

What is the process for conducting a power analysis for multiple regression?

Authored by
stats writer

June 29, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the process for conducting a power analysis for multiple regression?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=159467>

The process for conducting a power analysis for multiple regression involves carefully determining the required sample size and statistical power to detect a significant relationship between multiple independent variables and a dependent variable. This involves considering factors such as effect size, significance level, and desired power level. Once the necessary calculations have been made, appropriate statistical software can be used to determine the sample size needed for the study. This process is crucial in ensuring the accuracy and validity of the results obtained from a multiple regression analysis. It allows researchers to adequately plan and design their study, ensuring that they have a sufficient number of participants to detect meaningful relationships between variables. Overall, conducting a power analysis for multiple regression is an essential step in the research process that helps to ensure the reliability and validity of statistical findings.

Multiple Regression Power Analysis

Introduction

Power analysis is the name given to the process for determining the sample size for a research study. The technical definition of power is that it is the probability of detecting a "true" effect when it exists. Many students think that there is a simple formula for determining sample size for every research situation. However, the reality is that there are many research situations that are so complex that they almost defy rational power analysis. In most cases, power analysis involves a number of simplifying assumptions, in order to make the problem tractable, and running the analyses numerous times with different variations to cover all of the contingencies.

In this unit we will try to illustrate how to do a power analysis for multiple regression model that has two control variables, one continuous research variable and one categorical research variable (three levels).

Description of the Experiment

A school district is designing a multiple regression study looking at the effect of gender, family income, mother's education and language spoken in the home on the English language proficiency scores of Latino high school students. The variables gender and family income are control variables and not of primary research interest. Mother's education is a continuous research variable that measures the number of years that the mother attended school. The range of this variable is expected to be from 4 to 20. The variable language spoken in the home is a categorical research variable with three levels: 1) Spanish only, 2) both Spanish and English, and 3) English only. Since there are three levels, it will take two dummy variables to code language spoken in the home.

The full regression model will look something like this,

$$\text{engprof} = b_0 + b_1(\text{gender}) + b_2(\text{income}) + b_3(\text{momeduc}) + b_4(\text{homelang1}) + b_5(\text{homelang2})$$

Thus, the primary research hypotheses are the test of b_3 and the joint test of b_4 and b_5 . These tests are equivalent to testing the change in R^2 when momeduc (or homelang1 and homelang2) are added last to the regression equation.

The Power Analysis

We will make use of the SPSS command `power univariate linear` to do the power analysis. To begin with, we believe, from previous research, that the R^2 for the full-model (r^2_f) with five predictor variables (2 control, 1 continuous research, and two dummy variables for the categorical variable) will be about 0.48.

Let's start with the continuous predictor (momeduc). We think that it will add about 0.03 to the R^2 when it is added last to the model. This means that the R^2 for the model without the variable (the reduced model) would be about 0.45. The total number of variables is 5 and the number being tested is 1. We will use the values .7, .8

and .9 to get the estimated number of required subjects for each of these levels of power.

power univariate linear

/parameters model=fixed significance=0.05 power=.7 .8

.9 total_predictors=5 test_predictors=1

full_model=0.48 nested_model=0.45 intercept =true.

Power Analysis - Linear Regression

Power Analysis Table

Type III F-test ^a	N	Actual Power ^b	Predictors		Test Assumptions			
			Full	Nested	Power	Full ^c	Nested ^d	Sig.
1	110	.704	5	1	.700	.48	.45	.05
2	139	.803	5	1	.800	.48	.45	.05
3	185	.901	5	1	.900	.48	.45	.05

a. Intercept term is included.

b. Predictors are assumed to be fixed.

c. Squared multiple correlation coefficient in full model.

d. Squared multiple correlation coefficient in nested model.

This gives us a range of sample sizes ranging from 110 to 185 depending on power.

Let's see how this compares with the categorical predictor (homelang1 and homelang2), which uses two dummy variables in the model. We believe that the change in R² attributed to the two dummy variables will be about 0.025. This would give an r² of 0.455.

power univariate linear

/parameters model=fixed significance=0.05 power=.7 .8

.9 total_predictors=5 test_predictors=2

full_model=0.48 nested_model=0.455 intercept =true.

Power Analysis - Linear Regression

Power Analysis Table

Type III F-test ^a	N	Actual Power ^b	Predictors		Test Assumptions			
			Full	Nested	Power	Full ^c	Nested ^d	Sig.
1	164	.702	5	2	.700	.48	.455	.05
2	204	.801	5	2	.800	.48	.455	.05
3	267	.901	5	2	.900	.48	.455	.05

a. Intercept term is included.

b. Predictors are assumed to be fixed.

c. Squared multiple correlation coefficient in full model.

d. Squared multiple correlation coefficient in nested model.

This series of power analyses yielded sample sizes ranging from 164 to 267. These sample sizes are larger than those for the continuous research variable.

If it is the case that both of these research variables are important, we might want to take into that we are testing two separate hypotheses (one for the continuous variable and one for the categorical variables) by adjusting the alpha level. The simplest but most draconian method would be to use a Bonferroni adjustment, which is done by dividing the nominal

alpha level, 0.05, by the number of hypotheses, 2, yielding an alpha of 0.025. We will rerun the categorical variable power analysis using the new adjusted alpha level.

power univariate linear

/parameters model=fixed significance=0.025 power=.7 .8 .9 total_predictors=5 test_predictors=2 full_model=0.48 nested_model=0.455 intercept =true.

Power Analysis - Linear Regression

Power Analysis Table

Type III F-test ^a	N	Actual Power ^b	Predictors		Test Assumptions			
			Full	Nested	Power	Full ^c	Nested ^d	Sig.
1	199	.700	5	2	.700	.48	.455	.025
2	243	.800	5	2	.800	.48	.455	.025
3	311	.900	5	2	.900	.48	.455	.025

a. Intercept term is included.

b. Predictors are assumed to be fixed.

c. Squared multiple correlation coefficient in full model.

d. Squared multiple correlation coefficient in nested model.

The Bonferroni adjustment assumes that the tests of the two hypotheses are independent which is, in fact, not the case. The squared correlation between the two sets of predictors is about .2, which is equivalent to a correlation of approximately .45. Using an internet applet to compute a Bonferroni adjusted alpha taking into account the correlation gives us an adjusted alpha

value of 0.034 to use in the power analysis.

power univariate linear

/parameters model=fixed significance=0.034 power=.7 .8

.9 total_predictors=5 test_predictors=2

full_model=0.48 nested_model=0.455 intercept =true.

Power Analysis - Linear Regression

Power Analysis Table

Type III F-test ^a	N	Actual Power ^b	Predictors		Test Assumptions			
			Full	Nested	Power	Full ^c	Nested ^d	Sig.
1	184	.702	5	2	.700	.48	.455	.034
2	226	.801	5	2	.800	.48	.455	.034
3	292	.901	5	2	.900	.48	.455	.034

a. Intercept term is included.

b. Predictors are assumed to be fixed.

c. Squared multiple correlation coefficient in full model.

d. Squared multiple correlation coefficient in nested model.

Based on the series of power analyses the school district has decided to collect data on a sample of about 225 students. This sample size should yield a power of around 0.8 in testing hypotheses concerning both the continuous research (momeduc) variable and the categorical research variable language spoken in the home (homelang1 and homelang2). The nominal alpha level is 0.05 but has been adjusted to .034 to take into account the number of hypotheses tested and the

correlation between the predictors.

References

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, Second Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

ARABPSYCHOLOGY.COM