Chapter 841

Simple Linear Regression using R²

Introduction

This procedure computes power and sample size for a simple linear regression analysis in which the relationship between a dependent variable Y and an independent variable X is to be studied. Interest often focuses on the regression coefficient, however, since the X values are usually not available during the planning phase, little is known about the coefficient until after the analysis is run. Hence, this procedure uses the squared correlation coefficient, R^2, as the measure of effect size upon which the power analysis and sample size are based.

Gatsonis and Sampson (1989) present two power calculation formulas: *unconditional* and *conditional*. This procedure provides a calculation for both approaches.

Conditional Power Calculation

In conditional approach, X is assumed to be fixed (values known) and it is not treated as a random variable with a probability distribution. Hypotheses that are tested are conditional on the specific set of X values. The focus in this analysis is the size of R^2.

Define R^2 be the value of R^2 that occurs when Y is regressed on X.

Test Statistic in the Conditional Formula

You can construct an F-test that will test whether the regression coefficient is zero using

$$F_{1,N-2} = \frac{(R^2)(N-2)}{(1-R^2)}$$

This F-test is identical to the two-sided t-test of the regression coefficient.

Calculating the Power in the Conditional Formula

In this case, power calculations are based on the noncentral-F distribution. The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value $F_{1,N-2,\alpha}$ where α is the probability of a type-I error.

2. Calculate the noncentrality parameter λ using the formula:

$$\lambda = N\left(\frac{R^2}{1-R^2}\right)$$

3. Compute the power as the probability of being greater than $F_{u,v,\alpha}$ in a noncentral-F distribution with noncentrality parameter λ.

Unconditional Power Calculation

When using the unconditional power calculation, the X's and Y are assumed to have a joint bivariate normal distribution with a specified mean vector and covariance matrix given by

$$\begin{bmatrix} \sigma_Y^2 & \Sigma_{YX}' \\ \Sigma_{YX} & \Sigma_X \end{bmatrix}$$

The study-specific values of X are unknown at the design phase, so the sample size determination is based on a single, effect-size parameter which represents the expected variation in X, and its relationship with Y. This effect-size parameter is the *squared correlation coefficient* which is defined in terms of the covariance matrix as

$$\rho_{YX}^2 = \frac{\Sigma_{YX}' \Sigma_X^{-1} \Sigma_{YX}}{\sigma_Y^2}$$

If this coefficient is zero, the variables X provide no information about the linear prediction of Y. Note that we will use ρ^2 to represent ρ_{YX}^2 going forward.

The sample statistic corresponding to this parameter is R^2, the *coefficient of determination*.

Test Statistic when in the Unconditional Case

An *F*-test with $k = 1$ and *N-k-1* degrees of freedom can be constructed that will test whether the regression coefficient is zero as follows

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

The quantity R_{YX}^2 is the sample estimate of the population squared correlation coefficient .

Calculating the Power in the Unconditional Case

The statistical hypotheses is $H_0: \rho^2 = 0$ versus $H_1: \rho^2 > 0$.

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value r_α from the CDF such that $P(R^2 \leq r_\alpha | N, 1, 0) = 1 - \alpha$. Note that we use the value of ρ^2 specified in the null hypothesis.

2. Compute the power using Power $= 1 - P(R^2 \leq r_\alpha | N, 1, \rho_1^2)$.

Krishnamoorthy and Xia (2003) give the CDF of R^2 as

$$P(R^2 \leq x | N, 1, \rho^2) = \sum_{i=0}^{\infty} P(Y = i) \, I_x\left(i, \frac{N-1}{2}\right)$$

where

$$I_x(a, b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1 - t)^{b-1} dt$$

$$P(Y = i) = \frac{\Gamma\left(\frac{N+1}{2} + i\right)}{\Gamma(i + 1)\Gamma\left(\frac{N+1}{2}\right)} (\rho^2)^i (1 - \rho^2)^{\frac{N+1}{2}}$$

This formulation does not admit $\rho^2 = 0$, so when this occurs, the program inserts $\rho^2 = 0.000000000001$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power*, *Sample Size*, and *Effect Size*. Under most situations, you will select either *Power* or *Sample Size*.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level.

Select *Power* when you want to calculate the power of an experiment.

Power Calculation

Power Calculation Method

Select the method to be used to calculate power. Both methods assume that the same t-test is used for significance testing during the analysis of the data. However, they differ in the calculation of power and the parameter used for input.

The conditional method is the most commonly used and so this is the method we recommend unless you are certain that Y and X will follow the bivariate normal distribution.

Conditional (Recommended) - Uses R^2

This method uses R^2, the multiple correlation coefficient, for input.

The power calculations use the noncentral t distribution and are based on the assumption that the residuals are normally distributed with constant variance.

Regression analysis estimates the conditional expectation of Y given that the values of the X variables are fixed and known. This is the usual assumption made when performing significance tests in regression analysis.

The actual X values are not needed in the power calculations. Instead, R^2 is all that is needed.

This method anticipates that the conditional assumption will be used at the time of the analysis, which is usually the case.

Unconditional (Assumes Bivariate Normality) - Uses ρ^2

This method assumes that Y and X are from a bivariate normal distribution and that the X values will not be known until the study is completed. This reduces the power because of the uncertainty in the X values.

This method uses ρ^2, the squared correlation coefficient. This is the population parameter estimated by R^2.

This method is based on a special algorithm that integrates the bivariate normal distribution.

The method may sound appealing until one realizes that the assumption of bivariate normality is very restrictive and that the power calculations are not robust to departures from this assumption. For example, the X variable may have a non-normal distribution such as uniform, skewed right, or binary. Or, it may exhibit clumping or outliers.

So, unless the bivariate normality assumption is reasonable and you will not be making the conditional expectation assumption when the data are analyzed, you should NOT use this option.

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error (alpha). A type-I error occurs when you reject the null hypothesis when in fact it is true.

Values of alpha must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Sample Size)

This option specifies the value(s) for *N*, the sample size. Note that $1 < N$.

Effect Size – Conditional Case

R² (R-Squared | H1)

Enter one or more values of R². This is the value of R² assumed by the power calculation.

Note that the power calculation uses the Conditional (fixed X) model.

The range is $0 < R^2 < 1$.

You can enter a single value such as 0.1 or a series of values such as 0.1 0.2 0.3 or 0.1 to 0.5 by 0.1.

Effect Size – Unconditional Case

ρ² (ρ-Squared | H1)

Enter one or more values of ρ^2, the squared correlation coefficient. ρ^2 is also interpreted as the proportion of the variation in Y that is explained by the variation in X.

This is the value of ρ^2 assumed by the power calculation.

The range is $0 < \rho^2 < 1$.

You can enter a single value such as 0.1 or a series of values such as 0.1 0.2 0.3 or 0.1 to 0.5 by 0.1.

Example 1 – Finding Sample Size in the Conditional Case

Suppose researchers are planning a simple linear regression study to look at the significance of a certain independent variable. The researchers want to use the conditional power calculation.

They want to find the sample size necessary to detect an ρ^2 of 0.2, 0.3, or 0.4. They want the power at 0.9 and the significance level at 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Simple Linear Regression using R²** procedure. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**N (Sample Size)**
Power Calculation Method......................	**Conditional (Recommended) - Uses R²**
Power..	**0.90**
Alpha..	**0.05**
R² (R-Squared | H1)...............................	**0.2 0.3 0.4**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results ——
Power Method: Conditional (Recommended)
H0: R² = 0 versus H1: R² > 0
H0: B = 0 versus H1: B ≠ 0

Power	N	R²	Alpha
0.9063	45	0.200	0.050
0.9046	27	0.300	0.050
0.9015	18	0.400	0.050

References
Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
Gatsonis, C. and Sampson, A.R. 1989. 'Multiple Correlation: Exact Power and Sample Size Calculations.' Psychological Bulletin, Vol. 106, No. 3, Pages 516-524.

Report Definitions
The test assumes that the X values known constants and that the residuals are normally distributed.
Power is the probability of rejecting a false null hypothesis.
N is the number of observations on which the multiple regression is computed.
R² is the proportion of the variation in Y that is accounted for by the linear regression of Y on X. This is the value used in the power calculation.
Alpha is the probability of rejecting a true null hypothesis.

Summary Statements ——
A sample size of 45 achieves 91% power to detect a non-zero R² attributed to one independent variable using an F-Test (or two-sided t-test) at a significance level (alpha) of 0.050. The power is calculated assuming that the actual value of R² is 0.200. The sample X values are assumed to be fixed and known. That is, the test is conditional upon known X values.

This report shows the necessary sample sizes. The definitions of each of the columns is given in the Report Definitions section.

Plots Section
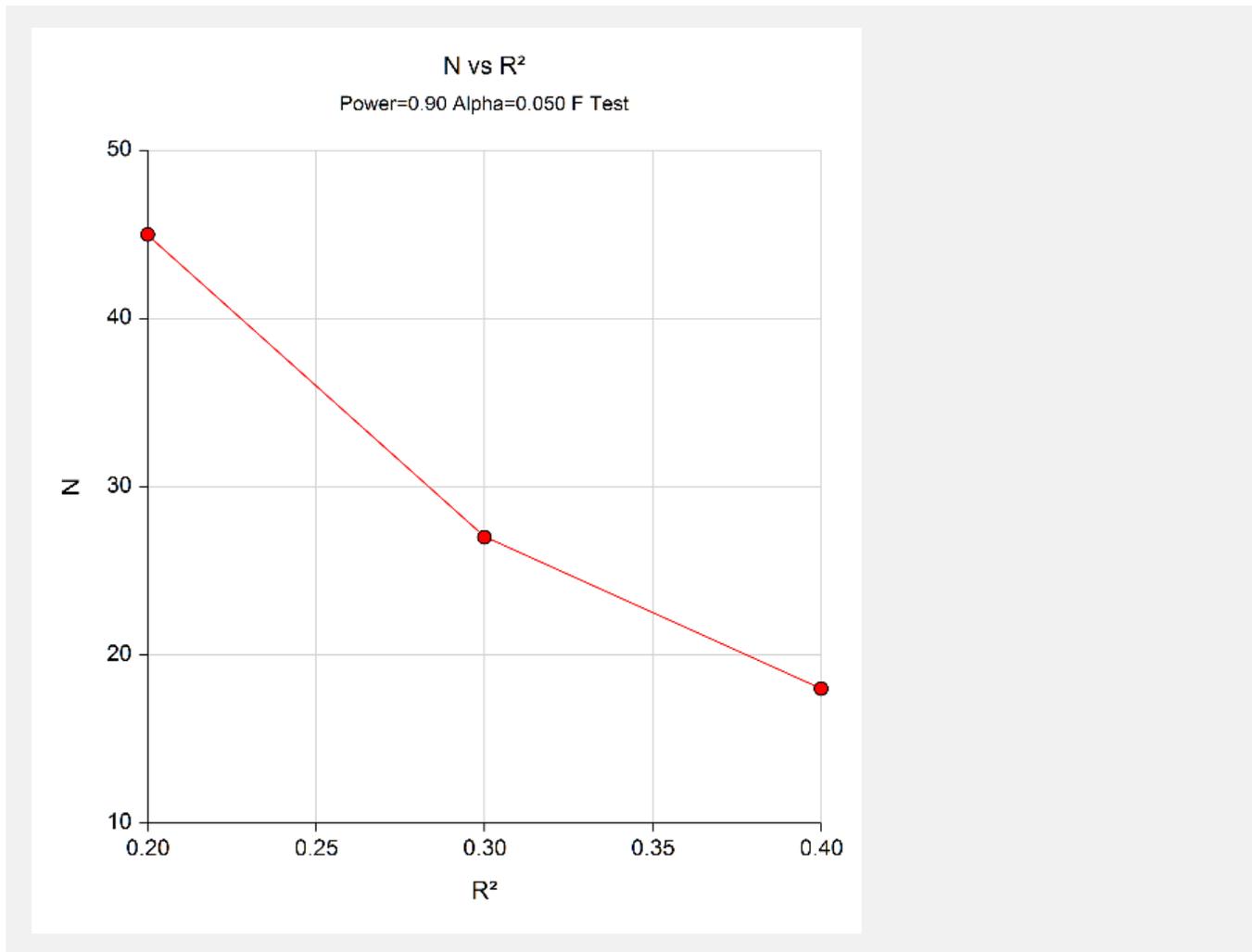
This plot shows the relationship between sample size and effect size.

Example 2 – Validation using another PASS Procedure

We have validated the **PASS** procedure entitled **Multiple Regression**. Since this procedure is a subset of that validated procedure, we can use it to validate this procedure. In that procedure, suppose we set alpha to 0.01, N is 90, k_T to 1, k_C to 0, and R² to 0.2. The power is calculated to be 0.9811.

To validate this procedure, we will set the parameters as given above.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Simple Linear Regression using R²** procedure. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Power**
Power Calculation Method......................	**Conditional (Recommended) - Uses R²**
Alpha..	**0.01**
N (Sample Size)......................................	**90**
R² (R-Squared | H1)................................	**0.2**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results ——
Power Method: Conditional (Recommended)
H0: R² = 0 versus H1: R² > 0
H0: B = 0 versus H1: B ≠ 0

Power	N	R²	Alpha
0.9811	90	0.200	0.010

PASS has again calculated the power to be 0.9811.

Example 3 – Minimum Detectable R²

Suppose the researchers in Example 1 can afford a sample size of only about 20. They want to know the minimum detectable R^2 that can be detected if the power is 80% and 90%. They want to look at a range of sample sizes from 15 to 25. They want a find the minimum the power is 0.8 or 0.9, and the significance level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Simple Linear Regression using R²** procedure. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**R²**
Power Calculation Method......................	**Conditional (Recommended) - Uses R²**
Power...	**0.8 0.9**
Alpha..	**0.05**
N (Sample Size)....................................	**15 20 25**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results ——
Power Method: Conditional (Recommended)
H0: R² = 0 versus H1: R² > 0
H0: B = 0 versus H1: B ≠ 0

Power	N	R²	Alpha
0.8000	15	0.380	0.050
0.8000	20	0.305	0.050
0.8000	25	0.255	0.050
0.9000	15	0.451	0.050
0.9000	20	0.370	0.050
0.9000	25	0.314	0.050

This report gives the value of R² for each scenario.

Example 4 – Finding Sample Size in the Unconditional Case

Suppose researchers are planning a simple linear regression study to look at the significance of a particular independent variable. The researchers will use the unconditional method to compute power.

They want a sample size large enough to detect significance when the actual value of ρ^2 is 0.25, 0.30, 0.35, or 0.4. They want a power of 0.9 and a significance level of 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Simple Linear Regression using R²** procedure. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option	**Value**
Design Tab	
Solve For	**N (Sample Size)**
Power Calculation Method......................	**Unconditional (Assumes Bivariate Normality) - Uses ρ^2**
Power..	**0.90**
Alpha..	**0.05**
ρ^2 (ρ-Squared | H1)................................	**0.25 0.3 0.35 0.4**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results ——
Power Method: Unconditional (Assumes Bivariate Normality)
H0: ρ^2 = 0 versus H1: ρ^2 > 0
H0: B = 0 versus H1: B ≠ 0

Power	**N**	**ρ^2**	**Alpha**
0.9011	37	0.250	0.050
0.9019	30	0.300	0.050
0.9028	25	0.350	0.050
0.9004	21	0.400	0.050

References
Gatsonis, C. and Sampson, A.R. 1989. 'Multiple Correlation: Exact Power and Sample Size Calculations.'
   Psychological Bulletin, Vol. 106, No. 3, Pages 516-524.
Benton, D. and Krishnamoorthy, K. 2003. 'Computing discrete mistures of continuous distributions: noncentral
   chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient.'
   Computational Statistics & Data Analysis, Vol. 43, Pages 249-267.
Krishnamoorthy, K. and Xia, Y. 2008. 'Sample Size Calculation for Estimating or Testing a Nonzero Squared
   Multiple Correlation Coefficient.' Multivariate Behavioral Research, Vol. 43, Pages 382-410.
Cohen, Jacob. 1988. Statistical Power Analysis for the Behavioral Sciences, Lawrence Erlbaum Associates,
   Hillsdale, New Jersey.

Report Definitions
Y and X are assumed to have a bivariate normal distribution.
Power is the probability of rejecting a false null hypothesis.
N is the number of observations on which the multiple regression is computed.
ρ^2 is the squared correlation at which the power is computed.
Alpha is the probability of rejecting a true null hypothesis.

Simple Linear Regression using R²

Summary Statements ——
A sample size of 37 achieves 90% power to detect a non-zero ρ^2 attributed to one independent
variable using an F-Test (or two-sided t-test) at a significance level (alpha) of 0.050. The
power is calculated assuming that the actual value of ρ^2 is 0.250. The power calculation
assumes a random model in which X and Y have a bivariate normal distribution.

This report shows the necessary sample sizes. The definitions of each of the columns is given in the Report Definitions section.

Plots Section
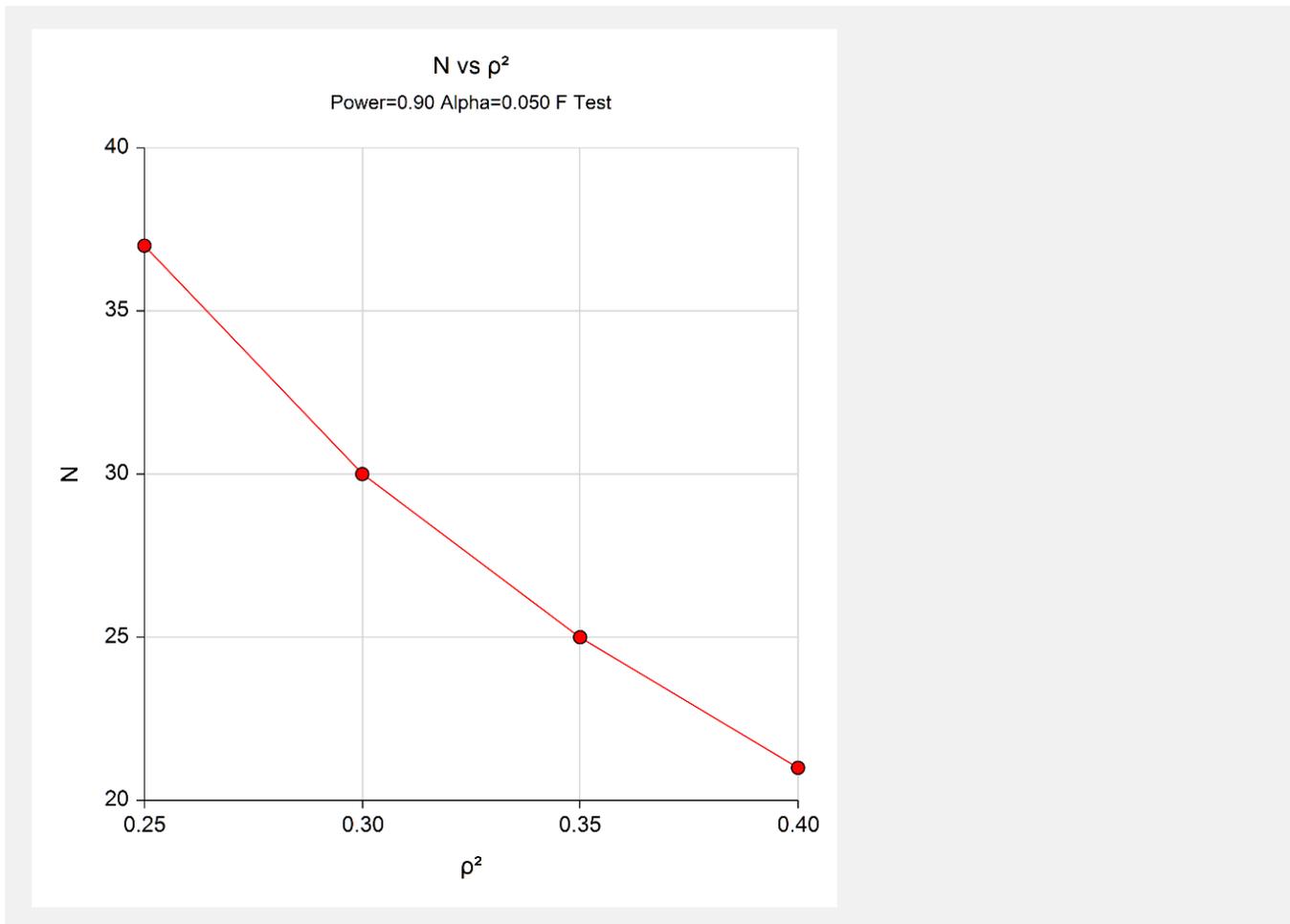
This plot shows the relationship between sample size and effect size.

Example 5 – Validation

We have validated the **PASS** procedure entitled **Multiple Regression**. Since the current procedure is a subset of that validated procedure, we can use it to validate this procedure. In that procedure, suppose we set alpha to 0.01, N is 40, L to 0, K to 1, ρ_0^2(Null) to 0, and ρ_1^2(Actual) to 0.35. The power is calculated to be 0.9446.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Simple Linear Regression using R²** procedure. You may then make the appropriate entries as listed below, or open **Example 5** by going to the **File** menu and choosing **Open Example Template**.

Design Tab

Solve For .... **Power**
Power Calculation Method...................... **Unconditional (Assumes Bivariate Normality) - Uses ρ²**
Alpha.. **0.01**
N (Sample Size)..................................... **40**
ρ² (ρ-Squared | H1)................................ **0.35**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results ——
Power Method: Unconditional (Assumes Bivariate Normality)
H0: ρ^2 = 0 versus H1: ρ^2 > 0
H0: B = 0 versus H1: B ≠ 0

Power	N	ρ^2	Alpha
0.9446	40	0.350	0.010

PASS has also calculated the power to be 0.9446.