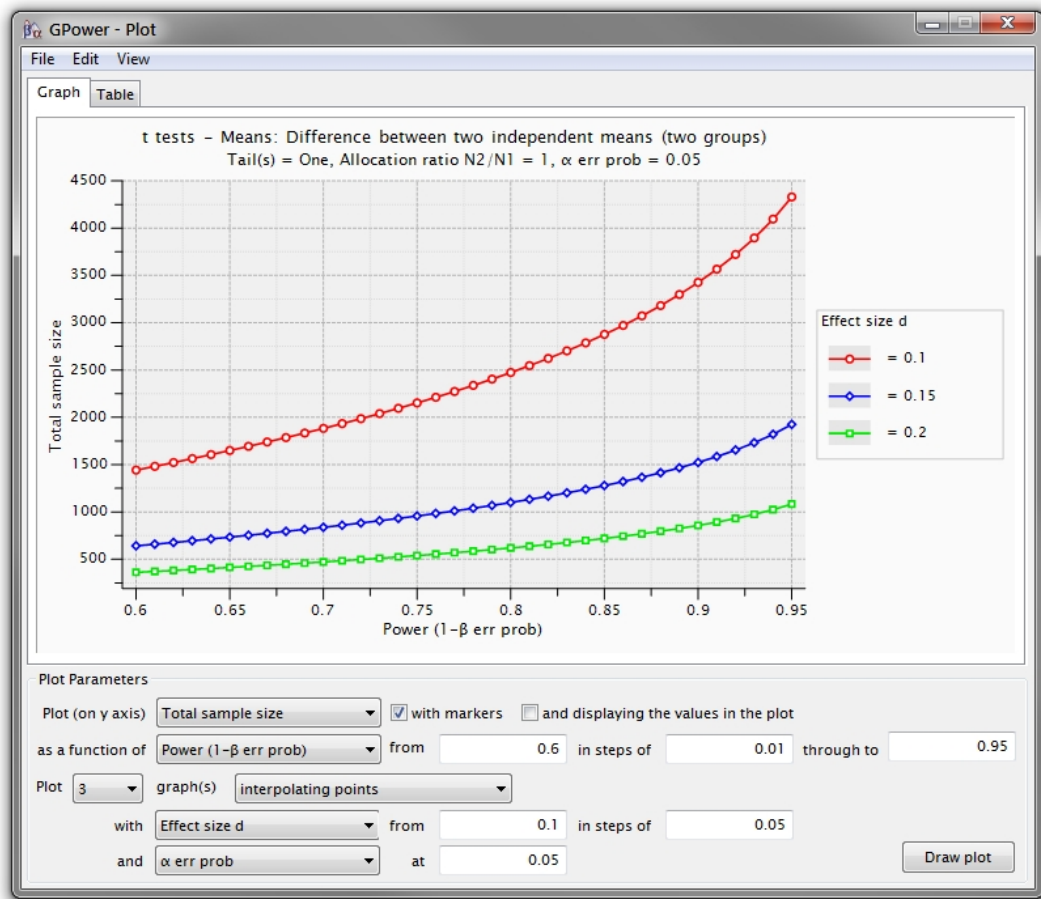


Figure 2: The plot window of G\*POWER

t tests - Means: Difference between two independent means (two groups)				
Tail(s) = One, Allocation ratio N2/N1 = 1, α err prob = 0.05				
		Effect size d = 0.1	Effect size d = 0.15	Effect size d = 0.2
#	Power (1-β err prob)	Total sample size	Total sample size	Total sample size
1	0.600000	1442.62	641.920	361.676
2	0.610000	1482.33	659.568	371.604
3	0.620000	1522.88	677.588	381.740
4	0.630000	1564.30	696.000	392.096
5	0.640000	1606.66	714.825	402.685
6	0.650000	1650.00	734.086	413.520
7	0.660000	1694.38	753.810	424.614
8	0.670000	1739.86	774.023	435.984
9	0.680000	1786.51	794.757	447.646
10	0.690000	1834.40	816.043	459.620
11	0.700000	1883.62	837.918	471.925
12	0.710000	1934.25	860.423	484.583
13	0.720000	1986.40	883.600	497.621
14	0.730000	2040.18	907.460	511.064

Figure 3: The table view of the data for the graphs shown in Fig. 2

## 2 The G \* Power calculator

G\*POWER contains a simple but powerful calculator that can be opened by selecting the menu label "Calculator" in the main window. Figure 4 shows an example session. This small example script calculates the power for the one-tailed t test for matched pairs and demonstrates most of the available features:

- There can be any number of expressions
- The result is set to the value of the last expression in the script
- Several expressions on a line are separated by a semicolon
- Expressions can be assigned to variables that can be used in following expressions
- The character # starts a comment. The rest of the line following # is ignored
- Many standard mathematical functions like square root, sin, cos etc are supported (for a list, see below)
- Many important statistical distributions are supported (see list below)
- The script can be easily saved and loaded. In this way a number of useful helper scripts can be created.

The calculator supports the following arithmetic operations (shown in descending precedence):

- Power:  $^$  ( $2^3 = 8$ )
- Multiply:  $*$  ( $2 * 2 = 4$ )
- Divide:  $/$  ( $6 / 2 = 3$ )
- Plus:  $+$  ( $2 + 3 = 5$ )
- Minus:  $-$  ( $3 - 2 = 1$ )

Supported general functions

- $\text{abs}(x)$  - Absolute value  $|x|$
- $\text{sin}(x)$  - Sine of  $x$
- $\text{asin}(x)$  - Arc sine of  $x$
- $\text{cos}(x)$  - Cosine of  $x$
- $\text{acos}(x)$  - Arc cosine of  $x$
- $\text{tan}(x)$  - Tangent of  $x$
- $\text{atan}(x)$  - Arc tangent of  $x$
- $\text{atan2}(x, y)$  - Arc tangent of  $y/x$
- $\text{exp}(x)$  - Exponential  $e^x$
- $\text{log}(x)$  - Natural logarithm  $\ln(x)$
- $\text{sqrt}(x)$  - Square root  $\sqrt{x}$
- $\text{sqr}(x)$  - Square  $x^2$

- $\text{sign}(x)$  - Sign of  $x$ :  $x < 0 \rightarrow -1, x = 0 \rightarrow 0, x > 0 \rightarrow 1$ .
- $\text{lngamma}(x)$  - Natural logarithm of the gamma function  $\ln(\Gamma(x))$
- $\text{frac}(x)$  - Fractional part of floating point  $x$ :  $\text{frac}(1.56)$  is 0.56.
- $\text{int}(x)$  - Integer part of float point  $x$ :  $\text{int}(1.56)$  is 1.
- $\text{min}(x, y)$  - Minimum of  $x$  and  $y$
- $\text{max}(x, y)$  - Maximum of  $x$  and  $y$
- $\text{uround}(x, m)$  - round  $x$  up to a multiple of  $m$   
 $\text{uround}(2.3, 1)$  is 3,  $\text{uround}(2.3, 2) = 4$ .

Supported distribution functions (CDF = cumulative distribution function, PDF = probability density function, Quantile = inverse of the CDF). For information about the properties of these distributions check <http://mathworld.wolfram.com/>.

- $\text{zcdf}(x)$  - CDF  
 $\text{zpdf}(x)$  - PDF  
 $\text{zinv}(p)$  - Quantile of the standard normal distribution.
- $\text{normcdf}(x, m, s)$  - CDF  
 $\text{normpdf}(x, m, s)$  - PDF  
 $\text{norminv}(p, m, s)$  - Quantile of the normal distribution with mean  $m$  and standard deviation  $s$ .
- $\text{chi2cdf}(x, df)$  - CDF  
 $\text{chi2pdf}(x, df)$  - PDF  
 $\text{chi2inv}(p, df)$  - Quantile of the chi square distribution with  $df$  degrees of freedom:  $\chi_{df}^2(x)$ .
- $\text{fcdf}(x, df1, df2)$  - CDF  
 $\text{fpdf}(x, df1, df2)$  - PDF  
 $\text{finv}(p, df1, df2)$  - Quantile of the F distribution with  $df1$  numerator and  $df2$  denominator degrees of freedom  $F_{df1, df2}(x)$ .
- $\text{tcdf}(x, df)$  - CDF  
 $\text{tpdf}(x, df)$  - PDF  
 $\text{tinv}(p, df)$  - Quantile of the Student t distribution with  $df$  degrees of freedom  $t_{df}(x)$ .
- $\text{ncx2cdf}(x, df, nc)$  - CDF  
 $\text{ncx2pdf}(x, df, nc)$  - PDF  
 $\text{ncx2inv}(p, df, nc)$  - Quantile of noncentral chi square distribution with  $df$  degrees of freedom and noncentrality parameter  $nc$ .
- $\text{ncfcdf}(x, df1, df2, nc)$  - CDF  
 $\text{ncfpdf}(x, df1, df2, nc)$  - PDF  
 $\text{ncfinv}(p, df1, df2, nc)$  - Quantile of noncentral F distribution with  $df1$  numerator and  $df2$  denominator degrees of freedom and noncentrality parameter  $nc$ .

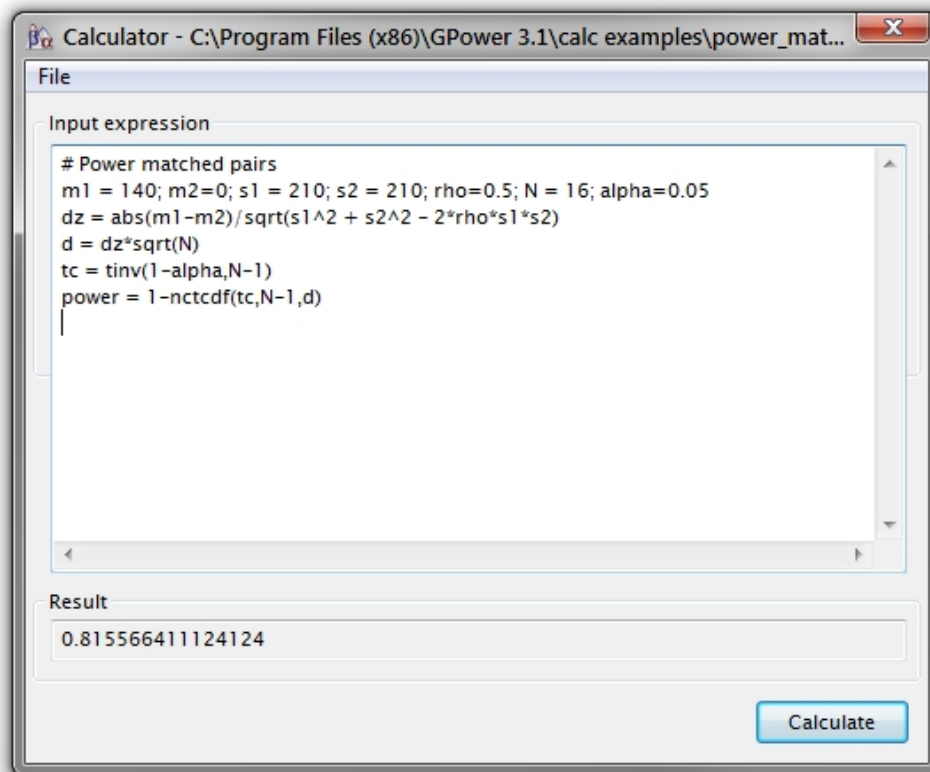


Figure 4: The G\*POWER calculator

- $nctcdf(x, df, nc)$  - CDF  
 $nctpdf(x, df, nc)$  - PDF  
 $nctinv(p, df, nc)$  - Quantile  
of noncentral Student t distribution with  $df$  degrees of freedom and noncentrality parameter  $nc$ .
- $betacdf(x, a, b)$  - CDF  
 $betapdf(x, a, b)$  - PDF  
 $betainv(p, a, b)$  - Quantile  
of the beta distribution with shape parameters  $a$  and  $b$ .
- $poisscdf(x, \lambda)$  - CDF  
 $poisspdf(x, \lambda)$  - PDF  
 $poissinv(p, \lambda)$  - Quantile  
 $poissmean(x, \lambda)$  - Mean  
of the poisson distribution with mean  $\lambda$ .
- $binocdf(x, N, \pi)$  - CDF  
 $binopdf(x, N, \pi)$  - PDF  
 $binoinv(p, N, \pi)$  - Quantile  
of the binomial distribution for sample size  $N$  and success probability  $\pi$ .
- $hygecdf(x, N, ns, nt)$  - CDF  
 $hygepdf(x, N, ns, nt)$  - PDF  
 $hygeinv(p, N, ns, nt)$  - Quantile  
of the hypergeometric distribution for samples of size  $N$  from a population of total size  $nt$  with  $ns$  successes.
- $corrpdf(r, \rho, N)$  - CDF  
 $corrpdf(r, \rho, N)$  - PDF  
 $corrinv(p, \rho, N)$  - Quantile  
of the distribution of the sample correlation coefficient  $r$  for population correlation  $\rho$  and samples of size  $N$ .
- $mr2cdf(R^2, \rho^2, k, N)$  - CDF  
 $mr2pdf(R^2, \rho^2, k, N)$  - PDF  
 $mr2inv(p, \rho^2, k, N)$  - Quantile  
of the distribution of the sample squared multiple correlation coefficient  $R^2$  for population squared multiple correlation coefficient  $\rho^2$ ,  $k - 1$  predictors, and samples of size  $N$ .
- $logncdf(x, m, s)$  - CDF  
 $lognpdf(x, m, s)$  - PDF  
 $logninv(p, m, s)$  - Quantile  
of the log-normal distribution, where  $m, s$  denote mean and standard deviation of the associated normal distribution.
- $laplcdf(x, m, s)$  - CDF  
 $laplpdf(x, m, s)$  - PDF  
 $laplinv(p, m, s)$  - Quantile  
of the Laplace distribution, where  $m, s$  denote location and scale parameter.
- $expcdf(x, \lambda)$  - CDF  
 $exppdf(x, \lambda)$  - PDF  
 $expinv(p, \lambda)$  - Quantile  
of the exponential distribution with parameter  $\lambda$ .
- $unicdf(x, a, b)$  - CDF  
 $unipdf(x, a, b)$  - PDF  
 $uniinv(p, a, b)$  - Quantile  
of the uniform distribution in the interval  $[a, b]$ .



### 3 Exact: Correlation - Difference from constant (one sample case)

The null hypothesis is that in the population the true correlation  $\rho$  between two bivariate normally distributed random variables has the fixed value  $\rho_0$ . The (two-sided) alternative hypothesis is that the correlation coefficient has a different value:  $\rho \neq \rho_0$ :

$$\begin{aligned} H_0 : \rho - \rho_0 &= 0 \\ H_1 : \rho - \rho_0 &\neq 0. \end{aligned}$$

A common special case is  $\rho_0 = 0$  (see e.g. Cohen, 1969, Chap. 3). The two-sided test (“two tails”) should be used if there is no restriction on the direction of the deviation of the sample  $r$  from  $\rho_0$ . Otherwise use the one-sided test (“one tail”).

#### 3.1 Effect size index

To specify the effect size, the conjectured alternative correlation coefficient  $\rho$  should be given.  $\rho$  must conform to the following restrictions:  $-1 + \varepsilon < \rho < 1 - \varepsilon$ , with  $\varepsilon = 10^{-6}$ . The proper effect size is the difference between  $\rho$  and  $\rho_0$ :  $\rho - \rho_0$ . Zero effect sizes are not allowed in a priori analyses. G\*POWER therefore imposes the additional restriction that  $|\rho - \rho_0| > \varepsilon$  in this case.

For the special case  $\rho_0 = 0$ , Cohen (1969, p.76) defines the following effect size conventions:

- **small**  $\rho = 0.1$
- **medium**  $\rho = 0.3$
- **large**  $\rho = 0.5$

Pressing the Determine button on the left side of the effect size label opens the effect size drawer (see Fig. 5). You can use it to calculate  $|\rho|$  from the coefficient of determination  $r^2$ .

**Figure 5:** Effect size dialog to determine the coefficient of determination from the correlation coefficient  $\rho$ .

#### 3.2 Options

The procedure uses either the exact distribution of the correlation coefficient or a large sample approximation based on the  $z$  distribution. The options dialog offers the following choices:

1. *Use exact distribution if  $N < x$ .* The computation time of the exact distribution increases with  $N$ , whereas that of the approximation does not. Both procedures are

asymptotically identical, that is, they produce essentially the same results if  $N$  is large. Therefore, a threshold value  $x$  for  $N$  can be specified that determines the transition between both procedures. The exact procedure is used if  $N < x$ , the approximation otherwise.

2. *Use large sample approximation (Fisher Z).* With this option you select always to use the approximation.

There are two properties of the output that can be used to discern which of the procedures was actually used: The option field of the output in the protocol, and the naming of the critical values in the main window, in the distribution plot, and in the protocol ( $r$  is used for the exact distribution and  $z$  for the approximation).

#### 3.3 Examples

In the null hypothesis we assume  $\rho_0 = 0.60$  to be the correlation coefficient in the population. We further assume that our treatment increases the correlation to  $\rho = 0.65$ . If we require  $\alpha = \beta = 0.05$ , how many subjects do we need in a two-sided test?

- **Select**  
*Type of power analysis:* A priori
- **Options**  
*Use exact distribution if  $N < :$*  10000
- **Input**  
*Tail(s):* Two  
*Correlation  $\rho$  H1:* 0.65  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95  
*Correlation  $\rho$  H0:* 0.60
- **Output**  
*Lower critical r:* 0.570748  
*Upper critical r:* 0.627920  
*Total sample size:* 1928  
*Actual power:* 0.950028

In this case we would reject the null hypothesis if we observed a sample correlation coefficient outside the interval  $[0.571, 0.627]$ . The total sample size required to ensure a power  $(1 - \beta) > 0.95$  is 1928; the actual power for this  $N$  is 0.950028.

In the example just discussed, using the large sample approximation leads to almost the same sample size  $N = 1929$ . Actually, the approximation is very good in most cases. We now consider a small sample case, where the deviation is more pronounced: In a post hoc analysis of a two-sided test with  $\rho_0 = 0.8$ ,  $\rho = 0.3$ , sample size 8, and  $\alpha = 0.05$  the exact power is 0.482927. The approximation gives the slightly lower value 0.422599.

#### 3.4 Related tests

Similar tests in G\*POWER 3.0:

- Correlation: Point biserial model
- Correlations: Two independent Pearson  $r$ 's (two samples)

### 3.5 Implementation notes

**Exact distribution.** The  $H_0$ -distribution is the sample correlation coefficient distribution  $sr(\rho_0, N)$ , the  $H_1$ -distribution is  $sr(\rho, N)$ , where  $N$  denotes the total sample size,  $\rho_0$  denotes the value of the baseline correlation assumed in the null hypothesis, and  $\rho$  denotes the ‘alternative correlation’. The (implicit) effect size is  $\rho - \rho_0$ . The algorithm described in [Barabesi and Greco \(2002\)](#) is used to calculate the CDF of the sample coefficient distribution.

**Large sample approximation.** The  $H_0$ -distribution is the standard normal distribution  $N(0, 1)$ , the  $H_1$ -distribution is  $N(F_z(\rho) - F_z(\rho_0))/\sigma, 1)$ , with  $F_z(r) = \ln((1+r)/(1-r))/2$  (Fisher  $z$  transformation) and  $\sigma = \sqrt{1/(N-3)}$ .

### 3.6 Validation

The results in the special case of  $\rho_0 = 0$  were compared with the tabulated values published in [Cohen \(1969\)](#). The results in the general case were checked against the values produced by PASS ([Hintze, 2006](#)).

## 4 Exact: Proportion - difference from constant (one sample case)

The problem considered in this case is whether the probability  $\pi$  of an event in a given population has the constant value  $\pi_0$  (null hypothesis). The null and the alternative hypothesis can be stated as:

$$\begin{aligned} H_0 : \pi - \pi_0 &= 0 \\ H_1 : \pi - \pi_0 &\neq 0. \end{aligned}$$

A two-tailed binomial tests should be performed to test this undirected hypothesis. If it is possible to predict a priori the direction of the deviation of sample proportions  $p$  from  $\pi_0$ , e.g.  $p - \pi_0 < 0$ , then a one-tailed binomial test should be chosen.

### 4.1 Effect size index

The effect size  $g$  is defined as the deviation from the constant probability  $\pi_0$ , that is,  $g = \pi - \pi_0$ .

The definition of  $g$  implies the following restriction:  $\varepsilon \leq (\pi_0 + g) \leq 1 - \varepsilon$ . In an a priori analysis we need to respect the additional restriction  $|g| > \varepsilon$  (this is in accordance with the general rule that zero effect hypotheses are undefined in a priori analyses). With respect to these constraints, G\*POWER sets  $\varepsilon = 10^{-6}$ .

Pressing the Determine button on the left side of the effect size label opens the effect size drawer:

You can use this dialog to calculate the effect size  $g$  from  $\pi_0$  (called  $P1$  in the dialog above) and  $\pi$  (called  $P2$  in the dialog above) or from several relations between them. If you open the effect dialog, the value of  $P1$  is set to the value in the constant proportion input field in the main window. There are four different ways to specify  $P2$ :

1. *Direct input*: Specify  $P2$  in the corresponding input field below  $P1$
2. *Difference*: Choose difference  $P2-P1$  and insert the difference into the text field on the left side (the difference is identical to  $g$ ).
3. *Ratio*: Choose ratio  $P2/P1$  and insert the ratio value into the text field on the left side
4. *Odds ratio*: Choose odds ratio and insert the odds ratio  $(P2/(1 - P2))/(P1/(1 - P1))$  between  $P1$  and  $P2$  into the text field on the left side.

The relational value given in the input field on the left side and the two proportions given in the two input fields on the right side are automatically synchronized if you leave one of the input fields. You may also press the Sync values button to synchronize manually.

Press the Calculate button to preview the effect size  $g$  resulting from your input values. Press the Transfer to main window button to (1) to calculate the effect size  $g = \pi - \pi_0 = P2 - P1$  and (2) to change, in the main window, the Constant proportion field to  $P1$  and the Effect size  $g$  field to  $g$  as calculated.

### 4.2 Options

The binomial distribution is discrete. It is thus not normally possible to arrive exactly at the nominal  $\alpha$ -level. For two-sided tests this leads to the problem how to “distribute”  $\alpha$  to the two sides. G\*POWER offers the three options listed here, the first option being selected by default:

1. *Assign  $\alpha/2$  to both sides*: Both sides are handled independently in exactly the same way as in a one-sided test. The only difference is that  $\alpha/2$  is used instead of  $\alpha$ . Of the three options offered by G\*POWER, this one leads to the greatest deviation from the actual  $\alpha$  (in post hoc analyses).
2. *Assign to minor tail  $\alpha/2$ , then rest to major tail ( $\alpha_2 = \alpha/2, \alpha_1 = \alpha - \alpha_2$ )*: First  $\alpha/2$  is applied to the side of the noncentral distribution that is farther away from the noncentral distribution (minor tail). The criterion used for the other side is then  $\alpha - \alpha_1$ , where  $\alpha_1$  is the actual  $\alpha$  found on the minor side. Since  $\alpha_1 \leq \alpha/2$  one can conclude that (in post hoc analyses) the sum of the actual values  $\alpha_1 + \alpha_2$  is in general closer to the nominal  $\alpha$ -level than it would be if  $\alpha/2$  were assigned to both side (see Option 1).
3. *Assign  $\alpha/2$  to both sides, then increase to minimize the difference of  $\alpha_1 + \alpha_2$  to  $\alpha$* : The first step is exactly the same as in Option 1. Then, in the second step, the critical values on both sides of the distribution are increased (using the lower of the two potential incremental  $\alpha$ -values) until the sum of both actual  $\alpha$  values is as close as possible to the nominal  $\alpha$ .

Press the Options button in the main window to select one of these options.

### 4.3 Examples

We assume a constant proportion  $\pi_0 = 0.65$  in the population and an effect size  $g = 0.15$ , i.e.  $\pi = 0.65 + 0.15 = 0.8$ . We want to know the power of a one-sided test given  $\alpha = .05$  and a total sample size of  $N = 20$ .

- **Select**  
*Type of power analysis*: Post hoc
- **Options**  
*Alpha balancing in two-sided tests*: Assign  $\alpha/2$  on both sides

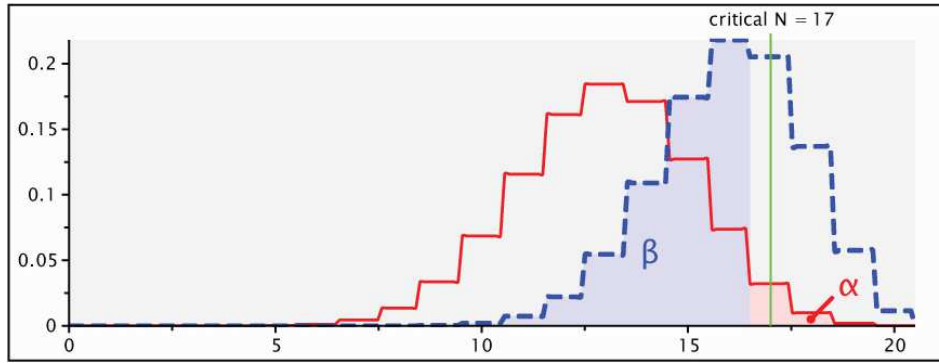


Figure 6: Distribution plot for the example (see text)

- **Input**

Tail(s): One  
 Effect size  $g$ : 0.15  
 $\alpha$  err prob: 0.05  
 Total sample size: 20  
 Constant proportion: 0.65

- **Output**

Lower critical  $N$ : 17  
 Upper critical  $N$ : 17  
 Power ( $1 - \beta$  err prob): 0.411449  
 Actual  $\alpha$ : 0.044376

The results show that we should reject the null hypothesis of  $\pi = 0.65$  if in 17 out of the 20 possible cases the relevant event is observed. Using this criterion, the actual  $\alpha$  is 0.044, that is, it is slightly lower than the requested  $\alpha$  of 5%. The power is 0.41.

Figure 6 shows the distribution plots for the example. The red and blue curves show the binomial distribution under  $H_0$  and  $H_1$ , respectively. The vertical line is positioned at the critical value  $N = 17$ . The horizontal portions of the graph should be interpreted as the top of bars ranging from  $N - 0.5$  to  $N + 0.5$  around an integer  $N$ , where the height of the bars correspond to  $p(N)$ .

We now use the graphics window to plot power values for a range of sample sizes. Press the X-Y plot for a range of values button at the bottom of the main window to open the Power Plot window. We select to plot the power as a function of total sample size. We choose a range of samples sizes from 10 in steps of 1 through to 50. Next, we select to plot just one graph with  $\alpha = 0.05$  and effect size  $g = 0.15$ . Pressing the Draw Plot button produces the plot shown in Fig. 7. It can be seen that the power does not increase monotonically but in a zig-zag fashion. This behavior is due to the discrete nature of the binomial distribution that prevents that arbitrary  $\alpha$  value can be realized. Thus, the curve should not be interpreted to show that the power for a fixed  $\alpha$  sometimes decreases with increasing sample size. The real reason for the non-monotonic behaviour is that the actual  $\alpha$  level that can be realized deviates more or less from the nominal  $\alpha$  level for different sample sizes.

This non-monotonic behavior of the power curve poses a problem if we want to determine, in an a priori analysis, the minimal sample size needed to achieve a certain power. In these cases G\*POWER always tries to find the lowest sample size for which the power is not less than the specified value. In the case depicted in Fig. 7, for instance, G\*POWER

would choose  $N = 16$  as the result of a search for the sample size that leads to a power of at least 0.3. All types of power analyses except post hoc are confronted with similar problems. To ensure that the intended result has been found, we recommend to check the results from these types of power analysis by a power vs. sample size plot.

#### 4.4 Related tests

Similar tests in G\*POWER 3.0:

- Proportions: Sign test.

#### 4.5 Implementation notes

The  $H_0$ -distribution is the Binomial distribution  $B(N, \pi_0)$ , the  $H_1$ -distribution the Binomial distribution  $B(N, g + \pi_0)$ .  $N$  denotes the total sample size,  $\pi_0$  the constant proportion assumed in the null hypothesis, and  $g$  the effect size index as defined above.

#### 4.6 Validation

The results of G\*POWER for the special case of the sign test, that is  $\pi_0 = 0.5$ , were checked against the tabulated values given in Cohen (1969, chapter 5). Cohen always chose from the realizable  $\alpha$  values the one that is closest to the nominal value even if it is larger then the nominal value. G\*POWER, in contrast, always requires the actual  $\alpha$  to be lower then the nominal value. In cases where the  $\alpha$  value chosen by Cohen happens to be lower then the nominal  $\alpha$ , the results computed with G\*POWER were very similar to the tabulated values. In the other cases, the power values computed by G\*POWER were lower then the tabulated ones.

In the general case ( $\pi_0 \neq 0.5$ ) the results of post hoc analyses for a number of parameters were checked against the results produced by PASS (Hintze, 2006). No differences were found in one-sided tests. The results for two-sided tests were also identical if the alpha balancing method "Assign  $\alpha/2$  to both sides" was chosen in G\*POWER.

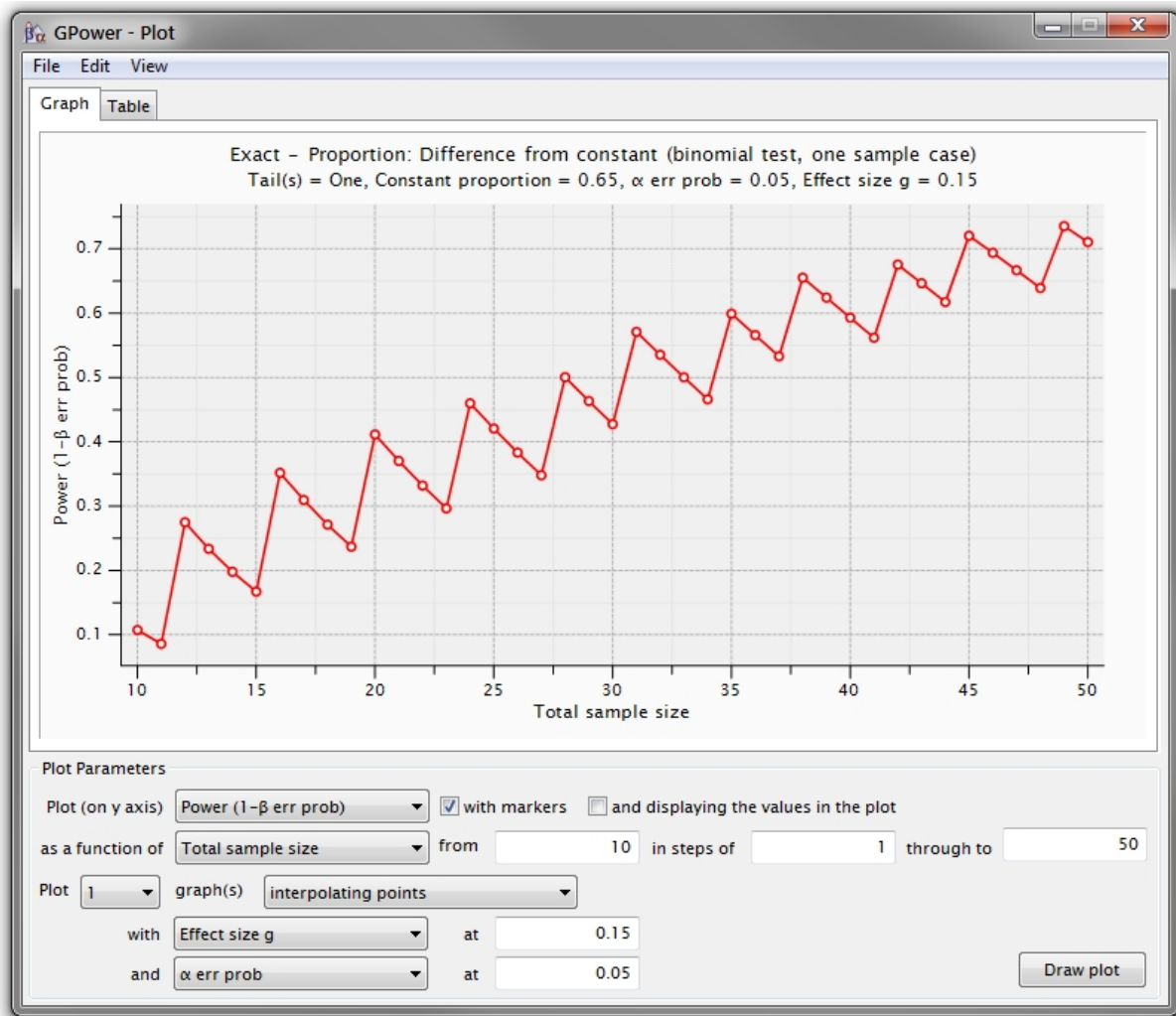


Figure 7: Plot of power vs. sample size in the binomial test (see text)

## 5 Exact: Proportion - inequality, two dependent groups (McNemar)

This procedure relates to tests of paired binary responses. Such data can be represented in a  $2 \times 2$  table:

Treatment	Standard		
	Yes	No	
Yes	$\pi_{11}$	$\pi_{12}$	$\pi_t$
No	$\pi_{21}$	$\pi_{22}$	$1 - \pi_t$
	$\pi_s$	$1 - \pi_s$	1

where  $\pi_{ij}$  denotes the probability of the respective response. The probability  $\pi_D$  of discordant pairs, that is, the probability of yes/no-response pairs, is given by  $\pi_D = \pi_{12} + \pi_{21}$ . The hypothesis of interest is that  $\pi_s = \pi_t$ , which is formally identical to the statement  $\pi_{12} = \pi_{21}$ .

Using this fact, the null hypothesis states (in a ratio notation) that  $\pi_{12}$  is identical to  $\pi_{21}$ , and the alternative hypothesis states that  $\pi_{12}$  and  $\pi_{21}$  are different:

$$H_0 : \pi_{12}/\pi_{21} = 1$$

$$H_1 : \pi_{12}/\pi_{21} \neq 1.$$

In the context of the McNemar test the term odds ratio (OR) denotes the ratio  $\pi_{12}/\pi_{21}$  that is used in the formulation of  $H_0$  and  $H_1$ .

### 5.1 Effect size index

The Odds ratio  $\pi_{12}/\pi_{21}$  is used to specify the effect size. The odds ratio must lie inside the interval  $[10^{-6}, 10^6]$ . An odds ratio of 1 corresponds to a null effect. Therefore this value must not be used in a priori analyses.

In addition to the odds ratio, the proportion of discordant pairs, i.e.  $\pi_D$ , must be given in the input parameter field called Prop discordant pairs. The values for this proportion must lie inside the interval  $[\varepsilon, 1 - \varepsilon]$ , with  $\varepsilon = 10^{-6}$ .

If  $\pi_D$  and  $d = \pi_{12} - \pi_{21}$  are given, then the odds ratio may be calculated as:  $OR = (d + \pi_D)/(d - \pi_D)$ .

### 5.2 Options

Press the Options button in the main window to select one of the following options.

#### 5.2.1 Alpha balancing in two-sided tests

The binomial distribution is discrete. It is therefore not normally possible to arrive at the exact nominal  $\alpha$ -level. For two-sided tests this leads to the problem how to "distribute"  $\alpha$  to the two sides. G\*POWER offers the three options listed here, the first option being selected by default:

1. *Assign  $\alpha/2$  to both sides*: Both sides are handled independently in exactly the same way as in a one-sided test. The only difference is that  $\alpha/2$  is used instead of  $\alpha$ . Of the three options offered by G\*POWER, this one leads to the greatest deviation from the actual  $\alpha$  (in post hoc analyses).
2. *Assign to minor tail  $\alpha/2$ , then rest to major tail ( $\alpha_2 = \alpha/2, \alpha_1 = \alpha - \alpha_2$ )*: First  $\alpha/2$  is applied to the side of

the central distribution that is farther away from the noncentral distribution (minor tail). The criterion used on the other side is then  $\alpha - \alpha_1$ , where  $\alpha_1$  is the actual  $\alpha$  found on the minor side. Since  $\alpha_1 \leq \alpha/2$  one can conclude that (in post hoc analyses) the sum of the actual values  $\alpha_1 + \alpha_2$  is in general closer to the nominal  $\alpha$ -level than it would be if  $\alpha/2$  were assigned to both sides (see Option 1).

3. *Assign  $\alpha/2$  to both sides, then increase to minimize the difference of  $\alpha_1 + \alpha_2$  to  $\alpha$* : The first step is exactly the same as in Option 1. Then, in the second step, the critical values on both sides of the distribution are increased (using the lower of the two potential incremental  $\alpha$ -values) until the sum of both actual  $\alpha$  values is as close as possible to the nominal  $\alpha$ .

#### 5.2.2 Computation

You may choose between an exact procedure and a faster approximation (see implementation notes for details):

1. *Exact (unconditional) power if  $N < x$* . The computation time of the exact procedure increases much faster with sample size  $N$  than that of the approximation. Given that both procedures usually produce very similar results for large sample sizes, a threshold value  $x$  for  $N$  can be specified which determines the transition between both procedures. The exact procedure is used if  $N < x$ ; the approximation is used otherwise.  
Note: G\*POWER does not show distribution plots for exact computations.
2. *Faster approximation (assumes number of discordant pairs to be constant)*. Choosing this option instructs G\*POWER to always use the approximation.

### 5.3 Examples

As an example we replicate the computations in O'Brien (2002, p. 161-163). The assumed table is:

Treatment	Standard		
	Yes	No	
Yes	.54	.08	.62
No	.32	.06	.38
	.86	.14	1

In this table the proportion of discordant pairs is  $\pi_D = .32 + .08 = 0.4$  and the Odds Ratio  $OR = \pi_{12}/\pi_{21} = 0.08/.32 = 0.25$ . We want to compute the exact power for a one-sided test. The sample size  $N$ , that is, the number of pairs, is 50 and  $\alpha = 0.05$ .

- **Select**  
Type of power analysis: Post hoc
- **Options**  
Computation: Exact
- **Input**  
Tail(s): One  
Odds ratio: 0.25  
 $\alpha$  err prob: 0.05  
Total sample size: 50  
Prop discordant pairs: 0.4

- **Output**

Power (1- $\beta$  err prob): 0.839343

Actual  $\alpha$ : 0.032578

Proportion p12: 0.08

Proportion p21: 0.32

The power calculated by G\*POWER (0.839343) corresponds within the given precision to the result computed by O'Brien (0.839). Now we use the Power Plot window to calculate the power for several other sample sizes and to generate a graph that gives us an overview of a section of the parameter space. The Power Plot window can be opened by pressing the X-Y plot for a range of values button in the lower part of the main window.

In the Power Plot window we choose to plot the power on the Y-Axis (with markers and displaying the values in the plot) as a function of total sample size. The sample sizes shall range from 50 in steps of 25 through to 150. We choose to draw a single plot. We specify  $\alpha = 0.05$  and Odds ratio = 0.25.

The results shown in figure 8 replicate exactly the values in the table in O'Brien (2002, p. 163)

To replicate the values for the two-sided case, we must decide how the  $\alpha$  error should be distributed to the two sides. The method chosen by O'Brien corresponds to Option 2 in G\*POWER ("Assign to minor tail  $\alpha/2$ , then rest to major tail", see above). In the main window, we select Tail(s) "Two" and set the other input parameters exactly as shown in the example above. For sample sizes 50, 75, 100, 125, 150 we get power values 0.798241, 0.930639, 0.980441, 0.994839, and 0.998658, respectively, which are again equal to the values given in O'Brien's table.

## 5.4 Related tests

## 5.5 Implementation notes

**Exact (unconditional) test** . In this case G\*POWER calculates the unconditional power for the exact conditional test: The number of discordant pairs  $N_D$  is a random variable with binomial distribution  $B(N, \pi_D)$ , where  $N$  denotes the total number of pairs, and  $\pi_D = \pi_{12} + \pi_{21}$  denotes the probability of discordant pairs. Thus  $P(N_D) = \binom{N}{N_D} (\pi_{11} + \pi_{22})^{N_D} (\pi_{12} + \pi_{21})^{N-N_D}$ . Conditional on  $N_D$ , the frequency  $f_{12}$  has a binomial distribution  $B(N_D, \pi_0 = \pi_{12}/\pi_D)$  and we test the  $H_0: \pi_0 = 0.5$ . Given the conditional 'binomial power'  $\text{Pow}(N_D, \pi_0 | N_D = i)$  the exact unconditional power is  $\sum_i P(N_D = i) \text{Pow}(N_D, \pi_0 | N_D = i)$ . The summation starts at the most probable value for  $N_D$  and then steps outward until the values are small enough to be ignored.

**Fast approximation** . In this case an ordinary one sample binomial power calculation with  $H_0$ -distribution  $B(N\pi_D, 0.5)$ , and  $H_1$ -Distribution  $B(N\pi_D, OR/(OR + 1))$  is performed.

## 5.6 Validation

The results of the exact procedure were checked against the values given on pages 161-163 in O'Brien (2002). Complete correspondence was found in the one-tailed case and

also in the two-tailed case when the alpha balancing Option 2 ("Assign to minor tail  $\alpha/2$ , then rest to major tail", see above) was chosen in G\*POWER .

We also compared the exact results of G\*POWER generated for a large range of parameters to the results produced by PASS (Hintze, 2006) for the same scenarios. We found complete correspondence in one-sided test. In two-sided tests PASS uses an alpha balancing strategy corresponding to Option 1 in G\*POWER ("Assign  $\alpha/2$  on both sides", see above). With two-sided tests we found small deviations between G\*POWER and PASS (about  $\pm 1$  in the third decimal place), especially for small sample sizes. These deviations were always much smaller than those resulting from a change of the balancing strategy. All comparisons with PASS were restricted to  $N < 2000$ , since for larger  $N$  the exact routine in PASS sometimes produced nonsensical values (this restriction is noted in the PASS manual).

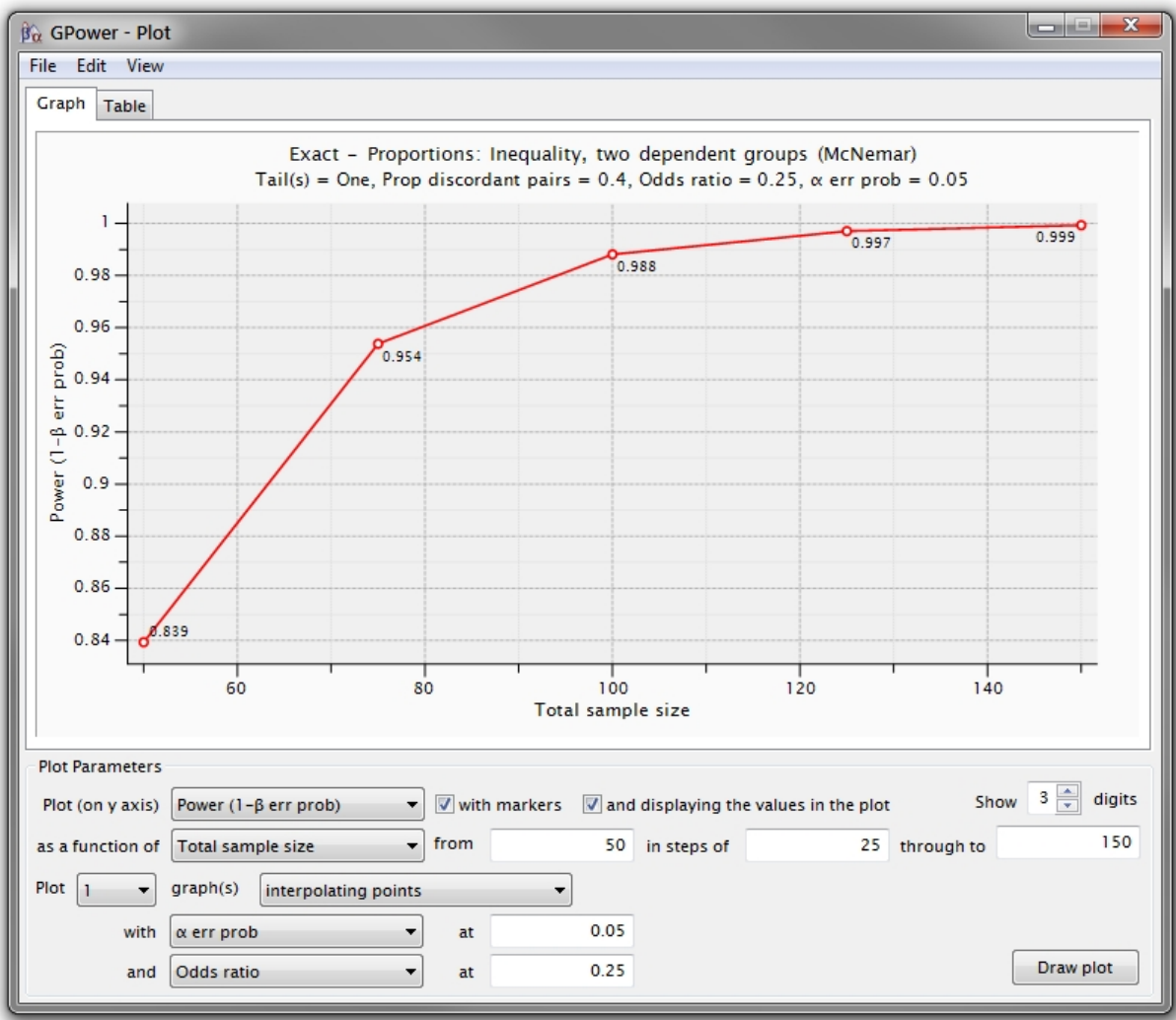


Figure 8: Result of the sample McNemar test (see text for details).



## 6 Exact: Proportions - inequality of two independent groups (Fisher's exact-test)

### 6.1 Introduction

This procedure calculates power and sample size for tests comparing two independent binomial populations with probabilities  $\pi_1$  and  $\pi_2$ , respectively. The results of sampling from these two populations can be given in a  $2 \times 2$  contingency table  $X$ :

	Group 1	Group 2	Total
Success	$x_1$	$x_2$	$m$
Failure	$n_1 - x_1$	$n_2 - x_2$	$N - m$
Total	$n_1$	$n_2$	$N$

Here,  $n_1, n_2$  are the sample sizes, and  $x_1, x_2$  the observed number of successes in the two populations.  $N = n_1 + n_2$  is the total sample size, and  $m = x_1 + x_2$  the total number of successes.

The null hypothesis states that  $\pi_1 = \pi_2$ , whereas the alternative hypothesis assumes different probabilities in both populations:

$$\begin{aligned} H_0 : \pi_1 - \pi_2 &= 0 \\ H_1 : \pi_1 - \pi_2 &\neq 0. \end{aligned}$$

### 6.2 Effect size index

The effect size is determined by directly specifying the two proportions  $\pi_1$  and  $\pi_2$ .

### 6.3 Options

This test has no options.

### 6.4 Examples

### 6.5 Related tests

### 6.6 Implementation notes

#### 6.6.1 Exact unconditional power

The procedure computes the exact unconditional power of the (conditional) test.

The exact probability of table  $X$  (see introduction) under  $H_0$ , conditional on  $x_1 + x_2 = m$ , is given by:

$$Pr(X|m, H_0) = \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{N}{m}}$$

Let  $T$  be a test statistic,  $t$  a possible value of  $T$ , and  $M$  the set of all tables  $X$  with a total number of successes equal to  $m$ . We define  $M_t = \{X \in M : T \geq t\}$ , i.e.  $M_t$  is the subset of  $M$  containing all tables for which the value of the test statistic is equal to or exceeds the value  $t$ . The exact null distribution of  $T$  is obtained by calculating  $Pr(T \geq t|m, H_0) = \sum_{X \in M_t} Pr(X|m, H_0)$  for all possible  $t$ . The critical value  $t_\alpha$  is the smallest value such that  $Pr(T \geq t_\alpha|m, H_0) \leq \alpha$ . The power is then defined as:

$$1 - \beta = \sum_{m=0}^N P(m) Pr(T \geq t_\alpha|m, H_1),$$

where

$$Pr(T \geq t_\alpha|m, H_1) = \sum_{X \in M_{t_\alpha}} \frac{B_{12}}{\sum_{X \in M} B_{12}},$$

$$P(m) = Pr(x_1 + x_2 = m|H_1) = B_{12}, \text{ and}$$

$$B_{12} = \binom{n_1}{x_1} \pi_1^{x_1} (1 - \pi_1)^{n_1 - x_1} \binom{n_2}{x_2} \pi_2^{x_2} (1 - \pi_2)^{n_2 - x_2}$$

For **two-sided tests** G\*POWER provides three common test statistics that are asymptotically equivalent:

1. Fisher's exact test:

$$T = -\ln \left[ \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{N}{m}} \right]$$

2. Persons's exact test:

$$T = \sum_{j=1}^2 \left( \frac{(x_j - mn_j/N)^2}{mn_j/N} + \frac{[(n_j - x_j) - (N - m)n_j/N]^2}{(N - m)n_j/N} \right)$$

3. Likelihood ratio exact test:

$$T = 2 \sum_{j=1}^2 \left( x_j \ln \left[ \frac{x_j}{mn_j/N} \right] + (n_j - x_j) \ln \left[ \frac{n_j - x_j}{(N - m)n_j/N} \right] \right)$$

The choice of the test statistics only influences the way in which  $\alpha$  is distributed on both sides of the null distribution.

For **one-sided tests** the test statistic is  $T = x_2$ .

#### 6.6.2 Large sample approximation

The large sample approximation is based on a continuity corrected  $\chi^2$  test with pooled variances. To permit a two-sided test, a  $z$  test version is used: The  $H_0$  distribution is the standard normal distribution  $N(0, 1)$ , and the  $H_1$  distribution given by the normal distribution  $N(m(k), \sigma)$ , with

$$\sigma = \frac{1}{\sigma_0} \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$$

$$m(k) = \frac{1}{\sigma_0} [p_2 - p_1 - k(1/n_1 + 1/n_2)/2], \text{ with}$$

$$\sigma_0 = \sqrt{\frac{n_1(1 - p_1) + n_2(1 - p_2)}{n_1 n_2} \cdot \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}}$$

$$k = \begin{cases} p_1 < p_2 : & -1 \\ p_1 \geq p_2 : & +1 \end{cases}$$

### 6.7 Validation

The results were checked against the values produced by GPower 2.0.

## 7 Exact test: Multiple Regression - random model

In multiple regression analyses, the relation of a dependent variable  $Y$  to  $m$  independent factors  $X = (X_1, \dots, X_m)$  is studied. The present procedure refers to the so-called unconditional or random factors model of multiple regression (Gatsonis & Sampson, 1989; Sampson, 1974), that is, it is assumed that  $Y$  and  $X_1, \dots, X_m$  are random variables, where  $(Y, X_1, \dots, X_m)$  have a joint multivariate normal distribution with a positive definite covariance matrix:

$$\begin{pmatrix} \sigma_Y^2 & \Sigma'_{YX} \\ \Sigma_{YX} & \Sigma_X \end{pmatrix}$$

and mean  $(\mu_Y, \mu_X)$ . Without loss of generality we may assume centered variables with  $\mu_Y = 0, \mu_{X_i} = 0$ .

The squared population multiple correlation coefficient between  $Y$  and  $X$  is given by:

$$\rho_{YX}^2 = \Sigma'_{YX} \Sigma_X^{-1} \Sigma_{YX} / \sigma_Y^2.$$

and the regression coefficient vector  $\gamma$ , the analog of  $\beta$  in the fixed model, by  $\gamma = \Sigma_X^{-1} \Sigma_{YX}$ . The maximum likelihood estimates of the regression coefficient vector and the residuals are the same under both the fixed and the random model (see theorem 1 in Sampson (1974)); the models differ, however, with respect to power.

The present procedure allows power analyses for the test that the population squared correlations coefficient  $\rho_{YX}^2$  has the value  $\rho_0^2$ . The null and alternate hypotheses are:

$$\begin{aligned} H_0: & \rho_{YX}^2 = \rho_0^2 \\ H_1: & \rho_{YX}^2 \neq \rho_0^2. \end{aligned}$$

An important special case is  $\rho_0 = 0$  (corresponding to the assumption  $\Sigma_{YX} = 0$ ). A commonly used test statistic for this case is  $F = [(N - m - 1) / p] R_{YX}^2 / (1 - R_{YX}^2)$ , which has a central  $F$  distribution with  $df_1 = m$ , and  $df_2 = N - m - 1$ . This is the same test statistic that is used in the fixed model. The power differs, however, in both cases.

### 7.1 Effect size index

The effect size is the population squared correlation coefficient  $H_1 \rho^2$  under the alternative hypothesis. To fully specify the effect size, you also need to give the population squared correlation coefficient  $H_0 \rho^2$  under the null hypothesis.

Pressing the button Determine on the left of the effect size label in the main window opens the effect size drawer (see Fig. 9) that may be used to calculate  $\rho^2$  either from the confidence interval for the population  $\rho_{YX}^2$  given an observed squared multiple correlation  $R_{YX}^2$  or from predictor correlations.

**Effect size from C.I.** Figure (9) shows an example of how the  $H_1 \rho^2$  can be determined from the confidence interval computed for an observed  $R^2$ . You have to input the sample size, the number of predictors, the observed  $R^2$  and the confidence level of the confidence interval. In the remaining input field a relative position inside the confidence interval can be given that determines the  $H_1 \rho^2$  value. The

Figure 9: Effect size drawer to calculate  $\rho^2$  from a confidence interval or via predictor correlations (see text).

value can range from 0 to 1, where 0, 0.5 and 1 corresponds to the left, central and right position inside the interval, respectively. The output fields C. I. lower  $\rho^2$  and C. I. upper  $\rho^2$  contain the left and right border of the two-sided  $100(1 - \alpha)$  percent confidence interval for  $\rho^2$ . The output fields Statistical lower bound and Statistical upper bound show the one-sided  $(0, R)$  and  $(L, 1)$  intervals, respectively.

**Effect size from predictor correlations** By choosing the option "From predictor correlation matrix" (see Fig. (9)) one may compute  $\rho^2$  from the matrix of correlation among the predictor variables and the correlations between predictors and the dependent variable  $Y$ . Pressing the "Insert/edit matrix"-button opens a window, in which one can specify (1) the row vector  $u$  containing the correlations between each of the  $m$  predictors  $X_i$  and the dependent variable  $Y$  and (2) the  $m \times m$  matrix  $B$  of correlations among the predictors. The squared multiple correlation coefficient is then given by  $\rho^2 = uB^{-1}u'$ . Each input correlation must lie in the interval  $[-1, 1]$ , the matrix  $B$  must be positive-definite, and the resulting  $\rho^2$  must lie in the interval  $[0, 1]$ . Pressing the Button "Calc  $\rho^2$ " tries to calculate  $\rho^2$  from the input and checks the positive-definiteness of matrix  $B$ .

**Relation of  $\rho^2$  to effect size  $f^2$**  The relation between  $\rho^2$  and effect size  $f^2$  used in the fixed factors model is:

$$f^2 = \frac{\rho^2}{1 - \rho^2}$$

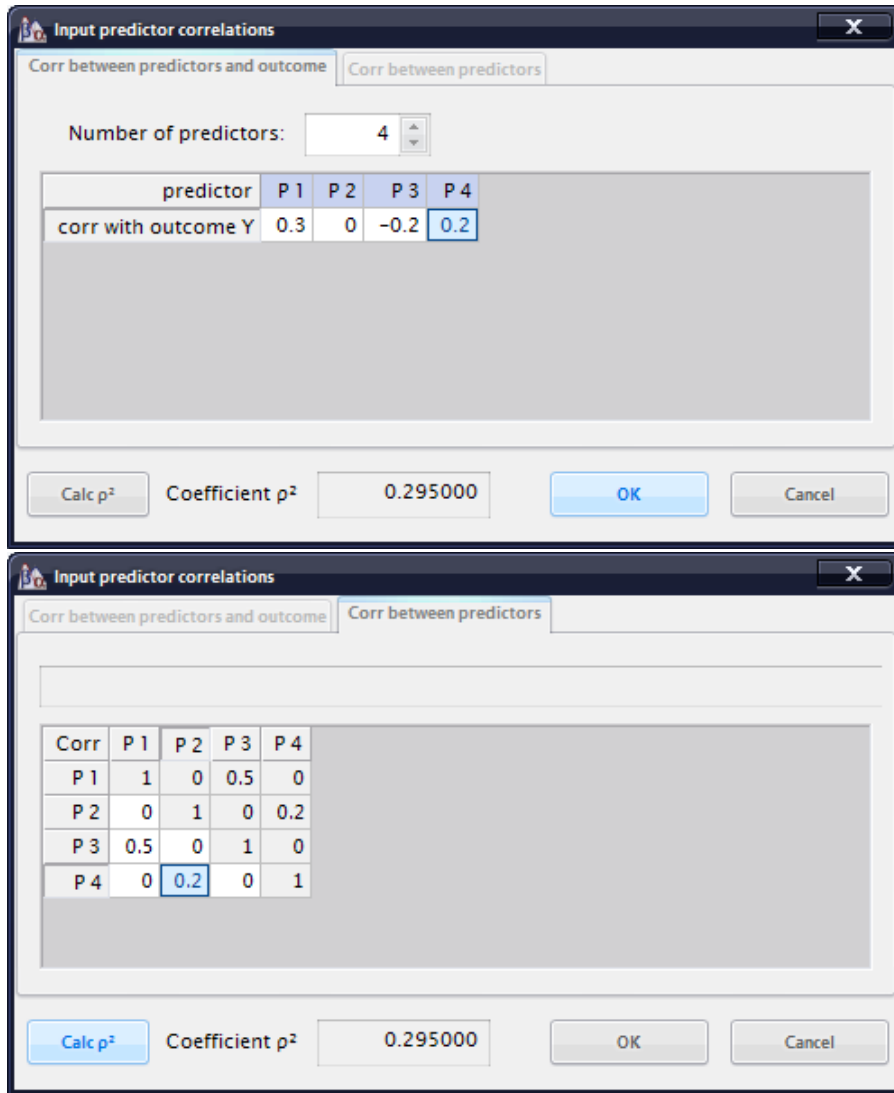


Figure 10: Input of correlations between predictors and Y (top) and the matrix of correlations among the predictors (bottom).

and conversely:

$$\rho^2 = \frac{f^2}{1 + f^2}$$

Cohen (1988, p. 412) defines the following conventional values for the effect size  $f^2$ :

- **small**  $f^2 = 0.02$
- **medium**  $f^2 = 0.15$
- **large**  $f^2 = 0.35$

which translate into the following values for  $\rho^2$ :

- **small**  $\rho^2 = 0.02$
- **medium**  $\rho^2 = 0.13$
- **large**  $\rho^2 = 0.26$

## 7.2 Options

You can switch between an exact procedure for the calculation of the distribution of the squared multiple correlation coefficient  $\rho^2$  and a three-moment  $F$  approximation suggested by Lee (1971, p.123). The latter is slightly faster and may be used to check the results of the exact routine.

## 7.3 Examples

### 7.3.1 Power and sample size

**Example 1** We replicate an example given for the procedure for the fixed model, but now under the assumption that the predictors are not fixed but random samples: We assume that a dependent variable  $Y$  is predicted by as set  $B$  of 5 predictors and that  $\rho_{YX}^2$  is 0.10, that is that the 5 predictors account for 10% of the variance of  $Y$ . The sample size is  $N = 95$  subjects. What is the power of the  $F$  test that  $\rho_{YX}^2 = 0$  at  $\alpha = 0.05$ ?

We choose the following settings in G\*POWER to calculate the power:

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* One  
*H1  $\rho^2$ :* 0.1  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 95  
*Number of predictors:* 5  
*H0  $\rho^2$ :* 0.0

- **Output**

Lower critical  $R^2$ : 0.115170  
 Upper critical  $R^2$ : 0.115170  
 Power (1-  $\beta$ ): 0.662627

The output shows that the power of this test is about 0.663 which is slightly lower than the power 0.674 found in the fixed model. This observation holds in general: The power in the random model is never larger than that found for the same scenario in the fixed model.

**Example 2** We now replicate the test of the hypotheses  $H_0 : \rho^2 \leq 0.3$  versus  $H_1 : \rho^2 > 0.3$  given in [Shieh and Kung \(2007, p.733\)](#), for  $N = 100$ ,  $\alpha = 0.05$ , and  $m = 5$  predictors. We assume that  $H_1 \rho^2 = 0.4$ . The settings and output in this case are:

- **Select**

Type of power analysis: Post hoc

- **Input**

Tail(s): One  
 $H_1 \rho^2$ : 0.4  
 $\alpha$  err prob: 0.05  
 Total sample size: 100  
 Number of predictors: 5  
 $H_0 \rho^2$ : 0.3

- **Output**

Lower critical  $R^2$ : 0.456625  
 Upper critical  $R^2$ : 0.456625  
 Power (1-  $\beta$ ): 0.346482

The results show, that  $H_0$  should be rejected if the observed  $R^2$  is larger than 0.457. The power of the test is about 0.346. Assume we observed  $R^2 = 0.5$ . To calculate the associated  $p$ -value we may use the G\*POWER-calculator. The syntax of the CDF of the squared sample multiple correlation coefficient is  $\text{mr2cdf}(R^2, \rho^2, m+1, N)$ . Thus for the present case we insert  $1-\text{mr2cdf}(0.5, 0.3, 6, 100)$  in the calculator and pressing Calculate gives 0.01278. These values replicate those given in [Shieh and Kung \(2007\)](#).

**Example 3** We now ask for the minimum sample size required for testing the hypothesis  $H_0 : \rho^2 \geq 0.2$  vs. the specific alternative hypothesis  $H_1 : \rho^2 = 0.05$  with 5 predictors to achieve power=0.9 and  $\alpha = 0.05$  (Example 2 in [Shieh and Kung \(2007\)](#)). The inputs and outputs are:

- **Select**

Type of power analysis: A priori

- **Input**

Tail(s): One  
 $H_1 \rho^2$ : 0.05  
 $\alpha$  err prob: 0.05  
 Power (1-  $\beta$ ): 0.9  
 Number of predictors: 5  
 $H_0 \rho^2$ : 0.2

- **Output**

Lower critical  $R^2$ : 0.132309  
 Upper critical  $R^2$ : 0.132309  
 Total sample size: 153  
 Actual power: 0.901051

The results show that  $N$  should not be less than 153. This confirms the results in [Shieh and Kung \(2007\)](#).

### 7.3.2 Using confidence intervals to determine the effect size

Suppose that in a regression analysis with 5 predictors and  $N = 50$  we observed a squared multiple correlation coefficient  $R^2 = 0.3$  and we want to use the lower boundary of the 95% confidence interval for  $\rho^2$  as  $H_1 \rho^2$ . Pressing the Determine-button next to the effect size field in the main window opens the effect size drawer. After selecting input mode "From confidence interval" we insert the above values (50, 5, 0.3, 0.95) in the corresponding input field and set Rel C.I. pos to use (0=left, 1=right) to 0 to select the left interval border. Pressing calculate computes the lower, upper, and 95% two-sided confidence intervals: [0, 4245], [0.0589, 1] and [0.0337, 0.4606]. The left boundary of the two-sided interval (0.0337) is transferred to the field  $H_1 \rho^2$ .

### 7.3.3 Using predictor correlations to determine effect size

We may use assumptions about the  $(m \times m)$  correlation matrix between a set of  $m$  predictors, and the  $m$  correlations between predictor variables and the dependent variable  $Y$  to determine  $\rho^2$ . Pressing the Determine-button next to the effect size field in the main window opens the effect size drawer. After selecting input mode "From predictor correlations" we insert the number of predictors in the corresponding field and press "Insert/edit matrix". This opens a input dialog (see Fig. (10)). Suppose that we have 4 predictors and that the 4 correlations between  $X_i$  and  $Y$  are  $u = (0.3, 0.1, -0.2, 0.2)$ . We insert this values in the tab "Corr between predictors and outcome". Assume further that the correlations between  $X_1$  and  $X_3$  and between  $X_2$  and  $X_4$  are 0.5 and 0.2, respectively, whereas all other predictor pairs are uncorrelated. We insert the correlation matrix

$$B = \begin{pmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.2 \\ 0.5 & 0 & 1 & 0 \\ 0 & 0.2 & 0 & 1 \end{pmatrix}$$

under the "Corr between predictors" tab. Pressing the "Calc  $\rho^2$ "-button computes  $\rho^2 = uB^{-1}u'$  = 0.297083, which also confirms that  $B$  is positive-definite and thus a correct correlation matrix.

## 7.4 Related tests

Similar tests in G\*POWER 3.0:

- Linear Multiple Regression: Deviation of  $R^2$  from zero.
- Linear Multiple Regression: Increase of  $R^2$ .

## 7.5 Implementation notes

The procedure uses the exact sampling distribution of the squared multiple correlation coefficient (MRC-distribution) [Lee \(1971, 1972\)](#). The parameters of this distribution are the population squared multiple correlation coefficient  $\rho^2$ , the

number of predictors  $m$ , and the sample size  $N$ . The only difference between the H0 and H1 distribution is that the population multiple correlation coefficient is set to "H0  $\rho^2$ " in the former and to "H1  $\rho^2$ " in the latter case.

Several algorithms for the computation of the exact or approximate CDF of the distribution have been proposed (Benton & Krishnamoorthy, 2003; Ding, 1996; Ding & Bargmann, 1991; Lee, 1971, 1972). Benton and Krishnamoorthy (2003) have shown, that the implementation proposed by Ding and Bargmann (1991) (that is used in Dunlap, Xin, and Myers (2004)) may produce grossly false results in some cases. The implementation of Ding (1996) has the disadvantage that it overflows for large sample sizes, because factorials occurring in ratios are explicitly evaluated. This can easily be avoided by using the log of the gamma function in the computation instead.

In G\*POWER we use the procedure of Benton and Krishnamoorthy (2003) to compute the exact CDF and a modified version of the procedure given in Ding (1996) to compute the exact PDF of the distribution. Optionally, one can choose to use the 3-moment noncentral F approximation proposed by Lee (1971) to compute the CDF. The latter procedure has also been used by Steiger and Fouladi (1992) in their R2 program, which provides similar functionality.

## 7.6 Validation

The power and sample size results were checked against the values produced by R2 (Steiger & Fouladi, 1992), the tables in Gatsonis and Sampson (1989), and results reported in Dunlap et al. (2004) and Shieh and Kung (2007). Slight deviations from the values computed with R2 were found, which are due to the approximation used in R2, whereas complete correspondence was found in all other tests made. The confidence intervals were checked against values computed in R2, the results reported in Shieh and Kung (2007), and the tables given in Mendoza and Stafford (2001).

## 7.7 References

See Chapter 9 in Cohen (1988) for a description of the fixed model. The random model is described in Gatsonis and Sampson (1989) and Sampson (1974).

## 8 Exact: Proportion - sign test

The sign test is equivalent to a test that the probability  $\pi$  of an event in the populations has the value  $\pi_0 = 0.5$ . It is identical to the special case  $\pi_0 = 0.5$  of the test Exact: Proportion - difference from constant (one sample case). For a more thorough description see the comments for that test.

### 8.1 Effect size index

The effect size index is  $g = \pi - 0.5$ .

(Cohen, 1969, p. 142) defines the following **effect size conventions**:

- **small**  $g = 0.05$
- **medium**  $g = 0.15$
- **large**  $g = 0.25$

### 8.2 Options

See comments for Exact: Proportion - difference from constant (one sample case) in chapter 4 (page 11).

### 8.3 Examples

### 8.4 Related tests

Similar tests in G\*POWER 3.0:

- Exact: Proportion - difference from constant (one sample case).

### 8.5 Implementation notes

See comments for Exact: Proportion - difference from constant (one sample case) in chapter 4 (page 11).

### 8.6 Validation

The results were checked against the tabulated values in Cohen (1969, chap. 5). For more information see comments for Exact: Proportion - difference from constant (one sample case) in chapter 4 (page 11).

## 9 Exact: Generic binomial test

### 9.1 Effect size index

### 9.2 Options

Since the binomial distribution is discrete, it is normally not possible to achieve exactly the nominal  $\alpha$ -level. For two-sided tests this leads to the problem how to “distribute”  $\alpha$  on the two sides. G\*POWER offers three options (case 1 is the default):

1. *Assign  $\alpha/2$  on both sides*: Both sides are handled independently in exactly the same way as in a one-sided test. The only difference is that here  $\alpha/2$  is used instead of  $\alpha$ . From the three options this one leads to the greatest deviation from the actual  $\alpha$  (in post hoc analyses).
2. *Assign to minor tail  $\alpha/2$ , then rest to major tail ( $\alpha_2 = \alpha/2, \alpha_1 = \alpha - \alpha_2$ )*: First  $\alpha/2$  is applied on the side of the central distribution that is farther away from the noncentral distribution (minor tail). The criterion used on the other side is then  $\alpha - \alpha_1$ , where  $\alpha_1$  is the actual  $\alpha$  found on the minor side. Since  $\alpha_1 \leq \alpha/2$  one can conclude that (in post hoc analyses) the sum of the actual values  $\alpha_1 + \alpha_2$  is in general closer to the nominal  $\alpha$ -level than in case 1.
3. *Assign  $\alpha/2$  on both sides, then increase to minimize the difference of  $\alpha_1 + \alpha_2$  to  $\alpha$* : The first step is exactly the same as in case 1. Then, in the second step, the critical values on both sides of the distribution are increased (using the lower of the two potential incremental  $\alpha$ -values) until the sum of both actual  $\alpha$ 's is as close as possible to the nominal  $\alpha$ .

### 9.3 Examples

### 9.4 Related tests

### 9.5 Implementation notes

### 9.6 Validation

The results were checked against the values produced by GPower 2.0.

### 9.7 References

Cohen...

# 10 F test: Fixed effects ANOVA - one way

The fixed effects one-way ANOVA tests whether there are any differences between the means  $\mu_i$  of  $k \geq 2$  normally distributed random variables with equal variance  $\sigma$ . The random variables represent measurements of a variable  $X$  in  $k$  fixed populations. The one-way ANOVA can be viewed as an extension of the two group t test for a difference of means to more than two groups.

The null hypothesis is that all  $k$  means are identical  $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ . The alternative hypothesis states that at least two of the  $k$  means differ.  $H_1 : \mu_i \neq \mu_j$ , for at least one pair  $i, j$  with  $1 \leq i, j \leq k$ .

## 10.1 Effect size index

The effect size  $f$  is defined as:  $f = \sigma_m / \sigma$ . In this equation  $\sigma_m$  is the standard deviation of the group means  $\mu_i$  and  $\sigma$  the common standard deviation within each of the  $k$  groups. The total variance is then  $\sigma_i^2 = \sigma_m^2 + \sigma^2$ . A different but equivalent way to specify the effect size is in terms of  $\eta^2$ , which is defined as  $\eta^2 = \sigma_m^2 / \sigma_i^2$ . That is,  $\eta^2$  is the ratio between the between-groups variance  $\sigma_m^2$  and the total variance  $\sigma_i^2$  and can be interpreted as "proportion of variance explained by group membership". The relationship between  $\eta^2$  and  $f$  is:  $\eta^2 = f^2 / (1 + f^2)$  or solved for  $f$ :  $f = \sqrt{\eta^2 / (1 - \eta^2)}$ .

Cohen (1969, p.348) defines the following effect size conventions:

- **small**  $f = 0.10$
- **medium**  $f = 0.25$
- **large**  $f = 0.40$

If the mean  $\mu_i$  and size  $n_i$  of all  $k$  groups are known then the standard deviation  $\sigma_m$  can be calculated in the following way:

$$\bar{\mu} = \sum_{i=1}^k w_i \mu_i, \quad (\text{grand mean}),$$

$$\sigma_m = \sqrt{\sum_{i=1}^k w_i (\mu_i - \bar{\mu})^2}.$$

where  $w_i = n_i / (n_1 + n_2 + \dots + n_k)$  stands for the relative size of group  $i$ .

Pressing the Determine button to the left of the effect size label opens the effect size drawer. You can use this drawer to calculate the effect size  $f$  from variances, from  $\eta^2$  or from the group means and group sizes. The drawer essentially contains two different dialogs and you can use the Select procedure selection field to choose one of them.

### 10.1.1 Effect size from means

In this dialog (see left side of Fig. 11) you normally start by setting the number of groups. G\*POWER then provides you with a mean and group size table of appropriate size. Insert the standard deviation  $\sigma$  common to all groups in the SD  $\sigma$  within each group field. Then you need to specify the mean  $\mu_i$  and size  $n_i$  for each group. If all group sizes are equal then you may insert the common group size in the input field to the right of the Equal n button. Clicking on

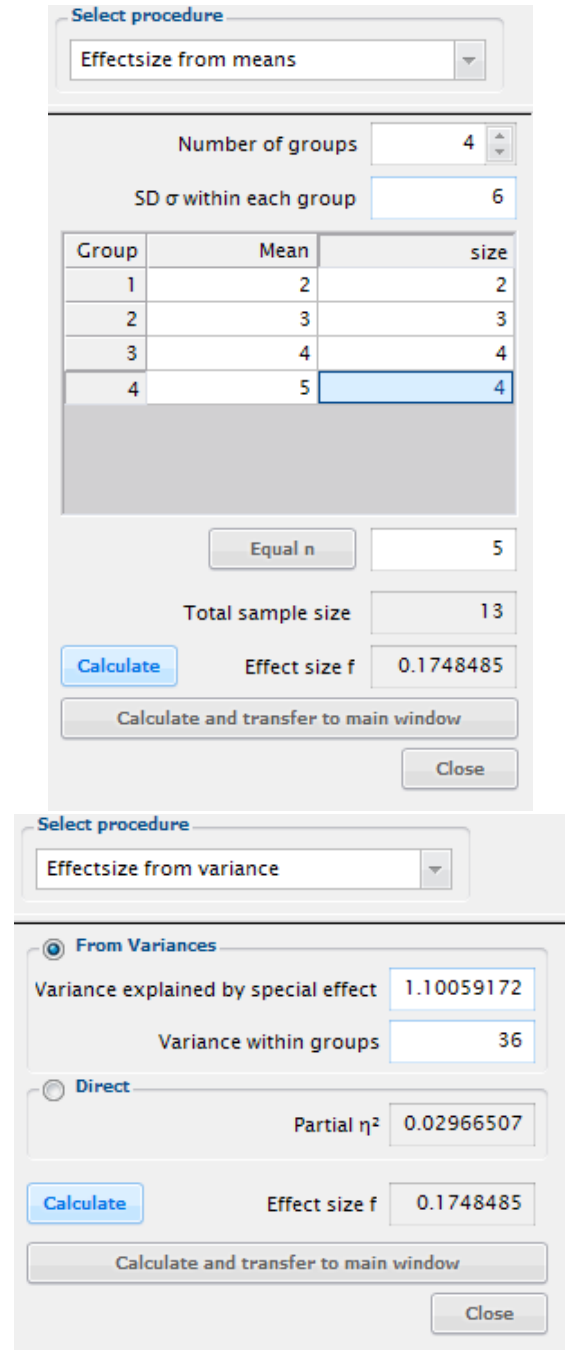


Figure 11: Effect size dialogs to calculate  $f$

this button fills the size column of the table with the chosen value.

Clicking on the Calculate button provides a preview of the effect size that results from your inputs. If you click on the Calculate and transfer to main window button then G\*POWER calculates the effect size and transfers the result into the effect size field in the main window. If the number of groups or the total sample size given in the effect size drawer differ from the corresponding values in the main window, you will be asked whether you want to adjust the values in the main window to the ones in the effect size drawer.



### 10.1.2 Effect size from variance

This dialog offers two ways to specify  $f$ . If you choose From Variances then you need to insert the variance of the group means, that is  $\sigma_m^2$ , into the Variance explained by special effect field, and the square of the common standard deviation within each group, that is  $\sigma^2$ , into the Variance within groups field. Alternatively, you may choose the option Direct and then specify the effect size  $f$  via  $\eta^2$ .

### 10.2 Options

This test has no options.

### 10.3 Examples

We compare 10 groups, and we have reason to expect a "medium" effect size ( $f = .25$ ). How many subjects do we need in a test with  $\alpha = 0.05$  to achieve a power of 0.95?

- **Select**

*Type of power analysis:* A priori

- **Input**

*Effect size  $f$ :* 0.25

*$\alpha$  err prob:* 0.05

*Power (1- $\beta$  err prob):* 0.95

*Number of groups:* 10

- **Output**

*Noncentrality parameter  $\lambda$ :* 24.375000

*Critical F:* 1.904538

*Numerator df:* 9

*Denominator df:* 380

*Total sample size:* 390

*Actual Power:* 0.952363

Thus, we need 39 subjects in each of the 10 groups. What if we had only 200 subjects available? Assuming that both  $\alpha$  and  $\beta$  error are equally costly (i.e., the ratio  $q := \beta/\alpha = 1$ ) which probably is the default in basic research, we can compute the following compromise power analysis:

- **Select**

*Type of power analysis:* Compromise

- **Input**

*Effect size  $f$ :* 0.25

*$\beta/\alpha$  ratio:* 1

*Total sample size:* 200

*Number of groups:* 10

- **Output**

*Noncentrality parameter  $\lambda$ :* 12.500000

*Critical F:* 1.476210

*Numerator df:* 9

*Denominator df:* 190

*$\alpha$  err prob:* 0.159194

*$\beta$  err prob:* 0.159194

*Power (1- $\beta$  err prob):* 0.840806

### 10.4 Related tests

- ANOVA: Fixed effects, special, main effects and interactions
- ANOVA: Repeated measures, between factors

### 10.5 Implementation notes

The distribution under  $H_0$  is the central  $F(k-1, N-k)$  distribution with numerator  $df_1 = k-1$  and denominator  $df_2 = N-k$ . The distribution under  $H_1$  is the noncentral  $F(k-1, N-k, \lambda)$  distribution with the same  $df$ 's and non-centrality parameter  $\lambda = f^2N$ . ( $k$  is the number of groups,  $N$  is the total sample size.)

### 10.6 Validation

The results were checked against the values produced by GPower 2.0.

## 11 F test: Fixed effects ANOVA - special, main effects and interactions

This procedure may be used to calculate the power of main effects and interactions in fixed effects ANOVAs with factorial designs. It can also be used to compute power for planned comparisons. We will discuss both applications in turn.

### 11.0.1 Main effects and interactions

To illustrate the concepts underlying tests of main effects and interactions we will consider the specific example of an  $A \times B \times C$  factorial design, with  $i = 3$  levels of  $A$ ,  $j = 3$  levels of  $B$ , and  $k = 4$  levels of  $C$ . This design has a total number of  $3 \times 3 \times 4 = 36$  groups. A general assumption is that all groups have the same size and that in each group the dependent variable is normally distributed with identical variance.

In a three factor design we may test three main effects of the factors  $A$ ,  $B$ ,  $C$ , three two-factor interactions  $A \times B$ ,  $A \times C$ ,  $B \times C$ , and one three-factor interaction  $A \times B \times C$ . We write  $\mu_{ijk}$  for the mean of group  $A = i, B = j, C = k$ . To indicate the mean of means across a dimension we write a star ( $\star$ ) in the corresponding index. Thus, in the example  $\mu_{ij\star}$  is the mean of the groups  $A = i, B = j, C = 1, 2, 3, 4$ . To simplify the discussion we assume that the grand mean  $\mu_{\star\star\star}$  over all groups is zero. This can always be achieved by subtracting a given non-zero grand mean from each group mean.

In testing the **main effects**, the null hypothesis is that all means of the corresponding factor are identical. For the main effect of factor  $A$  the hypotheses are, for instance:

$$H_0 : \mu_{1\star\star} = \mu_{2\star\star} = \mu_{3\star\star}$$

$$H_1 : \mu_{i\star\star} \neq \mu_{j\star\star} \text{ for at least one index pair } i, j.$$

The assumption that the grand mean is zero implies that  $\sum_i \mu_{i\star\star} = \sum_j \mu_{\star j\star} = \sum_k \mu_{\star\star k} = 0$ . The above hypotheses are therefore equivalent to

$$H_0 : \mu_{i\star\star} = 0 \text{ for all } i$$

$$H_1 : \mu_{i\star\star} \neq 0 \text{ for at least one } i.$$

In testing **two-factor interactions**, the residuals  $\delta_{ij\star}$ ,  $\delta_{i\star k}$ , and  $\delta_{\star ik}$  of the groups means after subtraction of the main effects are considered. For the  $A \times B$  interaction of the example, the  $3 \times 3 = 9$  relevant residuals are  $\delta_{ij\star} = \mu_{ij\star} - \mu_{i\star\star} - \mu_{\star j\star}$ . The null hypothesis of no interaction effect states that all residuals are identical. The hypotheses for the  $A \times B$  interaction are, for example:

$$H_0 : \delta_{ij\star} = \delta_{kl\star} \text{ for all index pairs } i, j \text{ and } k, l.$$

$$H_1 : \delta_{ij\star} \neq \delta_{kl\star} \text{ for at least one combination of } i, j \text{ and } k, l.$$

The assumption that the grand mean is zero implies that  $\sum_{i,j} \delta_{ij\star} = \sum_{i,k} \delta_{i\star k} = \sum_{j,k} \delta_{\star jk} = 0$ . The above hypotheses are therefore equivalent to

$$H_0 : \delta_{ij\star} = 0 \text{ for all } i, j$$

$$H_1 : \delta_{ij\star} \neq 0 \text{ for at least one } i, j.$$

In testing the **three-factor interactions**, the residuals  $\delta_{ijk}$  of the group means after subtraction of all main effects and all two-factor interactions are considered. In a three factor design there is only one possible three-factor interaction. The  $3 \times 3 \times 4 = 36$  residuals in the example are calculated as  $\delta_{ijk} = \mu_{ijk} - \mu_{i\star\star} - \mu_{\star j\star} - \mu_{\star\star k} - \delta_{ij\star} - \delta_{i\star k} - \delta_{\star jk}$ . The null hypothesis of no interaction states that all residuals are equal. Thus,

$$H_0 : \delta_{ijk} = \delta_{lmn} \text{ for all combinations of index triples } i, j, k \text{ and } l, m, n.$$

$$H_1 : \delta_{ijk} \neq \delta_{lmn} \text{ for at least one combination of index triples } i, j, k \text{ and } l, m, n.$$

The assumption that the grand mean is zero implies that  $\sum_{i,j,k} \delta_{ijk} = 0$ . The above hypotheses are therefore equivalent to

$$H_0 : \delta_{ijk} = 0 \text{ for all } i, j, k$$

$$H_1 : \delta_{ijk} \neq 0 \text{ for at least one } i, j, k.$$

It should be obvious how the reasoning outlined above can be generalized to designs with 4 and more factors.

### 11.0.2 Planned comparisons

Planned comparison are specific tests between levels of a factor planned before the experiment was conducted.

One application is the **comparison between two sets of levels**. The general idea is to subtract the means across two sets of levels that should be compared from each other and to test whether the difference is zero. Formally this is done by calculating the sum of the component-wise product of the mean vector  $\vec{\mu}$  and a nonzero contrast vector  $\vec{c}$  (i.e. the scalar product of  $\vec{\mu}$  and  $\vec{c}$ ):  $C = \sum_{i=1}^k c_i \mu_i$ . The contrast vector  $\vec{c}$  contains negative weights for levels on one side of the comparison, positive weights for the levels on the other side of the comparison and zero for levels that are not part of the comparison. The sum of weights is always zero. Assume, for instance, that we have a factor with 4 levels and mean vector  $\vec{\mu} = (2, 3, 1, 2)$  and that we want to test whether the means in the first two levels are identical to the means in the last two levels. In this case we define  $\vec{c} = (-1/2, -1/2, 1/2, 1/2)$  and get  $C = \sum_i \vec{\mu}_i \vec{c}_i = -1 - 3/2 + 1/2 + 1 = -1$ .

A second application is testing polygonal contrasts in a **trend analysis**. In this case it is normally assumed that the factor represents a quantitative variable and that the levels of the factor that correspond to specific values of this quantitative variable are equally spaced (for more details, see e.g. [Hays \(1988, p. 706ff\)](#)). In a factor with  $k$  levels  $k - 1$  orthogonal polynomial trends can be tested.

In planned comparisons the null hypothesis is:  $H_0 : C = 0$ , and the alternative hypothesis  $H_1 : C \neq 0$ .

### 11.1 Effect size index

The effect size  $f$  is defined as:  $f = \sigma_m / \sigma$ . In this equation  $\sigma_m$  is the standard deviation of the effects that we want to test and  $\sigma$  the common standard deviation within each of the groups in the design. The total variance is then  $\sigma_t^2 = \sigma_m^2 + \sigma^2$ . A different but equivalent way to specify the effect size

is in terms of  $\eta^2$ , which is defined as  $\eta^2 = \sigma_m^2 / \sigma_i^2$ . That is,  $\eta^2$  is the ratio between the between-groups variance  $\sigma_m^2$  and the total variance  $\sigma_i^2$  and can be interpreted as “proportion of variance explained by the effect under consideration”. The relationship between  $\eta^2$  and  $f$  is:  $\eta^2 = f^2 / (1 + f^2)$  or solved for  $f$ :  $f = \sqrt{\eta^2 / (1 - \eta^2)}$ .

Cohen (1969, p.348) defines the following effect size conventions:

- **small**  $f = 0.10$
- **medium**  $f = 0.25$
- **large**  $f = 0.40$

Figure 12: Effect size dialog to calculate  $f$

Clicking on the Determine button to the left of the effect size label opens the effect size drawer (see Fig. 12). You can use this drawer to calculate the effect size  $f$  from variances or from  $\eta^2$ . If you choose From Variances then you need to insert the variance explained by the effect under consideration, that is  $\sigma_m^2$ , into the Variance explained by special effect field, and the square of the common standard deviation within each group, that is  $\sigma^2$ , into the Variance within groups field. Alternatively, you may choose the option Direct and then specify the effect size  $f$  via  $\eta^2$ .

See examples section below for information on how to calculate the effect size  $f$  in tests of main effects and interactions and tests of planned comparisons.

## 11.2 Options

This test has no options.

## 11.3 Examples

### 11.3.1 Effect sizes from means and standard deviations

To illustrate the test of main effects and interaction we assume the specific values for our  $A \times B \times C$  example shown in Table 13. Table 14 shows the results of a SPSS analysis (GLM univariate) done for these data. In the following we will show how to reproduce the values in the Observed Power column in the SPSS output with G\*POWER.

As a first step we calculate the grand mean of the data. Since all groups have the same size ( $n = 3$ ) this is just the arithmetic mean of all 36 groups means:  $m_g = 3.1382$ . We then subtract this grand mean from all cells (this step

is not essential but makes the calculation and discussion easier). Next, we estimate the common variance  $\sigma$  within each group by calculating the mean variance of all cells, that is,  $\sigma^2 = 1/36 \sum_i s_i^2 = 1.71296$ .

**Main effects** To calculate the power for the  $A$ ,  $B$ , and  $C$  main effects we need to know the effect size  $f = \sigma_m / \sigma$ . We already know  $\sigma^2$  to be 1.71296 but need to calculate the variance of the means  $\sigma_m^2$  for each factor. The procedure is analogous for all three main effects. We therefore demonstrate only the calculations necessary for the main effect of factor  $A$ .

We first calculate the three means for factor  $A$ :  $\mu_{i^{**}} = \{-0.722231, 1.30556, -0.583331\}$ . Due to the fact that we have first subtracted the grand mean from each cell we have  $\sum_i \mu_{i^{**}} = 0$ , and we can easily compute the variance of these means as mean square:  $\sigma_m^2 = 1/3 \sum_i \mu_{i^{**}}^2 = 0.85546$ . With these values we calculate  $f = \sqrt{\sigma_m / \sigma} = \sqrt{0.85546 / 1.71296} = 0.7066856$ . The effect size drawer in G\*POWER can be used to do the last calculation: We choose From Variances and insert 0.85546 in the Variance explained by special effect and 1.71296 in the Error Variance field. Pressing the Calculate button gives the above value for  $f$  and a partial  $\eta^2$  of 0.3330686. Note that the partial  $\eta^2$  given by G\*POWER is calculated from  $f$  according to the formula  $\eta^2 = f^2 / (1 + f^2)$  and is not identical to the SPSS partial  $\eta^2$ , which is based on sample estimates. The relation between the two is “SPSS  $\eta_0^2$ ” =  $\eta^2 N / (N + k(\eta^2 - 1))$ , where  $N$  denotes the total sample size,  $k$  the total number of groups in the design and  $\eta^2$  the G\*POWER value. Thus  $\eta_0^2 = 0.33306806 \cdot 108 / (108 - 36 + 0.33306806 \cdot 36) = 0.42828$ , which is the value given in the SPSS output.

We now use G\*POWER to calculate the power for  $\alpha = 0.05$  and a total sample size  $3 \times 3 \times 4 \times 3 = 108$ . We set

- **Select**  
Type of power analysis: Post hoc
- **Input**  
Effect size  $f$ : 0.7066856  
 $\alpha$  err prob: 0.05  
Total sample size: 108  
Numerator df: 2 (number of factor levels - 1, A has 3 levels)  
Number of groups: 36 (total number of groups in the design)
- **Output**  
Noncentrality parameter  $\lambda$ : 53.935690  
Critical F: 3.123907  
Denominator df: 72  
Power (1- $\beta$  err prob): 0.99999

The value of the noncentrality parameter and the power computed by G\*POWER are identical to the values in the SPSS output.

**Two-factor interactions** To calculate the power for two-factor interactions  $A \times B$ ,  $A \times C$ , and  $B \times C$  we need to calculate the effect size  $f$  corresponding to the values given in table 13. The procedure is analogous for each of the three

A=1					A=2					A=3				
B	C	m	s	n	B	C	m	s	n	B	C	m	s	n
1	1	3.0000	1.0000	3	1	1	3.3333	1.5275	3	1	1	2.6667	0.5774	3
1	2	4.6667	1.5275	3	1	2	3.3333	1.5275	3	1	2	2.6667	2.0817	3
1	3	2.0000	1.0000	3	1	3	4.6667	1.5275	3	1	3	2.0000	1.0000	3
1	4	2.0000	1.0000	3	1	4	4.6667	1.5275	3	1	4	2.0000	1.0000	3
2	1	3.0000	1.0000	3	2	1	6.3333	0.5774	3	2	1	4.6667	2.8868	3
2	2	2.0000	1.0000	3	2	2	4.6667	1.5275	3	2	2	2.0000	1.0000	3
2	3	2.0000	1.0000	3	2	3	4.6667	1.5275	3	2	3	2.0000	1.0000	3
2	4	2.0000	1.0000	3	2	4	4.6667	1.5275	3	2	4	2.0000	1.0000	3
3	1	2.3333	0.5774	3	3	1	3.0000	0.0000	3	3	1	4.6667	2.0817	3
3	2	2.0000	1.0000	3	3	2	4.6667	1.5275	3	3	2	2.0000	1.0000	3
3	3	2.0000	1.0000	3	3	3	4.6667	1.5275	3	3	3	2.0000	1.0000	3
3	4	2.0000	1.0000	3	3	4	4.6667	1.5275	3	3	4	2.0000	1.0000	3

Figure 13: Hypothetical means (m) and standard deviations (s) of a  $3 \times 3 \times 4$  design.

Tests of Between-Subjects Effects

Dependent Variable: X

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Corrected Model	173.583 <sup>b</sup>	35	4.960	2.895	.000	.585	101.335	1.000
Intercept	1064.083	1	1064.083	621.195	.000	.896	621.195	1.000
A	92.389	2	46.194	26.968	.000	.428	53.935	1.000
B	2.167	2	1.083	.632	.534	.017	1.265	.152
C	10.917	3	3.639	2.124	.105	.081	6.373	.521
A * B	11.111	4	2.778	1.622	.178	.083	6.486	.476
A * C	22.500	6	3.750	2.189	.054	.154	13.135	.740
B * C	14.500	6	2.417	1.411	.222	.105	8.465	.517
A * B * C	20.000	12	1.667	.973	.482	.140	11.676	.513
Error	123.333	72	1.713					
Total	1361.000	108						
Corrected Total	296.917	107						

a. Computed using alpha = .05

b. R Squared = .585 (Adjusted R Squared = .383)

Figure 14: Results computed with SPSS for the values given in table 13

two-factor interactions and we thus restrict ourselves to the  $A \times B$  interaction.

The values needed to calculate  $\sigma_m^2$  are the  $3 \times 3 = 9$  residuals  $\delta_{ij*}$ . They are given by  $\delta_{ij*} = \mu_{ij*} - \mu_{i**} - \mu_{*j*} = \{0.555564, -0.361111, -0.194453, -0.388903, 0.444447, -0.0555444, -0.166661, -0.0833361, 0.249997\}$ . The mean of these values is zero (as a consequence of subtracting the grand mean). Thus, the variance  $\sigma_m$  is given by  $1/9 \sum_{i,j} \delta_{ij*}^2 = 0.102881$ . This results in an effect size  $f = \sqrt{0.102881/1.71296} = 0.2450722$  and a partial  $\eta^2 = 0.0557195$ . Using the formula given in the previous section on main effects it can be checked that this corresponds to a "SPSS  $\eta_0^2$ " of 0.0813, which is identical to that given in the SPSS output.

We use G\*POWER to calculate the power for  $\alpha = 0.05$  and a total sample size  $3 \times 3 \times 4 \times 3 = 108$ . We set:

• **Select**

Type of power analysis: Post hoc

• **Input**

Effect size f: 0.2450722

$\alpha$  err prob: 0.05

Total sample size: 108

Numerator df: 4 (#A-1)(#B-1) = (3-1)(3-1)

Number of groups: 36 (total number of groups in the design)

• **Output**

Noncentrality parameter  $\lambda$ : 6.486521

Critical F: 2.498919

Denominator df: 72

Power (1- $\beta$  err prob): 0.475635

(The notation #A in the comment above means number of levels in factor A). A check reveals that the value of the noncentrality parameter and the power computed by G\*POWER are identical to the values for (A \* B) in the SPSS output.

**Three-factor interactions** To calculate the effect size of the three-factor interaction corresponding to the values given in table 13 we need the variance of the 36 residuals  $\delta_{ijk} = \mu_{ijk} - \mu_{i**} - \mu_{*j*} - \mu_{**k} - \delta_{ij*} - \delta_{i*j} - \delta_{*jk} = \{0.333336, 0.777792, -0.555564, -0.555564, -0.416656, -0.305567, 0.361111, 0.361111, 0.0833194, -0.472225, 0.194453, 0.194453, 0.166669, -0.944475, 0.388903, 0.388903, 0.666653, 0.222242, -0.444447, -0.444447, -0.833322, 0.722233, 0.0555444, 0.0555444, -0.500006, 0.166683, 0.166661, 0.166661, -0.249997, 0.083325, 0.0833361, 0.0833361, 0.750003, -0.250008, -$

0.249997, -0.249997}. The mean of these values is zero (as a consequence of subtracting the grand mean). Thus, the variance  $\sigma_m$  is given by  $1/36 \sum_{i,j,k} \delta_{ijk}^2 = 0.185189$ . This results in an effect size  $f = \sqrt{0.185189/1.71296} = 0.3288016$  and a partial  $\eta^2 = 0.09756294$ . Using the formula given in the previous section on main effects it can be checked that this corresponds to a “SPSS  $\eta^2$ ” of 0.140, which is identical to that given in the SPSS output.

We use G\*POWER to calculate the power for  $\alpha = 0.05$  and a total sample size  $3 \times 3 \times 4 \times 3 = 108$ . We therefore choose

- **Select**

*Type of power analysis:* Post hoc

- **Input**

*Effect size f:* 0.3288016

*$\alpha$  err prob:* 0.05

*Total sample size:* 108

*Numerator df:* 12 (#A-1)(#B-1)(#C-1) = (3-1)(3-1)(4-1)

*Number of groups:* 36 (total number of groups in the design)

- **Output**

*Noncentrality parameter  $\lambda$ :* 11.675933

*Critical F:* 1.889242

*Denominator df:* 72

*Power (1- $\beta$  err prob):* 0.513442

(The notation #A in the comment above means number of levels in factor A). Again a check reveals that the value of the noncentrality parameter and the power computed by G\*POWER are identical to the values for (A \* B \* C) in the SPSS output.

### 11.3.2 Using conventional effect sizes

In the example given in the previous section, we assumed that we know the true values of the mean and variances in all groups. We are, however, seldom in that position. Instead, we usually only have rough estimates of the expected effect sizes. In these cases we may resort to the conventional effect sizes proposed by Cohen.

Assume that we want to calculate the total sample size needed to achieve a power of 0.95 in testing the  $A \times C$  two-factor interaction at  $\alpha$  level 0.05. Assume further that the total design in this scenario is  $A \times B \times C$  with  $3 \times 2 \times 5$  factor levels, that is, 30 groups. Theoretical considerations suggest that there should be a small interaction. We thus use the conventional value  $f = 0.1$  defined by Cohen (1969) as small effect. The inputs into and outputs of G\*POWER for this scenario are:

- **Select**

*Type of power analysis:* A priori

- **Input**

*Effect size f:* 0.1

*$\alpha$  err prob:* 0.05

*Power (1- $\beta$  err prob):* 0.95

*Numerator df:* 8 (#A-1)(#C-1) = (3-1)(5-1)

*Number of groups:* 30 (total number of groups in the design)

- **Output**

*Noncentrality parameter  $\lambda$ :* 22.830000

*Critical F:* 1.942507

*Denominator df:* 2253

*Total sample size:* 2283

*Actual power:* 0.950078

G\*POWER calculates a total sample size of 2283. Please note that this sample size is not a multiple of the group size 30 (2283/30 = 76.1)! If you want to ensure that you have equal group sizes, round this value up to a multiple of 30 by choosing a total sample size of 30\*77=2310. A post hoc analysis with this sample size reveals that this increases the power to 0.952674.

### 11.3.3 Power for planned comparisons

To calculate the effect size  $f = \sigma_m/\sigma$  for a given comparison  $C = \sum_{i=1}^k \mu_i c_i$  we need to know—besides the standard deviation  $\sigma$  within each group—the standard deviation  $\sigma_m$  of the effect. It is given by:

$$\sigma_m = \frac{|C|}{\sqrt{N \sum_{i=1}^k c_i^2 / n_i}}$$

where  $N$ ,  $n_i$  denote total sample size and sample size in group  $i$ , respectively.

Given the mean vector  $\mu = (1.5, 2, 3, 4)$ , sample size  $n_i = 5$  in each group, and standard deviation  $\sigma = 2$  within each group, we want to calculate the power for the following contrasts:

contrast	weights $c$				$\sigma_\mu$	$f$	$\eta^2$
1,2 vs. 3,4	$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0.875	0.438	0.161
lin. trend	-3	-1	1	3	0.950	0.475	0.184
quad. trend	1	-1	-1	1	0.125	0.063	0.004

Each contrast has a numerator  $df = 1$ . The denominator dfs are  $N - k$ , where  $k$  is the number of levels (4 in the example).

To calculate the power of the linear trend at  $\alpha = 0.05$  we specify:

- **Select**

*Type of power analysis:* A priori

- **Input**

*Effect size f:* 0.475164

*$\alpha$  err prob:* 0.05

*Total sample size:* 20

*Numerator df:* 1

*Number of groups:* 4

- **Output**

*Noncentrality parameter  $\lambda$ :* 4.515617

*Critical F:* 4.493998

*Denominator df:* 16

*Power (1- $\beta$  err prob):* 0.514736

Inserting the  $f$ 's for the other two contrasts yields a power of 0.451898 for the comparison of “1,2 vs. 3,4”, and a power of 0.057970 for the test of a quadratic trend.

## 11.4 Related tests

- ANOVA: One-way

## 11.5 Implementation notes

The distribution under  $H_0$  is the central  $F(df_1, N - k)$  distribution. The numerator  $df_1$  is specified in the input and the denominator df is  $df_2 = N - k$ , where  $N$  is the total sample size and  $k$  the total number of groups in the design. The distribution under  $H_1$  is the noncentral  $F(df_1, N - k, \lambda)$  distribution with the same df's and noncentrality parameter  $\lambda = f^2N$ .

## 11.6 Validation

The results were checked against the values produced by GPower 2.0.

## 12 t test: Linear Regression (size of slope, one group)

A linear regression is used to estimate the parameters  $a$ ,  $b$  of a linear relationship  $Y = a + bX$  between the dependent variable  $Y$  and the independent variable  $X$ .  $X$  is assumed to be a set of fixed values, whereas  $Y_i$  is modeled as a random variable:  $Y_i = a + bX_i + \varepsilon_i$ , where  $\varepsilon_i$  denotes normally distributed random errors with mean 0 and standard deviation  $\sigma_i$ . A common assumption also adopted here is that all  $\sigma_i$ 's are identical, that is  $\sigma_i = \sigma$ . The standard deviation of the error is also called the standard deviation of the residuals.

A common task in linear regression analysis is to test whether the slope  $b$  is identical to a fixed value  $b_0$  or not. The null and the two-sided alternative hypotheses are:

$$\begin{aligned} H_0: & b - b_0 = 0 \\ H_1: & b - b_0 \neq 0. \end{aligned}$$

### 12.1 Effect size index

Slope H1, the slope  $b$  of the linear relationship assumed under H1 is used as effect size measure. To fully specify the effect size, the following additional inputs must be given:

- Slope H0

This is the slope  $b_0$  assumed under H0.

- Std dev  $\sigma_x$

The standard deviation  $\sigma_x$  of the values in  $X$ :  $\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^n (X_i - \bar{X})^2}$ . The standard deviation must be  $> 0$ .

- Std dev  $\sigma_y$

The standard deviation  $\sigma_y > 0$  of the  $Y$ -values. Important relationships of  $\sigma_y$  to other relevant measures are:

$$\sigma_y = (b\sigma_x)/\rho \quad (1)$$

$$\sigma_y = \sigma / \sqrt{1 - \rho^2} \quad (2)$$

where  $\sigma$  denotes the standard deviation of the residuals  $Y_i - (aX + b)$  and  $\rho$  the correlation coefficient between  $X$  and  $Y$ .

The effect size dialog may be used to determine Std dev  $\sigma_y$  and/or Slope H1 from other values based on Eqns (1) and (2) given above.

Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 15).

The right panel in Fig 15 shows the combinations of input and output values in different input modes. The input variables stand on the left side of the arrow ' $\Rightarrow$ ', the output variables on the right side. The input values must conform to the usual restrictions, that is,  $\sigma > 0, \sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$ . In addition, Eqn. (1) together with the restriction on  $\rho$  implies the additional restriction  $-1 < b \cdot \sigma_x / \sigma_y < 1$ .

Clicking on the button Calculate and transfer to main window copies the values given in Slope H1, Std dev  $\sigma_y$  and Std dev  $\sigma_x$  to the corresponding input fields in the main window.

## 12.2 Options

This test has no options.

### 12.3 Examples

We replicate an example given on page 593 in Dupont and Plummer (1998). The question investigated in this example is, whether the actual average time spent per day exercising is related to the body mass index (BMI) after 6 month on a training program. The estimated standard deviation of exercise time of participants is  $\sigma_x = 7.5$  minutes. From a previous study the standard deviation of the BMI in the group of participants is estimated to be  $\sigma_y = 4$ . The sample size is  $N = 100$  and  $\alpha = 0.05$ . We want to determine the power with which a slope  $b_0 = -0.0667$  of the regression line (corresponding to a drop of BMI by 2 per 30 min/day exercise) can be detected.

- **Select**

*Type of power analysis:* Post hoc

- **Input**

*Tail(s):* Two

*Effect size Slope H1:* -0.0667

*$\alpha$  err prob:* 0.05

*Total sample size:* 100

*Slope H0:* 0

*Std dev  $\sigma_x$ :* 7.5

*Std dev  $\sigma_y$ :* 4

- **Output**

*Noncentrality parameter  $\delta$ :* 1.260522

*Critical t:* -1.984467

*Df:* 98

*Power (1-  $\beta$ ):* 0.238969

The output shows that the power of this test is about 0.24. This confirms the value estimated by Dupont and Plummer (1998, p. 596) for this example.

#### 12.3.1 Relation to Multiple Regression: Omnibus

The present procedure is a special case of Multiple regression, or better: a different interface to the same procedure using more convenient variables. To show this, we demonstrate how the MRC procedure can be used to compute the example above. First, we determine  $R^2 = \rho^2$  from the relation  $b = \rho \cdot \sigma_y / \sigma_x$ , which implies  $\rho^2 = (b \cdot \sigma_x / \sigma_y)^2$ . Entering  $(-0.0667 \cdot 7.5 / 4)^2$  into the G\*POWER calculator gives  $\rho^2 = 0.01564062$ . We enter this value in the effect size dialog of the MRC procedure and compute an effect size  $f^2 = 0.0158891$ . Selecting a post hoc analysis and setting  $\alpha$  err prob to 0.05, Total sample size to 100 and Number of predictors to 1, we get exactly the same power as given above.

## 12.4 Related tests

Similar tests in G\*POWER 3.0:

- Multiple Regression: Omnibus ( $R^2$  deviation from zero).
- Correlation: Point biserial model

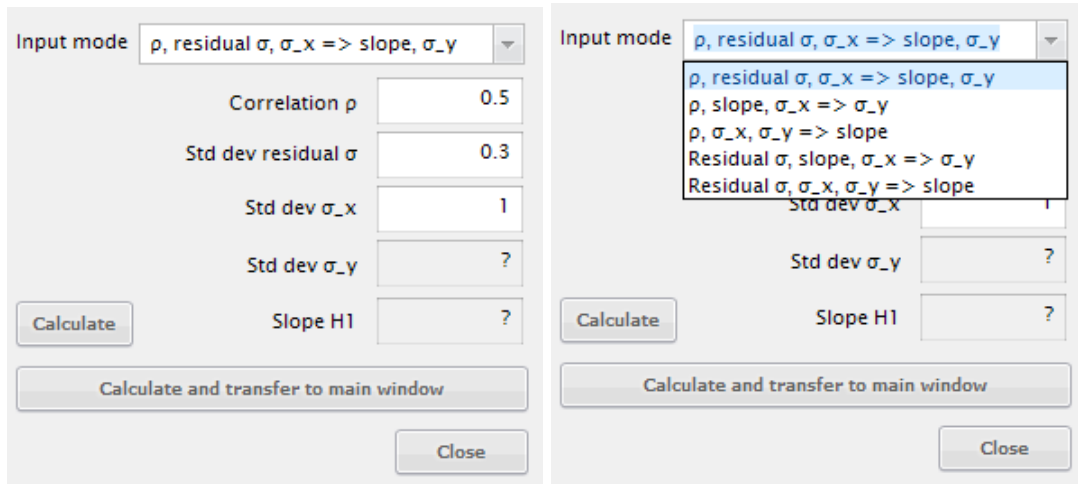


Figure 15: Effect size drawer to calculate  $\sigma_y$  and/or slope  $b$  from various inputs constellations (see right panel).

## 12.5 Implementation notes

The  $H_0$ -distribution is the central  $t$  distribution with  $df_2 = N - 2$  degrees of freedom, where  $N$  is the sample size. The  $H_1$ -distribution is the noncentral  $t$  distribution with the same degrees of freedom and noncentrality parameter  $\delta = \sqrt{N}[\sigma_x(b - b_0)]/\sigma_y$ .

## 12.6 Validation

The results were checked against the values produced by by PASS (Hintze, 2006) and perfect correspondance was found.



## 13 F test: Multiple Regression - omnibus (deviation of $R^2$ from zero), fixed model

In multiple regression analyses the relation of a dependent variable  $Y$  to  $p$  independent factors  $X_1, \dots, X_m$  is studied. The present procedure refers to the so-called conditional or fixed factors model of multiple regression (Gatsonis & Sampson, 1989; Sampson, 1974), that is, it is assumed that

$$Y = X\beta + \varepsilon$$

where  $X = (1X_1X_2 \dots X_m)$  is a  $N \times (m + 1)$  matrix of a constant term and fixed and known predictor variables  $X_i$ . The elements of the column vector  $\beta$  of length  $m + 1$  are the regression weights, and the column vector  $\varepsilon$  of length  $N$  contains error terms, with  $\varepsilon_i \sim N(0, \sigma)$ .

This procedure allows power analyses for the test that the proportion of variance of a dependent variable  $Y$  explained by a set of predictors  $B$ , that is  $R_{Y \cdot B}^2$ , is zero. The null and alternate hypotheses are:

$$\begin{aligned} H_0: & R_{Y \cdot B}^2 = 0 \\ H_1: & R_{Y \cdot B}^2 > 0. \end{aligned}$$

As will be shown in the examples section, the MRC procedure is quite flexible and can be used as a substitute for some other tests.

### 13.1 Effect size index

The general definition of the effect size index  $f^2$  used in this procedure is:  $f^2 = V_S/V_E$ , where  $V_S$  is the proportion of variance explained by a set of predictors, and  $V_E$  the residual or error variance ( $V_E + V_S = 1$ ). In the special case considered here (case 0 in Cohen (1988, p. 407ff.)) the proportion of variance explained is given by  $V_S = R_{Y \cdot B}^2$  and the residual variance by  $V_E = 1 - R_{Y \cdot B}^2$ . Thus:

$$f^2 = \frac{R_{Y \cdot B}^2}{1 - R_{Y \cdot B}^2}$$

and conversely:

$$R_{Y \cdot B}^2 = \frac{f^2}{1 + f^2}$$

Cohen (1988, p. 412) defines the following conventional values for the effect size  $f^2$ :

- **small**  $f^2 = 0.02$
- **medium**  $f^2 = 0.15$
- **large**  $f^2 = 0.35$

Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 16).

**Effect size from squared multiple correlation coefficient**  
Choosing input mode "From correlation coefficient" allows to calculate effect size  $f^2$  from the squared multiple correlation coefficient  $R_{Y \cdot B}^2$ .

Figure 16: Effect size drawer to calculate  $f^2$  from either  $R^2$  or from predictor correlations.

**Effect size from predictor correlations** By choosing the option "From predictor correlations" (see Fig. (17)) one may compute  $\rho^2$  from the matrix of correlations among the predictor variables and the correlations between predictors and the dependent variable  $Y$ . Pressing the "Insert/edit matrix"-button opens a window, in which one can specify (a) the row vector  $u$  containing the correlations between each of the  $m$  predictors  $X_i$  and the dependent variable  $Y$ , and (b) the  $m \times m$  matrix  $B$  of correlations among the predictors. The squared multiple correlation coefficient is then given by  $\rho^2 = uB^{-1}u'$ . Each input correlation must lie in the interval  $[-1, 1]$ , the matrix  $B$  must be positive-definite, and the resulting  $\rho^2$  must lie in the interval  $[0, 1]$ . Pressing the Button "Calc  $\rho^2$ " tries to calculate  $\rho^2$  from the input and checks the positive-definiteness of matrix  $B$  and the restriction on  $\rho^2$ .

### 13.2 Options

This test has no options.

### 13.3 Examples

#### 13.3.1 Basic example

We assume that a dependent variable  $Y$  is predicted by as set  $B$  of 5 predictors and that the population  $R_{Y \cdot B}^2$  is 0.10, that is that the 5 predictors account for 10% of the variance of  $Y$ . The sample size is  $N = 95$  subjects. What is the power of the F test at  $\alpha = 0.05$ ?

First, by inserting  $R^2 = 0.10$  in the effect size dialog we calculate the corresponding effect size  $f^2 = 0.1111111$ . We then use the following settings in G\*POWER to calculate the power:

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Effect size  $f^2$ :* 0.1111111  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 95

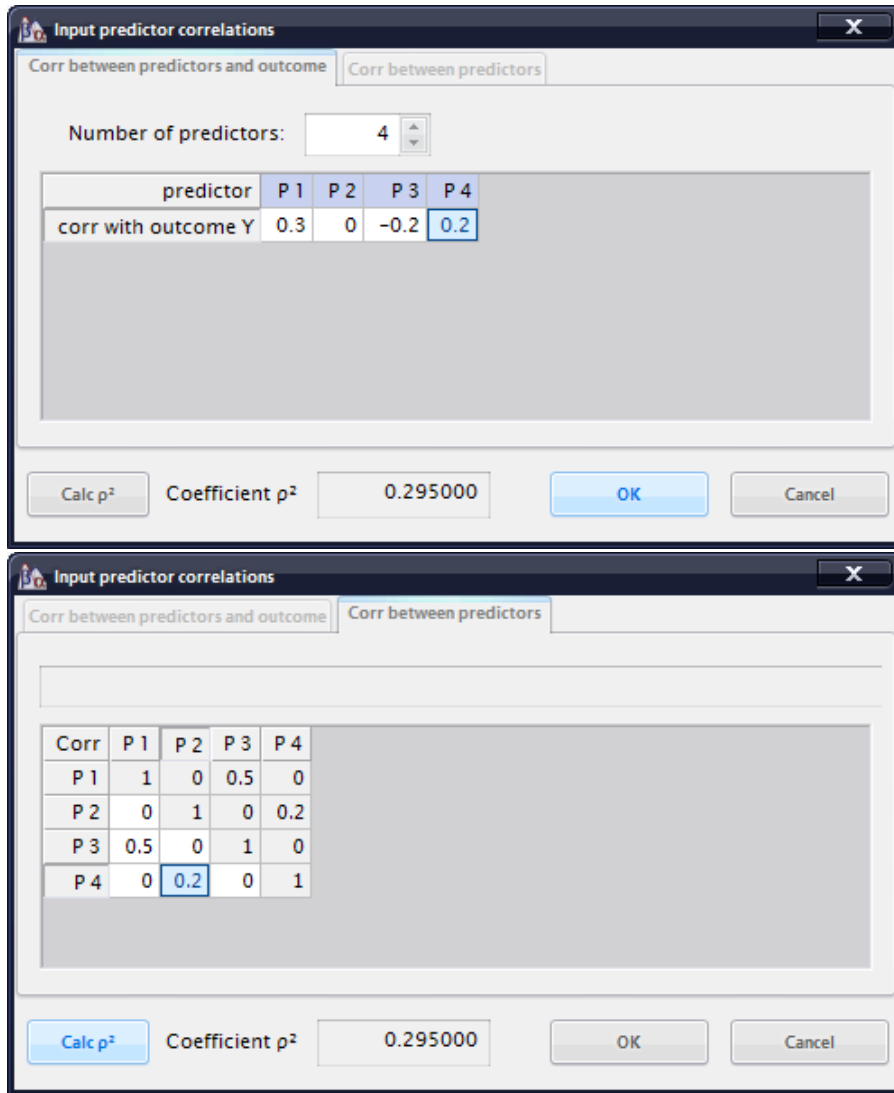


Figure 17: Input of correlations between predictors and Y (top) and the matrix of the correlations among predictors (see text).

Number of predictors: 5

• **Output**

- Noncentrality parameter  $\lambda$ : 10.555555
- Critical F: 2.316858
- Numerator df: 5
- Denominator df: 89
- Power (1-  $\beta$ ): 0.673586

The output shows that the power of this test is about 0.67. This confirms the value estimated by Cohen (1988, p. 424) in his example 9.1, which uses identical values.

**13.3.2 Example showing relations to a one-way ANOVA and the two-sample t-test**

We assume the means 2, 3, 2, 5 for the  $k = 4$  experimental groups in a one-factor design. The sample sizes in the group are 5, 6, 6, 5, respectively, and the common standard deviation is assumed to be  $\sigma = 2$ . Using the effect size dialog of the one-way ANOVA procedure we calculate from these values the effect size  $f = 0.5930904$ . With  $\alpha = 0.05$ , 4 groups, and a total sample size of 22, a power of 0.536011 is computed.

An equivalent analysis could be done using the MRC procedure. To this end we set the effect size of the MRC procedure to  $f^2 = 0.59309044^2 = 0.351756$  and the number of predictors to (number of groups -1) — in the example to  $k - 1 = 3$ . Choosing the remaining parameters  $\alpha$  and total sample size exactly as in the one-way ANOVA case leads to identical result.

From the fact that the two-sided t-tests for the difference in means of two independent groups is a special case of the one-way ANOVA, it can be concluded that this test can also be regarded as a special case of the MRC procedure. The relation between the effect size  $d$  of the t-test and  $f^2$  is as follows:  $f^2 = (d/2)^2$ .

**13.3.3 Example showing the relation to two-sided tests of point-biserial correlations**

For testing whether a point biserial correlation  $r$  is different from zero, using the special procedure provided in G\*POWER is recommended. But the power analysis of the two-sided test can also be done as well with the current MRC procedure. We just need to set  $R^2 = r^2$  and Number of predictor = 1.

Given the correlation  $r = 0.5$  ( $r^2 = 0.25$ ) we get  $f^2 =$

$0.25/(1 - 0.25) = 0.333$ . For  $\alpha = 0.05$  and total sample size  $N = 12$  a power of 0.439627 is computed from both procedures.

### 13.4 Related tests

Similar tests in G\*POWER 3.0:

- Multiple Regression: Special ( $R^2$  increase).
- ANOVA: Fixed effects, omnibus, one-way.
- Means: Difference between two independent means (two groups)
- Correlation: Point biserial model

### 13.5 Implementation notes

The  $H_0$ -distribution is the central  $F$  distribution with numerator degrees of freedom  $df_1 = m$ , and denominator degrees of freedom  $df_2 = N - m - 1$ , where  $N$  is the sample size and  $p$  the number of predictors in the set  $B$  explaining the proportion of variance given by  $R_{Y.B}^2$ . The  $H_1$ -distribution is the noncentral  $F$  distribution with the same degrees of freedom and noncentrality parameter  $\lambda = f^2N$ .

### 13.6 Validation

The results were checked against the values produced by GPower 2.0 and those produced by PASS (Hintze, 2006). Slight deviations were found to the values tabulated in Cohen (1988). This is due to an approximation used by Cohen (1988) that underestimates the noncentrality parameter  $\lambda$  and therefore also the power. This issue is discussed more thoroughly in Erdfelder, Faul, and Buchner (1996).

### 13.7 References

See Chapter 9 in Cohen (1988) .

## 14 F test: Multiple Regression - special (increase of $R^2$ ), fixed model

In multiple regression analyses the relation of a dependent variable  $Y$  to  $m$  independent factors  $X_1, \dots, X_m$  is studied. The present procedure refers to the so-called conditional or fixed factors model of multiple regression (Gatsonis & Sampson, 1989; Sampson, 1974), that is, it is assumed that

$$Y = X\beta + \varepsilon$$

where  $X = (1X_1X_2 \dots X_m)$  is a  $N \times (m + 1)$  matrix of a constant term and fixed and known predictor variables  $X_i$ . The elements of the column vector  $\beta$  of length  $m + 1$  are the regression weights, and the column vector  $\varepsilon$  of length  $N$  contains error terms, with  $\varepsilon_i \sim N(0, \sigma)$ .

This procedure allows power analyses for the test, whether the proportion of variance of variable  $Y$  explained by a set of predictors  $A$  is increased if an additional nonempty predictor set  $B$  is considered. The variance explained by predictor sets  $A$ ,  $B$ , and  $A \cup B$  is denoted by  $R_{Y \cdot A}^2$ ,  $R_{Y \cdot B}^2$ , and  $R_{Y \cdot A, B}^2$ , respectively.

Using this notation, the null and alternate hypotheses are:

$$\begin{aligned} H_0 &: R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2 = 0 \\ H_1 &: R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2 > 0. \end{aligned}$$

The directional form of  $H_1$  is due to the fact that  $R_{Y \cdot A, B}^2$  that is the proportion of variance explained by sets  $A$  and  $B$  combined, cannot be lower than the proportion  $R_{Y \cdot A}^2$  explained by  $A$  alone.

As will be shown in the examples section, the MRC procedure is quite flexible and can be used as a substitute for some other tests.

### 14.1 Effect size index

The general definition of the effect size index  $f^2$  used in this procedure is:  $f^2 = V_S / V_E$ , where  $V_S$  is the proportion of variance explained by a set of predictors, and  $V_E$  is the residual or error variance.

1. In the first special cases considered here (case 1 in Cohen (1988, p. 407ff.)), the proportion of variance explained by the additional predictor set  $B$  is given by  $V_S = R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2$  and the residual variance by  $V_E = 1 - R_{Y \cdot A, B}^2$ . Thus:

$$f^2 = \frac{R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2}{1 - R_{Y \cdot A, B}^2}$$

The quantity  $R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2$  is also called the semipartial multiple correlation and symbolized by  $R_{Y \cdot (B \cdot A)}^2$ . A further interesting quantity is the **partial multiple correlation coefficient**, which is defined as

$$R_{YB \cdot A}^2 := \frac{R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2}{1 - R_{Y \cdot A}^2} = \frac{R_{Y \cdot (B \cdot A)}^2}{1 - R_{Y \cdot A}^2}$$

Using this definition,  $f$  can alternatively be written in terms of the partial  $R^2$ :

$$f^2 = \frac{R_{YB \cdot A}^2}{1 - R_{YB \cdot A}^2}$$

2. In a second special case (case 2 in Cohen (1988, p. 407ff.)), the same effect variance  $V_S$  is considered, but it is assumed that there is a third set of predictors  $C$  that also accounts for parts of the variance of  $Y$  and thus reduces the error variance:  $V_E = 1 - R_{Y \cdot A, B, C}^2$ . In this case, the effect size is

$$f^2 = \frac{R_{Y \cdot A, B}^2 - R_{Y \cdot A}^2}{1 - R_{Y \cdot A, B, C}^2}$$

We may again define a partial  $R_x^2$  as:

$$R_x^2 := \frac{R_{YB \cdot A}^2}{1 - (R_{Y \cdot A, B, C}^2 - R_{YB \cdot A}^2)}$$

and with this quantity we get

$$f^2 = \frac{R_x^2}{1 - R_x^2}$$

Note: Case 1 is the special case of case 2, where  $C$  is the empty set.

Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 18) that may be used to calculate  $f^2$  from the variances  $V_S$  and  $V_E$ , or alternatively from the partial  $R^2$ .

Figure 18: Effect size drawer to calculate  $f^2$  from variances or from the partial  $R^2$ .

(Cohen, 1988, p. 412) defines the following conventional values for the effect size  $f^2$ :

- small  $f^2 = 0.02$
- medium  $f^2 = 0.15$
- large  $f^2 = 0.35$

### 14.2 Options

This test has no options.

### 14.3 Examples

#### 14.3.1 Basic example for case 1

We make the following assumptions: A dependent variable  $Y$  is predicted by two sets of predictors  $A$  and  $B$ . The 5

predictors in  $A$  alone account for 25% of the variation of  $Y$ , thus  $R_{Y \cdot A}^2 = 0.25$ . Including the 4 predictors in set  $B$  increases the proportion of variance explained to 0.3, thus  $R_{Y \cdot A,B}^2 = 0.3$ . We want to calculate the power of a test for the increase due to the inclusion of  $B$ , given  $\alpha = 0.01$  and a total sample size of 90.

First we use the option From variances in the effect size drawer to calculate the effect size. In the input field Variance explained by special effect we insert  $R_{Y \cdot A,B}^2 - R_{Y \cdot A}^2 = 0.3 - 0.25 = 0.05$ , and as Residual variance we insert  $1 - R_{Y \cdot A,B}^2 = 1 - 0.3 = 0.7$ . After clicking on Calculate and transfer to main window we see that this corresponds to a partial  $R^2$  of about 0.0666 and to an effect size  $f = 0.07142857$ . We then set the input field Numerator df in the main window to 4, the number of predictors in set  $B$ , and Number of predictors to the total number of predictors in  $A$  and  $B$ , that is to  $4 + 5 = 9$ .

This leads to the following analysis in G\*POWER :

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Effect size  $f^2$ :* 0.0714286  
 *$\alpha$  err prob:* 0.01  
*Total sample size:* 90  
*Number of tested predictors:* 4  
*Total number of predictors:* 9
- **Output**  
*Noncentrality parameter  $\lambda$ :* 6.428574  
*Critical F:* 3.563110  
*Numerator df:* 4  
*Denominator df:* 80  
*Power (1-  $\beta$ ):* 0.241297

We find that the power of this test is very low, namely about 0.24. This confirms the result estimated by Cohen (1988, p. 434) in his example 9.10, which uses identical values. It should be noted, however, that Cohen (1988) uses an approximation to the correct formula for the noncentrality parameter  $\lambda$  that in general underestimates the true  $\lambda$  and thus also the true power. In this particular case, Cohen estimates  $\lambda = 6.1$ , which is only slightly lower than the correct value 6.429 given in the output above.

By using an a priori analysis, we can compute how large the sample size must be to achieve a power of 0.80. We find that the required sample size is  $N = 242$ .

### 14.3.2 Basis example for case 2

Here we make the following assumptions: A dependent variable  $Y$  is predicted by three sets of predictors  $A$ ,  $B$  and  $C$ , which stand in the following causal relationship  $A \Rightarrow B \Rightarrow C$ . The 5 predictors in  $A$  alone account for 10% of the variation of  $Y$ , thus  $R_{Y \cdot A}^2 = 0.10$ . Including the 3 predictors in set  $B$  increases the proportion of variance explained to 0.16, thus  $R_{Y \cdot A,B}^2 = 0.16$ . Considering in addition the 4 predictors in set  $C$ , increases the explained variance further to 0.2, thus  $R_{Y \cdot A,B,C}^2 = 0.2$ . We want to calculate the power of a test for the increase in variance explained by the inclusion of  $B$  in addition to  $A$ , given  $\alpha = 0.01$  and a total

sample size of 200. This is a case 2 scenario, because the hypothesis only involves sets  $A$  and  $B$ , whereas set  $C$  should be included in the calculation of the residual variance.

We use the option From variances in the effect size drawer to calculate the effect size. In the input field Variance explained by special effect we insert  $R_{Y \cdot A,B}^2 - R_{Y \cdot A}^2 = 0.16 - 0.1 = 0.06$ , and as Residual variance we insert  $1 - R_{Y \cdot A,B,C}^2 = 1 - 0.2 = 0.8$ . Clicking on Calculate and transfer to main window shows that this corresponds to a partial  $R^2 = 0.06976744$  and to an effect size  $f = 0.075$ . We then set the input field Numerator df in the main window to 3, the number of predictors in set  $B$  (which are responsible to a potential increase in variance explained), and Number of predictors to the total number of predictors in  $A$ ,  $B$  and  $C$  (which all influence the residual variance), that is to  $5 + 3 + 4 = 12$ .

This leads to the following analysis in G\*POWER :

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Effect size  $f^2$ :* 0.075  
 *$\alpha$  err prob:* 0.01  
*Total sample size:* 200  
*Number of tested predictors:* 3  
*Total number of predictors:* 12
- **Output**  
*Noncentrality parameter  $\lambda$ :* 15.000000  
*Critical F:* 3.888052  
*Numerator df:* 3  
*Denominator df:* 187  
*Power (1-  $\beta$ ):* 0.766990

We find that the power of this test is about 0.767. In this case the power is slightly larger than the power value 0.74 estimated by Cohen (1988, p. 439) in his example 9.13, which uses identical values. This is due to the fact that his approximation for  $\lambda = 14.3$  underestimates the true value  $\lambda = 15$  given in the output above.

### 14.3.3 Example showing the relation to factorial ANOVA designs

We assume a  $2 \times 3 \times 4$  design with three factors  $U$ ,  $V$ ,  $W$ . We want to test main effects, two-way interactions ( $U \times V$ ,  $U \times W$ ,  $V \times W$ ) and the three-way interaction ( $U \times V \times W$ ). We may use the procedure "ANOVA: Fixed effects, special, main effects and interactions" in G\*POWER to do this analysis (see the corresponding entry in the manual for details). As an example, we consider the test of the  $V \times W$  interaction. Assuming that Variance explained by the interaction = 0.422 and Error variance = 6.75, leads to an effect size  $f = 0.25$  (a mean effect size according to Cohen (1988)). Numerator df = 6 corresponds to (levels of  $V - 1$ )(levels of  $W - 1$ ), and the Number of Groups = 24 is the total number of cells ( $2 \cdot 3 \cdot 4 = 24$ ) in the design. With  $\alpha = 0.05$  and total sample size 120 we compute a power of 0.470.

We now demonstrate, how these analyses can be done with the MRC procedure. A main factor with  $k$  levels corresponds to  $k - 1$  predictors in the MRC analysis, thus

the number of predictors is 1, 2, and 3 for the three factors  $U, V$ , and  $W$ . The number of predictors in interactions is the product of the number of predictors involved. The  $V \times W$  interaction, for instance, corresponds to a set of  $(3 - 1)(4 - 1) = (2)(3) = 6$  predictors.

To test an effect with MRC we need to isolate the relative contribution of this source to the total variance, that is we need to determine  $V_S$ . We illustrate this for the  $V \times W$  interaction. In this case we must find  $R_{Y.V \times W}^2$  by excluding from the proportion of variance that is explained by  $V, W, V \times W$  together, that is  $R_{Y.V,W,V \times W}^2$ , the contribution of main effects, that is:  $R_{Y.V \times W}^2 = R_{Y.V,W,V \times W}^2 - R_{Y.V,W}^2$ . The residual variance  $V_E$  is the variance of  $Y$  from which the variance of all sources in the design have been removed.

This is a case 2 scenario, in which  $V \times W$  corresponds to set  $B$  with  $2 \cdot 3 = 6$  predictors,  $V \cup W$  corresponds to set  $A$  with  $2 + 3 = 5$  predictors, and all other sources of variance, that is  $U, U \times V, U \times W, U \times V \times W$ , correspond to set  $C$  with  $(1 + (1 \cdot 2) + (1 \cdot 3) + (1 \cdot 2 \cdot 3)) = 1 + 2 + 3 + 6 = 12$  predictors. Thus, the total number of predictors is  $(6 + 5 + 12) = 23$ . Note: The total number of predictors is always (number of cells in the design -1).

We now specify these contributions numerically:  $R_{Y.A,B,C} = 0.325$ ,  $R_{Y.A,B} - R_{Y.A}^2 = R_{Y.V \times W}^2 = 0.0422$ . Inserting these values in the effect size dialog (Variance explained by special effect = 0.0422, Residual variance =  $1 - 0.325 = 0.675$ ) yields an effect size  $f^2 = 0.06251852$  (compare these values with those chosen in the ANOVA analysis and note that  $f^2 = 0.25^2 = 0.0625$ ). To calculate power, we set Number of tested predictors = 6 (= number of predictors in set  $B$ ), and Total number of predictors = 23. With  $\alpha = 0.05$  and  $N = 120$ , we get—as expected—the same power 0.470 as from the equivalent ANOVA analysis.

## 14.4 Related tests

Similar tests in G\*POWER 3.0:

- Multiple Regression: Omnibus ( $R^2$  deviation from zero).
- ANOVA: Fixed effects, special, main effect and interactions.

## 14.5 Implementation notes

The  $H_0$ -distribution is the central  $F$  distribution with numerator degrees of freedom  $df_1 = q$ , and denominator degrees of freedom  $df_2 = N - m - 1$ , where  $N$  is the sample size,  $q$  the number of tested predictors in set  $B$ , which is responsible for a potential increase in explained variance, and  $m$  is the total number of predictors. In case 1 (see section “effect size index”),  $m = q + w$ , in case 2,  $m = q + w + v$ , where  $w$  is the number of predictors in set  $A$  and  $v$  the number of predictors in set  $C$ . The  $H_1$ -distribution is the noncentral  $F$  distribution with the same degrees of freedom and noncentrality parameter  $\lambda = f^2 N$ .

## 14.6 Validation

The results were checked against the values produced by GPower 2.0 and those produced by PASS (Hintze, 2006).

Slight deviations were found to the values tabulated in Cohen (1988). This is due to an approximation used by Cohen (1988) that underestimates the noncentrality parameter  $\lambda$  and therefore also the power. This issue is discussed more thoroughly in Erdfelder et al. (1996).

## 14.7 References

See Chapter 9 in Cohen (1988).

## 15 F test: Inequality of two Variances

This procedure allows power analyses for the test that the population variances  $\sigma_0^2$  and  $\sigma_1^2$  of two normally distributed random variables are identical. The null and (two-sided) alternate hypothesis of this test are:

$$\begin{aligned} H_0 &: \sigma_1 - \sigma_0 = 0 \\ H_1 &: \sigma_1 - \sigma_0 \neq 0. \end{aligned}$$

The two-sided test (“two tails”) should be used if there is no a priori restriction on the sign of the deviation assumed in the alternate hypothesis. Otherwise use the one-sided test (“one tail”).

### 15.1 Effect size index

The ratio  $\sigma_1^2/\sigma_0^2$  of the two variances is used as effect size measure. This ratio is 1 if  $H_0$  is true, that is, if both variances are identical. In an a priori analysis a ratio close or even identical to 1 would imply an exceedingly large sample size. Thus, G\*POWER prohibits inputs in the range [0.999, 1.001] in this case.

Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 19) that may be used to calculate the ratio from two variances. Insert the variances  $\sigma_0^2$  and  $\sigma_1^2$  in the corresponding input fields.

variance V0	1.5
variance V1	1
ratio V1/V0	0.6666667

Buttons: Calculate, Calculate and transfer to main window, Close

Figure 19: Effect size drawer to calculate variance ratios.

### 15.2 Options

This test has no options.

### 15.3 Examples

We want to test whether the variance  $\sigma_1^2$  in population B is different from the variance  $\sigma_0^2$  in population A. We here regard a ratio  $\sigma_1^2/\sigma_0^2 > 1.5$  (or, the other way around,  $< 1/1.5 = 0.6666$ ) as a substantial difference.

We want to realize identical sample sizes in both groups. How many subjects are needed to achieve the error levels  $\alpha = 0.05$  and  $\beta = 0.2$  in this test? This question can be answered by using the following settings in G\*POWER :

- **Select**  
Type of power analysis: A priori
- **Input**  
Tail(s): Two  
Ratio var1/var0: 1.5  
 $\alpha$  err prob: 0.05

Power (1-  $\beta$ ): 0.80  
Allocation ratio N2/N1: 1

- **Output**  
Lower critical F: 0.752964  
Upper critical F: 1.328085  
Numerator df: 192  
Denominator df: 192  
Sample size group 1: 193  
Sample size group 2: 193  
Actual power : 0.800105

The output shows that we need at least 386 subjects (193 in each group) in order to achieve the desired level of the  $\alpha$  and  $\beta$  error. To apply the test, we would estimate both variances  $s_1^2$  and  $s_0^2$  from samples of size  $N_1$  and  $N_0$ , respectively. The two-sided test would be significant at  $\alpha = 0.05$  if the statistic  $x = s_1^2/s_0^2$  were either smaller than the lower critical value 0.753 or greater than the upper critical value 1.328.

By setting “Allocation ratio N2/N1 = 2”, we can easily check that a much larger total sample size, namely  $N = 443$  (148 and 295 in group 1 and 2, respectively), would be required if the sample sizes in both groups are clearly different.

### 15.4 Related tests

Similar tests in G\*POWER 3.0:

- Variance: Difference from constant (two sample case).

### 15.5 Implementation notes

It is assumed that both populations are normally distributed and that the means are not known in advance but estimated from samples of size  $N_1$  and  $N_0$ , respectively. Under these assumptions, the  $H_0$ -distribution of  $s_1^2/s_0^2$  is the central  $F$  distribution with  $N_1 - 1$  numerator and  $N_0 - 1$  denominator degrees of freedom ( $F_{N_1-1, N_0-1}$ ), and the  $H_1$ -distribution is the same central  $F$  distribution scaled with the variance ratio, that is,  $(\sigma_1^2/\sigma_0^2) \cdot F_{N_1-1, N_0-1}$ .

### 15.6 Validation

The results were successfully checked against values produced by PASS (Hintze, 2006) and in a Monte Carlo simulation.

## 16 t test: Correlation - point biserial model

The point biserial correlation is a measure of association between a continuous variable  $X$  and a binary variable  $Y$ , the latter of which takes on values 0 and 1. It is assumed that the continuous variables  $X$  at  $Y = 0$  and  $Y = 1$  are normally distributed with means  $\mu_0, \mu_1$  and equal variance  $\sigma$ . If  $\pi$  is the proportion of values with  $Y = 1$  then the point biserial correlation coefficient is defined as:

$$\rho = \frac{(\mu_1 - \mu_0)\sqrt{\pi(1-\pi)}}{\sigma_x}$$

where  $\sigma_x = \sigma + (\mu_1 - \mu_0)^2/4$ .

The point biserial correlation is identical to a Pearson correlation between two vectors  $x, y$ , where  $x_i$  contains a value from  $X$  at  $Y = j$ , and  $y_i = j$  codes the group from which the  $X$  was taken.

The statistical model is the same as that underlying a test for a differences in means  $\mu_0$  and  $\mu_1$  in two independent groups. The relation between the effect size  $d = (\mu_1 - \mu_0)/\sigma$  used in that test and the point biserial correlation  $\rho$  considered here is given by:

$$\rho = \frac{d}{\sqrt{d^2 + \frac{N^2}{n_0 n_1}}}$$

where  $n_0, n_1$  denote the sizes of the two groups and  $N = n_0 + n_1$ .

The power procedure refers to a t-test used to evaluate the null hypothesis that there is no (point-biserial) correlation in the population ( $\rho = 0$ ). The alternative hypothesis is that the correlation coefficient has a non-zero value  $r$ .

$$\begin{aligned} H_0 : \rho &= 0 \\ H_1 : \rho &= r. \end{aligned}$$

The two-sided (“two tailed”) test should be used if there is no restriction on the sign of  $\rho$  under the alternative hypothesis. Otherwise use the one-sided (“one tailed”) test.

### 16.1 Effect size index

The effect size index  $|\rho|$  is the absolute value of the correlation coefficient in the population as postulated in the alternative hypothesis. From this definition it follows that  $0 \leq |\rho| < 1$ .

Cohen (1969, p.79) defines the following effect size conventions for  $|\rho|$ :

- **small**  $\rho = 0.1$
- **medium**  $\rho = 0.3$
- **large**  $\rho = 0.5$

Pressing the Determine button on the left side of the effect size label opens the effect size drawer (see Fig. 20). You can use it to calculate  $|\rho|$  from the coefficient of determination  $\rho^2$ .

### 16.2 Options

This test has no options.

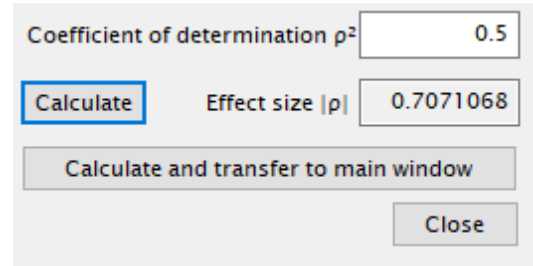


Figure 20: Effect size dialog to compute  $\rho$  from the coefficient of determination  $\rho^2$ .

### 16.3 Examples

We want to know how many subjects it takes to detect  $r = .25$  in the population, given  $\alpha = \beta = .05$ . Thus,  $H_0: \rho = 0$ ,  $H_1: \rho = 0.25$ .

- **Select**  
*Type of power analysis:* A priori
- **Input**  
*Tail(s):* One  
*Effect size  $|\rho|$ :* 0.25  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95
- **Output**  
*noncentrality parameter  $\delta$ :* 3.306559  
*Critical t:* 1.654314  
*df:* 162  
*Total sample size:* 164  
*Actual power:* 0.950308

The results indicate that we need at least  $N = 164$  subjects to ensure a power  $> 0.95$ . The actual power achieved with this  $N$  (0.950308) is slightly higher than the requested power.

To illustrate the connection to the two groups t test, we calculate the corresponding effect size  $d$  for equal sample sizes  $n_0 = n_1 = 82$ :

$$d = \frac{Nr}{\sqrt{n_0 n_1 (1 - r^2)}} = \frac{164 \cdot 0.25}{\sqrt{82 \cdot 82 \cdot (1 - 0.25^2)}} = 0.51639778$$

Performing a power analysis for the one-tailed two group t test with this  $d$ ,  $n_0 = n_1 = 82$ , and  $\alpha = 0.05$  leads to exactly the same power 0.930308 as in the example above. If we assume unequal sample sizes in both groups, for example  $n_0 = 64, n_1 = 100$ , then we would compute a different value for  $d$ :

$$d = \frac{Nr}{\sqrt{n_0 n_1 (1 - r^2)}} = \frac{164 \cdot 0.25}{\sqrt{100 \cdot 64 \cdot (1 - 0.25^2)}} = 0.52930772$$

but we would again arrive at the same power. It thus poses no restriction of generality that we only input the total sample size and not the individual group sizes in the t test for correlation procedure.

### 16.4 Related tests

Similar tests in G\*POWER 3.0:

- Exact test for the difference of one (Pearson) correlation from a constant
- Test for the difference of two (Pearson) correlations



## 16.5 Implementation notes

The  $H_0$ -distribution is the central  $t$ -distribution with  $df = N - 2$  degrees of freedom. The  $H_1$ -distribution is the non-central  $t$ -distribution with  $df = N - 2$  and noncentrality parameter  $\delta$  where

$$\delta = \sqrt{\frac{|\rho|^2 N}{1.0 - |\rho|^2}}$$

$N$  represents the total sample size and  $|\rho|$  represents the effect size index as defined above.

## 16.6 Validation

The results were checked against the values produced by GPower 2.0 [Faul and Erdfelder \(1992\)](#).

## 17 t test: Linear Regression (two groups)

A linear regression is used to estimate the parameters  $a, b$  of a linear relationship  $Y = a + bX$  between the dependent variable  $Y$  and the independent variable  $X$ .  $X$  is assumed to be a set of fixed values, whereas  $Y_i$  is modeled as a random variable:  $Y_i = a + bX_i + \varepsilon_i$ , where  $\varepsilon_i$  denotes normally distributed random errors with mean 0 and standard deviation  $\sigma_i$ . A common assumption also adopted here is that all  $\sigma_i$ 's are identical, that is  $\sigma_i = \sigma$ . The standard deviation of the error is also called the standard deviation of the residuals.

If we have determined the linear relationships between  $X$  and  $Y$  in two groups:  $Y_1 = a_1 + b_1X_1$ ,  $Y_2 = a_2 + b_2X_2$ , we may ask whether the slopes  $b_1, b_2$  are identical.

The null and the two-sided alternative hypotheses are

$$\begin{aligned} H_0: & b_1 - b_2 = 0 \\ H_1: & b_1 - b_2 \neq 0. \end{aligned}$$

### 17.1 Effect size index

The absolute value of the difference between the slopes  $|\Delta \text{slope}| = |b_1 - b_2|$  is used as effect size. To fully specify the effect size, the following additional inputs must be given:

- Std dev residual  $\sigma$

The standard deviation  $\sigma > 0$  of the residuals in the combined data set (i.e. the square root of the weighted sum of the residual variances in the two data sets): If  $\sigma_{r_1}^2$  and  $\sigma_{r_2}^2$  denote the variance of the residuals  $r_1 = (a_1 + b_1X_1) - Y_1$  and  $r_2 = (a_2 + b_2X_2) - Y_2$  in the two groups, and  $n_1, n_2$  the respective sample sizes, then

$$\sigma = \sqrt{\frac{n_1\sigma_{r_1}^2 + n_2\sigma_{r_2}^2}{n_1 + n_2}} \quad (3)$$

- Std dev  $\sigma_{x1}$

The standard deviation  $\sigma_{x_1} > 0$  of the  $X$ -values in group 1.

- Std dev  $\sigma_{x2}$

The standard deviation  $\sigma_{x_2} > 0$  of the  $X$ -values in group 2.

Important relationships between the standard deviations  $\sigma_{x_i}$  of  $X_i$ ,  $\sigma_{y_i}$  of  $Y_i$ , the slopes  $b_i$  of the regression lines, and the correlation coefficient  $\rho_i$  between  $X_i$  and  $Y_i$  are:

$$\sigma_{y_i} = (b_i\sigma_{x_i})/\rho_i \quad (4)$$

$$\sigma_{y_i} = \sigma_{r_i}/\sqrt{1 - \rho_i^2} \quad (5)$$

where  $\sigma_i$  denotes the standard deviation of the residuals  $Y_i - (b_iX + a_i)$ .

The **effect size dialog** may be used to determine Std dev residual  $\sigma$  and  $|\Delta \text{slope}|$  from other values based on Eqns (7), (8) and (9) given above. Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 23).

The left panel in Fig 23 shows the combinations of input and output values in different input modes. The input variables stand on the left side of the arrow '=>', the

output variables on the right side. The input values must conform to the usual restrictions, that is,  $\sigma_{x_i} > 0, \sigma_{y_i} > 0, -1 < \rho_i < 1$ . In addition, Eq (8) together with the restriction on  $\rho_i$  implies the additional restriction  $-1 < b \cdot \sigma_{x_i}/\sigma_{y_i} < 1$ .

Clicking on the button Calculate and transfer to main window copies the values given in Std dev  $\sigma_{x1}$ , Std dev  $\sigma_{x2}$ , Std dev residual  $\sigma$ , Allocation ration  $N2/N1$ , and  $|\Delta \text{slope}|$  to the corresponding input fields in the main window.

### 17.2 Options

This test has no options.

### 17.3 Examples

We replicate an example given on page 594 in Dupont and Plummer (1998) that refers to an example in Armitage, Berry, and Matthews (2002, p. 325). The data and relevant statistics are shown in Fig. (24). Note: Contrary to Dupont and Plummer (1998), we here consider the data as hypothesized true values and normalize the variance by  $N$  not  $(N - 1)$ .

The relation of age and vital capacity for two groups of men working in the cadmium industry is investigated. Group 1 includes  $n_1 = 28$  worker with less than 10 years of cadmium exposure, and Group 2  $n_2 = 44$  workers never exposed to cadmium. The standard deviation of the ages in both groups are  $\sigma_{x1} = 9.029$  and  $\sigma_{x2} = 11.87$ . Regressing vital capacity on age gives the following slopes of the regression lines  $\beta_1 = -0.04653$  and  $\beta_2 = -0.03061$ . To calculate the pooled standard deviation of the residuals we use the effect size dialog: We use input mode " $\sigma_x, \sigma_y, \text{slope} \Rightarrow \text{residual } \sigma, \rho$ " and insert the values given above, the standard deviations  $\sigma_{y1}, \sigma_{y2}$  of  $y$  (capacity) as given in Fig. 24, and the allocation ratio  $n_2/n_1 = 44/28 = 1.571428$ . This results in an pooled standard deviation of the residuals of  $\sigma = 0.5578413$  (compare the right panel in Fig. 23).

We want to recruit enough workers to detect a true difference in slope of  $|(-0.03061) - (-0.04653)| = 0.01592$  with 80% power,  $\alpha = 0.05$  and the same allocation ratio to the two groups as in the sample data.

- **Select**

Type of power analysis: A priori

- **Input**

Tail(s): Two

$|\Delta \text{slope}|$ : 0.01592

$\alpha \text{ err prob}$ : 0.05

Power (1-  $\beta$ ): 0.80

Allocation ratio  $N2/N1$ : 1.571428

Std dev residual  $\sigma$ : 0.5578413

Std dev  $\sigma_{x1}$ : 9.02914

Std dev  $\sigma_{x2}$ : 11.86779

- **Output**

Noncentrality parameter  $\delta$ : 2.811598

Critical  $t$ : 1.965697

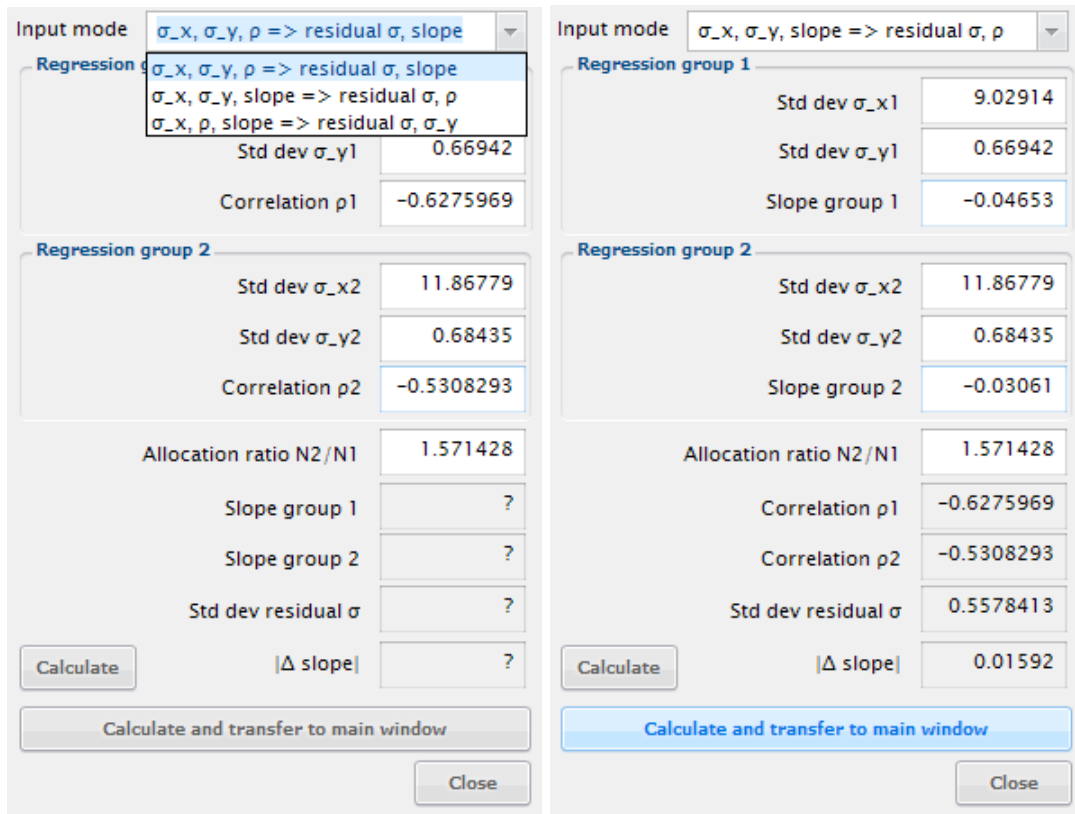
Df: 415

Sample size group 1: 163

Sample size group 2: 256

Total sample size: 419

Actual power: 0.800980



**Figure 21:** Effect size drawer to calculate the pooled standard deviation of the residuals  $\sigma$  and the effect size  $|\Delta \text{slope}|$  from various inputs constellations (see left panel). The right panel shows the inputs for the example discussed below.

Group 1: exposure				Group 2: no exposure			
age	capacity	age	capacity	age	capacity	age	capacity
29	5.21	50	3.50	27	5.29	41	3.77
29	5.17	45	5.06	25	3.67	41	4.22
33	4.88	48	4.06	24	5.82	37	4.94
32	4.50	51	4.51	32	4.77	42	4.04
31	4.47	46	4.66	23	5.71	39	4.51
29	5.12	58	2.88	25	4.47	41	4.06
29	4.51			32	4.55	43	4.02
30	4.85			18	4.61	41	4.99
21	5.22			19	5.86	48	3.86
28	4.62			26	5.20	47	4.68
23	5.07			33	4.44	53	4.74
35	3.64			27	5.52	49	3.76
38	3.64			33	4.97	54	3.98
38	5.09			25	4.99	48	5.00
43	4.61			42	4.89	49	3.31
39	4.73			35	4.09	47	3.11
38	4.58			35	4.24	52	4.76
42	5.12			41	3.88	58	3.95
43	3.89			38	4.85	62	4.60
43	4.62			41	4.79	65	4.83
37	4.30			36	4.36	62	3.18
50	2.70			36	4.02	59	3.03
n1 =	28	n2 =	44				
sx1 =	9.029147	sx2 =	11.867794				
sy1 =	0.669424	sy2 =	0.684350				
b1 =	-0.046532	b2 =	-0.030613				
a1 =	6.230031	a2 =	5.680291				
r1 =	-0.627620	r2 =	-0.530876				

**Figure 22:** Data for the example discussed Dupont and Plumer

The output shows that we need 419 workers in total, with 163 in group 1 and 256 in group 2. These values are close to those reported in [Dupont and Plummer \(1998, p. 596\)](#) for this example ( $166 + 261 = 427$ ). The slight difference is due to the fact that they normalize the variances by  $N - 1$ , and use shifted central  $t$ -distributions instead of non-central  $t$  distributions.

### 17.3.1 Relation to Multiple Regression: Special

The present procedure is essentially a special case of Multiple regression, but provides a more convenient interface. To show this, we demonstrate how the MRC procedure can be used to compute the example above (see also [Dupont and Plummer \(1998, p.597\)](#)).

First, the data are combined into a data set of size  $n_1 + n_2 = 28 + 44 = 72$ . With respect to this combined data set, we define the following variables (vectors of length 72):

- $y$  contains the measured vital capacity
- $x_1$  contains the age data
- $x_2$  codes group membership (0 = not exposed, 1 = exposed)
- $x_3$  contains the element-wise product of  $x_1$  and  $x_2$

The multiple regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i$$

reduces to  $y = \beta_0 + \beta_1 x_1$ , and  $y = (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3) x_1$ , for unexposed and exposed workers, respectively. In this model  $\beta_3$  represents the difference in slope between both groups, which is assumed to be zero under the null hypothesis. Thus, the above model reduces to

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

if the null hypothesis is true.

Performing a multiple regression analysis with the full model leads to  $\beta_1 = -0.01592$  and  $R_1^2 = 0.3243$ . With the reduced model assumed in the null hypothesis one finds  $R_0^2 = 0.3115$ . From these values we compute the following effect size:

$$f^2 = \frac{R_1^2 - R_0^2}{1 - R_1^2} = \frac{0.3243 - 0.3115}{1 - 0.3243} = 0.018870$$

Selecting an a priori analysis and setting  $\alpha$  err prob to 0.05, Power ( $1 - \beta$  err prob) to 0.80, Numerator df to 1 and Number of predictors to 3, we get  $N = 418$ , that is almost the same result as in the example above.

## 17.4 Related tests

Similar tests in G\*POWER 3.0:

- Multiple Regression: Omnibus ( $R^2$  deviation from zero).
- Correlation: Point biserial model

## 17.5 Implementation notes

The procedure implements a slight variant of the algorithm proposed in [Dupont and Plummer \(1998\)](#). The only difference is, that we replaced the approximation by shifted central  $t$ -distributions used in their paper with noncentral  $t$ -distributions. In most cases this makes no great difference.

The  $H_0$ -distribution is the central  $t$  distribution with  $df = n_1 + n_2 - 4$  degrees of freedom, where  $n_1$  and  $n_2$  denote the sample size in the two groups, and the  $H_1$  distribution is the non-central  $t$  distribution with the same degrees of freedom and the noncentrality parameter  $\delta = \Delta\sqrt{n_2}$ .

**Statistical test.** The power is calculated for the  $t$ -test for equal slopes as described in [Armitage et al. \(2002\)](#) in chapter 11. The test statistic is (see Eqn 11.18, 11.19, 11.20):

$$t = \frac{\hat{b}_1 - \hat{b}_2}{s_{\hat{b}_1 - \hat{b}_2}}$$

with  $df = n_1 + n_2 - 4$  degrees of freedom.

Let for group  $i \in \{1, 2\}$ ,  $S_{xi}$ ,  $S_{yi}$ , denote the sum of squares in  $X$  and  $Y$  (i.e. the variance times  $n_i$ ). Then a pooled estimate of the residual variance be obtained by  $s_r^2 = (S_{y1} + S_{y2}) / (n_1 + n_2 - 4)$ . The standard error of the difference of slopes is

$$s_{\hat{b}_1 - \hat{b}_2} = \sqrt{s_r^2 \left( \frac{1}{S_{x1}} + \frac{1}{S_{x2}} \right)}$$

**Power of the test:** In the procedure for **equal slopes** the noncentrality parameter is  $\delta = \Delta\sqrt{n}$ , with  $\Delta = |\Delta\text{slope}| / \sigma_R$  and

$$\sigma_R = \sqrt{\sigma^2 \left( 1 + \frac{1}{m\sigma_{x1}^2} + \frac{1}{\sigma_{x2}^2} \right)} \quad (6)$$

where  $m = n_1/n_2$ ,  $\sigma_{x1}$  and  $\sigma_{x2}$  are the standard deviations of  $X$  in group 1 and 2, respectively, and  $\sigma$  the common standard deviation of the residuals.

## 17.6 Validation

The results were checked for a range of input scenarios against the values produced by the program PS published by [Dupont and Plummer \(1998\)](#). Only slight deviations were found that are probably due to the use of the non-central  $t$ -distribution in G\*POWER instead of the shifted central  $t$ -distributions that are used in PS.

## 18 t test: Linear Regression (two groups)

A linear regression is used to estimate the parameters  $a, b$  of a linear relationship  $Y = a + bX$  between the dependent variable  $Y$  and the independent variable  $X$ .  $X$  is assumed to be a set of fixed values, whereas  $Y_i$  is modeled as a random variable:  $Y_i = a + bX_i + \varepsilon_i$ , where  $\varepsilon_i$  denotes normally distributed random errors with mean 0 and standard deviation  $\sigma_i$ . A common assumption also adopted here is that all  $\sigma_i$ 's are identical, that is  $\sigma_i = \sigma$ . The standard deviation of the error is also called the standard deviation of the residuals.

If we have determined the linear relationships between  $X$  and  $Y$  in two groups:  $Y_1 = a_1 + b_1X_1$ ,  $Y_2 = a_2 + b_2X_2$ , we may ask whether the slopes  $b_1, b_2$  are identical.

The null and the two-sided alternative hypotheses are

$$H_0 : b_1 - b_2 = 0$$

$$H_1 : b_1 - b_2 \neq 0.$$

### 18.1 Effect size index

The absolute value of the difference between the slopes  $|\Delta \text{slope}| = |b_1 - b_2|$  is used as effect size. To fully specify the effect size, the following additional inputs must be given:

- Std dev residual  $\sigma$

The standard deviation  $\sigma > 0$  of the residuals in the combined data set (i.e. the square root of the weighted sum of the residual variances in the two data sets): If  $\sigma_{r_1}^2$  and  $\sigma_{r_2}^2$  denote the variance of the residuals  $r_1 = (a_1 + b_1X_1) - Y_1$  and  $r_2 = (a_2 + b_2X_2) - Y_2$  in the two groups, and  $n_1, n_2$  the respective sample sizes, then

$$\sigma = \sqrt{\frac{n_1\sigma_{r_1}^2 + n_2\sigma_{r_2}^2}{n_1 + n_2}} \quad (7)$$

- Std dev  $\sigma_{x1}$

The standard deviation  $\sigma_{x_1} > 0$  of the  $X$ -values in group 1.

- Std dev  $\sigma_{x2}$

The standard deviation  $\sigma_{x_2} > 0$  of the  $X$ -values in group 2.

Important relationships between the standard deviations  $\sigma_{x_i}$  of  $X_i$ ,  $\sigma_{y_i}$  of  $Y_i$ , the slopes  $b_i$  of the regression lines, and the correlation coefficient  $\rho_i$  between  $X_i$  and  $Y_i$  are:

$$\sigma_{y_i} = (b_i\sigma_{x_i})/\rho_i \quad (8)$$

$$\sigma_{y_i} = \sigma_{r_i}/\sqrt{1 - \rho_i^2} \quad (9)$$

where  $\sigma_i$  denotes the standard deviation of the residuals  $Y_i - (b_iX + a_i)$ .

The **effect size dialog** may be used to determine Std dev residual  $\sigma$  and  $|\Delta \text{slope}|$  from other values based on Eqns (7), (8) and (9) given above. Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 23).

The left panel in Fig 23 shows the combinations of input and output values in different input modes. The input variables stand on the left side of the arrow ' $\Rightarrow$ ', the

output variables on the right side. The input values must conform to the usual restrictions, that is,  $\sigma_{x_i} > 0, \sigma_{y_i} > 0, -1 < \rho_i < 1$ . In addition, Eq (8) together with the restriction on  $\rho_i$  implies the additional restriction  $-1 < b \cdot \sigma_{x_i}/\sigma_{y_i} < 1$ .

Clicking on the button Calculate and transfer to main window copies the values given in Std dev  $\sigma_{x1}$ , Std dev  $\sigma_{x2}$ , Std dev residual  $\sigma$ , Allocation ration  $N2/N1$ , and  $|\Delta \text{slope}|$  to the corresponding input fields in the main window.

### 18.2 Options

This test has no options.

### 18.3 Examples

We replicate an example given on page 594 in Dupont and Plummer (1998) that refers to an example in Armitage et al. (2002, p. 325). The data and relevant statistics are shown in Fig. (24). Note: Contrary to Dupont and Plummer (1998), we here consider the data as hypothesized true values and normalize the variance by  $N$  not  $(N - 1)$ .

The relation of age and vital capacity for two groups of men working in the cadmium industry is investigated. Group 1 includes  $n_1 = 28$  worker with less than 10 years of cadmium exposure, and Group 2  $n_2 = 44$  workers never exposed to cadmium. The standard deviation of the ages in both groups are  $\sigma_{x1} = 9.029$  and  $\sigma_{x2} = 11.87$ . Regressing vital capacity on age gives the following slopes of the regression lines  $\beta_1 = -0.04653$  and  $\beta_2 = -0.03061$ . To calculate the pooled standard deviation of the residuals we use the effect size dialog: We use input mode " $\sigma_x, \sigma_y, \text{slope} \Rightarrow \text{residual } \sigma, \rho$ " and insert the values given above, the standard deviations  $\sigma_{y1}, \sigma_{y2}$  of  $y$  (capacity) as given in Fig. 24, and the allocation ratio  $n_2/n_1 = 44/28 = 1.571428$ . This results in an pooled standard deviation of the residuals of  $\sigma = 0.5578413$  (compare the right panel in Fig. 23).

We want to recruit enough workers to detect a true difference in slope of  $|(-0.03061) - (-0.04653)| = 0.01592$  with 80% power,  $\alpha = 0.05$  and the same allocation ratio to the two groups as in the sample data.

- **Select**

*Type of power analysis:* A priori

- **Input**

*Tail(s):* Two

*$|\Delta \text{slope}|$ :* 0.01592

*$\alpha \text{ err prob}$ :* 0.05

*Power (1-  $\beta$ ):* 0.80

*Allocation ratio  $N2/N1$ :* 1.571428

*Std dev residual  $\sigma$ :* 0.5578413

*Std dev  $\sigma_{x1}$ :* 9.02914

*Std dev  $\sigma_{x2}$ :* 11.86779

- **Output**

*Noncentrality parameter  $\delta$ :* 2.811598

*Critical  $t$ :* 1.965697

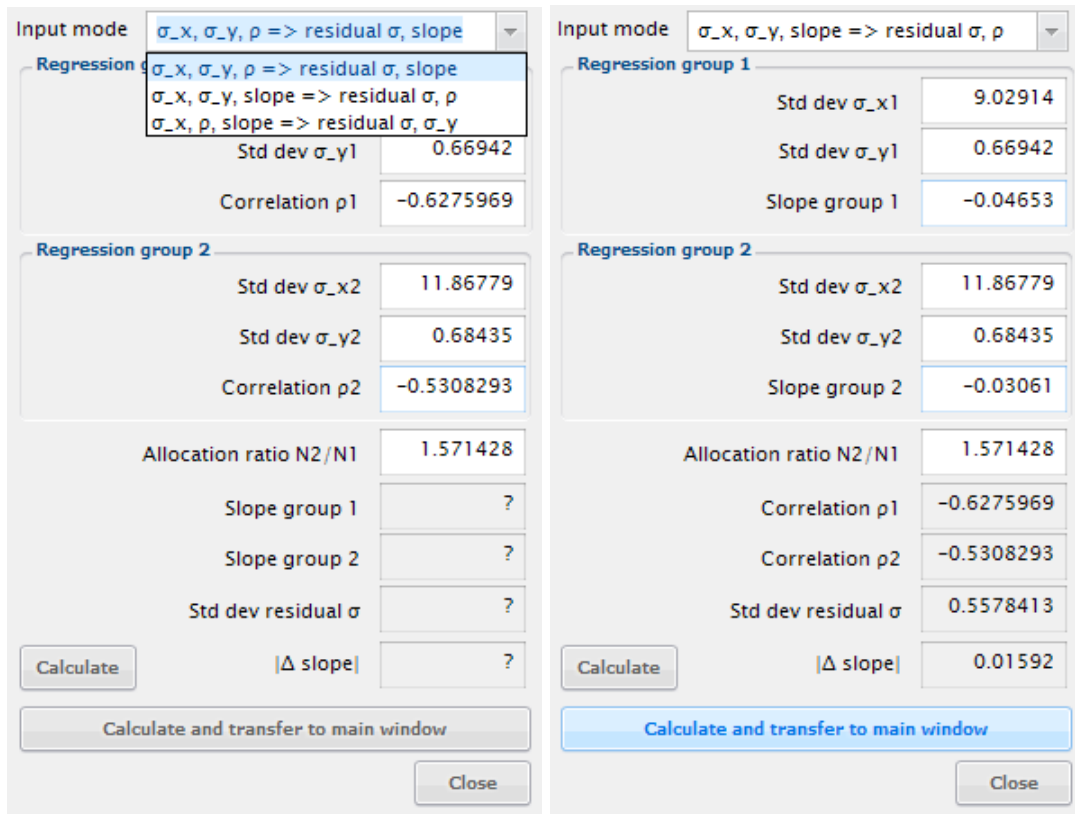
*Df:* 415

*Sample size group 1:* 163

*Sample size group 2:* 256

*Total sample size:* 419

*Actual power:* 0.800980



**Figure 23:** Effect size drawer to calculate the pooled standard deviation of the residuals  $\sigma$  and the effect size  $|\Delta \text{slope}|$  from various inputs constellations (see left panel). The right panel shows the inputs for the example discussed below.

Group 1: exposure				Group 2: no exposure			
age	capacity	age	capacity	age	capacity	age	capacity
29	5.21	50	3.50	27	5.29	41	3.77
29	5.17	45	5.06	25	3.67	41	4.22
33	4.88	48	4.06	24	5.82	37	4.94
32	4.50	51	4.51	32	4.77	42	4.04
31	4.47	46	4.66	23	5.71	39	4.51
29	5.12	58	2.88	25	4.47	41	4.06
29	4.51			32	4.55	43	4.02
30	4.85			18	4.61	41	4.99
21	5.22			19	5.86	48	3.86
28	4.62			26	5.20	47	4.68
23	5.07			33	4.44	53	4.74
35	3.64			27	5.52	49	3.76
38	3.64			33	4.97	54	3.98
38	5.09			25	4.99	48	5.00
43	4.61			42	4.89	49	3.31
39	4.73			35	4.09	47	3.11
38	4.58			35	4.24	52	4.76
42	5.12			41	3.88	58	3.95
43	3.89			38	4.85	62	4.60
43	4.62			41	4.79	65	4.83
37	4.30			36	4.36	62	3.18
50	2.70			36	4.02	59	3.03
n1 =		28		n2 =		44	
sx1 =	9.029147			sx2 =	11.867794		
sy1 =	0.669424			sy2 =	0.684350		
b1 =	-0.046532			b2 =	-0.030613		
a1 =	6.230031			a2 =	5.680291		
r1 =	-0.627620			r2 =	-0.530876		

**Figure 24:** Data for the example discussed Dupont and Plumer

The output shows that we need 419 workers in total, with 163 in group 1 and 256 in group 2. These values are close to those reported in Dupont and Plummer (1998, p. 596) for this example (166 + 261 = 427). The slight difference is due to the fact that they normalize the variances by  $N - 1$ , and use shifted central  $t$ -distributions instead of non-central  $t$  distributions.

### 18.3.1 Relation to Multiple Regression: Special

The present procedure is essentially a special case of Multiple regression, but provides a more convenient interface. To show this, we demonstrate how the MRC procedure can be used to compute the example above (see also Dupont and Plummer (1998, p.597)).

First, the data are combined into a data set of size  $n_1 + n_2 = 28 + 44 = 72$ . With respect to this combined data set, we define the following variables (vectors of length 72):

- $y$  contains the measured vital capacity
- $x_1$  contains the age data
- $x_2$  codes group membership (0 = not exposed, 1 = exposed)
- $x_3$  contains the element-wise product of  $x_1$  and  $x_2$

The multiple regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i$$

reduces to  $y = \beta_0 + \beta_1 x_1$ , and  $y = (\beta_0 + \beta_2 x_2) + (\beta_1 + \beta_3) x_1$ , for unexposed and exposed workers, respectively. In this model  $\beta_3$  represents the difference in slope between both groups, which is assumed to be zero under the null hypothesis. Thus, the above model reduces to

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

if the null hypothesis is true.

Performing a multiple regression analysis with the full model leads to  $\beta_1 = -0.01592$  and  $R_1^2 = 0.3243$ . With the reduced model assumed in the null hypothesis one finds  $R_0^2 = 0.3115$ . From these values we compute the following effect size:

$$f^2 = \frac{R_1^2 - R_0^2}{1 - R_1^2} = \frac{0.3243 - 0.3115}{1 - 0.3243} = 0.018870$$

Selecting an a priori analysis and setting  $\alpha$  err prob to 0.05, Power (1- $\beta$  err prob) to 0.80, Numerator df to 1 and Number of predictors to 3, we get  $N = 418$ , that is almost the same result as in the example above.

## 18.4 Related tests

Similar tests in G\*POWER 3.0:

- Multiple Regression: Omnibus ( $R^2$  deviation from zero).
- Correlation: Point biserial model

## 18.5 Implementation notes

The procedure implements a slight variant of the algorithm proposed in Dupont and Plummer (1998). The only difference is, that we replaced the approximation by shifted central  $t$ -distributions used in their paper with noncentral  $t$ -distributions. In most cases this makes no great difference.

The  $H_0$ -distribution is the central  $t$  distribution with  $df = n_1 + n_2 - 4$  degrees of freedom, where  $n_1$  and  $n_2$  denote the sample size in the two groups, and the  $H_1$  distribution is the non-central  $t$  distribution with the same degrees of freedom and the noncentrality parameter  $\delta = \Delta\sqrt{n_2}$ .

**Statistical test.** The power is calculated for the  $t$ -test for equal slopes as described in Armitage et al. (2002) in chapter 11. The test statistic is (see Eqn 11.18, 11.19, 11.20):

$$t = \frac{\hat{b}_1 - \hat{b}_2}{s_{\hat{b}_1 - \hat{b}_2}}$$

with  $df = n_1 + n_2 - 4$  degrees of freedom.

Let for group  $i \in \{1, 2\}$ ,  $S_{xi}$ ,  $S_{yi}$ , denote the sum of squares in  $X$  and  $Y$  (i.e. the variance times  $n_i$ ). Then a pooled estimate of the residual variance be obtained by  $s_r^2 = (S_{y1} + S_{y2}) / (n_1 + n_2 - 4)$ . The standard error of the difference of slopes is

$$s_{\hat{b}_1 - \hat{b}_2} = \sqrt{s_r^2 \left( \frac{1}{S_{x1}} + \frac{1}{S_{x2}} \right)}$$

**Power of the test:** In the procedure for **equal slopes** the noncentrality parameter is  $\delta = \Delta\sqrt{n}$ , with  $\Delta = |\Delta\text{slope}| / \sigma_R$  and

$$\sigma_R = \sqrt{\sigma^2 \left( 1 + \frac{1}{m\sigma_{x1}^2} + \frac{1}{\sigma_{x2}^2} \right)} \quad (10)$$

where  $m = n_1/n_2$ ,  $\sigma_{x1}$  and  $\sigma_{x2}$  are the standard deviations of  $X$  in group 1 and 2, respectively, and  $\sigma$  the common standard deviation of the residuals.

## 18.6 Validation

The results were checked for a range of input scenarios against the values produced by the program PS published by Dupont and Plummer (1998). Only slight deviations were found that are probably due to the use of the non-central  $t$ -distribution in G\*POWER instead of the shifted central  $t$ -distributions that are used in PS.

## 19 t test: Means - difference between two dependent means (matched pairs)

The null hypothesis of this test is that the population means  $\mu_x, \mu_y$  of two matched samples  $x, y$  are identical. The sampling method leads to  $N$  pairs  $(x_i, y_i)$  of matched observations.

The null hypothesis that  $\mu_x = \mu_y$  can be reformulated in terms of the difference  $z_i = x_i - y_i$ . The null hypothesis is then given by  $\mu_z = 0$ . The alternative hypothesis states that  $\mu_z$  has a value different from zero:

$$\begin{aligned} H_0 : \mu_z &= 0 \\ H_1 : \mu_z &\neq 0. \end{aligned}$$

If the sign of  $\mu_z$  cannot be predicted a priori then a two-sided test should be used. Otherwise use the one-sided test.

### 19.1 Effect size index

The effect size index  $d_z$  is defined as:

$$d_z = \frac{|\mu_z|}{\sigma_z} = \frac{|\mu_x - \mu_y|}{\sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho_{xy} \sigma_x \sigma_y}}$$

where  $\mu_x, \mu_y$  denote the population means,  $\sigma_x$  and  $\sigma_y$  denote the standard deviation in either population, and  $\rho_{xy}$  denotes the correlation between the two random variables.  $\mu_z$  and  $\sigma_z$  are the population mean and standard deviation of the difference  $z$ .

A click on the Determine button to the left of the effect size label in the main window opens the effect size drawer (see Fig. 25). You can use this drawer to calculate  $d_z$  from the mean and standard deviation of the differences  $z_i$ . Alternatively, you can calculate  $d_z$  from the means  $\mu_x$  and  $\mu_y$  and the standard deviations  $\sigma_x$  and  $\sigma_y$  of the two random variables  $x, y$  as well as the correlation between the random variables  $x$  and  $y$ .

### 19.2 Options

This test has no options.

### 19.3 Examples

Let us try to replicate the example in Cohen (1969, p. 48). The effect of two teaching methods on algebra achievements are compared between 50 IQ matched pairs of pupils (i.e. 100 pupils). The effect size that should be detected is  $d = (m_0 - m_1)/\sigma = 0.4$ . Note that this is the effect size index representing differences between two independent means (two groups). We want to use this effect size as a basis for a matched-pairs study. A sample estimate of the correlation between IQ-matched pairs in the population has been calculated to be  $r = 0.55$ . We thus assume  $\rho_{xy} = 0.55$ . What is the power of a two-sided test at an  $\alpha$  level of 0.05?

To compute the effect size  $d_z$  we open the effect size drawer and choose "from group parameters". We only know the ratio  $d = (\mu_x - \mu_y)\Delta m/\sigma = 0.4$ . We are thus free to choose any values for the means and (equal) standard deviations that lead to this ratio. We set "Mean group 1 = 0",

Figure 25: Effect size dialog to calculate effect size  $d$  from the parameters of two correlated random variables

"Mean group 2 = 0.4", "SD group 1 = 1", and "SD group 2 = 1". Finally, we set the "Correlation between groups" to be 0.55. Pressing the "Calculate and transfer to main window" button copies the resulting effect size  $d_z = 0.421637$  to the main window. We supply the remaining input values in the main window and press "Calculate"

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* Two  
*Effect size dz:* 0.421637  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 50
- **Output**  
*Noncentrality parameter  $\delta$ :* 2.981424  
*Critical t:* 2.009575  
*df:* 49  
*Power (1-  $\beta$ ):* 0.832114

The computed power of 0.832114 is close to the value 0.84 estimated by Cohen using his tables. To estimate the increase in power due to the correlation between pairs (i.e., due to the shifting from a two-group design to a matched-pairs design), we enter "Correlation between groups = 0" in the effect size drawer. This leads to  $d_z = 0.2828427$ . Repeating the above analysis with this effect size leads to a power of only 0.500352.

How many subjects would we need to arrive at a power of about 0.832114 in a two-group design? We click X-Y plot for a range of values to open the Power Plot window. Let us plot (on the y axis) the power (with markers and displaying the values in the plot) as a function of the total sample size. We want to plot just 1 graph with the err prob set to 0.05 and effect size  $d_z$  fixed at 0.2828427.



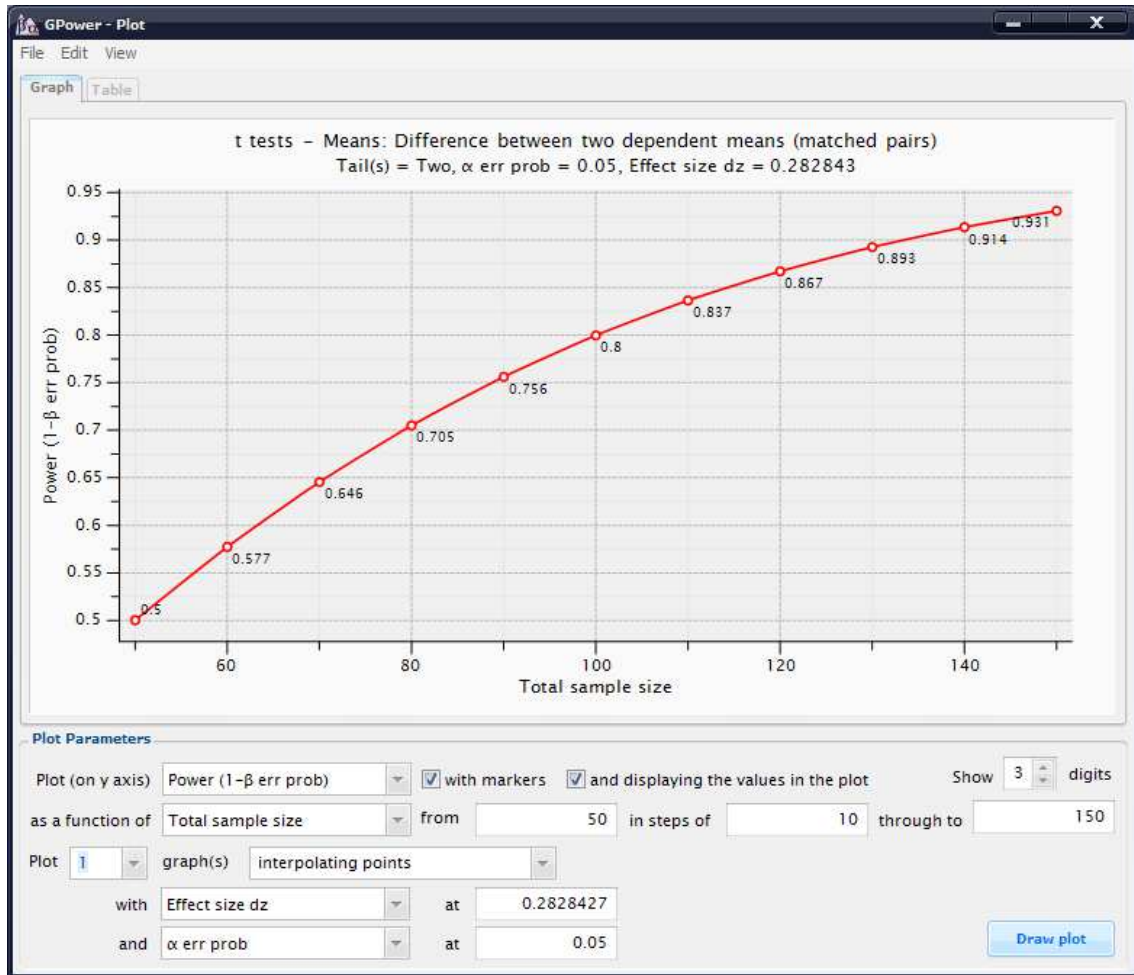


Figure 26: Power vs. sample size plot for the example.

Clicking Draw plot yields a graph in which we can see that we would need about 110 pairs, that is, 220 subjects (see Fig 26 on page 49). Thus, by matching pupils we cut down the required size of the sample by more than 50%.

## 19.4 Related tests

## 19.5 Implementation notes

The  $H_0$  distribution is the central Student  $t$  distribution  $t(N-1, 0)$ , the  $H_1$  distribution is the noncentral Student  $t$  distribution  $t(N-1, \delta)$ , with noncentrality  $\delta = d\sqrt{N}$ .

## 19.6 Validation

The results were checked against the values produced by GPower 2.0.

## 20 t test: Means - difference from constant (one sample case)

The one-sample t test is used to determine whether the population mean  $\mu$  equals some specified value  $\mu_0$ . The data are from a random sample of size  $N$  drawn from a normally distributed population. The true standard deviation in the population is unknown and must be estimated from the data. The null and alternate hypothesis of the t test state:

$$\begin{aligned} H_0 : \mu - \mu_0 &= 0 \\ H_1 : \mu - \mu_0 &\neq 0. \end{aligned}$$

The two-sided (“two tailed”) test should be used if there is no restriction on the sign of the deviation from  $\mu_0$  assumed in the alternate hypothesis. Otherwise use the one-sided (“one tailed”) test .

### 20.1 Effect size index

The effect size index  $d$  is defined as:

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

where  $\sigma$  denotes the (unknown) standard deviation in the population. Thus, if  $\mu$  and  $\mu_0$  deviate by one standard deviation then  $d = 1$ .

Cohen (1969, p. 38) defines the following conventional values for  $d$ :

- **small**  $d = 0.2$
- **medium**  $d = 0.5$
- **large**  $d = 0.8$

Pressing the Determine button on the left side of the effect size label opens the effect size drawer (see Fig. 27). You can use this dialog to calculate  $d$  from  $\mu, \mu_0$  and the standard deviation  $\sigma$ .

Figure 27: Effect size dialog to calculate effect size  $d$  means and standard deviation.

### 20.2 Options

This test has no options.

## 20.3 Examples

We want to test the null hypothesis that the population mean is  $\mu = \mu_0 = 10$  against the alternative hypothesis that  $\mu = 15$ . The standard deviation in the population is estimated to be  $\sigma = 8$ . We enter these values in the effect size dialog: Mean H0 = 10, Mean H1 = 15, SD  $\sigma = 8$  to calculate the effect size  $d = 0.625$ .

Next we want to know how many subjects it takes to detect the effect  $d = 0.625$ , given  $\alpha = \beta = .05$ . We are only interested in increases in the mean and thus choose a one-tailed test.

- **Select**  
*Type of power analysis:* A priori
- **Input**  
*Tail(s):* One  
*Effect size d:* 0.625  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95
- **Output**  
*Noncentrality parameter  $\delta$ :* 3.423266  
*Critical t:* 1.699127  
*df:* 29  
*Total sample size:* 30  
*Actual power:* 0.955144

The results indicates that we need at least  $N = 30$  subjects to ensure a power  $> 0.95$ . The actual power achieved with this  $N$  (0.955144) is slightly higher than the requested one.

Cohen (1969, p.59) calculates the sample size needed in a two-tailed test that the departure from the population mean is at least 10% of the standard deviation, that is  $d = 0.1$ , given  $\alpha = 0.01$  and  $\beta \leq 0.1$ . The input and output values for this analysis are:

- **Select**  
*Type of power analysis:* A priori
- **Input**  
*Tail(s):* Two  
*Effect size d:* 0.1  
 *$\alpha$  err prob:* 0.01  
*Power (1- $\beta$  err prob):* 0.90
- **Output**  
*Noncentrality parameter  $\delta$ :* 3.862642  
*Critical t:* 2.579131  
*df:* 1491  
*Total sample size:* 1492  
*Actual power:* 0.900169

G\*POWER outputs a sample size of  $n = 1492$  which is slightly higher than the value 1490 estimated by Cohen using his tables.

### 20.4 Related tests

### 20.5 Implementation notes

The  $H_0$  distribution is the central Student  $t$  distribution  $t(N - 1, 0)$ , the  $H_1$  distribution is the noncentral Student  $t$  distribution  $t(N - 1, \delta)$ , with noncentrality  $\delta = d\sqrt{N}$ .

## 20.6 Validation

The results were checked against the values produced by GPower 2.0.

## 21 t test: Means - difference between two independent means (two groups)

The two-sample t test is used to determine if two population means  $\mu_1, \mu_2$  are equal. The data are two samples of size  $n_1$  and  $n_2$  from two independent and normally distributed populations. The true standard deviations in the two populations are unknown and must be estimated from the data. The null and alternate hypothesis of this t test are:

$$\begin{aligned} H_0 : \mu_1 - \mu_2 &= 0 \\ H_1 : \mu_1 - \mu_2 &\neq 0. \end{aligned}$$

The two-sided test (“two tails”) should be used if there is no restriction on the sign of the deviation assumed in the alternate hypothesis. Otherwise use the one-sided test (“one tail”).

### 21.1 Effect size index

The effect size index  $d$  is defined as:

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

Cohen (1969, p. 38) defines the following conventional values for  $d$ :

- small  $d = 0.2$
- medium  $d = 0.5$
- large  $d = 0.8$

Pressing the button Determine on the left side of the effect size label opens the effect size drawer (see Fig. 28). You can use it to calculate  $d$  from the means and standard deviations in the two populations. The  $t$ -test assumes the variances in both populations to be equal. However, the test is relatively robust against violations of this assumption if the sample sizes are equal ( $n_1 = n_2$ ). In this case a mean  $\sigma'$  may be used as the common within-population  $\sigma$  (Cohen, 1969, p.42):

$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

In the case of substantially different sample sizes  $n_1 \neq n_2$  this correction should not be used because it may lead to power values that differ greatly from the true values (Cohen, 1969, p.42).

### 21.2 Options

This test has no options.

### 21.3 Examples

### 21.4 Related tests

### 21.5 Implementation notes

The  $H_0$  distribution is the central Student  $t$  distribution  $t(N - 2, 0)$ ; the  $H_1$  distribution is the noncentral Student

Figure 28: Effect size dialog to calculate effect size  $d$  from the parameters of two independent random variables.

$t$  distribution  $t(N - 2, \delta)$ , where  $N = n_1 + n_2$  and  $\delta$  the noncentrality parameter, which is defined as:

$$\delta = d \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

### 21.6 Validation

The results were checked against the values produced by GPower 2.0.

## 22 Wilcoxon signed-rank test: Means - difference from constant (one sample case)

This test is essentially identical to the Wilcoxon signed-rank test for matched pairs described in chapter 23. The only difference is in the calculation of the effect size  $d$  (or the corresponding moment  $p_1$ ). The effect size used here is analogous to the one sample t-test.

### 22.1 Examples

We replicate some power values from table 4 given in Shieh, Jan, and Randles (2007) (see table 29).

To compare these values with the power computed by G\*POWER, we choose the Lehmann method without continuity correction. For effect size  $d = 0.1$  and a normal distribution as parent we get:

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* One  
*Parent distribution:* Normal  
*Effect size d:* 0.1  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 649
- **Output**  
*critical z:* 1.6448536  
*Power (1-  $\beta$ ):* 0.8000589  
*Moment p1:* 0.5398278  
*Moment p2:* 0.5562315  
*Moment p3:* 0.3913924

For  $d = 0.8$  and the Laplace parent we get

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* One  
*Parent distribution:* Laplace  
*Effect size dz:* 0.8  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 11
- **Output**  
*critical z:* 1.6448536  
*Power (1-  $\beta$ ):* 0.8308341  
*Moment p1:* 0.8387046  
*Moment p2:* 0.8890997  
*Moment p3:* 0.8206572

In both cases, the values computed with G\*POWER using the "Lehmann method without continuity correction" correspond closely to those given in the table for the "Exact variance" method and both values correspond closely to simulated power.

In the latter case we would get a power of 0.923 With the A.R.E-method if the Laplace-parent is chosen. The lower

bound for the power that results if min ARE is chosen is 0.7319.

### 22.2 Related tests

Related tests in G\*POWER 3.0:

- Means: Difference from constant.
- Proportions: Sign test.
- Means: Wilcoxon rank sum test.

### 22.3 Validation

The results were checked against the results given in Shieh et al. (2007), analogous computations in PASS (Hintze, 2006) and results produced by unifyPow O'Brien (1998). There was complete correspondence with the values given in O'Brien, while there were slight differences to those produced by PASS. The reason of these differences seems to be that PASS truncates the weighted sample sizes to integer values.

d =	0.1	0.2	0.4	0.6	0.8	1.0
Standard normal						
Sample size	649	164	42	20	12	9
Nominal power						
Exact variance	0.8001	0.8015	0.8001	0.8088	0.8074	0.8478
Lehmann	0.8005	0.8031	0.8074	0.8251	0.8395	0.8823
Noether	0.7985	0.7953	0.7762	0.7561	0.7195	0.7070
Simulated power						
Error	0.8054	0.7970	0.7990	0.8084	0.8235	0.8468
Exact variance	-0.0053	0.0045	0.0011	0.0004	-0.0161	0.0010
Lehmann	-0.0049	0.0061	0.0084	0.0167	0.0160	0.0355
Noether	-0.0069	-0.0017	-0.0228	-0.0523	-0.1040	-0.1398
Laplace (0, 1)						
Sample size	419	109	31	16	11	8
Nominal power						
Exact variance	0.8002	0.8003	0.8047	0.8079	0.8308	0.8185
Lehmann	0.8061	0.8208	0.8653	0.9097	0.9530	0.9726
Noether	0.7976	0.7907	0.7708	0.7409	0.7225	0.6755
Simulated power						
Error	0.8050	0.7954	0.8099	0.8117	0.8265	0.8196
Exact variance	-0.0048	0.0049	-0.0052	-0.0038	0.0043	-0.0011
Lehmann	0.0011	0.0254	0.0554	0.0980	0.1265	0.1530
Noether	-0.0074	-0.0047	-0.0391	-0.0708	-0.1040	-0.1441

**Figure 29:** A part of table 4 in Shieh et al. (2007) that compares nominal power values computed using different methods with simulated power

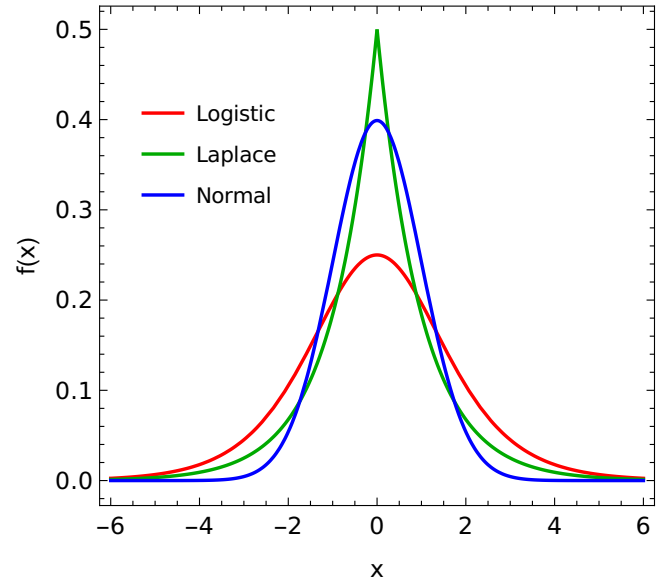
## 23 Wilcoxon signed-rank test: (matched pairs)

The Wilcoxon signed-rank test is a nonparametric alternative to the one sample t test. Its use is mainly motivated by uncertainty concerning the assumption of normality made in the t test.

The Wilcoxon signed-rank test can be used to test whether a given distribution  $H$  is symmetric about zero. The power routines implemented in *G\*POWER* refer to the important special case of a “shift model”, which states that  $H$  is obtained by subtracting two symmetric distributions  $F$  and  $G$ , where  $G$  is obtained by shifting  $F$  by an amount  $\Delta$ :  $G(x) = F(x - \Delta)$  for all  $x$ . The relation of this shift model to the one sample t test is obvious if we assume that  $F$  is the fixed distribution with mean  $\mu_0$  stated in the null hypothesis and  $G$  the distribution of the test group with mean  $\mu$ . Under this assumptions  $H(x) = F(x) - G(x)$  is symmetric about zero under  $H_0$ , that is if  $\Delta = \mu - \mu_0 = 0$  or, equivalently  $F(x) = G(x)$ , and asymmetric under  $H_1$ , that is if  $\Delta \neq 0$ .

**The Wilcoxon signed-rank test** The signed-rank test is based on ranks. Assume that a sample of size  $N$  is drawn from a distribution  $H(x)$ . To each sample value  $x_i$  a rank  $S$  between 1 and  $N$  is assigned that corresponds to the position of  $|x_i|$  in a increasingly ordered list of all absolute sample values. The general idea of the test is to calculate the sum of the ranks assigned to positive sample values ( $x > 0$ ) and the sum of the ranks assigned to negative sample values ( $x < 0$ ) and to reject the hypothesis that  $H$  is symmetric if these two rank sums are clearly different.

The actual procedure is as follows: Since the rank sum of negative values is known if that of positive values is given, it suffices to consider the rank sum  $V_s$  of positive values. The positive ranks can be specified by a  $n$ -tuple  $(S_1, \dots, S_n)$ , where  $0 \leq n \leq N$ . There are  $\binom{N}{n}$  possible  $n$ -tuples for a given  $n$ . Since  $n$  can take on the values  $0, 1, \dots, N$ , the total number of possible choices for the  $S$ 's is:  $\sum_{i=0}^N \binom{N}{i} = 2^N$ . (We here assume a continuous distribution  $H$  for which the probabilities for ties, that is the occurrence of two identical  $|x|$ , is zero.) Therefore, if the null hypothesis is true, then the probability to observe a particular  $n$  and a certain  $n$ -tuple is:  $P(N_+ = n; S_1 = s_1, \dots, S_n = s_n) = 1/2^N$ . To calculate the probability to observe (under  $H_0$ ) a particular positive rank sum  $V_s = S_1 + \dots + S_n$  we just need to count the number  $k$  of all tuples with rank sum  $V_s$  and to add their probabilities, thus  $P(V_s = v) = k/2^N$ . Doing this for all possible  $V_s$  between the minimal value 0 corresponding to the case  $n = 0$ , and the maximal value  $N(N + 1)/2$ , corresponding to the  $n = N$  tuple  $(1, 2, \dots, N)$ , gives the discrete probability distribution of  $V_s$  under  $H_0$ . This distribution is symmetric about  $N(N + 1)/4$ . Referring to this probability distribution we choose in a one-sided test a critical value  $c$  with  $P(V_s \geq c) \leq \alpha$  and reject the null hypothesis if a rank sum  $V_s > c$  is observed. With increasing sample size the exact distribution converges rapidly to the normal distribution with mean  $E(V_s) = N(N + 1)/4$  and variance  $Var(V_s) = N(N + 1)(2N + 1)/24$ .



**Figure 30:** Densities of the Normal, Laplace, and Logistic distributions (with  $\mu = 0, \sigma = 1$ )

**Power of the Wilcoxon rank-sum test** The signed-rank test as described above is distribution free in the sense that its validity does not depend on the specific form of the response distribution  $H$ . This distribution independence does no longer hold, however, if one wants to estimate numerical values for the power of the test. The reason is that the effect of a certain shift  $\Delta$  on the deviation from symmetry and therefore the distribution of  $V_s$  depends on the specific form of  $F$  (and  $G$ ). For power calculations it is therefore necessary to specify the response distribution  $F$ . *G\*POWER* provides three predefined continuous and symmetric response functions that differ with respect to kurtosis, that is the “peakedness” of the distribution:

- Normal distribution  $N(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Laplace or Double Exponential distribution:

$$p(x) = \frac{1}{2} e^{-|x|}$$

- Logistic distribution

$$p(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

Scaled and/or shifted versions of the Laplace and Logistic densities that can be calculated by applying the transformation  $\frac{1}{a} p((x - b)/a)$ ,  $a > 0$ , are again probability densities and are referred to by the same name.

**Approaches to the power analysis** *G\*POWER* implements two different methods to estimate the power for the signed-rank Wilcoxon test: A) The asymptotic relative efficiency (A.R.E.) method that defines power relative to the one sample t test, and B) a normal approximation to the power proposed by [Lehmann \(1975, pp. 164-166\)](#). We describe the general idea of both methods in turn. More specific information can be found in the implementation section below.

- **A.R.E-method:** The A.R.E method assumes the shift model described in the introduction. It relates normal approximations to the power of the one-sample t-test (Lehmann, 1975, Eq. (4.44), p. 172) and the Wilcoxon test for a specified  $H$  (Lehmann, 1975, Eq. (4.15), p. 160). If for a model with fixed  $H$  and  $\Delta$  the sample size  $N$  are required to achieve a specified power for the Wilcoxon signed-rank test and a samples size  $N'$  is required in the t test to achieve the same power, then the ratio  $N'/N$  is called the efficiency of the Wilcoxon signed-rank test relative to the one-sample t test. The limiting efficiency as sample size  $N$  tends to infinity is called the asymptotic relative efficiency (A.R.E. or Pitman efficiency) of the Wilcoxon signed rank test relative to the t test. It is given by (Hettmansperger, 1984, p. 71):

$$12\sigma_H^2 \left[ \int_{-\infty}^{+\infty} H^2(x) dx \right]^2 = 12 \int_{-\infty}^{+\infty} x^2 H(x) dx \left[ \int_{-\infty}^{+\infty} H^2(x) dx \right]^2$$

Note, that the A.R.E. of the Wilcoxon signed-rank test to the one-sample t test is identical to the A.R.E of the Wilcoxon rank-sum test to the two-sample t test (if  $H = F$ ; for the meaning of  $F$  see the documentation of the Wilcoxon rank sum test).

If  $H$  is a normal distribution, then the A.R.E. is  $3/\pi \approx 0.955$ . This shows, that the efficiency of the Wilcoxon test relative to the t test is rather high even if the assumption of normality made in the t test is true. It can be shown that the minimal A.R.E. (for  $H$  with finite variance) is 0.864. For non-normal distributions the Wilcoxon test can be much more efficient than the t test. The A.R.E.s for some specific distributions are given in the implementation notes. To estimate the power of the Wilcoxon test for a given  $H$  with the A.R.E. method one basically scales the sample size with the corresponding A.R.E. value and then performs the procedure for the t test for two independent means.

- **Lehmann method:** The computation of the power requires the distribution of  $V_s$  for the non-null case, that is for cases where  $H$  is not symmetric about zero. The Lehmann method uses the fact that

$$\frac{V_s - E(V_s)}{\sqrt{\text{Var}V_s}}$$

tends to the standard normal distribution as  $N$  tends to infinity for any fixed distributions  $H$  for which  $0 < P(X < 0) < 1$ . The problem is then to compute expectation and variance of  $V_s$ . These values depend on three "moments"  $p_1, p_2, p_3$ , which are defined as:

- $p_1 = P(X < 0)$ .
- $p_2 = P(X + Y > 0)$ .
- $p_3 = P(X + Y > 0 \text{ and } X + Z > 0)$ .

where  $X, Y, Z$  are independent random variables with distribution  $H$ . The expectation and variance are given as:

$$E(V_s) = N(N-1)p_2/2 + Np_1$$

$$\begin{aligned} \text{Var}(V_s) &= N(N-1)(N-2)(p_3 - p_1^2) \\ &+ N(N-1)[2(p_1 - p_2)^2 \\ &+ 3p_2(1 - p_2)]/2 + Np_1(1 - p_1) \end{aligned}$$

The value  $p_1$  is easy to interpret: If  $H$  is continuous and shifted by an amount  $\Delta > 0$  to larger values, then  $p_1$  is the probability to observe a negative value. For a null shift (no treatment effect,  $\Delta = 0$ , i.e.  $H$  symmetric about zero) we get  $p_1 = 1/2$ .

If  $c$  denotes the critical value of a level  $\alpha$  test and  $\Phi$  the CDF of the standard normal distribution, then the normal approximation of the power of the (one-sided) test is given by:

$$\Pi(H) \approx 1 - \Phi \left[ \frac{c - a - E(V_s)}{\sqrt{\text{Var}(V_s)}} \right]$$

where  $a = 0.5$  if a continuity correction is applied, and  $a = 0$  otherwise.

The formulas for  $p_1, p_2, p_3$  for the predefined distributions are given in the implementation section.

## 23.1 Effect size index

The conventional values proposed by Cohen (1969, p. 38) for the t-test are applicable. He defines the following conventional values for  $d$ :

- **small**  $d = 0.2$
- **medium**  $d = 0.5$
- **large**  $d = 0.8$

Pressing the button Determine on the left side of the effect size label opens the effect size dialog (see Fig. 31). You can use this dialog to calculate  $d$  from the means and a common standard deviations in the two populations.

If  $N_1 = N_2$  but  $\sigma_1 \neq \sigma_2$  you may use a mean  $\sigma'$  as common within-population  $\sigma$  (Cohen, 1969, p.42):

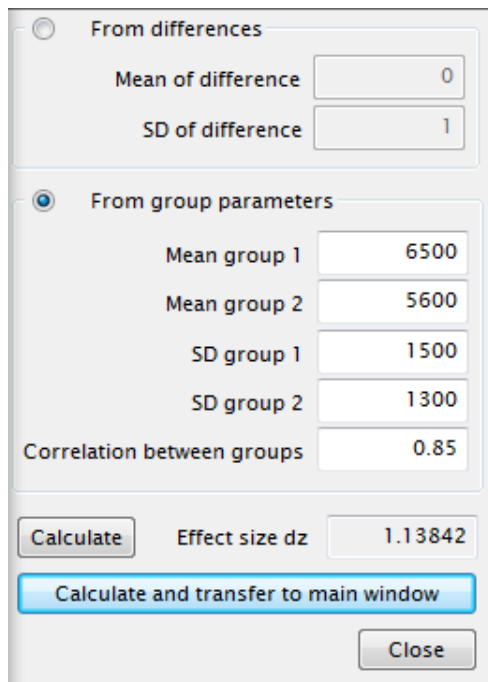
$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

If  $N_1 \neq N_2$  you should not use this correction, since this may lead to power values that differ greatly from the true values (Cohen, 1969, p.42).

## 23.2 Options

This test has several options (see Fig. 32). First, you may choose between the A.R.E-method and the Lehmann method. Given the A.R.E method you may either choose the value for  $k$  (see implementation notes) implicitly via the corresponding parent distribution or specify it manually. Analogously, you may specify the three moments by choosing the distribution or "manually". There are further options to choose between two different types of effect size parameter ( $d$  vs. Moment  $p_1$ ), the type of distribution shown in the plot window, whether both  $d$  and  $p_2$  are given in the output, and finally, whether to use a continuity-correction.





**Figure 31:** Effect size dialog to calculate effect size  $d$  from the parameters of two matched random variables.

### 23.3 Examples

We replicate the example in O'Brien (2002, p. 141): The question is, whether women report eating different amounts of food on pre- vs. post-menstrual days? The null hypothesis is that there is no difference, that is 50% report to eat more in pre-menstrual days and 50% to eat more in post-menstrual days. In the alternative hypothesis the assumption is that 80% of the women report to eat more in the pre-menstrual days.

We want to know the power of a two-sided test under the assumptions that the responses are distributed according to a Laplace parent distribution, that  $N = 10$ , and  $\alpha = 0.05$ . We use the effect size dialog zu compute  $dz$ . In the options we select the Lehmann method, the use of predefined parent distributions, the effect size measure  $d$  and to use the continuity correction.

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* Two  
*Parent distribution:* Laplace  
*Effect size dz:* 1.13842  
 *$\alpha$  err prob:* 0.05  
*Total sample size:* 10
- **Output**  
*critical z:* 1.9599640  
*Power (1-  $\beta$ ):* 0.8534958  
*Moment p1:* 0.9000531  
*Moment p2:* 0.9478560  
*Moment p3:* 0.9122280

### 23.4 Related tests

Related tests in G\*POWER 3.0:

- Means: Difference from contrast.
- Proportions: Sign test.
- Means: Wilcoxon rank sum test.

### 23.5 Implementation notes

The  $H_0$  distribution is the central Student  $t$  distribution  $t(Nk - 2, 0)$ ; the  $H_1$  distribution is the noncentral Student  $t$  distribution  $t(Nk - 2, \delta)$ , where the noncentrality parameter  $\delta$  is given by:

$$\delta = d \sqrt{\frac{N_1 N_2 k}{N_1 + N_2}}$$

The parameter  $k$  represents the asymptotic relative efficiency vs. corresponding  $t$  tests (Lehmann, 1975, p. 371ff) and depends in the following way on the parent distribution:

Parent	Value of k (ARE)
Uniform:	1.0
Normal:	$3/\pi$
Logistic:	$\pi^2/9$
Laplace:	$3/2$
min ARE:	0.864

Min ARE is a limiting case that gives a theoretic al minimum of the power for the Wilcoxon-Mann-Whitney test.

### 23.6 Validation

The results were checked against the values produced by PASS (Hintze, 2006) and those produced by unifyPow O'Brien (1998). There was complete correspondence with the values given in O'Brien, while there were slight differences to those produced by PASS. The reason of these differences seems to be that PASS truncates the weighted sample sizes to integer values.

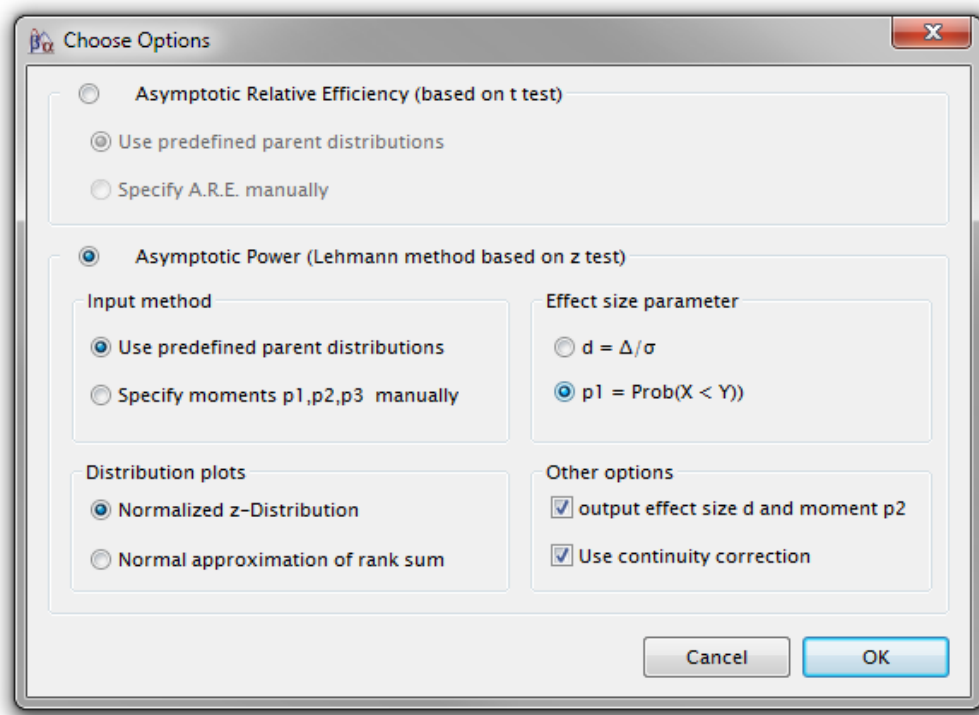


Figure 32: Options for Wilcoxon tests

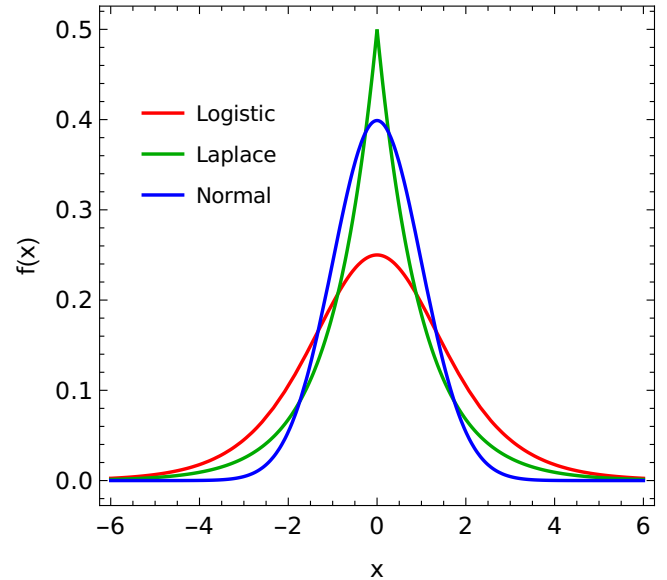
## 24 Wilcoxon-Mann-Whitney test of a difference between two independent means

The Wilcoxon-Mann-Whitney (WMW) test (or U-test) is a nonparametric alternative to the two-group t test. Its use is mainly motivated by uncertainty concerning the assumption of normality made in the t test.

It refers to a general two sample model, in which  $F$  and  $G$  characterize response distributions under two different conditions. The null effects hypothesis states  $F = G$ , while the alternative is  $F \neq G$ . The power routines implemented in *G\*POWER* refer to the important special case of a “shift model”, which states that  $G$  is obtained by shifting  $F$  by an amount  $\Delta$ :  $G(x) = F(x - \Delta)$  for all  $x$ . The shift model expresses the assumption that the treatment adds a certain amount  $\Delta$  to the response  $x$  (Additivity).

**The WMW test** The WMW-test is based on ranks. Assume  $m$  subjects are randomly assigned to control group  $A$  and  $n$  other subjects to treatment group  $B$ . After the experiment, all  $N = n + m$  subjects together are ranked according to a some response measure, i.e. each subject is assigned a unique number between 1 and  $N$ . The general idea of the test is to calculate the sum of the ranks assigned to subjects in either group and to reject the “no effect hypothesis” if the rank sums in the two groups are clearly different. The actual procedure is as follows: Since the  $m$  ranks of control group  $A$  are known if those of treatment group  $B$  are given, it suffices to consider the  $n$  ranks of group  $B$ . They can be specified by a  $n$ -tuple  $(S_1, \dots, S_n)$ . There are  $\binom{N}{n}$  possible  $n$ -tuples. Therefore, if the null hypothesis is true, then the probability to observe a particular  $n$ -tuple is:  $P(S_1 = s_1, \dots, S_n = s_n) = 1/\binom{N}{n}$ . To calculate the probability to observe (under  $H_0$ ) a particular rank sum  $W_s = S_1 + \dots + S_n$  we just need to count the number  $k$  of all tuples with rank sum  $W_s$  and to add their probabilities, thus  $P(W_s = w) = k/\binom{N}{n}$ . Doing this for all possible  $W_s$  between the minimal value  $n(n+1)/2$  corresponding to the tuple  $(1, 2, \dots, n)$ , and the maximal value  $n(1+2m+n)/2$ , corresponding to the tuple  $(N-n+1, N-n+2, \dots, N)$ , gives the discrete probability distribution of  $W_s$  under  $H_0$ . This distribution is symmetric about  $n(N+1)/2$ . Referring to this probability distribution we choose a critical value  $c$  with  $P(W_s \geq c) \leq \alpha$  and reject the null hypothesis if a rank sum  $W_s > c$  is observed. With increasing sample sizes the exact distribution converges rapidly to the normal distribution with mean  $E(W_s) = n(N+1)/2$  and variance  $Var(W_s) = mn(N+1)/12$ .

A drawback of using  $W_s$  is that the minimal value of  $W_s$  depends on  $n$ . It is often more convenient to subtract the minimal value and use  $W_{XY} = W_s - n(n+1)/2$  instead. The statistic  $W_{XY}$  (also known as the Mann-Whitney statistic) can also be derived from a slightly different perspective: If  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$  denote the observations in the control and treatment group, respectively, than  $W_{XY}$  is the number of pairs  $(X_i, Y_j)$  with  $X_i < Y_j$ . The approximating normal distribution of  $W_{XY}$  has mean  $E(W_{XY}) = mn/2$  and  $Var(W_{XY}) = Var(W_s) = mn(N+1)/12$ .



**Figure 33:** Densities of the Normal, Laplace, and Logistic distributions (with  $\mu = 0, \sigma = 1$ )

**Power of the WMW test** The WMW test as described above is distribution free in the sense that its validity does not depend on the specific form of the response distribution  $F$ . This distribution independence does not longer hold, however, if one wants to estimate numerical values for the power of the test. The reason is that the effect of a certain shift  $\Delta$  on the rank distribution depends on the specific form of  $F$  (and  $G$ ). For power calculations it is therefore necessary to specify the response distribution  $F$ . *G\*POWER* provides three predefined continuous and symmetric response functions that differ with respect to kurtosis, that is the “peakedness” of the distribution:

- Normal distribution  $N(\mu, \sigma^2)$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

- Laplace or Double Exponential distribution:

$$p(x) = \frac{1}{2} e^{-|x|}$$

- Logistic distribution

$$p(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

Scaled and/or shifted versions of the Laplace and Logistic densities that can be calculated by applying the transformation  $\frac{1}{a} p((x-b)/a), a > 0$ , are again probability densities and are referred to by the same name.

**Approaches to the power analysis** *G\*POWER* implements two different methods to estimate the power for the WMW test: A) The asymptotic relative efficiency (A.R.E.) method that defines power relative to the two groups t test, and B) a normal approximation to the power proposed by [Lehmann \(1975, pp. 69-71\)](#). We describe the general idea of both methods in turn. More specific information can be found in the implementation section below.

- **A.R.E-method:** The A.R.E method assumes the shift model described in the introduction. It relates normal approximations to the power of the t-test (Lehmann, 1975, Eq. (2.42), p. 78) and the Wilcoxon test for a specified  $F$  (Lehmann, 1975, Eq. (2.29), p. 72). If for a model with fixed  $F$  and  $\Delta$  the sample sizes  $m = n$  are required to achieve a specified power for the Wilcoxon test and samples sizes  $m' = n'$  are required in the t test to achieve the same power, then the ratio  $n'/n$  is called the efficiency of the Wilcoxon test relative to the t test. The limiting efficiency as sample sizes  $m$  and  $n$  tend to infinity is called the asymptotic relative efficiency (A.R.E. or Pitman efficiency) of the Wilcoxon test relative to the t test. It is given by (Hettmansperger, 1984, p. 71):

$$12\sigma_F^2 \left[ \int_{-\infty}^{+\infty} F^2(x)dx \right]^2 = 12 \int_{-\infty}^{+\infty} x^2 F(x)dx \left[ \int_{-\infty}^{+\infty} F^2(x)dx \right]^2$$

If  $F$  is a normal distribution, then the A.R.E. is  $3/\pi \approx 0.955$ . This shows, that the efficiency of the Wilcoxon test relative to the t test is rather high even if the assumption of normality made in the t test is true. It can be shown that the minimal A.R.E. (for  $F$  with finite variance) is 0.864. For non-normal distributions the Wilcoxon test can be much more efficient than the t test. The A.R.E.s for some specific distributions are given in the implementation notes. To estimate the power of the Wilcoxon test for a given  $F$  with the A.R.E. method one basically scales the sample size with the corresponding A.R.E. value and then performs the procedure for the t test for two independent means.

- **Lehmann method:** The computation of the power requires the distribution of  $W_{XY}$  for the non-null case  $F \neq G$ . The Lehmann method uses the fact that

$$\frac{W_{XY} - E(W_{XY})}{\sqrt{\text{Var}W_{XY}}}$$

tends to the standard normal distribution as  $m$  and  $n$  tend to infinity for any fixed distributions  $F$  and  $G$  for  $X$  and  $Y$  for which  $0 < P(X < Y) < 1$ . The problem is then to compute expectation and variance of  $W_{XY}$ . These values depend on three "moments"  $p_1, p_2, p_3$ , which are defined as:

- $p_1 = P(X < Y)$ .
- $p_2 = P(X < Y \text{ and } X < Y')$ .
- $p_3 = P(X < Y \text{ and } X' < Y)$ .

Note that  $p_2 = p_3$  for symmetric  $F$  in the shift model. The expectation and variance are given as:

$$\begin{aligned} E(W_{XY}) &= mnp_1 \\ \text{Var}(W_{XY}) &= mnp_1(1-p_1) \dots \\ &\quad + mn(n-1)(p_2-p_1^2) \dots \\ &\quad + nm(m-1)(p_3-p_1^2) \end{aligned}$$

The value  $p_1$  is easy to interpret: If the response distribution  $G$  of the treatment group is shifted to larger

values, then  $p_1$  is the probability to observe a lower value in the control condition than in the test condition. For a null shift (no treatment effect,  $\Delta = 0$ ) we get  $p_1 = 1/2$ .

If  $c$  denotes the critical value of a level  $\alpha$  test and  $\Phi$  the CDF of the standard normal distribution, then the normal approximation of the power of the (one-sided) test is given by:

$$\Pi(F, G) \approx 1 - \Phi \left[ \frac{c - a - mnp_1}{\sqrt{\text{Var}(W_{XY})}} \right]$$

where  $a = 0.5$  if a continuity correction is applied, and  $a = 0$  otherwise.

The formulas for  $p_1, p_2, p_3$  for the predefined distributions are given in the implementation section.

## 24.1 Effect size index

You may choose between the A.R.E. and the Lehmann-Method in the options section.

**A.R.E. method** In the A.R.E. method the effect size  $d$  is defined as:

$$d = \frac{\mu_1 - \mu_2}{\sigma} = \frac{\Delta}{\sigma}$$

where  $\mu_1, \mu_2$  are the means of the response functions  $F$  and  $G$  and  $\sigma$  the standard deviation of the response distribution  $F$ .

In addition to the effect size, you have to specify the A.R.E. for  $F$ . If you want to use the predefined distributions, you may choose the Normal, Logistic, or Laplace transformation or the minimal A.R.E. From this selection  $G^*$ POWER determines the A.R.E. automatically. Alternatively, you can choose the option to determine the A.R.E. value by hand. In this case you must first calculate the A.R.E. for your  $F$  using the formula given above.

The conventional values proposed by Cohen (1969, p. 38) for the t-test are applicable. He defines the following conventional values for  $d$ :

- **small**  $d = 0.2$
- **medium**  $d = 0.5$
- **large**  $d = 0.8$

Pressing the button Determine on the left side of the effect size label opens the effect size dialog (see Fig. 34). You can use this dialog to calculate  $d$  from the means and a common standard deviations in the two populations.

If  $N_1 = N_2$  but  $\sigma_1 \neq \sigma_2$  you may use a mean  $\sigma'$  as common within-population  $\sigma$  (Cohen, 1969, p.42):

$$\sigma' = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

If  $N_1 \neq N_2$  you should not use this correction, since this may lead to power values that differ greatly from the true values (Cohen, 1969, p.42).

**Figure 34:** Effect size dialog to calculate effect size  $d$  from the parameters of two independent random variables with equal variance.

**Lehmann method** If the Lehmann method is chosen then the effect size is given by the first moment  $p_1$  and the relative SD. If one of the three pre-defined parent distributions is chosen, then the remaining moments  $p_2$  and  $p_3$  are computed automatically, otherwise they must also be given.

## 24.2 Options

This test has several options (see Fig. 32). First, you may choose between the A.R.E-method and the Lehmann method. Given the A.R.E method you may either choose the value for  $k$  (see implementation notes) implicitly via the corresponding parent distribution or specify it manually. Analogously, you may specify the three moments by choosing the distribution or by entering all three values manually. There are further options to choose between two different types of effect size parameter ( $d$  vs. Moment  $p_1$ ), the type of distribution shown in the plot window, whether both  $d$  and  $p_2$  are given in the output, and finally, whether to use a continuity-correction.

## 24.3 Examples

**Example 1** We replicate the example in O'Brien (2002, p. 130ff): The question is, whether the injury rates in two types of shoes differ. The null hypothesis is that there is no difference. The means 0.3 and 0.27 and a common standard deviation  $sd = 0.08$  are assumed in group 1 and 2, respectively.

We want to know the power of a two-sided test under the assumptions that the responses are distributed according to a Laplace parent distribution, that  $N_1 = 67, N_2 = 134$  (i.e.  $N_1 + N_2 = 201$ ) and  $\alpha = 0.05$ . We use the effect size dialog to compute the effect size  $d = 0.375$ . In the options we select the Lehmann method, the use of predefined parent distri-

butions, the effect size measure  $d$  and to use the continuity correction.

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* Two  
*Parent distribution:* Laplace  
*Effect size d:* 0.375  
 *$\alpha$  err prob:* 0.05  
*Sample size group 1:* 67  
*Sample size group 2:* 134
- **Output**  
*critical z:* 1.9599640  
*Power (1-  $\beta$ ):* 0.8472103  
*Moment p1:* 0.6277817  
*Moment p2:* 0.6277817

The computed power of 0.847 corresponds exactly to the value given in O'Brien for this case. Changing the parent distribution to normal reduces the power to 0.68.

**Example 2** This replicates the example described in O'Brien (2002, p. 134ff) and illustrates the specification of the effect size via the first moment  $p_1$ . To enable this input method we select in the options the Lehmann method, the use of predefined parent distributions, the effect size measure  $p_1$  and to use the continuity correction.

The first moment is defined as the probability that a random  $Y$  from the first group is larger than a random  $Y$  from the second group. The parent distribution is assumed to be "logistic" and we choose  $p_1 = 0.6$  relative to this parent distribution. An increase in SD in the second group can also be specified. Here a value of rel. SD = 1.15 is used.

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* One  
*Parent distribution:* Logistic  
*Relative SD:* 1.15  
*Moment p1:* 0.6  
 *$\alpha$  err prob:* 0.05  
*Sample size group 1:* 67  
*Sample size group 2:* 134
- **Output**  
*critical z:* 1.6448536  
*Power (1-  $\beta$ ):* 0.6461276  
*Effect size d:* 0.2911838  
*Moment p2:* 0.4251206

The computed power 0.646 is exactly the value given in O'Brien.

If the parent distribution is changed to "Laplace" then a very similar power 0.647 is computed. This deviates considerably from the value 0.755 given in O'Brien for this case in his table. The reason is that G\*POWER always uses a single parent, that is, if the parent is changed to "Laplace" then the given  $p_1$  is interpreted relative to this parent. For the table in O'Brien, however, the effect size was always computed

from the Logistic parent. To replicate this behavior, one can change to the effect size  $d$  input mode, use the effect size  $d$  given in the output for the logistic parent ( $d = 0.291183$ ), and choose Laplace as the new parent:

- **Select**  
*Type of power analysis:* Post hoc
- **Input**  
*Tail(s):* One  
*Parent distribution:* Laplace  
*Relative SD:* 1.15  
*Effect size d:* 0.2911838  
 *$\alpha$  err prob:* 0.05  
*Sample size group 1:* 67  
*Sample size group 2:* 134
- **Output**  
*critical z:* 1.6448536  
*Power (1-  $\beta$ ):* 0.7551564  
*Moment p1:* 0.6005708  
*Moment p2:* 0.4405953

## 24.4 Related tests

## 24.5 Implementation notes

The procedure uses the The  $H_0$  distribution is the central Student  $t$  distribution  $t(Nk - 2, 0)$ ; the  $H_1$  distribution is the noncentral Student  $t$  distribution  $t(Nk - 2, \delta)$ , where the noncentrality parameter  $\delta$  is given by:

$$\delta = d \sqrt{\frac{N_1 N_2 k}{N_1 + N_2}}$$

The parameter  $k$  represents the asymptotic relative efficiency vs. correspondig  $t$  tests (Lehmann, 1975, p. 371ff) and depends in the following way on the parent distribution:

Parent	Value of k (ARE)
Uniform:	1.0
Normal:	$3/\pi$
Logistic:	$\pi^2/9$
Laplace:	$3/2$
min ARE:	0.864

Min ARE is a limiting case that gives a theoretical minimum of the power for the Wilcoxon-Mann-Whitney test.

### 24.5.1 Conversion from effect size d to p1

Given effect size  $d$  the first moment  $p_1$  is

- Normal distribution:

$$p_1 = \Phi\left(d/\sqrt{2}\right)$$

- Logistic distribution:

$$p_1 = \frac{1 - (1 + u) \exp(-u)}{(1 - \exp(-u))^2}$$

with  $u = d\pi/\sqrt{3}$

- Laplace distribution

$$p_1 = 1 - \exp(-u)(1 + u/2)/2$$

with  $u = d\sqrt{2}$

## 24.6 Validation

The results were checked against the values produced by PASS (Hintze, 2006) and those produced by unifyPow O'Brien (1998). There was complete correspondence with the values given in O'Brien, while there were slight differences to those produced by PASS. The reason of these differences seems to be that PASS truncates the weighted sample sizes to integer values.

## 25 t test: Generic case

With generic  $t$  tests you can perform power analyses for any test that depends on the  $t$  distribution. All parameters of the noncentral  $t$ -distribution can be manipulated independently. Note that with Generic  $t$ -Tests you cannot do a priori power analyses, the reason being that there is no definite association between  $N$  and  $df$  (the degrees of freedom). You need to tell G\*POWER the values of both  $N$  and  $df$  explicitly.

### 25.1 Effect size index

In the generic case, the noncentrality parameter  $\delta$  of the noncentral  $t$  distribution may be interpreted as effect size.

### 25.2 Options

This test has no options.

### 25.3 Examples

To illustrate the usage we calculate with the generic  $t$  test the power of a one sample  $t$  test. We assume  $N = 25$ ,  $\mu_0 = 0$ ,  $\mu_1 = 1$ , and  $\sigma = 4$  and get the effect size  $d = (\mu_0 - \mu_1)/\sigma = (0 - 1)/4 = -0.25$ , the noncentrality parameter  $\delta = d\sqrt{N} = -0.25\sqrt{25} = -1.25$ , and the degrees of freedom  $df = N - 1 = 24$ . We choose a post hoc analysis and a two-sided test. As result we get the power  $(1 - \beta) = 0.224525$ ; this is exactly the same value we get from the specialized routine for this case in G\*POWER .

### 25.4 Related tests

Related tests in G\*POWER are:

- Correlation: Point biserial model
- Means: Difference between two independent means (two groups)
- Means: Difference between two dependent means (matched pairs)
- Means: Difference from constant (one sample case)

### 25.5 Implementation notes

One distribution is fixed to the central Student  $t$  distribution  $t(df)$ . The other distribution is a noncentral  $t$  distribution  $t(df, \delta)$  with noncentrality parameter  $\delta$ .

In an a priori analysis the specified  $\alpha$  determines the critical  $t_c$ . In the case of a one-sided test the critical value  $t_c$  has the same sign as the noncentrality parameter, otherwise there are two critical values  $t_c^1 = -t_c^2$ .

### 25.6 Validation

The results were checked against the values produced by GPower 2.0.

## 26 $\chi^2$ test: Variance - difference from constant (one sample case)

This procedure allows power analyses for the test that the population variance  $\sigma^2$  of a normally distributed random variable has the specific value  $\sigma_0^2$ . The null and (two-sided) alternative hypothesis of this test are:

$$\begin{aligned} H_0 : \sigma - \sigma_0 &= 0 \\ H_1 : \sigma - \sigma_0 &\neq 0. \end{aligned}$$

The two-sided test (“two tails”) should be used if there is no restriction on the sign of the deviation assumed in the alternative hypothesis. Otherwise use the one-sided test (“one tail”).

### 26.1 Effect size index

The ratio  $\sigma^2/\sigma_0^2$  of the variance assumed in H1 to the base line variance is used as effect size measure. This ratio is 1 if  $H_0$  is true, that is, if both variances are identical. In an a priori analysis a ratio close or even identical to 1 would imply an exceedingly large sample size. Thus, G\*POWER prohibits inputs in the range [0.999, 1.001] in this case.

Pressing the button Determine on the left side of the effect size label in the main window opens the effect size drawer (see Fig. 35) that may be used to calculate the ratio from the two variances. Insert the baseline variance  $\sigma_0^2$  in the field variance V0 and the alternate variance in the field variance V1.

variance V0	1.5
variance V1	1
ratio V1/V0	0.6666667

Buttons: Calculate, Calculate and transfer to main window, Close

Figure 35: Effect size drawer to calculate variance ratios.

### 26.2 Options

This test has no options.

### 26.3 Examples

We want to test whether the variance in a given population is clearly lower than  $\sigma_0 = 1.5$ . In this application we use “ $\sigma^2$  is less than 1” as criterion for “clearly lower”. Inserting variance V0 = 1.5 and variance V1 = 1 in the effect size drawer we calculate as effect size a Ratio var1/var0 of 0.6666667.

How many subjects are needed to achieve the error levels  $\alpha = 0.05$  and  $\beta = 0.2$  in this test? This question can be answered by using the following settings in G\*POWER :

- **Select**

Type of power analysis: A priori

- **Input**

Tail(s): One  
Ratio var1/var0: 0.6666667  
 $\alpha$  err prob: 0.05  
Power (1-  $\beta$ ): 0.80

- **Output**

Lower critical  $\chi^2$ : 60.391478  
Upper critical  $\chi^2$ : 60.391478  
Df: 80  
Total sample size: 81  
Actual power : 0.803686

The output shows that using a one-sided test we need at least 81 subjects in order to achieve the desired level of the  $\alpha$  and  $\beta$  error. To apply the test, we would estimate the variance  $s^2$  from the sample of size  $N$ . The one-sided test would be significant at  $\alpha = 0.05$  if the statistic  $x = (N - 1) \cdot s^2/\sigma_0^2$  were lower than the critical value 60.39.

By setting “Tail(s) = Two”, we can easily check that a two-sided test under the same conditions would have required a much larger sample size, namely  $N = 103$ , to achieve the error criteria.

### 26.4 Related tests

Similar tests in G\*POWER 3.0:

- Variance: Test of equality (two sample case).

### 26.5 Implementation notes

It is assumed that the population is normally distributed and that the mean is not known in advance but estimated from a sample of size  $N$ . Under these assumptions the  $H_0$ -distribution of  $s^2(N - 1)/\sigma_0^2$  is the central  $\chi^2$  distribution with  $N - 1$  degrees of freedom ( $\chi_{N-1}^2$ ), and the  $H_1$ -distribution is the same central  $\chi^2$  distribution scaled with the variance ratio, that is,  $(\sigma^2/\sigma_0^2) \cdot \chi_{N-1}^2$ .

### 26.6 Validation

The correctness of the results were checked against values produced by PASS (Hintze, 2006) and in a monte carlo simulation.



## 27 z test: Correlation - inequality of two independent Pearson r's

This procedure refers to tests of hypotheses concerning differences between two independent population correlation coefficients. The null hypothesis states that both correlation coefficients are identical  $\rho_1 = \rho_2$ . The (two-sided) alternative hypothesis is that the correlation coefficients are different:  $\rho_1 \neq \rho_2$ :

$$H_0 : \rho_1 - \rho_2 = 0$$

$$H_1 : \rho_1 - \rho_2 \neq 0.$$

If the direction of the deviation  $\rho_1 - \rho_2$  cannot be predicted a priori, a two-sided ('two-tailed') test should be used. Otherwise use a one-sided test.

### 27.1 Effect size index

The effect size index  $q$  is defined as a difference between two 'Fisher  $z'$ -transformed correlation coefficients:  $q = z_1 - z_2$ , with  $z_1 = \ln((1 + r_1)/(1 - r_1))/2$ ,  $z_2 = \ln((1 + r_2)/(1 - r_2))/2$ . G\*POWER requires  $q$  to lie in the interval  $[-10, 10]$ .

Cohen (1969, p. 109ff) defines the following **effect size conventions** for  $q$ :

- **small**  $q = 0.1$
- **medium**  $q = 0.3$
- **large**  $q = 0.5$

Pressing the button Determine on the left side of the effect size label opens the effect size dialog:

It can be used to calculate  $q$  from two correlations coefficients.

If  $q$  and  $r_2$  are given, then the formula  $r_1 = (a - 1)/(a + 1)$ , with  $a = \exp[2q + \ln((1 + r_2)/(1 - r_2))]$  may be used to calculate  $r_1$ .

G\*POWER outputs critical values for the  $z_c$  distribution. To transform these values to critical values  $q_c$  related to the effect size measure  $q$  use the formula:  $q_c = z_c \sqrt{((N_1 + N_2 - 6)/((N_1 - 3)(N_2 - 3)))}$  (see Cohen, 1969, p. 135).

### 27.2 Options

This test has no options.

### 27.3 Examples

Assume it is known that test A correlates with a criterion with  $r_1 = 0.75$  and that we want to test whether an alternative test B correlates higher, say at least with  $r_2 = 0.88$ .

We have two data sets one using A with  $N_1 = 51$ , and a second using B with  $N_2 = 206$ . What is the power of a two-sided test for a difference between these correlations, if we set  $\alpha = 0.05$ ?

We use the effect size drawer to calculate the effect size  $q$  from both correlations. Setting correlation coefficient  $r_1 = 0.75$  and correlation coefficient  $r_2 = 0.88$  yields  $q = -0.4028126$ . The input and outputs are as follows:

- **Select**  
Type of power analysis: Post hoc
- **Input**  
Tail(s): two  
Effect size  $q$ : -0.4028126  
 $\alpha$  err prob: 0.05  
Sample size: 260  
Sample size: 51
- **Output**  
Critical  $z$ : 60.391478  
Upper critical  $\chi^2$ : -1.959964  
Power (1-  $\beta$ ): 0.726352

The output shows that the power is about 0.726. This is very close to the result 0.72 given in Example 4.3 in Cohen (1988, p. 131), which uses the same input values. The small deviations are due to rounding errors in Cohen's analysis.

If we instead assume  $N_1 = N_2$ , how many subjects do we then need to achieve the same power? To answer this question we use an a priori analysis with the power calculated above as input and an Allocation ratio  $N_2/N_1 = 1$  to enforce equal sample sizes. All other parameters are chosen identical to that used in the above case. The result is, that we now need 84 cases in each group. Thus choosing equal sample sizes reduces the totally needed sample size considerably, from  $(260+51) = 311$  to 168 (84 +84).

### 27.4 Related tests

- Correlations: Difference from constant (one sample case)
- Correlations: Point biserial model

### 27.5 Implementation notes

The  $H_0$ -distribution is the standard normal distribution  $N(0,1)$ , the  $H_1$ -distribution is normal distribution  $N(q/s, 1)$ , where  $q$  denotes the effect size as defined above,  $N_1$  and  $N_2$  the sample size in each groups, and  $s = \sqrt{1/(N_1 - 3) + 1/(N_2 - 3)}$ .

### 27.6 Validation

The results were checked against the table in Cohen (1969, chap. 4).

## 28 z test: Correlation - inequality of two dependent Pearson r's

This procedure provides power analyses for tests of the hypothesis that two dependent Pearson correlation coefficients  $\rho_{a,b}$  and  $\rho_{c,d}$  are identical. The corresponding test statistics Z1\* and Z2\* were proposed by [Dunn and Clark \(1969\)](#) and are described in Eqns. (11) and (12) in [Steiger \(1980\)](#) (for more details on these tests, see implementation notes below).

Two correlation coefficients  $\rho_{a,b}$  and  $\rho_{c,d}$  are dependent, if at least one of the four possible correlation coefficients  $\rho_{a,c}$ ,  $\rho_{a,d}$ ,  $\rho_{b,c}$  and  $\rho_{b,d}$  between other pairs of the four data sets  $a, b, c, d$  is non-zero. Thus, in the general case where  $a, b, c$ , and  $d$  are different data sets, we do not only have to consider the two correlations under scrutiny but also four additional correlations.

In the special case, where two of the data sets are identical, the two correlations are obviously *always* dependent, because at least one of the four additional correlations mentioned above is exactly 1. Two other of the four additional correlations are identical to the two correlations under test. Thus, there remains only one additional correlation that can be freely specified. In this special case we denote the H0 correlation  $\rho_{a,b}$ , the H1 correlation  $\rho_{a,c}$  and the additional correlation  $\rho_{b,c}$ .

It is convenient to describe these two cases by two corresponding correlations matrices (that may be sub-matrices of a larger correlations matrix): A  $4 \times 4$ -matrix in the general case of four different data sets ('no common index'):

$$C_1 = \begin{pmatrix} 1 & \rho_{a,b} & \rho_{a,c} & \rho_{a,d} \\ \rho_{a,b} & 1 & \rho_{b,c} & \rho_{b,d} \\ \rho_{a,c} & \rho_{b,c} & 1 & \rho_{c,d} \\ \rho_{a,d} & \rho_{b,d} & \rho_{c,d} & 1 \end{pmatrix},$$

and a  $3 \times 3$  matrix in the special case, where one of the data sets is identical in both correlations ('common index'):

$$C_2 = \begin{pmatrix} 1 & \rho_{a,b} & \rho_{a,c} \\ \rho_{a,b} & 1 & \rho_{b,c} \\ \rho_{a,c} & \rho_{b,c} & 1 \end{pmatrix}$$

**Note:** The values for  $\rho_{x,y}$  in matrices  $C_1$  and  $C_2$  cannot be chosen arbitrarily between -1 and 1. This is easily illustrated by considering the matrix  $C_2$ : It should be obvious that we cannot, for instance, choose  $\rho_{a,b} = -\rho_{a,c}$  and  $\rho_{b,c} = 1.0$ , because the latter choice implies that the other two correlations are identical. It is, however, generally not that easy to decide whether a given matrix is a valid correlation matrix. In more complex cases the following formal criterion can be used: A given symmetric matrix is a valid correlation matrix, if and only if the matrix is positive semi-definite, that is, if all eigenvalues of the matrix are non-negative.

The null hypothesis in the general case with 'no common index' states that  $\rho_{a,b} = \rho_{c,d}$ . The (two-sided) alternative hypothesis is that these correlation coefficients are different:  $\rho_{a,b} \neq \rho_{c,d}$ :

$$\begin{aligned} H_0 : & \rho_{a,b} - \rho_{c,d} = 0 \\ H_1 : & \rho_{a,b} - \rho_{c,d} \neq 0. \end{aligned}$$

Here, G\*POWER refers to the test Z2\* described in Eq. (12) in [Steiger \(1980\)](#).

The null hypothesis in the special case of a 'common index' states that  $\rho_{a,b} = \rho_{a,c}$ . The (two-sided) alternative hypothesis is that these correlation coefficients are different:  $\rho_{a,b} \neq \rho_{a,c}$ :

$$\begin{aligned} H_0 : & \rho_{a,b} - \rho_{a,c} = 0 \\ H_1 : & \rho_{a,b} - \rho_{a,c} \neq 0. \end{aligned}$$

Here, G\*POWER refers to the test Z1\* described in Eq. (11) in [Steiger \(1980\)](#).

If the direction of the deviation  $\rho_{a,b} - \rho_{c,d}$  (or  $\rho_{a,b} - \rho_{a,c}$  in the 'common index' case) cannot be predicted a priori, a two-sided ('two-tailed') test should be used. Otherwise use a one-sided test.

### 28.1 Effect size index

In this procedure the correlation coefficient assumed under H1 is used as effect size, that is  $\rho_{c,d}$  in the general case of 'no common index' and  $\rho_{a,c}$  in the special case of a 'common index'.

To fully specify the effect size, the following additional inputs are required:

- $\rho_{a,b}$ , the correlation coefficient assumed under H0, and
- all other relevant correlation coefficients that specify the dependency between the correlations assumed in H0 and H1:  $\rho_{b,c}$  in the 'common index' case, and  $\rho_{a,c}, \rho_{a,d}, \rho_{b,c}, \rho_{b,d}$  in the general case of 'no common index'.

G\*POWER requires the correlations assumed under H0 and H1 to lie in the interval  $[-1 + \varepsilon, 1 - \varepsilon]$ , with  $\varepsilon = 10^{-6}$ , and the additional correlations to lie in the interval  $[-1, 1]$ . In a priori analyses zero effect sizes are not allowed, because this would imply an infinite sample size. In this case the additional restriction  $|\rho_{a,b} - \rho_{c,d}| > 10^{-6}$  (or  $|\rho_{a,b} - \rho_{a,c}| > 10^{-6}$ ) holds.

Why do we not use  $q$ , the effect size proposed by [Cohen \(1988\)](#) for the case of two independent correlations? The effect size  $q$  is defined as a difference between two 'Fisher z'-transformed correlation coefficients:  $q = z_1 - z_2$ , with  $z_1 = \ln((1 + r_1)/(1 - r_1))/2$ ,  $z_2 = \ln((1 + r_2)/(1 - r_2))/2$ . The choice of  $q$  as effect size is sensible for tests of independent correlations, because in this case the power of the test does not depend on the absolute value of the correlation coefficient assumed under H0, but only on the *difference*  $q$  between the transformed correlations under H0 and H1. This is no longer true for dependent correlations and we therefore used the effect size described above. (See the implementation section for a more thorough discussion of these issues.)

Although the power is not strictly independent of the value of the correlation coefficient under H0, the deviations are usually relatively small and it may therefore be convenient to use the definition of  $q$  to specify the correlation under H1 for a given correlation under H0. In this way, one can relate to the effect size conventions for  $q$  defined by [Cohen \(1969, p. 109ff\)](#) for independent correlations:

- **small**  $q = 0.1$
- **medium**  $q = 0.3$
- **large**  $q = 0.5$

The effect size drawer, which can be opened by pressing the button Determine on the left side of the effect size label, can be used to do this calculation:

The dialog performs the following transformation:  $r_2 = (a - 1)/(a + 1)$ , with  $a = \exp[-2q + \ln((1 + r_1)/(1 - r_1))]$ , where  $q$ ,  $r_1$  and  $r_2$  denote the effect size  $q$ , and the correlations coefficient assumed under H0 and H1, respectively.

## 28.2 Options

This test has no options.

## 28.3 Examples

We assume the following correlation matrix in the population:

$$C_p = \begin{pmatrix} 1 & \rho_{1,2} & \rho_{1,3} & \rho_{1,4} \\ \rho_{1,2} & 1 & \rho_{2,3} & \rho_{2,4} \\ \rho_{1,3} & \rho_{2,3} & 1 & \rho_{3,4} \\ \rho_{1,4} & \rho_{2,4} & \rho_{3,4} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.5 & 0.4 & 0.1 \\ 0.5 & 1 & 0.2 & -0.4 \\ 0.4 & 0.2 & 1 & 0.8 \\ 0.1 & -0.4 & 0.8 & 1 \end{pmatrix}$$

### 28.3.1 General case: No common index

We want to perform an a priori analysis for a one-sided test whether  $\rho_{1,4} = \rho_{2,3}$  or  $\rho_{1,4} < \rho_{2,3}$  holds.

With respect to the notation used in G\*POWER we have the following identities  $a = 1, b = 4, c = 2, d = 3$ . Thus we get: H0 correlation  $\rho_{a,b} = \rho_{1,4} = 0.1$ , H1 correlation  $\rho_{c,d} = \rho_{2,3} = 0.2$ , and  $\rho_{a,c} = \rho_{1,2} = 0.5, \rho_{a,d} = \rho_{1,3} = 0.4, \rho_{b,c} = \rho_{4,2} = \rho_{2,4} = -0.4, \rho_{b,d} = \rho_{4,3} = \rho_{3,4} = 0.8$ .

We want to know how large our samples need to be in order to achieve the error levels  $\alpha = 0.05$  and  $\beta = 0.2$ . We choose the procedure 'Correlations: Two independent Pearson r's (no common index)' and set:

- **Select**

Type of power analysis: A priori

- **Input**

Tail(s): one

H1 corr  $\rho_{cd}$ : 0.2

$\alpha$  err prob: 0.05

Power (1- $\beta$  err prob): 0.8

H0 Corr  $\rho_{ab}$ : 0.1

Corr  $\rho_{ac}$ : 0.5

Corr  $\rho_{ad}$ : 0.4

Corr  $\rho_{bc}$ : -0.4

Corr  $\rho_{bd}$ : 0.8

- **Output**

Critical z: 1.644854

Sample Size: 886

Actual Power: 0.800093

We find that the sample size in each group needs to be  $N = 886$ . How large would  $N$  be, if we instead assume that  $\rho_{1,4}$  and  $\rho_{2,3}$  are independent, that is, that  $\rho_{ac} = \rho_{ad} = \rho_{bc} = \rho_{bd} = 0$ ? To calculate this value, we may set the corresponding input fields to 0, or, alternatively, use the procedure for independent correlations with an allocation ratio  $N2/N1 = 1$ . In either case, we find a considerably larger sample size of  $N = 1183$  per data set (i.e. we correlate data vectors of length  $N = 1183$ ). This shows that the power of the test increases considerably if we take dependencies between correlations coefficients into account.

If we try to change the correlation  $\rho_{bd}$  from 0.8 to 0.9, G\*POWER shows an error message stating: 'The correlation matrix is not valid, that is, not positive semi-definite'. This indicates that the matrix  $C_p$  with  $\rho_{3,4}$  changed to 0.9 is not a possible correlation matrix.

### 28.3.2 Special case: Common index

Assuming again the population correlation matrix  $C_p$  shown above, we want to do an a priori analysis for the test whether  $\rho_{1,3} = \rho_{2,3}$  or  $\rho_{1,3} > \rho_{2,3}$  holds. With respect to the notation used in G\*POWER we have the following identities:  $a = 3$  (the common index),  $b = 1$  (the index of the second data set entering in the correlation assumed under H0, here  $\rho_{a,b} = \rho_{3,1} = \rho_{1,3}$ ), and  $c = 2$  (the index of the remaining data set).

Thus, we have: H0 correlation  $\rho_{a,b} = \rho_{3,1} = \rho_{1,3} = 0.4$ , H1 correlation  $\rho_{a,c} = \rho_{3,2} = \rho_{2,3} = 0.2$ , and  $\rho_{b,c} = \rho_{1,2} = 0.5$

For this effect size we want to calculate how large our sample size needs to be, in order to achieve the error levels  $\alpha = 0.05$  and  $\beta = 0.2$ . We choose the procedure 'Correlations: Two independent Pearson r's (common index)' and set:

- **Select**

Type of power analysis: A priori

- **Input**

Tail(s): one

H1 corr  $\rho_{ac}$ : 0.2

$\alpha$  err prob: 0.05

Power (1- $\beta$  err prob): 0.8

H0 Corr  $\rho_{ab}$ : 0.4

Corr  $\rho_{bc}$ : 0.5

- **Output**

Critical z: 1.644854

Sample Size: 144

Actual Power: 0.801161

The answer is that we need sample sizes of 144 in each group (i.e. the correlations are calculated between data vectors of length  $N = 144$ ).

### 28.3.3 Sensitivity analyses

We now assume a scenario that is identical to that described above, with the exception that  $\rho_{b,c} = -0.6$ . We want to know, what the minimal H1 correlation  $\rho_{a,c}$  is that we can detect with  $\alpha = 0.05$  and  $\beta = 0.2$  given a sample size  $N = 144$ . In this sensitivity analysis, we have in general two possible solutions, namely one for  $\rho_{a,c} \leq \rho_{a,b}$  and one for  $\rho_{a,c} \geq \rho_{a,b}$ . The relevant settings for the former case are:

- **Select**

Type of power analysis: Sensitivity

- **Input**

Tail(s): one

Effect direction:  $\rho_{ac} \leq \rho_{ab}$

$\alpha$  err prob: 0.05

Power (1- $\beta$  err prob): 0.8

Sample Size: 144

H0 Corr  $\rho_{ab}$ : 0.4

Corr  $\rho_{bc}$ : -0.6

- **Output**

Critical z: -1.644854

H1 corr  $\rho_{ac}$ : 0.047702

The result is that the error levels are as requested or lower if the H1 correlation  $\rho_{a,c}$  is equal to or lower than 0.047702.

We now try to find the corresponding H1 correlation that is larger than  $\rho_{a,b} = 0.4$ . To this end, we change the effect size direction in the settings shown above, that is, we choose  $\rho_{a,c} \geq \rho_{a,b}$ . In this case, however, G\*POWER shows an error message, indicating that no solution was found. The reason is, that there is no H1 correlation  $\rho_{a,c} \geq \rho_{a,b}$  that leads to a valid (i.e. positive semi-definite) correlation matrix and simultaneously ensures the requested error levels. To indicate a missing result the output for the H1 correlation is set to the nonsensical value 2.

G\*POWER checks in both the 'common index' case and the general case with no common index, whether the correlation matrix is valid and outputs a H1 correlation of 2 if no solution is found. This is also true in the XY-plot if the H1 correlation is the dependent variable: Combinations of input values for which no solution can be found show up with a nonsensical value of 2.

### 28.3.4 Using the effect size dialog

## 28.4 Related tests

- Correlations: Difference from constant (one sample case)
- Correlations: Point biserial model
- Correlations: Two independent Pearson r's (two sample case)

## 28.5 Implementation notes

### 28.5.1 Background

Let  $X_1, \dots, X_K$  denote multinormally distributed random variables with mean vector  $\mu$  and covariance matrix  $C$ . A sample of size  $N$  from this  $K$ -dimensional distribution

leads to a  $N \times K$  data matrix, and pair-wise correlation of all columns to a  $K \times K$  correlation matrix. By drawing  $M$  samples of size  $N$  one can compute  $M$  such correlation matrices, and one can determine the variances  $\sigma_{a,b}^2$  of the sample of  $M$  correlation coefficients  $r_{a,b}$ , and the covariances  $\sigma_{a,b;c,d}$  between samples of size  $M$  of two different correlations  $r_{a,b}, r_{c,d}$ . For  $M \rightarrow \infty$ , the elements of  $\Psi$ , which denotes the asymptotic variance-covariance matrix of the correlations times  $N$ , are given by [see Eqns (1) and (2) in Steiger (1980)]:

$$\Psi_{a,b;a,b} = N\sigma_{a,b}^2 = (1 - \rho_{a,b}^2)^2 \quad (11)$$

$$\begin{aligned} \Psi_{a,b;c,d} &= N\sigma_{a,b;c,d} \\ &= [(\rho_{a,c} - \rho_{a,b}\rho_{b,c})(\rho_{b,d} - \rho_{b,c}\rho_{c,d}) + \dots \\ &\quad (\rho_{a,d} - \rho_{a,c}\rho_{c,d})(\rho_{b,c} - \rho_{a,b}\rho_{a,c}) + \dots \\ &\quad (\rho_{a,c} - \rho_{a,d}\rho_{c,d})(\rho_{b,d} - \rho_{a,b}\rho_{a,d}) + \dots \\ &\quad (\rho_{a,d} - \rho_{a,b}\rho_{b,d})(\rho_{b,c} - \rho_{b,d}\rho_{c,d})]/2 \end{aligned} \quad (12)$$

When two correlations have an index in common, the expression given in Eq (12) simplifies to [see Eq (3) in Steiger (1980)]:

$$\begin{aligned} \Psi_{a,b;a,c} &= N\sigma_{a,b;a,c} \\ &= \rho_{b,c}(1 - \rho_{a,b}^2 - \rho_{a,c}^2) - \\ &\quad - \rho_{a,b}\rho_{a,c}(1 - \rho_{a,c}^2 - \rho_{a,b}^2 - \rho_{b,c}^2)/2 \end{aligned} \quad (13)$$

If the raw sample correlations  $r_{a,b}$  are transformed by the Fisher r-to-z transform:

$$z_{a,b} = \frac{1}{2} \ln \left( \frac{1 + r_{a,b}}{1 - r_{a,b}} \right)$$

then the elements of the variance-covariance matrix times  $(N-3)$  of the transformed raw correlations are [see Eqs. (9)-(11) in Steiger (1980)]:

$$\begin{aligned} c_{a,b;a,b} &= (N-3)\sigma_{z_{a,b};z_{a,b}} = 1 \\ c_{a,b;c,d} &= (N-3)\sigma_{z_{a,b};z_{c,d}} = \frac{\Psi_{a,b;c,d}}{(1 - \rho_{a,b}^2)(1 - \rho_{c,d}^2)} \\ c_{a,b;a,c} &= (N-3)\sigma_{z_{a,b};z_{a,c}} = \frac{\Psi_{a,b;a,c}}{(1 - \rho_{a,b}^2)(1 - \rho_{a,c}^2)} \end{aligned}$$

### 28.5.2 Test statistics

The test statistics proposed by Dunn and Clark (1969) are [see Eqns (12) and (13) in Steiger (1980)]:

$$Z1^* = (z_{a,b} - z_{a,c}) / \sqrt{\frac{2 - 2s_{a,b;a,c}}{N-3}}$$

$$Z2^* = (z_{a,b} - z_{c,d}) / \sqrt{\frac{2 - 2s_{a,b;c,d}}{N-3}}$$

where  $s_{a,b;a,c}$  and  $s_{a,b;c,d}$  denote sample estimates of the covariances  $c_{a,b;a,c}$  and  $c_{a,b;c,d}$  between the transformed correlations, respectively.

**Note:** 1. The SD of the difference of  $z_{a,b} - z_{a,c}$  given in the denominator of the formula for  $Z1^*$  depends on the value of  $\rho_{a,b}$  and  $\rho_{a,c}$ ; the same holds analogously for  $Z2^*$ . 2. The only difference between  $Z2^*$  and the z-statistic used for independent correlations is that in the latter the covariance  $s_{a,b;c,d}$  is assumed to be zero.

### 28.5.3 Central and noncentral distributions in power calculations

For the special case with a 'common index' the H0-distribution is the standard normal distribution  $N(0,1)$  and the H1 distribution the normal distribution  $N(m_1, s_1)$ , with

$$\begin{aligned} s_0 &= \sqrt{(2 - 2c_0)/(N - 3)}, \\ &\quad \text{with } c_0 = c_{a,b;a,c} \text{ for H0, i.e. } \rho_{a,c} = \rho_{a,b} \\ m_1 &= (z_{a,b} - z_{a,c})/s_0 \\ s_1 &= \left[ \sqrt{(2 - 2c_1)/(N - 3)} \right] / s_0, \\ &\quad \text{with } c_1 = c_{a,b;a,c} \text{ for H1.} \end{aligned}$$

In the general case 'without a common index' the H0-distribution is also the standard normal distribution  $N(0,1)$  and the H1 distribution the normal distribution  $N(m_1, s_1)$ , with

$$\begin{aligned} s_0 &= \sqrt{(2 - 2c_0)/(N - 3)}, \\ &\quad \text{with } c_0 = c_{a,b;c,d} \text{ for H0, i.e. } \rho_{c,d} = \rho_{a,b} \\ m_1 &= (z_{a,b} - z_{c,d})/s_0 \\ s_1 &= \left[ \sqrt{(2 - 2c_1)/(N - 3)} \right] / s_0, \\ &\quad \text{with } c_1 = c_{a,b;a,c} \text{ for H1.} \end{aligned}$$

## 28.6 Validation

The results were checked against Monte-Carlo simulations.

## 29 Z test: Multiple Logistic Regression

A logistic regression model describes the relationship between a binary response variable  $Y$  (with  $Y = 0$  and  $Y = 1$  denoting non-occurrence and occurrence of an event, respectively) and one or more independent variables (covariates or predictors)  $X_i$ . The variables  $X_i$  are themselves random variables with probability density function  $f_X(x)$  (or probability distribution  $f_X(x)$  for discrete  $X$ ).

In a simple logistic regression with one covariate  $X$  the assumption is that the probability of an event  $P = Pr(Y = 1)$  depends on  $X$  in the following way:

$$P = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

For  $\beta_1 \neq 0$  and continuous  $X$  this formula describes a smooth S-shaped transition of the probability for  $Y = 1$  from 0 to 1 ( $\beta_1 > 0$ ) or from 1 to 0 ( $\beta_1 < 0$ ) with increasing  $x$ . This transition gets steeper with increasing  $\beta_1$ . Rearranging the formula leads to:  $\log(P/(1 - P)) = \beta_0 + \beta_1 X$ . This shows that the logarithm of the odds  $P/(1 - P)$ , also called a logit, on the left side of the equation is linear in  $X$ . Here,  $\beta_1$  is the slope of this linear relationship.

The interesting question is whether covariate  $X_i$  is related to  $Y$  or not. Thus, in a *simple logistic regression model*, the null and alternative hypothesis for a two-sided test are:

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ H_1 : \beta_1 &\neq 0. \end{aligned}$$

The procedures implemented in G\*POWER for this case estimates the power of the Wald test. The standard normally distributed test statistic of the Wald test is:

$$z = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)/N}} = \frac{\hat{\beta}_1}{SE(\beta_1)}$$

where  $\hat{\beta}_1$  is the maximum likelihood estimator for parameter  $\beta_1$  and  $\text{var}(\hat{\beta}_1)$  the variance of this estimate.

In a *multiple logistic regression model*  $\log(P/(1 - P)) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$  the effect of a specific covariate in the presence of other covariates is tested. In this case the null hypothesis is  $H_0 : [\beta_1, \beta_2, \dots, \beta_p] = [0, \beta_2, \dots, \beta_p]$  and the alternative  $H_1 : [\beta_1, \beta_2, \dots, \beta_p] = [\hat{\beta}_1, \beta_2, \dots, \beta_p]$ , where  $\hat{\beta}_1 \neq 0$ .

### 29.1 Effect size index

In the simple logistic model the effect of  $X$  on  $Y$  is given by the size of the parameter  $\beta_1$ . Let  $p_1$  denote the probability of an event under  $H_0$ , that is  $\exp(\beta_0) = p_1/(1 - p_1)$ , and  $p_2$  the probability of an event under  $H_1$  at  $X = 1$ , that is  $\exp(\beta_0 + \beta_1) = p_2/(1 - p_2)$ . Then  $\exp(\beta_0 + \beta_1)/\exp(\beta_0) = \exp(\beta_1) = [p_2/(1 - p_2)]/[p_1/(1 - p_1)] := \text{odds ratio } OR$ , which implies  $\beta_1 = \log[OR]$ .

Given the probability  $p_1$  (input field  $Pr(Y=1|X=1)$   $H_0$ ) the effect size is specified either directly by  $p_2$  (input field  $Pr(Y=1|X=1)$   $H_1$ ) or optionally by the odds ratio (OR) (input field Odds ratio). Setting  $p_2 = p_1$  or equivalently  $OR = 1$  implies  $\beta_1 = 0$  and thus an effect size of zero. An effect size of zero must not be used in a priori analyses.

Besides these values the following additional inputs are needed

- $R^2$  other  $X$ .

In models with more than one covariate, the influence of the other covariates  $X_2, \dots, X_p$  on the power of the test can be taken into account by using a correction factor. This factor depends on the proportion  $R^2 = \rho_{1.23\dots p}^2$  of the variance of  $X_1$  explained by the regression relationship with  $X_2, \dots, X_p$ . If  $N$  is the sample size considering  $X_1$  alone, then the sample size in a setting with additional covariates is:  $N' = N/(1 - R^2)$ . This correction for the influence of other covariates has been proposed by Hsieh, Bloch, and Larsen (1998).  $R^2$  must lie in the interval  $[0, 1]$ .

- $X$  distribution:

Distribution of the  $X_i$ . There are 7 options:

1. Binomial [ $P(k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$ , where  $k$  is the number of successes ( $X = 1$ ) in  $N$  trials of a Bernoulli process with probability of success  $\pi$ ,  $0 < \pi < 1$ ]
2. Exponential [ $f(x) = (1/\lambda)e^{-x/\lambda}$ , exponential distribution with parameter  $\lambda > 0$ ]
3. Lognormal [ $f(x) = 1/(x\sigma\sqrt{2\pi}) \exp[-(\ln x - \mu)^2/(2\sigma^2)]$ , lognormal distribution with parameters  $\mu$  and  $\sigma > 0$ ]
4. Normal [ $f(x) = 1/(\sigma\sqrt{2\pi}) \exp[-(x - \mu)^2/(2\sigma^2)]$ , normal distribution with parameters  $\mu$  and  $\sigma > 0$ ]
5. Poisson [ $P(X = k) = (\lambda^k/k!)e^{-\lambda}$ , Poisson distribution with parameter  $\lambda > 0$ ]
6. Uniform [ $f(x) = 1/(b - a)$  for  $a \leq x \leq b$ ,  $f(x) = 0$  otherwise, continuous uniform distribution in the interval  $[a, b]$ ,  $a < b$ ]
7. Manual (Allows to manually specify the variance of  $\hat{\beta}$  under  $H_0$  and  $H_1$ )

G\*POWER provides two different types of procedure to calculate power: An enumeration procedure and large sample approximations. The Manual mode is only available in the large sample procedures.

### 29.2 Options

**Input mode** You can choose between two input modes for the effect size: The effect size may be given by either specifying the two probabilities  $p_1, p_2$  defined above, or instead by specifying  $p_1$  and the odds ratio  $OR$ .

**Procedure** G\*POWER provides two different types of procedure to estimate power. An "enumeration procedure" proposed by Lyles, Lin, and Williamson (2007) and large sample approximations. The enumeration procedure seems to provide reasonable accurate results over a wide range of situations, but it can be rather slow and may need large amounts of memory. The large sample approximations are much faster. Results of Monte-Carlo simulations indicate that the accuracy of the procedures proposed by Demidenko (2007) and Hsieh et al. (1998) are comparable to that of the enumeration procedure for  $N > 200$ . The procedure base on the work of Demidenko (2007) is more general and

slightly more accurate than that proposed by Hsieh et al. (1998). We thus recommend to use the procedure proposed by Demidenko (2007) as standard procedure. The enumeration procedure of Lyles et al. (2007) may be used to validate the results (if the sample size is not too large). It must also be used, if one wants to compute the power for likelihood ratio tests.

1. The enumeration procedure provides power analyses for the Wald-test and the Likelihood ratio test. The general idea is to construct an exemplary data set with weights that represent response probabilities given the assumed values of the parameters of the  $X$ -distribution. Then a fit procedure for the generalized linear model is used to estimate the variance of the regression weights (for Wald tests) or the likelihood ratio under  $H_0$  and  $H_1$  (for likelihood ratio tests). The size of the exemplary data set increases with  $N$  and the enumeration procedure may thus be rather slow (and may need large amounts of computer memory) for large sample sizes. The procedure is especially slow for analysis types other than "post hoc", which internally call the power routine several times. By specifying a threshold sample size  $N$  you can restrict the use of the enumeration procedure to sample sizes  $< N$ . For sample sizes  $\geq N$  the large sample approximation selected in the option dialog is used. Note: If a computation takes too long you can abort it by pressing the ESC key.
2. G\*POWER provides two different large sample approximations for a Wald-type test. Both rely on the asymptotic normal distribution of the maximum likelihood estimator for parameter  $\beta_1$  and are related to the method described by Whittemore (1981). The accuracy of these approximation increases with sample size, but the deviation from the true power may be quite noticeable for small and moderate sample sizes. This is especially true for  $X$ -distributions that are not symmetric about the mean, i.e. the lognormal, exponential, and poisson distribution, and the binomial distribution with  $\pi \neq 1/2$ . The approach of Hsieh et al. (1998) is restricted to binary covariates and covariates with standard normal distribution. The approach based on Demidenko (2007) is more general and usually more accurate and is recommended as standard procedure. For this test, a variance correction option can be selected that compensates for variance distortions that may occur in skewed  $X$  distributions (see implementation notes). If the Hsieh procedure is selected, the program automatically switches to the procedure of Demidenko if a distribution other than the standard normal or the binomial distribution is selected.

### 29.3 Possible problems

As illustrated in Fig. 36, the power of the test does not always increase monotonically with effect size, and the maximum attainable power is sometimes less than 1. In particular, this implies that in a sensitivity analysis the requested power cannot always be reached. From version 3.1.8 on G\*POWER returns in these cases the effect size which maximizes the power in the selected direction (output field "Actual power"). For an overview about possible problems, we

recommend to check the dependence of power on effect size in the plot window.

Covariates with a Lognormal distribution are especially problematic, because this distribution may have a very long tail (large positive skew) for larger values of  $m$  and  $s$  and may thus easily lead to numerical problems. In version 3.1.8 the numerical stability of the procedure has been considerably improved. In addition the power with highly skewed distributions may behave in an unintuitive manner and you should therefore check such cases carefully.

### 29.4 Examples

We first consider a model with a single predictor  $X$ , which is normally distributed with  $m = 0$  and  $\sigma = 1$ . We assume that the event rate under  $H_0$  is  $p_1 = 0.5$  and that the event rate under  $H_1$  is  $p_2 = 0.6$  for  $X = 1$ . The odds ratio is then  $OR = (0.6/0.4)/(0.5/0.5) = 1.5$ , and we have  $\beta_1 = \log(OR) \approx 0.405$ . We want to estimate the sample size necessary to achieve in a two-sided test with  $\alpha = 0.05$  a power of at least 0.95. We want to specify the effect size in terms of the odds ratio. When using the procedure of Hsieh et al. (1998) the input and output is as follows:

- **Select**  
*Statistical test:* Logistic Regression  
*Type of power analysis:* A priori
- **Options:**  
*Effect size input mode:* Odds ratio  
*Procedure:* Hsieh et al. (1998)
- **Input**  
*Tail(s):* Two  
*Odds ratio:* 1.5  
*Pr(Y=1) H0:* 0.5  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95  
*R<sup>2</sup> other X:* 0  
*X distribution:* Normal  
*X parm  $\mu$ :* 0  
*X parm  $\sigma$ :* 1
- **Output**  
*Critical z:* 1.959964  
*Total sample size:* 317  
*Actual power:* 0.950486

The results indicate that the necessary sample size is 317. This result replicates the value in Table II in Hsieh et al. (1998) for the same scenario. Using the other large sample approximation proposed by Demidenko (2007) we instead get  $N = 337$  with variance correction and  $N = 355$  without.

In the enumeration procedure proposed by Lyles et al. (2007) the  $\chi^2$ -statistic is used and the output is

- **Output**  
*Noncentrality parameter  $\lambda$ :* 13.029675  
*Critical  $\chi^2$ :* 3.841459  
*Df:* 1  
*Total sample size:* 358  
*Actual power:* 0.950498

Thus, this routine estimates the minimum sample size in this case to be  $N = 358$ .

In a Monte-Carlo simulation of the Wald test in the above scenario with 50000 independent cases we found a mean power of 0.940, 0.953, 0.962, and 0.963 for samples sizes 317, 337, 355, and 358, respectively. This indicates that in this case the method based on Demidenko (2007) with variance correction yields the best approximation.

We now assume that we have additional covariates and estimate the squared multiple correlation with these others covariates to be  $R^2 = 0.1$ . All other conditions are identical. The only change we need to make is to set the input field  $R^2$  other  $X$  to 0.1. Under this condition the necessary sample size increases from 337 to a value of 395 when using the procedure of Demidenko (2007) with variance correction.

As an example for a model with one binary covariate  $X$  we choose the values of the fourth example in Table I in Hsieh et al. (1998). That is, we assume that the event rate under  $H_0$  is  $p_1 = 0.05$ , and the event rate under  $H_0$  with  $X = 1$  is  $p_2 = 0.1$ . We further assume a balanced design ( $\pi = 0.5$ ) with equal sample frequencies for  $X = 0$  and  $X = 1$ . Again we want to estimate the sample size necessary to achieve in a two-sided test with  $\alpha = 0.05$  a power of at least 0.95. We want to specify the effect size directly in terms of  $p_1$  and  $p_2$ :

- **Select**  
*Statistical test:* Logistic Regression  
*Type of power analysis:* A priori
- **Options:**  
*Effect size input mode:* Two probabilities
- **Input**  
*Tail(s):* Two  
*Pr(Y=1|X=1) H1:* 0.1  
*Pr(Y=1|X=1) H0:* 0.05  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95  
 *$R^2$  other X:* 0  
*X Distribution:* Binomial  
*X parm  $\pi$ :* 0.5
- **Output**  
*Critical z:* 1.959964  
*Total sample size:* 1437  
*Actual power:* 0.950068

According to these results the necessary sample size is 1437. This replicates the sample size given in Table I in Hsieh et al. (1998) for this example. This result is confirmed by the procedure proposed by Demidenko (2007) with variance correction. The procedure without variance correction and the procedure of Lyles et al. (2007) for the Wald-test yield  $N = 1498$ . In Monte-Carlo simulations of the Wald test (50000 independent cases) we found mean power values of 0.953, and 0.961 for sample sizes 1437, and 1498, respectively. According to these results, the procedure of Demidenko (2007) with variance correction again yields the best power estimate for the tested scenario.

Changing just the parameter of the binomial distribution (the prevalence rate) to a lower value  $\pi = 0.2$ , increases the sample size to a value of 2158. Changing  $\pi$  to an equally unbalanced but higher value, 0.8, increases the required

sample size further to 2368 (in both cases the Demidenko procedure with variance correction was used). These examples demonstrate the fact that a balanced design requires a smaller sample size than an unbalanced design, and a low prevalence rate requires a smaller sample size than a high prevalence rate (Hsieh et al., 1998, p. 1625).

## 29.5 Related tests

- Poisson regression

## 29.6 Implementation notes

### 29.6.1 Enumeration procedure

The procedures for the Wald- and Likelihood ratio tests are implemented exactly as described in Lyles et al. (2007).

### 29.6.2 Large sample approximations

The large sample procedures for the univariate case are both related to the approach outlined in Whittemore (1981). The correction for additional covariates has been proposed by Hsieh et al. (1998). As large sample approximations they get more accurate for larger sample sizes.

**Demidenko-procedure** In the procedure based on Demidenko (2007, 2008), the  $H_0$  distribution is the standard normal distribution  $N(0, 1)$ , the  $H_1$  distribution the normal distribution  $N(m_1, s_1)$  with:

$$m_1 = \left[ \sqrt{N(1 - R^2)/v_1} \right] \beta_1 \quad (14)$$

$$s_1 = \sqrt{(av_0 + (1 - a)v_1)/v_1} \quad (15)$$

where  $N$  denotes the sample size,  $R^2$  the squared multiple correlation coefficient of the covariate of interest on the other covariates, and  $v_1$  the variance of  $\hat{\beta}_1$  under  $H_1$ , whereas  $v_0$  is the variance of  $\hat{\beta}_1$  for  $H_0$  with  $b_0^* = \ln(\mu/(1 - \mu))$ , with  $\mu = \int f_X(x) \exp(b_0 + b_1 x)/(1 + \exp(b_0 + b_1 x)) dx$ . For the procedure without variance correction  $a = 0$ , that is,  $s_1 = 1$ . In the procedure with variance correction  $a = 0.75$  for the lognormal distribution,  $a = 0.85$  for the binomial distribution, and  $a = 1$ , that is  $s_1 = \sqrt{v_0/v_1}$ , for all other distributions.

The motivation of the above setting is as follows: Under  $H_0$  the Wald-statistic has asymptotically a standard normal distribution, whereas under  $H_1$ , the asymptotic normal distribution of the test statistic has mean  $\beta_1/se(\hat{\beta}_1) = \beta_1/\sqrt{v_1/N}$ , and standard deviation 1. The variance for finite  $n$  is, however, biased (the degree depending on the  $X$ -distribution) and  $\sqrt{(av_0 + (1 - a)v_1)/v_1}$ , which is close to 1 for symmetric  $X$  distributions but deviate from 1 for skewed ones, gives often a much better estimate of the actual standard deviation of the distribution of  $\hat{\beta}_1$  under  $H_1$ . This was confirmed in extensive Monte-Carlo simulations with a wide range of parameters and  $X$  distributions.

The procedure uses the result that the  $(m + 1)$  maximum likelihood estimators  $\beta_0, \beta_1, \dots, \beta_m$  are asymptotically normally distributed, where the variance-covariance matrix is



given by the inverse of the  $(m + 1) \times (m + 1)$  Fisher information matrix  $I$ . The  $(i, j)$ th element of  $I$  is given by

$$I_{ij} = -E \left[ \frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j} \right]$$

$$= NE[X_i X_j \frac{\exp(\beta_0 + \beta_1 X_i + \dots + \beta_m X_m)}{1 + \exp(\beta_0 + \beta_1 X_i + \dots + \beta_m X_m)}^2]$$

Thus, in the case of one continuous predictor,  $I$  is a  $4 \times 4$  matrix with elements:

$$I_{00} = \int_{-\infty}^{\infty} G_X(x) dx$$

$$I_{10} = I_{01} = \int_{-\infty}^{\infty} x G_X(x) dx$$

$$I_{11} = \int_{-\infty}^{\infty} x^2 G_X(x) dx$$

with

$$G_X(x) := f_X(x) \frac{\exp(\beta_0 + \beta_1 x)}{(1 + \exp(\beta_0 + \beta_1 x))^2},$$

where  $f_X(x)$  is the PDF of the  $X$  distribution (for discrete predictors, the integrals must be replaced by corresponding sums). The element  $M_{11}$  of the inverse of this matrix ( $M = I^{-1}$ ), that is the variance of  $\beta_1$ , is given by:  $M_{11} = Var(\beta) = I_{00}/(I_{00}I_{11} - I_{01}^2)$ . In G\*POWER, numerical integration is used to compute these integrals.

To estimate the variance of  $\hat{\beta}_1$  under H1, the parameter  $\beta_0$  and  $\beta_1$  in the equations for  $I_{ij}$  are chosen as implied by the input, that is  $\beta_0 = \log[p_1/(1 - p_1)]$ ,  $\beta_1 = \log[OR]$ . To estimate the variance under H0, one chooses  $\beta_1 = 0$  and  $\beta_0 = \beta_0^*$ , where  $\beta_0^*$  is chosen as defined above.

**Hsieh et al.-procedure** The procedures proposed in Hsieh et al. (1998) are used. The samples size formula for continuous, normally distributed covariates is [Eqn (1) in Hsieh et al. (1998)]:

$$N = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{p_1(1 - p_1)\hat{\beta}^2}$$

where  $\hat{\beta} = \log([p_1/(1 - p_1)]/[p_2/(1 - p_2)])$  is the tested effect size, and  $p_1, p_2$  are the event rates at the mean of  $X$  and one  $SD$  above the mean, respectively.

For binary covariates the sample size formula is [Eqn (2) in Hsieh et al. (1998)]:

$$N = \frac{[z_{1-\alpha} \sqrt{pqB} + z_{1-\beta} \sqrt{p_1 q_1 + p_2 q_2 (1 - B)/B}]^2}{(p_1 - p_2)^2 (1 - B)}$$

with  $q_i := 1 - p_i$ ,  $q := (1 - p)$ , where  $p = (1 - B)p_1 + Bp_2$  is the overall event rate,  $B$  is the proportion of the sample with  $X = 1$ , and  $p_1, p_2$  are the event rates at  $X = 0$  and  $X = 1$  respectively.

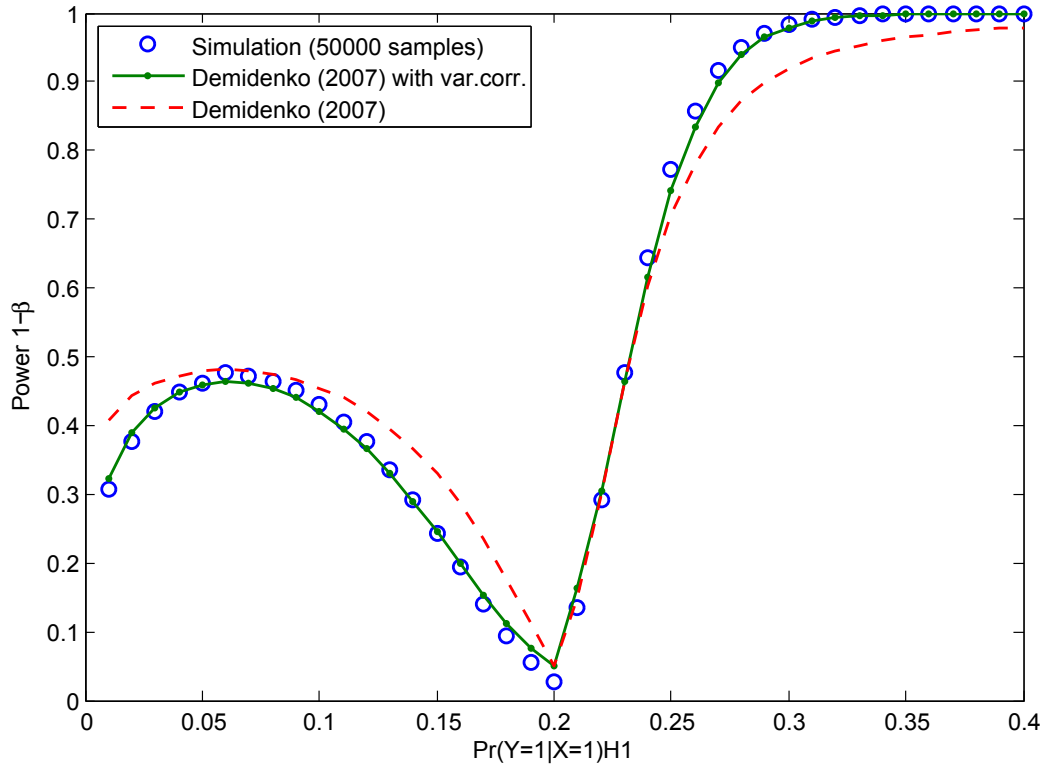
## 29.7 Validation

To check the correct implementation of the procedure proposed by Hsieh et al. (1998), we replicated all examples presented in Tables I and II in Hsieh et al. (1998). The single deviation found for the first example in Table I on p. 1626 (sample size of 1281 instead of 1282) is probably due to

rounding errors. Further checks were made against the corresponding routine in PASS Hintze (2006) and we usually found complete correspondence. For multiple logistic regression models with  $R^2$  other  $X > 0$ , however, our values deviated slightly from the result of PASS. We believe that our results are correct. There are some indications that the reason for these deviations is that PASS internally rounds or truncates sample sizes to integer values.

To validate the procedures of Demidenko (2007) and Lyles et al. (2007) we conducted Monte-Carlo simulations of Wald tests for a range of scenarios. In each case 150000 independent cases were used. This large number of cases is necessary to get about 3 digits precision. In our experience, the common praxis to use only 5000, 2000 or even 1000 independent cases in simulations (Hsieh, 1989; Lyles et al., 2007; Shieh, 2001) may lead to rather imprecise and thus misleading power estimates.

Table (2) shows the errors in the power estimates for different procedures. The labels "Dem(c)" and "Dem" denote the procedure of Demidenko (2007) with and without variance correction, the labels "LLW(W)" and "LLW(L)" the procedure of Lyles et al. (2007) for the Wald test and the likelihood ratio test, respectively. All six predefined distributions were tested (the parameters are given in the table head). The following 6 combinations of  $Pr(Y=1|X=1)$  H0 and sample size were used: (0.5,200),(0.2,300),(0.1,400),(0.05,600),(0.02,1000). These values were fully crossed with four odds ratios (1.3, 1.5, 1.7, 2.0), and two alpha values (0.01, 0.05). Max and mean errors were calculated for all power values  $< 0.999$ . The results show that the precision of the procedures depend on the  $X$  distribution. The procedure of Demidenko (2007) with the variance correction proposed here, predicted the simulated power values best.



**Figure 36:** Results of a simulation study investigating the power as a function of effect size for Logistic regression with covariate  $X \sim \text{Lognormal}(0,1)$ ,  $p_1 = 0.2$ ,  $N = 100$ ,  $\alpha = 0.05$ , one-sided. The plot demonstrates that the method of Demidenko (2007), especially if combined with variance correction, predicts the power values in the simulation rather well. Note also that here the power does not increase monotonically with effect size (on the left side from  $\Pr(Y = 1|X = 0) = p_1 = 0.2$ ) and that the maximum attainable power may be restricted to values clearly below one.

max error							
procedure	tails	bin(0.3)	exp(1)	lognorm(0,1)	norm(0,1)	poisson(1)	uni(0,1)
Dem(c)	1	0.0132	0.0279	0.0309	0.0125	0.0185	0.0103
Dem(c)	2	0.0125	0.0326	0.0340	0.0149	0.0199	0.0101
Dem	1	0.0140	0.0879	0.1273	0.0314	0.0472	0.0109
Dem	2	0.0145	0.0929	0.1414	0.0358	0.0554	0.0106
LLW(W)	1	0.0144	0.0878	0.1267	0.0346	0.0448	0.0315
LLW(W)	2	0.0145	0.0927	0.1407	0.0399	0.0541	0.0259
LLW(L)	1	0.0174	0.0790	0.0946	0.0142	0.0359	0.0283
LLW(L)	2	0.0197	0.0828	0.1483	0.0155	0.0424	0.0232
mean error							
procedure	tails	binomial	exp	lognorm	norm	poisson	uni
Dem(c)	1	0.0045	0.0113	0.0120	0.0049	0.0072	0.0031
Dem(c)	2	0.0064	0.0137	0.0156	0.0061	0.0083	0.0039
Dem	1	0.0052	0.0258	0.0465	0.0069	0.0155	0.0035
Dem	2	0.0049	0.0265	0.0521	0.0080	0.0154	0.0045
LLW(W)	1	0.0052	0.0246	0.0469	0.0078	0.0131	0.0111
LLW(W)	2	0.0049	0.0253	0.0522	0.0092	0.0139	0.0070
LLW(L)	1	0.0074	0.0174	0.0126	0.0040	0.0100	0.0115
LLW(L)	2	0.0079	0.0238	0.0209	0.0047	0.0131	0.0073

**Table 1:** Results of simulation. Shown are the maximum and mean error in power for different procedures (see text).

## 30 Z test: Poisson Regression

A Poisson regression model describes the relationship between a Poisson distributed response variable  $Y$  (a count) and one or more independent variables (covariates or predictors)  $X_i$ , which are themselves random variables with probability density  $f_X(x)$ .

The probability of  $y$  events during a fixed 'exposure time'  $t$  is:

$$Pr(Y = y|\lambda, t) = \frac{e^{-\lambda t}(\lambda t)^y}{y!}$$

It is assumed that the parameter  $\lambda$  of the Poisson distribution, the mean incidence rate of an event during exposure time  $t$ , is a function of the  $X_i$ 's. In the Poisson regression model considered here,  $\lambda$  depends on the covariates  $X_i$  in the following way:

$$\lambda = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m)$$

where  $\beta_0, \dots, \beta_m$  denote regression coefficients that are estimated from the data [Frome \(1986\)](#).

In a simple Poisson regression with just one covariate  $X_1$ , the procedure implemented in G\*POWER estimates the power of the Wald test, applied to decide whether covariate  $X_1$  has an influence on the event rate or not. That is, the null and alternative hypothesis for a two-sided test are:

$$\begin{aligned} H_0 : \beta_1 &= 0 \\ H_1 : \beta_1 &\neq 0. \end{aligned}$$

The standard normally distributed test statistic of the Wald test is:

$$z = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)/N}} = \frac{\hat{\beta}_1}{SE(\beta_1)}$$

where  $\hat{\beta}_1$  is the maximum likelihood estimator for parameter  $\beta_1$  and  $\text{var}(\hat{\beta}_1)$  the variance of this estimate.

In a *multiple Poisson regression model*:  $\lambda = \exp(\beta_0 + \beta_1 X_1 + \dots + \beta_m X_m)$ ,  $m > 1$ , the effect of a specific covariate in the presence of other covariates is tested. In this case the null and alternative hypotheses are:

$$\begin{aligned} H_0 : [\beta_1, \beta_2, \dots, \beta_m] &= [0, \beta_2, \dots, \beta_m] \\ H_1 : [\beta_1, \beta_2, \dots, \beta_m] &= [\beta_1^*, \beta_2, \dots, \beta_m] \end{aligned}$$

where  $\beta_1^* > 0$ .

### 30.1 Effect size index

The effect size is specified by the ratio  $R$  of  $\lambda$  under  $H_1$  to  $\lambda$  under  $H_0$ :

$$R = \frac{\exp(\beta_0 + \beta_1 X_1)}{\exp(\beta_0)} = \exp \beta_1 X_1$$

The following additional inputs are needed

- $\text{Exp}(\beta_1)$ .

This is the value of the  $\lambda$ -ratio  $R$  defined above for  $X = 1$ , that is the relative increase of the event rate over the base event rate  $\exp(\beta_0)$  assumed under  $H_0$ , if  $X$  is increased one unit. If, for instance, a 10% increase over

the base rate is assumed if  $X$  is increased by one unit, this value is set to  $(100+10)/100 = 1.1$ .

An input of  $\exp(\beta_1) = 1$  corresponds to "no effect" and must not be used in a priori calculations.

- Base rate  $\exp(\beta_0)$ .  
This is the mean event rate assumed under  $H_0$ . It must be greater than 0.
- Mean exposure.  
This is the time unit during which the events are counted. It must be greater than 0.
- $R^2$  other  $X$ .

In models with more than one covariate, the influence of the other covariates  $X_2, \dots, X_p$  on the power of the test can be taken into account by using a correction factor. This factor depends on the proportion  $R^2 = \rho_{1.23\dots p}^2$  of the variance of  $X_1$  explained by the regression relationship with  $X_2, \dots, X_p$ . If  $N$  is the sample size considering  $X_1$  alone, then the sample size in a setting with additional covariates is:  $N' = N/(1 - R^2)$ . This correction for the influence of other covariates is identical to that proposed by [Hsieh et al. \(1998\)](#) for the logistic regression.

In line with the interpretation of  $R^2$  as squared correlation it must lie in the interval  $[0, 1]$ .

- X distribution:  
Distribution of the  $X_i$ . There are 7 options:
  1. Binomial [ $P(k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$ , where  $k$  is the number of successes ( $X = 1$ ) in  $N$  trials of a Bernoulli process with probability of success  $\pi$ ,  $0 < \pi < 1$ ]
  2. Exponential [ $f(x) = (1/\lambda)e^{-1/\lambda}$ , exponential distribution with parameter  $\lambda > 0$ ]
  3. Lognormal [ $f(x) = 1/(x\sigma\sqrt{2\pi}) \exp[-(\ln x - \mu)^2/(2\sigma^2)]$ , lognormal distribution with parameters  $\mu$  and  $\sigma > 0$ .]
  4. Normal [ $f(x) = 1/(\sigma\sqrt{2\pi}) \exp[-(x - \mu)^2/(2\sigma^2)]$ , normal distribution with parameters  $\mu$  and  $\sigma > 0$ ]
  5. Poisson ( $P(X = k) = (\lambda^k/k!)e^{-\lambda}$ , Poisson distribution with parameter  $\lambda > 0$ )
  6. Uniform ( $f(x) = 1/(b - a)$  for  $a \leq x \leq b$ ,  $f(x) = 0$  otherwise, continuous uniform distribution in the interval  $[a, b]$ ,  $a < b$ )
  7. Manual (Allows to manually specify the variance of  $\hat{\beta}$  under  $H_0$  and  $H_1$ )

G\*POWER provides two different procedures—an enumeration procedure and a large sample approximation—to calculate power. The manual mode is only available in the large sample procedure.

## 30.2 Options

G\*POWER provides two different types of procedure to estimate power. An "enumeration procedure" proposed by Lyles et al. (2007) and large sample approximations. The enumeration procedure seems to provide reasonable accurate results over a wide range of situations, but it can be rather slow and may need large amounts of memory. The large sample approximations are much faster. Results of Monte-Carlo simulations indicate that the accuracy of the procedure based on the work of Demidenko (2007) is comparable to that of the enumeration procedure for  $N > 200$ , whereas errors of the procedure proposed by Signorini (1991) can be quite large. We thus recommend to use the procedure based on Demidenko (2007) as the standard procedure. The enumeration procedure of Lyles et al. (2007) may be used for small sample sizes and to validate the results of the large sample procedure using an a priori analysis (if the sample size is not too large). It must also be used, if one wants to compute the power for likelihood ratio tests. The procedure of Signorini (1991) is problematic and should not be used; it is only included to allow checks of published results referring to this widespread procedure.

1. The enumeration procedure provides power analyses for the Wald-test and the Likelihood ratio test. The general idea is to construct an exemplary data set with weights that represent response probabilities given the assumed values of the parameters of the  $X$ -distribution. Then a fit procedure for the generalized linear model is used to estimate the variance of the regression weights (for Wald tests) or the likelihood ratio under  $H_0$  and  $H_1$  (for likelihood ratio tests). The size of the exemplary data set increases with  $N$  and the enumeration procedure may thus be rather slow (and may need large amounts of computer memory) for large sample sizes. The procedure is especially slow for analysis types other than "post hoc", which internally call the power routine several times. By specifying a threshold sample size  $N$  you can restrict the use of the enumeration procedure to sample sizes  $< N$ . For sample sizes  $\geq N$  the large sample approximation selected in the option dialog is used. Note: If a computation takes too long you can abort it by pressing the ESC key.
2. G\*POWER provides two different large sample approximations for a Wald-type test. Both rely on the asymptotic normal distribution of the maximum likelihood estimator  $\hat{\beta}$ . The accuracy of these approximation increases with sample size, but the deviation from the true power may be quite noticeable for small and moderate sample sizes. This is especially true for  $X$ -distributions that are not symmetric about the mean, i.e. the lognormal, exponential, and poisson distribution, and the binomial distribution with  $\pi \neq 1/2$ . The procedure proposed by Signorini (1991) and variants of it Shieh (2001, 2005) use the "null variance formula" which is not correct for the test statistic assumed here (and that is used in existing software) Demidenko (2007, 2008). The other procedure which is based on the work of Demidenko (2007) on logistic regression is usually more accurate. For this test, a variance correction option can be selected that compensates for vari-

ance distortions that may occur in skewed  $X$  distributions (see implementation notes).

## 30.3 Examples

We replicate the example given on page 449 in Signorini (1991). The number of infections  $Y$  of swimmers ( $X = 1$ ) vs. non-swimmers ( $X=0$ ) during a swimming season (exposure time = 1) is tested. The infection rate is modeled as a Poisson distributed random variable.  $X$  is assumed to be binomially distributed with  $\pi = 0.5$  (equal numbers of swimmers and non-swimmers are sampled). The base rate, that is, the infection rate in non-swimmers, is estimated to be 0.85. The significance level is  $\alpha = 0.05$ . We want to know the sample size needed to detect a 30% or greater increase in infection rate with a power of 0.95. A 30% increase implies a relative rate of 1.3 ( $[100\%+30\%]/100\%$ ).

We first choose to use the procedure of Signorini (1991)

- **Select**  
*Statistical test:* Regression: Poisson Regression  
*Type of power analysis:* A priori
- **Input**  
*Tail(s):* One  
*Exp( $\beta_1$ ):* 1.3  
 *$\alpha$  err prob:* 0.05  
*Power (1- $\beta$  err prob):* 0.95  
*Base rate exp( $\beta_0$ ):* 0.85  
*Mean exposure:* 1.0  
*R<sup>2</sup> other X:* 0  
*X distribution:* Binomial  
*X parm  $\pi$ :* 0.5
- **Output**  
*Critical z:* 1.644854  
*Total sample size:* 697  
*Actual power:* 0.950121

The result  $N = 697$  replicates the value given in Signorini (1991). The other procedures yield  $N = 649$  (Demidenko (2007) with variance correction, and Lyles et al. (2007) for Likelihood-Ratio tests),  $N = 655$  (Demidenko (2007) without variance correction, and Lyles et al. (2007) for Wald tests). In Monte-Carlo simulations of the Wald test for this scenario with 150000 independent cases per test a mean power of 0.94997, 0.95183, and 0.96207 was found for sample sizes of 649, 655, and 697, respectively. This simulation results suggest that in this specific case the procedure based on Demidenko (2007) with variance correction gives the most accurate estimate.

We now assume that we have additional covariates and estimate the squared multiple correlation with these others covariates to be  $R^2 = 0.1$ . All other conditions are identical. The only change needed is to set the input field  $R^2$  other  $X$  to 0.1. Under this condition the necessary sample size increases to a value of 774 (when we use the procedure of Signorini (1991)).

**Comparison between procedures** To compare the accuracy of the Wald test procedures we replicated a number of test cases presented in table III in Lyles et al. (2007) and conducted several additional tests for  $X$ -distributions not

considered in Lyles et al. (2007). In all cases a two-sided test with  $N = 200$ ,  $\beta_0 = 0.5$ ,  $\alpha = 0.05$  is assumed.

If we use the procedure of Demidenko (2007), then the complete G\*POWER input and output for the test in the first row of table (2) below would be:

- **Select**  
Statistical test: Regression: Poisson Regression  
Type of power analysis: Post hoc
- **Input**  
Tail(s): Two  
Exp( $\beta_1$ ): =exp(-0.1)  
 $\alpha$  err prob: 0.05  
Total sample size: 200  
Base rate exp( $\beta_0$ ): =exp(0.5)  
Mean exposure: 1.0  
 $R^2$  other X: 0  
X distribution: Normal  
X parm  $\mu$ : 0  
X parm  $\sigma$ : 1
- **Output**  
Critical z: -1.959964  
Power (1- $\beta$  err prob): 0.444593

When using the enumeration procedure, the input is exactly the same, but in this case the test is based on the  $\chi^2$ -distribution, which leads to a different output.

- **Output**  
Noncentrality parameter  $\lambda$ : 3.254068  
Critical  $\chi^2$ : 3.841459  
Df: 1  
Power (1- $\beta$  err prob): 0.438076

The rows in table (2) show power values for different test scenarios. The column "Dist. X" indicates the distribution of the predictor  $X$  and the parameters used, the column " $\beta_1$ " contains the values of the regression weight  $\beta_1$  under H1, "Sim LLW" contains the simulation results reported by Lyles et al. (2007) for the test (if available), "Sim" contains the results of a simulation done by us with considerable more cases (150000 instead of 2000). The following columns contain the results of different procedures: "LLW" = Lyles et al. (2007), "Demi" = Demidenko (2007), "Demi(c)" = Demidenko (2007) with variance correction, and "Signorini" Signorini (1991).

### 30.4 Related tests

- Logistic regression

### 30.5 Implementation notes

#### 30.5.1 Enumeration procedure

The procedures for the Wald- and Likelihood ratio tests are implemented exactly as described in Lyles et al. (2007).

#### 30.5.2 Large sample approximations

The large sample procedures for the univariate case use the general approach of Signorini (1991), which is based on the approximate method for the logistic regression described in

Whittemore (1981). They get more accurate for larger sample sizes. The correction for additional covariates has been proposed by Hsieh et al. (1998).

The H0 distribution is the standard normal distribution  $N(0,1)$ , the H1 distribution the normal distribution  $N(m_1, s_1)$  with:

$$m_1 = \beta_1 \sqrt{N(1 - R^2) \cdot t / v_0} \quad (16)$$

$$s_1 = \sqrt{v_1 / v_0} \quad (17)$$

in the Signorini (1991) procedure, and

$$m_1 = \beta_1 \sqrt{N(1 - R^2) \cdot t / v_1} \quad (18)$$

$$s_1 = s^* \quad (19)$$

in the procedure based on Demidenko (2007). In these equations  $t$  denotes the mean exposure time,  $N$  the sample size,  $R^2$  the squared multiple correlation coefficient of the covariate of interest on the other covariates, and  $v_0$  and  $v_1$  the variance of  $\hat{\beta}_1$  under H0 and H1, respectively. Without variance correction  $s^* = 1$ , with variance correction  $s^* = \sqrt{(av_0^* + (1 - a)v_1) / v_1}$ , where  $v_0^*$  is the variance under H0 for  $\beta_0^* = \log(\mu)$ , with  $\mu = \int f_X(x) \exp(\beta_0 + \beta_1 x) dx$ . For the lognormal distribution  $a = 0.75$ ; in all other cases  $a = 1$ . With variance correction, the value of  $s^*$  is often close to 1, but deviates from 1 for  $X$ -distributions that are not symmetric about the mean. Simulations showed that this compensates for variance inflation/deflation in the distribution of  $\hat{\beta}_1$  under H1 that occurs with such distributions finite sample sizes.

The large sample approximations use the result that the  $(m + 1)$  maximum likelihood estimators  $\beta_0, \beta_1, \dots, \beta_m$  are asymptotically (multivariate) normal distributed, where the variance-covariance matrix is given by the inverse of the  $(m + 1) \times (m + 1)$  Fisher information matrix  $I$ . The  $(i, j)$ th element of  $I$  is given by

$$I_{ij} = -E \left[ \frac{\partial^2 \log L}{\partial \beta_i \partial \beta_j} \right] = NE[X_i X_j e^{\beta_0 + \beta_1 X_i + \dots + \beta_m X_m}]$$

Thus, in the case of one continuous predictor  $I$  is a  $4 \times 4$  matrix with elements:

$$\begin{aligned} I_{00} &= \int_{-\infty}^{\infty} f(x) \exp(\beta_0 + \beta_1 x) dx \\ I_{10} = I_{01} &= \int_{-\infty}^{\infty} f(x) x \exp(\beta_0 + \beta_1 x) dx \\ I_{11} &= \int_{-\infty}^{\infty} f(x) x^2 \exp(\beta_0 + \beta_1 x) dx \end{aligned}$$

where  $f(x)$  is the PDF of the  $X$  distribution (for discrete predictors the integrals must be replaced by corresponding sums). The element  $M_{11}$  of the inverse of this matrix ( $M = I^{-1}$ ), that is the variance of  $\beta_1$ , is given by:  $M_{11} = \text{Var}(\beta_1) = I_{00} / (I_{00} I_{11} - I_{01}^2)$ . In G\*POWER, numerical integration is used to compute these integrals.

To estimate the variance of  $\beta_1$  under H1, the parameter  $\beta_0$  and  $\beta_1$  in the equations for  $I_{ij}$  are chosen as specified in the input. To estimate the variance under H0, one chooses  $\beta_1 = 0$  and  $\beta_0 = \beta^*$ . In the procedures proposed by Signorini (1991)  $\beta^* = \beta_0$ . In the procedure with variance correction,  $\beta^*$  is chosen as defined above.

Dist. X	$\beta_1$	Sim LLW	Sim	LLW	Demi	Demi(c)	Signorini
N(0,1)	-0.10	0.449	0.440	0.438	0.445	0.445	0.443
N(0,1)	-0.15	0.796	0.777	0.774	0.782	0.782	0.779
N(0,1)	-0.20	0.950	0.952	0.953	0.956	0.956	0.954
Unif(0,1)	-0.20	0.167	0.167	0.169	0.169	0.169	0.195
Unif(0,1)	-0.40	0.478	0.472	0.474	0.475	0.474	0.549
Unif(0,1)	-0.80	0.923	0.928	0.928	0.928	0.932	0.966
Logn(0,1)	-0.05	0.275	0.298	0.305	0.320	0.291	0.501
Logn(0,1)	-0.10	0.750	0.748	0.690	0.695	0.746	0.892
Logn(0,1)	-0.15	0.947	0.947	0.890	0.890	0.955	0.996
Poiss(0.5)	-0.20	-	0.614	0.599	0.603	0.613	0.701
Poiss(0.5)	-0.40	-	0.986	0.971	0.972	0.990	0.992
Bin(0.2)	-0.20	-	0.254	0.268	0.268	0.254	0.321
Bin(0.2)	-0.40	-	0.723	0.692	0.692	0.716	0.788
Exp(3)	-0.40	-	0.524	0.511	0.518	0.521	0.649
Exp(3)	-0.70	-	0.905	0.868	0.871	0.919	0.952

**Table 2:** Power for different test scenarios as estimated with different procedures (see text)

The integrals given above happen to be proportional to derivatives of the moment generating function  $g(x) := E(e^{tX})$  of the  $X$  distribution, which is often known in closed form. Given  $g(x)$  and its derivatives the variance of  $\beta$  is calculated as:

$$Var(\beta) = \frac{g(\beta)}{g(\beta)g''(\beta) - g'(\beta)^2} \cdot \frac{1}{\exp(\beta_0)}$$

With this definition, the variance of  $\hat{\beta}_1$  under  $H_0$  is  $v_0 = \lim_{\beta \rightarrow 0} Var(\beta)$  and the variance of  $\hat{\beta}_1$  under  $H_1$  is  $v_1 = Var(\beta_1)$ .

The moment generating function for the lognormal distribution does not exist. For the other five predefined distributions this leads to:

1. Binomial distribution with parameter  $\pi$

$$\begin{aligned} g(x) &= (1 - \pi + \pi e^x)^n \\ v_0 \exp(\beta_0) &= \pi / (1 - \pi) \\ v_1 \exp(\beta_0) &= 1 / (1 - \pi) + 1 / (\pi \exp(\beta_1)) \end{aligned}$$

2. Exponential distribution with parameter  $\lambda$

$$\begin{aligned} g(x) &= (1 - x/\lambda)^{-1} \\ v_0 \exp(\beta_0) &= \lambda^2 \\ v_1 \exp(\beta_0) &= (\lambda - \beta_1)^3 / \lambda \end{aligned}$$

3. Normal distribution with parameters  $(\mu, \sigma)$

$$\begin{aligned} g(x) &= \exp(\mu x + \frac{\sigma^2 x^2}{2}) \\ v_0 \exp(\beta_0) &= 1/\sigma^2 \\ v_1 \exp(\beta_0) &= \exp(-[\beta_1 \mu + (\beta_1 \sigma)^2 / 2]) \end{aligned}$$

4. Poisson distribution with parameter  $\lambda$

$$\begin{aligned} g(x) &= \exp(\lambda(e^x - 1)) \\ v_0 \exp(\beta_0) &= 1/\lambda \\ v_1 \exp(\beta_0) &= \exp(-\beta_1 + \lambda - \exp(\beta_1)\lambda) / \lambda \end{aligned}$$

5. Uniform distribution with parameters  $(u, v)$  corresponding to the interval borders.

$$\begin{aligned} g(x) &= \frac{e^{xv} - e^{xu}}{x(v - u)} \\ v_0 \exp(\beta_0) &= 12 / (u - v)^2 \\ v_1 \exp(\beta_0) &= \frac{\beta_1^3 (h_u - h_v)(u - v)}{h_{2u} + h_{2v} - h_{u+v}(2 + \beta_1^2(u - v)^2)} \\ h_x &= \exp(\beta_1 x) \end{aligned}$$

In the manual mode you have to calculate the variances  $v_0$  and  $v_1$  yourself. To illustrate the necessary steps, we show the calculation for the exponential distribution (given above) in more detail:

$$\begin{aligned} g(\beta) &= (1 - \beta/\lambda)^{-1} = \frac{\lambda}{\lambda - \beta} \\ g'(\beta) &= \frac{\lambda}{(\lambda - \beta)^2} \\ g''(\beta) &= \frac{2\lambda}{(\lambda - \beta)^3} \\ Var(\beta) / \exp(\beta_0) &= \frac{g(\beta)}{g(\beta)g''(\beta) - g'(\beta)^2} \\ &= \frac{(\lambda - \beta)^3}{\lambda} \\ v_0 \exp(\beta_0) &= Var(0) \exp(\beta_0) \\ &= \lambda^2 \\ v_1 \exp(\beta_0) &= Var(\beta_1) \exp(\beta_0) \\ &= (\lambda - \beta_1)^3 / \lambda \end{aligned}$$

Note, however, that sensitivity analyses in which the effect size is estimated, are not possible in the manual mode. This is so, because in this case the argument  $\beta$  of the function  $Var(\beta)$  is not constant under  $H_1$  but the target of the search.

### 30.6 Validation

The procedure proposed by Signorini (1991) was checked against the corresponding routine in PASS (Hintze, 2006) and we found perfect agreement in one-sided tests. (The results of PASS 06 are wrong for normally distributed  $X$

with  $\sigma \neq 1$ ; this error has been corrected in the newest version PASS 2008). In two-sided tests we found small deviations from the results calculated with PASS. The reason for these deviations is that the (usually small) contribution to the power from the distant tail is ignored in PASS but not in G\*POWER. Given the generally low accuracy of the [Signorini \(1991\)](#) procedure, these small deviations are of no practical consequence.

The results of the other two procedures were compared to the results of Monte-Carlo simulations and with each other. The results given in table (2) are quite representative of these more extensive test. These results indicate that the deviation of the computed power values may deviate from the true values by  $\pm 0.05$ .

## 31 Z test: Tetrachoric Correlation

### 31.0.1 Background

In the model underlying tetrachoric correlation it is assumed that the frequency data in a  $2 \times 2$ -table stem from dichotomizing two continuous random variables  $X$  and  $Y$  that are bivariate normally distributed with mean  $m = (0, 0)$  and covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The value  $\rho$  in this (normalized) covariance matrix is the tetrachoric correlation coefficient. The frequency data in the table depend on the criterion used to dichotomize the marginal distributions of  $X$  and  $Y$  and on the value of the tetrachoric correlation coefficient.

From a theoretical perspective, a specific scenario is completely characterized by a  $2 \times 2$  probability matrix:

	$X = 0$	$X = 1$	
$Y = 0$	$p_{11}$	$p_{12}$	$p_{1*}$
$Y = 1$	$p_{21}$	$p_{22}$	$p_{2*}$
	$p_{*1}$	$p_{*2}$	1

where the marginal probabilities are regarded as fixed. The whole matrix is completely specified, if the marginal probabilities  $p_{*2} = Pr(X = 1)$ ,  $p_{2*} = Pr(Y = 1)$  and the table probability  $p_{11}$  are given. If  $z_1$  and  $z_2$  denote the quantiles of the standard normal distribution corresponding to the marginal probabilities  $p_{*1}$  and  $p_{1*}$ , that is  $\Phi(z_x) = p_{*1}$  and  $\Phi(z_y) = p_{1*}$ , then  $p_{11}$  is the CDF of the bivariate normal distribution described above, with the upper limits  $z_x$  and  $z_y$ :

$$p_{11} = \int_{-\infty}^{z_y} \int_{-\infty}^{z_x} \Phi(x, y, r) dx dy$$

where  $\Phi(x, y, r)$  denotes the density of the bivariate normal distribution. It should be obvious that the upper limits  $z_x$  and  $z_y$  are the values at which the variables  $X$  and  $Y$  are dichotomized.

Observed frequency data are assumed to be random samples from this theoretical distribution. Thus, it is assumed that random vectors  $(x_i, y_i)$  have been drawn from the bivariate normal distribution described above that are afterwards assigned to one of the four cells according to the 'column' criterion  $x_i \leq z_x$  vs  $x_i > z_x$  and the 'row' criterion  $y_i \leq z_y$  vs  $y_i > z_y$ . Given a frequency table:

	$X = 0$	$X = 1$	
$Y = 0$	$a$	$b$	$a + b$
$Y = 1$	$c$	$d$	$c + d$
	$a + c$	$b + d$	$N$

the central task in testing specific hypotheses about tetrachoric correlation coefficients is to estimate the correlation coefficient and its standard error from this frequency data. Two approaches have been proposed. One is to estimate the exact correlation coefficient (e.g. [Brown & Benedetti, 1977](#)), the other to use simple approximations  $\rho^*$  of  $\rho$  that are easier to compute (e.g. [Bonnett & Price, 2005](#)). G\*POWER provides power analyses for both approaches. (See, however, the implementation notes for a qualification of the term 'exact' used to distinguish between both approaches.)

The exact computation of the tetrachoric correlation coefficient is difficult. One reason is of a computational nature (see implementation notes). A more principal problem is, however, that frequency data are discrete, which implies that the estimation of a cell probability can be no more accurate than  $1/(2N)$ . The inaccuracies in estimating the true correlation  $\rho$  are especially severe when there are cell frequencies less than 5. In these cases caution is necessary in interpreting the estimated  $r$ . For a more thorough discussion of these issues see [Brown and Benedetti \(1977\)](#) and [Bonnett and Price \(2005\)](#).

### 31.0.2 Testing the tetrachoric correlation coefficient

The implemented routines estimate the power of a test that the tetrachoric correlation  $\rho$  has a fixed value  $\rho_0$ . That is, the null and alternative hypothesis for a two-sided test are:

$$\begin{aligned} H_0 &: \rho - \rho_0 = 0 \\ H_1 &: \rho - \rho_0 \neq 0. \end{aligned}$$

The hypotheses are identical for both the exact and the approximation mode.

In the power procedures the use of the Wald test statistic:  $W = (r - \rho_0)/se_0(r)$  is assumed, where  $se_0(r)$  is the standard error computed at  $\rho = \rho_0$ . As will be illustrated in the example section, the output of G\*POWER may be also be used to perform the statistical test.

### 31.1 Effect size index

The correlation coefficient assumed under  $H_1$  ( $H_1$  corr  $\rho$ ) is used as effect size.

The following additional inputs are needed to fully specify the effect size

- $H_0$  corr  $\rho$ .  
This is the tetrachoric correlation coefficient assumed under  $H_0$ . An input  $H_1$  corr  $\rho = H_0$  corr  $\rho$  corresponds to "no effect" and must not be used in a priori calculations.
- Marginal prop  $x$ .  
This is the marginal probability that  $X > z_x$  (i.e.  $p_{*2}$ )
- Marginal prop  $y$ .  
This is the marginal probability that  $Y > z_y$  (i.e.  $p_{2*}$ )

The correlations must lie in the interval  $] -1, 1[$ , and the probabilities in the interval  $]0, 1[$ .

**Effect size calculation** The effect size dialog may be used to determine  $H_1$  corr  $\rho$  in two different ways:

- A first possibility is to specify for each cell of the  $2 \times 2$  table the probability of the corresponding event assumed under  $H_1$ . Pressing the "Calculate" button calculates the exact (Correlation  $\rho$ ) and approximate (Approx. correlation  $\rho^*$ ) tetrachoric correlation coefficient, and the marginal probabilities Marginal prob  $x = p_{12} + p_{22}$ , and Marginal prob  $y = p_{21} + p_{22}$ . The exact correlation coefficient is used as  $H_1$  corr  $\rho$ . (see left panel in [Fig. 37](#)).



Note: The four cell probabilities must sum to 1. It therefore suffices to specify three of them explicitly. If you leave one of the four cells empty, G\*POWER computes the fourth value as: (1 - sum of three  $p$ ).

- A second possibility is to compute a confidence interval for the tetrachoric correlation in the population from the results of a previous investigation, and to choose a value from this interval as H1 corr  $\rho$ . In this case you specify four observed frequencies, the relative position  $0 < k < 1$  inside the confidence interval (0, 0.5, 1 corresponding to the left, central, and right position, respectively), and the confidence level  $(1 - \alpha)$  of the confidence interval (see right panel in Fig. 37).

From this data G\*POWER computes the total sample size  $N = f_{11} + f_{12} + f_{21} + f_{22}$  and estimates the cell probabilities  $p_{ij}$  by:  $p_{ij} = (f_{ij} + 0.5)/(N + 2)$ . These are used to compute the sample correlation coefficient  $r$ , the estimated marginal probabilities, the borders ( $L, R$ ) of the  $(1 - \alpha)$  confidence interval for the population correlation coefficient  $\rho$ , and the standard error of  $r$ . The value  $L + (R - L) * k$  is used as H1 corr  $\rho$ . The computed correlation coefficient, the confidence interval and the standard error of  $r$  depend on the choice for the exact [Brown and Benedetti \(1977\)](#) vs. the approximate [Bonnett and Price \(2005\)](#) computation mode, made in the option dialog. In the exact mode, the labels of the output fields are Correlation  $r$ , C.I.  $\rho$  lwr, C.I.  $\rho$  upr, and Std. error of  $r$ , in the approximate mode an asterisk \* is appended after  $r$  and  $\rho$ .

Clicking on the button Calculate and transfer to main window copies the values given in H1 corr  $\rho$ , Margin prob  $x$ , Margin prob  $y$ , and - in frequency mode - Total sample size to the corresponding input fields in the main window.

## 31.2 Options

You can choose between the exact approach in which the procedure proposed by [Brown and Benedetti \(1977\)](#) is used and the approximation suggested by [Bonnett and Price \(2005\)](#).

## 31.3 Examples

To illustrate the application of the procedure we refer to example 1 in [Bonnett and Price \(2005\)](#): The Yes or No answers of 930 respondents to two questions in a personality inventory are recorded in a  $2 \times 2$ -table with the following result:  $f_{11} = 203, f_{12} = 186, f_{21} = 167, f_{22} = 374$ .

First we use the effect size dialog to compute from these data the confidence interval for the tetrachoric correlation in the population. We choose in the effect size drawer, From C.I. calculated from observed freq. Next, we insert the above values in the corresponding fields and press Calculate. Using the 'exact' computation mode (selected in the Options dialog in the main window), we get an estimated correlation  $r = 0.334$ , a standard error of  $r = 0.0482$ , and the 95% confidence interval [0.240, 0.429] for the population  $\rho$ . We choose the left border of the C.I. (i.e. relative position 0, corresponding to 0.240) as the value of the tetrachoric correlation coefficient  $\rho$  under H0.

We now want to know, how many subjects we need to achieve a power of 0.95 in a one-sided test of the H0 that  $\rho = 0$  vs. the H1  $\rho = 0.24$ , given the same marginal probabilities and  $\alpha = 0.05$ .

Clicking on 'Calculate and transfer to main window' copies the computed H1 corr  $\rho = 0.2399846$  and the marginal probabilities  $p_x = 0.602$  and  $p_y = 0.582$  to the corresponding input fields in the main window. The complete input and output is as follows:

- **Select**  
Statistical test: Correlation: Tetrachoric model  
Type of power analysis: A priori
- **Input**  
Tail(s): One  
H1 corr  $\rho$ : 0.2399846  
 $\alpha$  err prob: 0.05  
Power (1- $\beta$  err prob): 0.95  
H0 corr  $\rho$ : 0  
Marginal prob  $x$ : 0.6019313  
Marginal prob  $y$ : 0.5815451
- **Output**  
Critical  $z$ : 1.644854  
Total sample size: 463  
Actual power: 0.950370  
H1 corr  $\rho$ : 0.239985  
H0 corr  $\rho$ : 0.0  
Critical  $r$  lwr: 0.122484  
Critical  $r$  upr: 0.122484  
Std err  $r$ : 0.074465

This shows that we need at least a sample size of 463 in this case (the Actual power output field shows the power for a sample size rounded to an integer value).

The output also contains the values for  $\rho$  under H0 and H1 used in the internal computation procedure. In the exact computation mode a deviation from the input values would indicate that the internal estimation procedure did not work correctly for the input values (this should only occur at extreme values of  $r$  or marginal probabilities). In the approximate mode, the output values correspond to the  $r$  values resulting from the approximation formula.

The remaining outputs show the critical value(s) for  $r$  under H0: In the Wald test assumed here,  $z = (r - \rho_0)/se_0(r)$  is approximately standard normally distributed under H0. The critical values of  $r$  under H0 are given (a) as a quantile  $z_{1-\alpha/2}$  of the standard normal distribution, and (b) in the form of critical correlation coefficients  $r$  and standard error  $se_0(r)$ . (In one-sided tests, the single critical value is reported twice in Critical  $r$  lwr and Critical  $r$  upr). In the example given above, the standard error of  $r$  under H0 is 0.074465, and the critical value for  $r$  is 0.122484. Thus,  $(r - \rho_0)/se(r) = (0.122484 - 0)/0.074465 = 1.64485 = z_{1-\alpha}$  as expected.

**Using G\*Power to perform the statistical test of H0**  
G\*POWER may also be used to perform the statistical test of H0. Assume that we want to test the H0:  $\rho = \rho_0 = 0.4$  vs the two-sided alternative H1:  $\rho \neq 0.4$  for  $\alpha = 0.05$ . Assume further that we observed the frequencies  $f_{11} = 120$ ,

The figure displays two screenshots of the G\*POWER software interface, specifically the 'Effect size drawer' used for calculating H1 correlation  $\rho$  and marginal probabilities. Both screenshots are for a confidence level of  $1 - \alpha = 0.95$ .

**Left Screenshot: From C.I. calculated from observed freq**

- Inputs:  $f_{11} = 25$ ,  $f_{12} = 34$ ,  $f_{21} = 12$ ,  $f_{22} = 25$ , Rel. pos in CI (0=left, 1=right) = 0, Confidence level (1- $\alpha$ ) = 0.95, Total sample size = ?, Correlation r = ?, Est. marginal prob x = ?, Est. marginal prob y = ?, C.I. p lwr = ?, C.I. p upr = ?, Std. error of r = ?.
- Calculated values: H1 corr  $\rho = -0.1380485$ .

**Right Screenshot: From cell probabilities**

- Inputs:  $p_{11} = 0.25$ ,  $p_{12} = 0.34$ ,  $p_{21} = 0.25$ ,  $p_{22} = 0.2$ , Confidence level (1- $\alpha$ ) = 0.95, Total sample size = 96, Correlation r = 0.1558819, Est. marginal prob x = 0.6122449, Est. marginal prob y = 0.3877551, C.I. p lwr = -0.1631428, C.I. p upr = 0.4749066, Std. error of r = 0.1627707.
- Calculated values: H1 corr  $\rho = -0.1631428$ .

Figure 37: Effect size drawer to calculate H1  $\rho$ , and the marginal probabilities (see text).

$f_{12} = 45$ ,  $f_{21} = 56$ , and  $f_{22} = 89$ . To perform the test we first use the option "From C.I. calculated from observed freq" in the effect size dialog to compute from the observed frequencies the correlation coefficient  $r$  and the estimated marginal probabilities. In the exact mode we find  $r = 0.513$ , "Est. marginal prob x" = 0.433, and "Est. marginal prob y" = 0.468. In the main window we then choose a "post hoc" analysis. Clicking on "Calculate and transfer to main window" in the effect size dialog copies the values for marginal x, marginal y, and the sample size 310 to the main window. We now set "H0 corr  $\rho$ " to 0.4 and " $\alpha$  err prob" to 0.05. After clicking on "Calculate" in the main window, the output section shows the critical values for the correlation coefficient ([0.244, 0.555]) and the standard error under H0 (0.079366). These values show that the test is not significant for the chosen  $\alpha$ -level, because the observed  $r = 0.513$  lies inside the interval [0.244, 0.555]. We then use the G\*POWER calculator to compute the  $p$ -value. Inserting  $z = (0.513 - 0.4) / 0.0794$ ;  $1 - \text{normcdf}(z, 0, 1)$  and clicking on the "Calculate" button yields  $p = 0.077$ .

If we instead want to use the approximate mode, we first choose this option in the "Options" dialog and then proceed in essentially the same way. In this case we find a very similar value for the correlation coefficient  $r^* = 0.509$ .

The critical values for  $r^*$  given in the output section of the main window are [0.233, 0.541] and the standard error for  $r^*$  is 0.0788. Note: To compute the  $p$ -Value in the approximate mode, we should use H0 corr  $\rho^*$  given in the output and not H0 corr  $\rho$  specified in the input. Accordingly, using the following input in the G\*POWER calculator  $z = (0.509 - 0.397) / 0.0788$ ;  $1 - \text{normcdf}(z, 0, 1)$  yields  $p = 0.0776$ , a value very close to that given above for the exact mode.

### 31.4 Related tests

### 31.5 Implementation notes

Given  $\rho$  and the marginal probabilities  $p_x$  and  $p_y$ , the following procedures are used to calculate the value of  $\rho$  (in the exact mode) or  $\rho^*$  (in the approximate mode) and to estimate the standard error of  $r$  and  $r^*$ .

#### 31.5.1 Exact mode

In the exact mode the algorithms proposed by Brown and Benedetti (1977) are used to calculate  $r$  and to estimate the standard error  $s(r)$ . Note that the latter is *not* the expected

standard error  $\sigma(r)$ ! To compute  $\sigma(r)$  would require to enumerate all possible tables  $T_i$  for the given  $N$ . If  $p(T_i)$  and  $r_i$  denote the probability and the correlation coefficient of table  $i$ , then  $\sigma^2(r) = \sum_i (r_i - \rho)^2 p(T_i)$  (see [Brown & Benedetti, 1977](#), p. 349) for details]. The number of possible tables increases rapidly with  $N$  and it is therefore in general computationally too expensive to compute this exact value. Thus, 'exact' does *not* mean, that the exact standard error is used in the power calculations.

In the exact mode it is not necessary to estimate  $r$  in order to calculate power, because it is already given in the input. We nevertheless report the  $r$  calculated by the routine in the output to indicate possible limitations in the precision of the routine for  $|r|$  near 1. Thus, if the  $r$ 's reported in the output section deviate markedly from those given in the input, all results should be interpreted with care.

To estimate  $s(r)$  the formula based on asymptotic theory proposed by Pearson in 1901 is used:

$$s(r) = \frac{1}{N^{3/2} \phi(z_x, z_y, r)} \{ (a+d)(b+d)/4 + (a+c)(b+d)\Phi_2^2 + (a+b)(c+d)\Phi_1^2 + 2(ad-bc)\Phi_1\Phi_2 - (ab-cd)\Phi_2 - (ac-bd)\Phi_1 \}^{1/2}$$

or with respect to cell probabilities:

$$s(r) = \frac{1}{N^{1/2} \phi(z_x, z_y, r)} \{ (p_{11} + p_{22})(p_{12} + p_{21})/4 + (p_{11} + p_{21})(p_{12} + p_{22})\Phi_2^2 + (p_{11} + p_{12})(p_{21} + p_{22})\Phi_1^2 + 2(p_{11}p_{22} - p_{12}p_{21})\Phi_1\Phi_2 - (p_{11}p_{12} - p_{21}p_{22})\Phi_2 - (p_{11}p_{21} - p_{12}p_{22})\Phi_1 \}^{1/2}$$

where

$$\Phi_1 = \phi\left(\frac{z_x - rz_y}{(1-r^2)^{1/2}}\right) - 0.5,$$

$$\Phi_2 = \phi\left(\frac{z_y - rz_x}{(1-r^2)^{1/2}}\right) - 0.5, \text{ and}$$

$$\Phi(z_x, z_y, r) = \frac{1}{2\pi(1-r^2)^{1/2}} \times \exp\left[-\frac{z_x^2 - 2rz_xz_y + z_y^2}{2(1-r^2)}\right].$$

[Brown and Benedetti \(1977\)](#) show that this approximation is quite good if the minimal cell frequency is at least 5 (see tables 1 and 2 in [Brown et al.](#)).

### 31.5.2 Approximation mode

[Bonett and Price \(2005\)](#) propose the following approximations.

**Correlation coefficient** Their approximation  $\rho^*$  of the tetrachoric correlation coefficient is:

$$\rho^* = \cos(\pi/(1 + \omega^c)),$$

where  $c = (1 - |p_{1*} - p_{*1}|/5 - (1/2 - p_m)^2)/2$ , with  $p_{1*} = p_{11} + p_{12}$ ,  $p_{*1} = p_{11} + p_{21}$ ,  $p_m =$  smallest marginal proportion, and  $\omega = p_{11}p_{22}/(p_{12}p_{21})$ .

The same formulas are used to compute an estimator  $r^*$  from frequency data  $f_{ij}$ . The only difference is that estimates  $\hat{p}_{ij} = (f_{ij} + 0.5)/N$  of the true probabilities are used.

**Confidence Interval** The  $100 * (1 - \alpha)$  confidence interval for  $r^*$  is computed as follows:

$$CI = [\cos(\pi/(1 + L^{\hat{c}})), \cos(\pi/(1 + U^{\hat{c}}))],$$

where

$$L = \exp(\ln \hat{\omega} + z_{\alpha/2} s(\ln \hat{\omega}))$$

$$U = \exp(\ln \hat{\omega} - z_{\alpha/2} s(\ln \hat{\omega}))$$

$$s(\ln \hat{\omega}) = \left\{ \frac{1}{N} \left( \frac{1}{\hat{p}_{11}} + \frac{1}{\hat{p}_{12}} + \frac{1}{\hat{p}_{21}} + \frac{1}{\hat{p}_{22}} \right) \right\}^{1/2}$$

and  $z_{\alpha/2}$  is the  $\alpha/2$  quartile of the standard normal distribution.

**Asymptotic standard error for  $r^*$**  The standard error is given by:

$$s(r^*) = k \left\{ \frac{1}{N} \left( \frac{1}{\hat{p}_{11}} + \frac{1}{\hat{p}_{12}} + \frac{1}{\hat{p}_{21}} + \frac{1}{\hat{p}_{22}} \right) \right\}^{1/2}$$

with

$$k = \pi \hat{c} w \frac{\sin[\pi/(1+w)]}{(1+w)^2}$$

where  $w = \hat{\omega}^{\hat{c}}$ .

### 31.5.3 Power calculation

The  $H_0$  distribution is the standard normal distribution  $N(0, 1)$ . The  $H_1$  distribution is the normal distribution with mean  $N(m_1, s_1)$ , where:

$$m_1 = (\rho - \rho_0)/sr_0$$

$$s_1 = sr_1/sr_0$$

The values  $sr_0$  and  $sr_1$  denote the standard error (approximate or exact) under  $H_0$  and  $H_1$ .

### 31.6 Validation

The correctness of the procedures to calc  $r, s(r)$  and  $r^*, s(r^*)$  was checked by reproducing the examples in [Brown and Benedetti \(1977\)](#) and [Bonett and Price \(2005\)](#), respectively. The soundness of the power routines were checked by Monte-Carlo simulations, in which we found good agreement between simulated and predicted power.

## References

- Armitage, P., Berry, G., & Matthews, J. (2002). *Statistical methods in medical research* (4th edition ed.). Blackwell Science Ltd.
- Barabesi, L., & Greco, L. (2002). A note on the exact computation of the student t snedecor f and sample correlation coefficient distribution functions. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51, 105-110.
- 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 method for the tetrachoric correlation coefficient. *Journal of Educational and Behavioral Statistics*, 30, 213-225.
- Brown, M. B., & Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika*, 42, 347-355.
- Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. New York: Academic Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- 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.
- Ding, C. G. (1996). On the computation of the distribution of the square of the sample multiple correlation coefficient. *Computational statistics & data analysis*, 22, 345-350.
- Ding, C. G., & Bargmann, R. E. (1991). Algorithm as 260: Evaluation of the distribution of the square of the sample multiple correlation coefficient. *Applied Statistics*, 40, 195-198.
- 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., Faul, F., & Buchner, A. (1996). Gpower: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28, 1-11.
- Faul, F., & Erdfelder, E. (1992). *Gpower 2.0*. Bonn, Germany: Universität Bonn.
- Frome, E. L. (1986). Multiple regression analysis: Applications in the health sciences. In D. Herbert & R. Myers (Eds.), (p. 84-123). The American Institute of Physics.
- Gatsonis, C., & Sampson, A. R. (1989). Multiple correlation: Exact power and sample size calculations. *Psychological Bulletin*, 106, 516-524.
- Hays, W. (1988). *Statistics* (4th ed.). Orlando, FL: Holt, Rinehart and Winston.
- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.
- Hintze, J. (2006). *NCSS, PASS, and GESS*. Kaysville, Utah: NCSS.
- Hsieh, F. Y. (1989). Sample size tables for logistic regression. *Statistics in medicine*, 8, 795-802.
- 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. (1971). Some results on the sampling distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33, 117-130.
- Lee, Y. (1972). Tables of the upper percentage points of the multiple correlation coefficient. *Biometrika*, 59, 179-189.
- Lehmann, E. (1975). *Nonparameterics: Statistical methods based on ranks*. New York: McGraw-Hill.
- 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., & Stafford, K. (2001). Confidence interval, 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.
- O'Brien, R. (1998). A tour of unifypow: A sas module/macro for sample-size analysis. *Proceedings of the 23rd SAS Users Group International Conference, Cary, NC, SAS Institute*, 1346-1355.
- O'Brien, R. (2002). Sample size analysis in study planning (using unifypow.sas).  
(available on the WWW:  
<http://www.bio.ri.ccf.org/UnifyPow.all/UnifyPowNotes02>
- Sampson, A. R. (1974). A tale of two regressions. *American Statistical Association*, 69, 682-689.
- Shieh, G. (2001). Sample size calculations for logistic and poisson regression models. *Biometrika*, 88, 1193-1199.
- Shieh, G. (2005). On power and sample size calculations for Wald tests in generalized linear models. *Journal of Statistical planning and inference*, 128, 43-59.
- Shieh, G., Jan, S.-L., & Randles, R. H. (2007). Power and sample size determinations for the wilcoxon signed-rank test. *Journal of Statistical Computation and Simulation*, 77(8), 717-724. doi: 10.1080/10629360600635245
- Shieh, G., & Kung, C.-F. (2007). Methodological and computational considerations for multiple correlation analysis. *Behavior Research Methods, Instruments, & Computers*, 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.
- Whittemore, A. S. (1981). Sample size for logistic regression with small response probabilities. *Journal of the*

*American Statistical Association, 76, 27-32.*