**Using SPSS to Obtain a Confidence Interval for *R2* from Regression**

The necessary scripts, originally obtained from [M. J. Smithson](http://psychology.anu.edu.au/_people/people_details.asp?recId=130), are now available at: the following sites:

* <http://dl.dropbox.com/u/1857674/CIstuff/CI.html>
* <http://core.ecu.edu/psyc/wuenschk/SPSS/SPSS-Programs.htm> .

Suppose you have conducted an independent samples *t* test and you wish to estimate the proportion of variance in the population which is explained by the grouping variable. For example, I have compared grade point averages of boys girls and found that girls’ GPA (*M* = 2.82, *SD* = .83, *N* = 33) was significantly higher than boys’ GPA (*M* = 2.24, *SD* = .81, *N* = 55), *t*(65.9) = 3.24, *p* = .002, *η2* = .11. Note that I have reported a separate variances *t*. To get *η2* and a confidence interval about *η2* I use a pooled *t*, which was, for these data, *t*(86) = 3.267. Of course, *η2* here is just the squared point biserial correlation coefficient.

* Double click on the **NoncF.sav** file to bring it into SPSS.
* Enter in the **fval** column the squared value of the obtained *t*. 3.2672 = 10.673.
* Enter 1 in the **df1** column.
* Enter the pooled *t* df in the **df2** column, *N* – 2 = 86.
* Enter .95 in the **conf** column, for a 95% confidence interval.
* File, Open Syntax, and open the NoncF3.sps file.
* On the command line, click **Run** and select **All**.

|  |  |
| --- | --- |
|  |  |

* Look back at **NoncF.sav**. In the **r2** column is the point value for *η2*. In the **lr2** and **ur2** columns are the lower and upper limits for the confidence interval on *η2*. For our data, the 90% confidence interval runs from .027 to .218.

Why 90% confidence instead of 95%? Well, if you use 95% it is possible that the ANOVA or the *t* test will indicate significance but the confidence interval will include 0. If you use 90%, the confidence interval will always be consistent with the results from the ANOVA or *t* test using the .05 criterion of statistical significance. This is related to the fact that *r* can be negative or positive, but *r2* cannot.

If your predictor variable(s) is(are) random rather than fixed, that is you have done a correlation analysis rather than a regression analysis, you should not use the procedures described above to put a confidence interval on *R2*. In that case you should use [Steiger and Fouladi’s R2 program](http://core.ecu.edu/psyc/wuenschk/StatHelp/CI-R2.htm).

The SPSS syntax here can also be used to put a confidence interval on R2 and pr2 from a multiple regression. Here I have used verbal and quantitative GRE scores to predict graduate grade point averages.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables Entered/Removeda** | | | |
| Model | Variables Entered | Variables Removed | Method |
| 1 | GRE\_V, GRE\_Qb | . | Enter |

|  |
| --- |
| a. Dependent Variable: GPA |
| b. All requested variables entered. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .697a | .485 | .447 | .4460 |

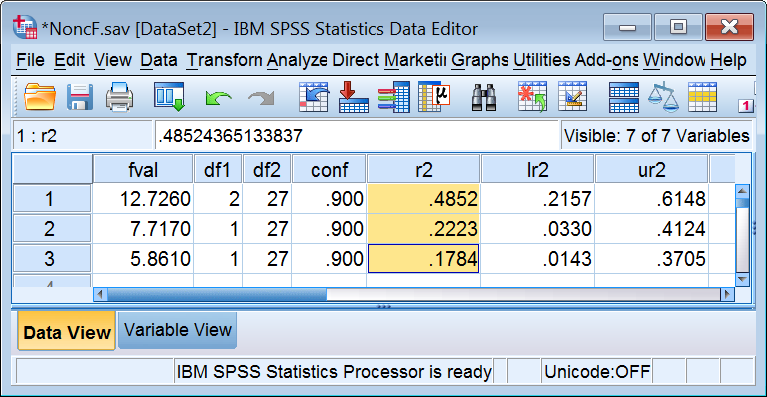
|  |
| --- |
| a. Predictors: (Constant), GRE\_V, GRE\_Q |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 5.063 | 2 | 2.532 | 12.726 | .000b |
| Residual | 5.371 | 27 | .199 |  |  |
| Total | 10.435 | 29 |  |  |  |

|  |
| --- |
| a. Dependent Variable: GPA |
| b. Predictors: (Constant), GRE\_V, GRE\_Q |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Correlations |
| B | Std. Error | Beta | Zero-order |
| 1 | (Constant) | -1.287 | .977 |  | -1.318 | .199 |  |
| GRE\_Q | .005 | .002 | .434 | 2.778 | .010 | .611 |
| GRE\_V | .003 | .001 | .378 | 2.421 | .022 | .581 |

The *t* values here, expressed as *F* values, are 2.7782 = 7.717 and 2.4212 = 5.861. For the multiple *R2* and each of the predictors, I enter into the .sav file the values for *F* and degrees of freedom and ran the syntax.



In the r2 column, first row, is the *R2* for the overall model. For each predictor in that same column is the *pr2* for that predictor.

|  |  |  |  |
| --- | --- | --- | --- |
| **Coefficientsa** | | | |
| Model | | Correlations | |
| Partial | Part |
| 1 | (Constant) |  |  |
| GRE\_Q | .472 | .384 |
| GRE\_V | .422 | .334 |

The squared partial correlations are .4722 = .2228 for GRE\_Q and .4222 = .1781 for GRE\_V, within rounding error of the values in the r2 column in the .sav file. The confidence intervals for the predictors in that file are confidence intervals for *pr2*.

Jiah Yoo, a doctoral student in Social Psychology and Personality at the University of Wisconsin, wrote “, I think eta squared (aka the squared semipartial correlation coefficient) might be a better indicator of the effect size. Would there be a way to obtain a CI for squared semipartial correlation? I reported the squared semipartial correlation as the effect size in my paper and one of reviewers asked for its CI.

I agree with Jiah, I generally prefer the semipartial to the partial. I probably could modify Smithson’s SPSS syntax to get confidence intervals for the semipartial, but I am not motivated to do so since the solution is already available with SAS’ GLM procedure, and the syntax is very easy. Here I bring into SAS the same SPSS data set used above. Then I submit this code:

**proc** **GLM**;

model GPA =GRE\_Q GRE\_V / ss3 EFFECTSIZE alpha=**0.1**; **run**; **quit**;

Here is part of the output:

| **Proportion of Variation Accounted for** | |
| --- | --- |
| **Eta-Square** | 0.49 |
| **Omega-Square** | 0.44 |
| **90% Confidence Limits** | (0.22,0.61) |

The total model *R2* is .49, with a confidence interval running from .22 to .61

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** | **Total Variation Accounted For** | | | | **Partial Variation Accounted For** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Semipartial Eta-Square** | **Semipartial Omega-Square** | **Conservative 90% Confidence Limits** | | **Partial Eta-Square** | **Partial Omega-Square** | **90% Confidence Limits** | |
| **GRE\_Q** | 1 | 1.53550504 | 1.53550504 | 7.72 | 0.0098 | 0.1472 | 0.1257 | 0.0051 | 0.3356 | 0.2223 | 0.1830 | 0.0319 | 0.4043 |
| **GRE\_V** | 1 | 1.16596421 | 1.16596421 | 5.86 | 0.0225 | 0.1117 | 0.0909 | 0.0000 | 0.2965 | 0.1784 | 0.1394 | 0.0139 | 0.3626 |

Notice that for GRE\_V, the confidence interval for the *sr2*, but not for the *pr2* contains zero, even though the *F* test for GRE\_V is significant at *p* = .023. The reason for this discrepancy is that the *F* test excludes from the denominator of the *F* ratio (the error variance) variance that is explained by any of the predictors, just like the *pr2* excludes from its denominator variance that is explained by other predictors in the model. The denominator of the *sr2*, however, includes all of the variance in Y.

[Karl L. Wuensch](http://core.ecu.edu/psyc/WuenschK/KLW.htm), East Carolina University, Greenville, NC. September, 2016.