Chapter 848

Non-Zero Null Tests for Simple Linear Regression using ρ²

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, ρ^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 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 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^2_{YX} 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_α 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 ρ^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 $\rho^2 = 0.000000000001$.

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 ρ_0^2 of 0.2 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ....**Sample Size**
Power..**0.9**
Alpha..**0.05**
ρ_0^2 ....**0.2**
ρ_1^2 ....**0.25 0.3 0.35 0.4**

Output

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

Numeric Reports

Numeric Results

Solve For: N (Sample Size)
Hypotheses: H0: $\rho^2 \le \rho_0^2$ vs.   H1: $\rho^2 > \rho_0^2$ if $\rho_0^2 < \rho_1^2$ or   H0: $\rho^2 \ge \rho_0^2$ vs.   H1: $\rho^2 < \rho_0^2$ if $\rho_0^2 > \rho_1^2$

| | | Squared Correlation | | |
| | Sample Size | Given H0 | Given H1 | |
Power	N	ρ_0^2	ρ_1^2	Alpha
0.9000	1849	0.2	0.25	0.05
0.9002	480	0.2	0.30	0.05
0.9007	219	0.2	0.35	0.05
0.9008	125	0.2	0.40	0.05

Y and X are assumed to have a bivariate normal distribution.

Power   The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N   The number of observations on which the multiple regression is computed.
ρ_0^2 The squared correlation assumed by the null hypothesis.
ρ_1^2 The squared correlation at which the power is computed.
Alpha   The probability of rejecting a true null hypothesis.

Summary Statements

A simple linear regression (single group, Y versus X) design will be used to test whether ρ^2 is greater than 0.2 (H0: $\rho^2 \le 0.2$ versus H1: $\rho^2 > 0.2$). The comparison will be made using an R^2-based F-test with a Type I error rate (α) of 0.05. The model is assumed to be a random model in which X and Y have a bivariate normal distribution. To detect a squared correlation (ρ^2) of 0.25 with 90% power, the number of needed subjects will be 1849.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	1849	2312	463
20%	480	600	120
20%	219	274	55
20%	125	157	32

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N	The evaluable sample size at which power is computed. If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N'	The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 2312 subjects should be enrolled to obtain a final sample size of 1849 subjects.

## 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.

This report shows the necessary sample sizes.

Plots Section

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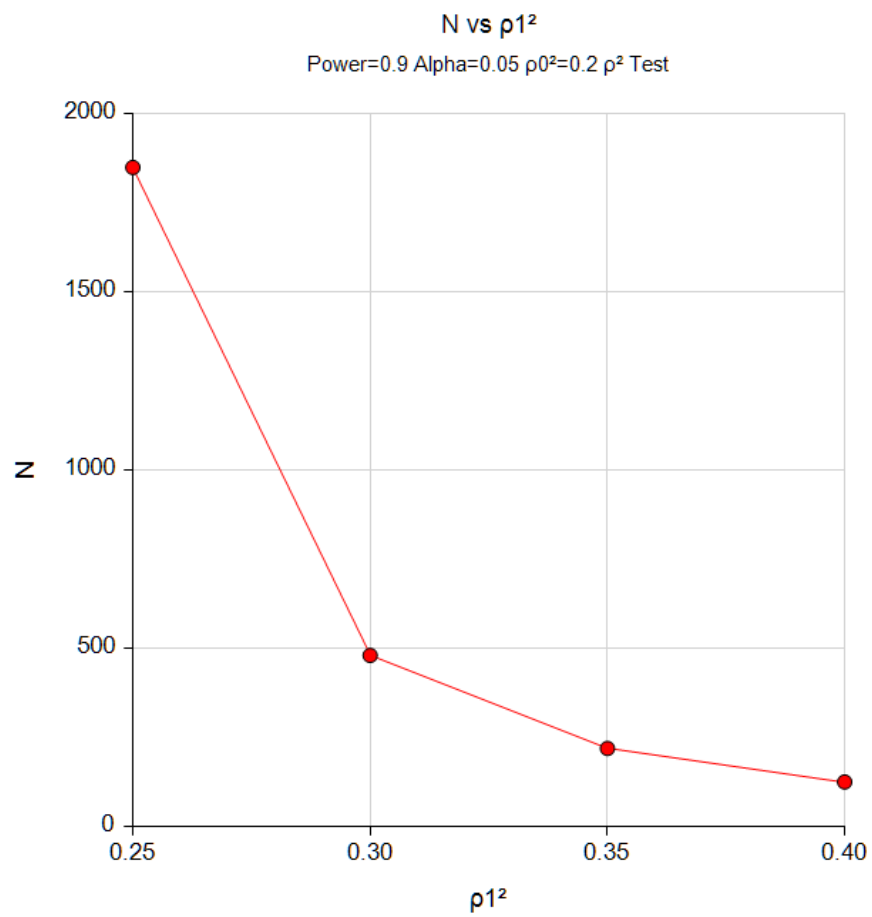
This plot shows the relationship between sample size and effect size.

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 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

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

Solve For: Power
Hypotheses: H0: $\rho^2 \le \rho 0^2$ vs.   H1: $\rho^2 > \rho 0^2$ if $\rho 0^2 < \rho 1^2$ or   H0: $\rho^2 \ge \rho 0^2$ vs.   H1: $\rho^2 < \rho 0^2$ if $\rho 0^2 > \rho 1^2$

	Sample Size	Squared Correlation		
		Given H0	Given H1	
Power	**N**	$\rho 0^2$	$\rho 1^2$	**Alpha**
0.8656	300	0.2	0.35	0.01

Y and X are assumed to have a bivariate normal distribution.

PASS has also calculated the power to be 0.8656.