Chapter 841

Simple Linear Regression using R^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, 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 $\mathrm{H}_0\colon \rho^2 = 0$ versus $\mathrm{H}_1\colon \rho^2 > 0$.

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

1. Determine the critical value r_α from the CDF such that $\mathrm{P}(R^2 \le r_\alpha | N, 1, 0) = 1 - \alpha$. Note that we use the value of ρ^2 specified in the null hypothesis.
2. Compute the power using $\text{Power} = 1 - \mathrm{P}(R^2 \le r_\alpha | N, 1, \rho_1^2)$.

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

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

where

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

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

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

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

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 .... **N (Sample Size)**
Power Calculation Method .... **Conditional (Recommended) - Uses R^2**
Power... **0.90**
Alpha.. **0.05**
R^2 (R-Squared | H1)...................................... **0.2 0.3 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)
Power Method: Conditional (Recommended)
Hypotheses: H0: R^2 = 0 versus H1: R^2 > 0
H0: B = 0 versus H1: B ≠ 0

Power	Sample Size N	R-Squared | H1 R^2	Alpha
0.9063	45	0.2	0.05
0.9046	27	0.3	0.05
0.9015	18	0.4	0.05

The test assumes that the X values are known constants and that the residuals are normally distributed.

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.
R^2 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 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 the slope (B) is different from 0 (H0: B = 0 versus H1: B ≠ 0, or, equivalently, H0: $R^2 = 0$ versus H1: $R^2 > 0$). The comparison will be made using a simple linear regression slope F-test (or equivalent two-sided t-test) with a Type I error rate (α) of 0.05. The sample X values are assumed to be fixed and known (the test is conditional upon known X values). To detect an R^2 of 0.2 with 90% power, the number of needed subjects will be 45.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	45	57	12
20%	27	34	7
20%	18	23	5

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, 57 subjects should be enrolled to obtain a final sample size of 45 subjects.

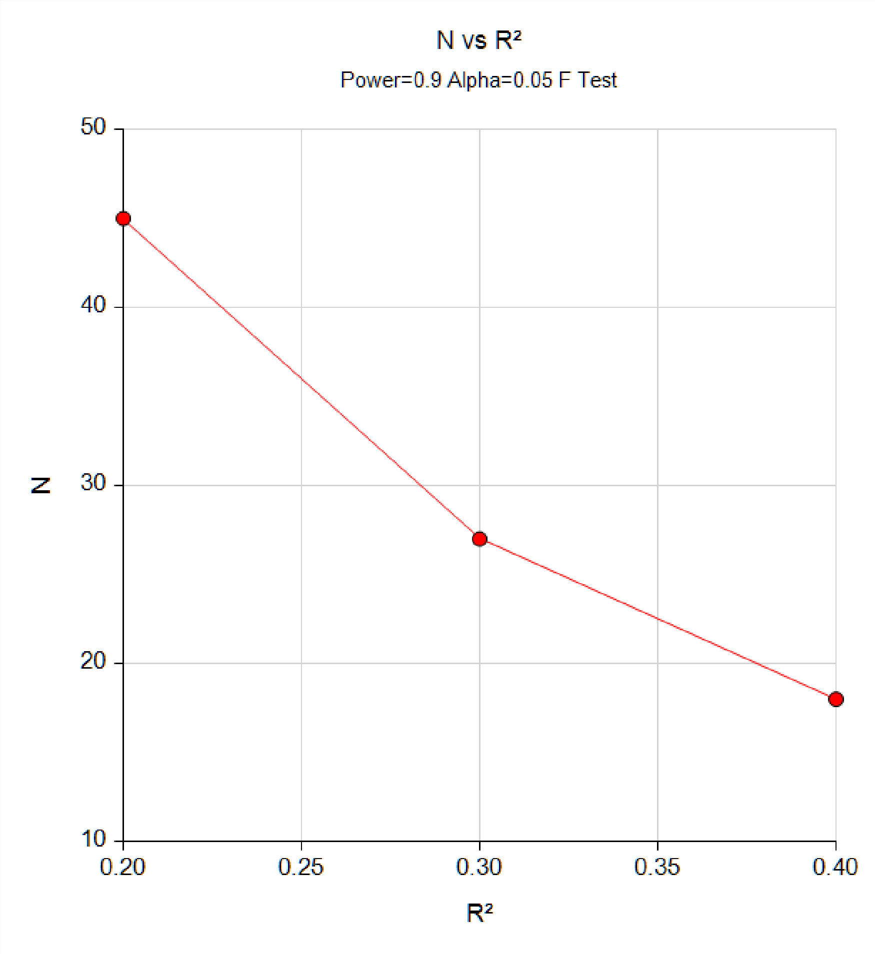
References

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.

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

Plots Section

Plots



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 R^2 to 0.2. The power is calculated to be 0.9811.

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

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**
Power Calculation Method .... **Conditional (Recommended) - Uses R^2**
Alpha.. **0.01**
N (Sample Size)... **90**
R^2 (R-Squared | H1)...................................... **0.2**

Output

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

Numeric Results

Solve For: Power
Power Method: Conditional (Recommended)
Hypotheses: H0: $R^2 = 0$ versus H1: $R^2 > 0$
H0: B = 0 versus H1: B ≠ 0

Power	Sample Size N	R-Squared | H1 R^2	Alpha
0.9811	90	0.2	0.01

The test assumes that the X values are known constants and that the residuals are normally distributed.

PASS has again calculated the power to be 0.9811.

Example 3 – Minimum Detectable R^2

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

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 3** 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 .... **R^2**
Power Calculation Method .... **Conditional (Recommended) - Uses R^2**
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

Solve For: R^2
Power Method: Conditional (Recommended)
Hypotheses: H0: $R^2 = 0$ versus H1: $R^2 > 0$
H0: $B = 0$ versus H1: $B \neq 0$

Power	Sample Size N	R-Squared | H1 R^2	Alpha
0.8	15	0.380	0.05
0.8	20	0.305	0.05
0.8	25	0.255	0.05
0.9	15	0.451	0.05
0.9	20	0.370	0.05
0.9	25	0.314	0.05

The test assumes that the X values are known constants and that the residuals are normally distributed.

This report gives the value of R^2 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

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 4** 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 .... **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**

Output

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

Numeric Reports

Numeric Results

Solve For: N (Sample Size)
Power Method: Unconditional (Assumes Bivariate Normality)
Hypotheses: H0: $\rho^2 = 0$ versus H1: $\rho^2 > 0$
H0: B = 0 versus H1: B ≠ 0

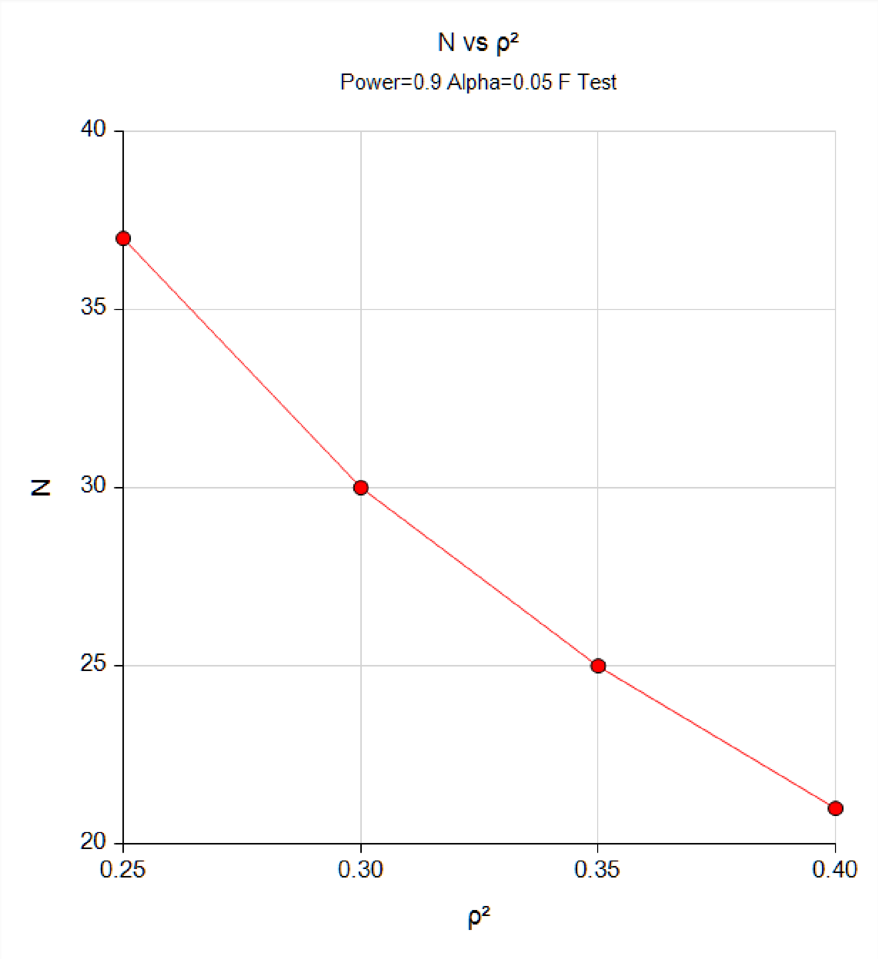
Power	Sample Size N	ρ-Squared | H1 ρ^2	Alpha
0.9011	37	0.25	0.05
0.9019	30	0.30	0.05
0.9028	25	0.35	0.05
0.9004	21	0.40	0.05

Y and X are assumed to 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

Plots



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 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, $\rho 0^2$(Null) to 0, and $\rho 1^2$(Actual) to 0.35. The power is calculated to be 0.9446.

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 5** 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**
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

Solve For: Power
Power Method: Unconditional (Assumes Bivariate Normality)
Hypotheses: H0: ρ² = 0 versus H1: ρ² > 0
H0: B = 0 versus H1: B ≠ 0

Power	Sample Size N	ρ-Squared | H1 ρ²	Alpha
0.9446	40	0.35	0.01

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

PASS has also calculated the power to be 0.9446.