How can I conduct a multiple regression power analysis using SAS software for data analysis?

Authored by

stats writer

June 29, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How can I conduct a multiple regression power analysis using SAS software for data analysis?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=159444

Multiple regression power analysis is a statistical method used to determine the appropriate sample size for a multiple regression analysis. It helps researchers to assess the power of their study, which refers to the probability of detecting a true effect if it exists. In order to conduct a multiple regression power analysis using SAS software, the following steps should be followed:

1. Input the relevant data into SAS software, including the independent and dependent variables.
2. Specify the power level, effect size, and significance level for the analysis.
3. Use the PROC POWER procedure in SAS to calculate the required sample size for the desired power level.
4. Check the assumptions of the multiple regression model and make any necessary adjustments.
5. Run the multiple regression analysis using the determined sample size.

This process allows researchers to determine the appropriate sample size for their multiple regression analysis, ensuring that their study has sufficient power to detect any true effects. Overall, conducting a multiple regression power analysis using SAS software is a crucial step in ensuring the validity and accuracy of statistical findings in research studies.

Multiple Regression Power Analysis | SAS Data Analysis Examples

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 it that there are many research situations that are so complex that they almost defy

ARABPSYCHOLOGY.COM

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

ARABPSYCHOLOGY.COM

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,

engprof = b0 + b1(gender) + b2(income) + b3(momeduc) + b4(homelang1) + b5(homelang2)

Thus, the primary research hypotheses are the test of b3 and the joint test of b4 and b5. These tests are equivalent the testing the change in R2 when momeduc (or homelang1 & homelang2) are added last to the regression equation.

The Power Analysis

We will make use of the SAS proc power to do the power analysis. To begin with, we believe, from previous

ARABPSYCHOLOGY.COM

research, that the R2 for the full-model (r2f) with five predictor variables (2 control, 1 continuous research, and 2 dummy variables for the categorical variable) will be will be about 0.48.

Let's start with the continuous predictor (momeduc). We think that it will add about 0.03 to the R2 when it is added last to the model. This means that the R2 for the model without the variable (the reduced model) would be about 0.45, which leads to the difference in R2 (rsquarediff) of .03. The total number of variables (nfullpredictors) is 5 and the number being tested (ntestpredictors) is one. We will run proc power for powers equal to .7, .8 and .9.

ARABPSYCHOLOGY.COM

```
proc power;
multreg
model = fixed
nfullpredictors = 5
ntestpredictors = 1
rsquarefull = 0.48
rsquarediff = 0.03
ntotal = .
```

power = 0.7 to .9 by .1;

run;

The POWER Procedure

Type III F Test in Multiple Regression

Fixed Scenario Elements

Method Exact

Model Fixed X

Number of Predictors in Full Model 5

Number of Test Predictors 1

R-square of Full Model 0.48

Difference in R-square 0.03

Alpha 0.05

Computed N Total

Nominal Actual N

Index Power Power Total

1 0.7 0.704 110

2 0.8 0.803 139

3 0.9 0.901 185

ARABPSYCHOLOGY.COM

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 & 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. The
nfullpredictors stays at 5 while the ntestpredictors is now 2.**

```
proc power;
multreg
model = fixed
nfullpredictors = 5
ntestpredictors = 2
rsquarefull = 0.48
rsquarediff = 0.025
ntotal = .
power = 0.7 to .9 by .1;
run;
```

ARABPSYCHOLOGY.COM

Type III F Test in Multiple Regression

Fixed Scenario Elements

Method Exact
Model Fixed X
Number of Predictors in Full Model 5
Number of Test Predictors 2
R-square of Full Model 0.48
Difference in R-square 0.025
Alpha 0.05

Computed N Total

Nominal Actual N
Index Power Power Total

1 0.7 0.702 164
2 0.8 0.801 204
3 0.9 0.901 267

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.

ARABPSYCHOLOGY.COM

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 and one for the categorical) by adjusting the alpha level. The simplest but most draconian method would be to use a bonferroni adjustment 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.

ARABPSYCHOLOGY.COM

```
proc power;
multreg
model = fixed
nfullpredictors = 5
ntestpredictors = 2
rsquarefull = 0.48
rsquarediff = 0.025
ntotal = .
alpha = .025
power = 0.7 to .9 by .1;
run;
```

Type III F Test in Multiple Regression

Fixed Scenario Elements

Method Exact
Model Fixed X
Number of Predictors in Full Model 5
Number of Test Predictors 2
Alpha 0.025
R-square of Full Model 0.48
Difference in R-square 0.025

Computed N Total

Nominal Actual N
Index Power Power Total

1 0.7 0.700 199
2 0.8 0.800 243
3 0.9 0.900 311

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

ARABPSYCHOLOGY.COM

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.

```
proc power;
multreg
model = fixed
nfullpredictors = 5
ntestpredictors = 2
rsquarefull = 0.48
rsquarediff = 0.025
ntotal = .
alpha = .034
power = 0.7 to .9 by .1;
run;
```

Type III F Test in Multiple Regression

Fixed Scenario Elements

Method Exact
Model Fixed X
Number of Predictors in Full Model 5

ARABPSYCHOLOGY.COM

Number of Test Predictors 2
Alpha 0.034
R-square of Full Model 0.48
Difference in R-square 0.025

Computed N Total

Nominal Actual N
Index Power Power Total

1 0.7 0.702 184
2 0.8 0.801 226
3 0.9 0.901 292

Based on the series of power analyses the school district has decided to collect data on a sample of about 226 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 & 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

ARABPSYCHOLOGY.COM

predictors.

See Also

Accuracy in Parameter Estimation

Cohen, J. 1988. Statistical Power Analysis for the Behavioral Sciences, Second Edition.

Mahwah, NJ: Lawrence Erlbaum Associates.

ARABPSYCHOLOGY.COM