

## Chapter 848

# Non-Zero Null Tests for Simple Linear Regression using $\rho^2$

### 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,  $\rho^2$ , as the measure of effect size upon which the power analysis and sample size are based.

Gatsonis and Sampson (1989) present power analysis results for two methods of computing power: *unconditional* and *conditional*. This procedure provides a calculation for the *unconditional* method. The *conditional* method is used in another procedure.

### 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  $\rho^2$  to represent  $\rho_{YX}^2$  going forward.

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

## Non-Zero Null Tests for Simple Linear Regression using $\rho^2$

### Test Statistic in the Unconditional Case

An  $F$ -test with  $k = 1$  and  $N-k-1$  degrees of freedom can be constructed that will test whether all the regression coefficients simultaneously 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

In the unconditional model approach, the statistical hypotheses that is usually of most interest is the set  $H_0: \rho^2 \leq \rho_0^2$  versus  $H_1: \rho^2 > \rho_0^2$  where  $\rho_0^2 > 0$ , because you want to establish a lower bound for the value, not just that it is greater than zero.

However, the hypothesis  $H_0: \rho^2 \geq \rho_0^2$  versus  $H_1: \rho^2 < \rho_0^2$  is also valid. In the program, when  $\rho_1^2 > \rho_0^2$  the former hypothesis set is assumed. Otherwise, the later set is assumed.

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

1. Determine the critical value  $r_\alpha$  from the CDF such that  $P(R^2 \leq r_\alpha | N, 1, \rho_0^2) = 1 - \alpha$ . Note that we use the value of  $\rho^2$  specified in the null hypothesis.
2. Compute the power using  $\text{Power} = 1 - P(R^2 \leq r_\alpha | N, 1, \rho_1^2)$ .

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 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  $2 < N$ .

Non-Zero Null Tests for Simple Linear Regression using  $\rho^2$ 

---

**Effect Size** **$\rho_0^2$  (Squared Correlation | H0)**

Enter one or more values of  $\rho_0^2$ , the *squared correlation coefficient*, used in the null hypothesis.  $\rho^2$  is the proportion of the variation in Y that is explained by the variation in the independent variable.

Often, this value is zero. When it is greater than zero, it is called a *non-zero null*. Values less than this value are assumed to be ignorable.

**The statistical hypotheses tested are**

If  $\rho_0^2 < \rho_1^2$ , the hypotheses tested is  $H_0: \rho^2 \leq \rho_0^2$  vs.  $H_1: \rho^2 > \rho_0^2$ .

If  $\rho_0^2 > \rho_1^2$ , the hypotheses tested is  $H_0: \rho^2 \geq \rho_0^2$  vs.  $H_1: \rho^2 < \rho_0^2$ .

**Range**

$0 \leq \rho_0^2 < \rho_1^2 < 1$  or  $0 \leq \rho_1^2 < \rho_0^2 < 1$ .

 **$\rho_1^2$  (Squared Correlation | H1)**

Enter one or more values of  $\rho_1^2$ , the *squared correlation coefficient* at which the power is calculated.  $\rho^2$  is the proportion of the variation in Y that is explained by the variation in the independent variable.

**The statistical hypotheses tested are**

If  $\rho_0^2 < \rho_1^2$ , the hypotheses tested is  $H_0: \rho^2 \leq \rho_0^2$  vs.  $H_1: \rho^2 > \rho_0^2$ .

If  $\rho_0^2 > \rho_1^2$ , the hypotheses tested is  $H_0: \rho^2 \geq \rho_0^2$  vs.  $H_1: \rho^2 < \rho_0^2$ .

**Range**

$0 \leq \rho_0^2 < \rho_1^2 < 1$  or  $0 \leq \rho_1^2 < \rho_0^2 < 1$ .

Non-Zero Null Tests for Simple Linear Regression using  $\rho^2$ **Example 1 – Finding Sample Size**

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

They want a sample size large enough to detect a  $\rho^2$  of 0.2 when the actual value of  $\rho^2$  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 **Non-Zero Null Tests for Simple Linear Regression using  $\rho^2$**  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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.9</b>
Alpha.....	<b>0.05</b>
$\rho^2$ .....	<b>0.2</b>
$\rho_1^2$ .....	<b>0.25 0.3 0.35 0.4</b>

**Annotated Output**

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

**Numeric Results****Numeric Results**

<b>Power</b>	<b>N</b>	<b>Squared Correlation Given H0 <math>\rho^2</math></b>	<b>Squared Correlation Given H1 <math>\rho_1^2</math></b>	<b>Alpha</b>
0.9000	1849	0.200	0.250	0.050
0.9002	480	0.200	0.300	0.050
0.9007	219	0.200	0.350	0.050
0.9008	125	0.200	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 mixtures 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**

Hypotheses:  $H_0: \rho^2 \leq \rho_0^2$  versus  $H_1: \rho^2 > \rho_0^2$  if  $\rho_0^2 < \rho_1^2$  or  $H_0: \rho^2 \geq \rho_0^2$  versus  $H_1: \rho^2 < \rho_0^2$  if  $\rho_0^2 > \rho_1^2$ .  
 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.  
 $\rho^2$  is the squared correlation assumed by the null hypothesis.  
 $\rho_1^2$  is the squared correlation at which the power is computed.  
 Alpha is the probability of rejecting a true null hypothesis. It should be small.

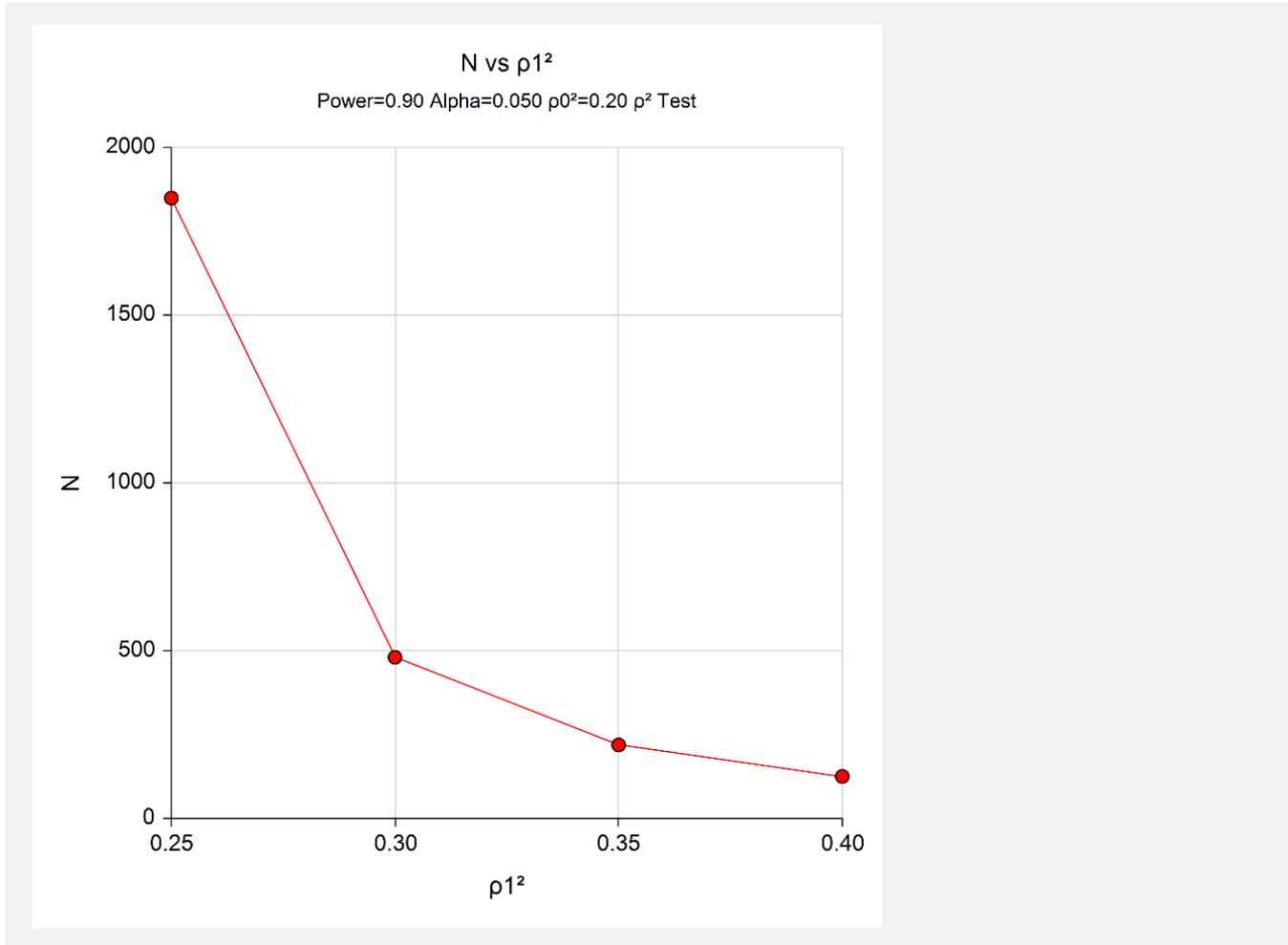
## Non-Zero Null Tests for Simple Linear Regression using $\rho^2$

### Summary Statements

A sample size of 1849 achieves 90% power to detect a squared correlation of at least 0.200 attributed to a single independent variable when the significance level (alpha) is 0.050 and the actual value of the squared correlation is 0.250.

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

### Plots Section



These plots show the relationship between sample size, effect size, and power.

## Non-Zero Null Tests for Simple Linear Regression using $\rho^2$

### Example 2 – Validation using Shieh and Kung (2007)

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 300, L to 0, K to 1,  $\rho_0^2$ (Null) to 0.2, and  $\rho_1^2$ (Actual) to 0.35. The power is calculated to be 0.8656.

#### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Zero Null Tests for Simple Linear Regression using  $\rho^2$**  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**.

#### Design Tab

Solve For ..... **Power**  
 Alpha..... **0.01**  
 N (Sample Size)..... **300**  
 $\rho_0^2$  (Squared Correlation | H0) ..... **0.2**  
 $\rho_1^2$  (Squared Correlation | H1) ..... **0.35**

#### Output

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

#### Numeric Results

Numeric Results				
Power	N	Squared Correlation Given H0 $\rho_0^2$	Squared Correlation Given H1 $\rho_1^2$	Alpha
0.8656	300	0.200	0.350	0.010

PASS has also calculated the power to be 0.8656.