This page is archived and no longer maintained.

REGRESSION WITH SPSS FOR MULTIPLE REGRESSION ANALYSIS | SPSS ANNOTATED OUTPUT

This page shows an example multiple regression analysis with footnotes explaining the output. The analysis uses a data file about scores obtained by elementary schools, predicting **api00** from **ell**, **meals**, **yr_rnd**, **mobility**, **acs_k3**, **acs_46**, **full**, **emer** and **enroll** using the following SPSS commands. This example uses the [elemapi2 (https://stats.idre.ucla.edu/wp-content/uploads/2016/02/elemapi2-2.sav)](https://stats.idre.ucla.edu/wp-content/uploads/2016/02/elemapi2-2.sav) dataset.

```
regression
  /dependent api00
  /method=enter ell meals yr_rnd mobility acs_k3 acs_46 full emer enroll.
```

Variables Entered/Removed(b)[a]

Model	Variables Entered	Variables Removed	Method
1	ENROLL, ACS_46, MOBILITY, ACS_K3, EMER, ELL, YR_RND, MEALS, FULL(a)	.	Enter
a All requested variables entered.			
b Dependent Variable: API00			

Model Summary

Model[b]	R[c]	R Square[d]	Adjusted R Square[e]	Std. Error of the Estimate[f]
1	.919(a)	.845	.841	56.768
a Predictors: (Constant), ENROLL, ACS_46, MOBILITY, ACS_K3, EMER, ELL, YR_RND, MEALS, FULL				

ANOVA(b)

Model[g]		Sum of Squares[h]	df[i]	Mean Square[j]	F[k]	Sig.[k]
1	Regression	6740702.006	9	748966.890	232.409	.000(a)
	Residual	1240707.781	385	3222.618		
	Total	7981409.787	394			
a Predictors: (Constant), ENROLL, ACS_46, MOBILITY, ACS_K3, EMER, ELL, YR_RND, MEALS, FULL						
b Dependent Variable: API00						

Coefficients(a)

Model[l]	Unstandardized Coefficients		Standardized Coefficients	t[p]	Sig.[p]
	B[m]	Std. Error[n]	Beta[o]		
1 (Constant)	758.942	62.286		12.185	.000
ELL	-.860	.211	-.150	-4.083	.000
MEALS	-2.948	.170	-.661	-17.307	.000
YR_RND	-19.889	9.258	-.059	-2.148	.032
MOBILITY	-1.301	.436	-.069	-2.983	.003
ACS_K3	1.319	2.253	.013	.585	.559
ACS_46	2.032	.798	.055	2.546	.011
FULL	.610	.476	.064	1.281	.201
EMER	-.707	.605	-.058	-1.167	.244
ENROLL	-1.216E-02	.017	-.019	-.724	.469

a Dependent Variable: API00

a. This is a summary of the regression analysis performed. It lists the predictor variables and the outcome variable. It indicates that there was only one model tested and that all of the predictor variables were entered for that model.

b. This is a list of the models that were tested. In this case, there was only one model used.

c. R is the square root of R Square (shown in the next column).

d. R-Square is the proportion of variance in the dependent variable (api00) which can be predicted from the independent variables (**ell, meals, yr_rnd, mobility, acs_k3, acs_46, full, emer** and **enroll**). This value indicates that 84% of the variance in api00 can be predicted from the variables **ell, meals, yr_rnd, mobility, acs_k3, acs_46, full, emer** and **enroll**. Note that this is an overall measure of the strength of association, and does not reflect the extent to which any particular independent variable is associated with the dependent variable.

e. Adjusted R-square. As predictors are added to the model, each predictor will explain some of the variance in the dependent variable simply due to chance. One could continue to add predictors to the model which would continue to improve the ability of the predictors to explain the dependent variable, although some of this increase in R-square would be simply due to chance variation in that particular sample. The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population.   The value of R-square was .8446, while the value of Adjusted R-square was .8409. Adjusted R-squared is computed using the formula $1 - ((1-R\text{-}sq)(N\text{-}1 / N - k - 1))$. From this formula, you can see that when the number of observations is small and the number of predictors is large, there will be a much greater difference between R-square and adjusted R-square (because the ratio of $(N\text{-}1 / N - k - 1)$ will be much less than 1. By contrast, when the number of observations is very large compared to the number of predictors, the value of R-square and adjusted R-square will be much closer because the ratio of $(N\text{-}1)/(N\text{-}k\text{-}1)$ will approach 1.

f. Root MSE is the standard deviation of the error term, and is the square root of the Mean Square Error

g. This shows the model number (in this case we ran only one model, so it is model #1). Also, this column shows the source of variance, Regression, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Regression) and the variance which is not explained by the independent variables (Residual).   Note that the Sums of Squares for the Regression and Residual add up to the Total Variance, reflecting the fact that the Total Variance is partitioned into Regression and Residual variance.

h. These are the Sum of Squares associated with the three sources of variance, Total, Regression & Residual. These can be computed in many ways. Conceptually, these formulas can be expressed as:

   SSTotal.   The total variability around the mean. $(Y - Ybar)^2$.
   SSResidual. The sum of squared errors in prediction. $(Y - Ypredicted)^2$.
   SSRegression.   The improvement in prediction by using the predicted value of Y over just using the mean of Y. Hence, this would be the squared differences between the predicted value of Y and the mean of Y, $(Ypredicted - Ybar)^2$. Another way to think of this is the SSRegression is SSTotal – SSResidual. Note that the SSTotal = SSRegression + SSResidual. Note that SSRegression / SSTotal is equal to .10, the value of R-Square. This is because R-Square is the proportion of the variance explained by the independent variables, hence can be computed by SSRegression / SSTotal.

i. These are the degrees of freedom associated with the sources of variance. The Total variance has N-1 degrees of freedom (DF). In this case, there were N=395 observations, so the DF for total is 394. The Regression degrees of freedom corresponds to the number of predictors minus 1 (K-1). You may think this would be 9-1 (since there were 9 independent variables in the model: **ell**, **meals**, **yr_rnd**, **mobility**, **acs_k3**, **acs_46**, **full**, **emer** and **enroll**). But, the intercept is automatically included in the model (unless you explicitly omit the intercept). Including the intercept, there are 10 predictors, so the Regression has 10-1=9 degrees of freedom. The Residual degrees of freedom is the DFTotal minus the DFModel, 394 – 9 is 385.

j. These are the Mean Squares, the Sum of Squares divided by their respective DF. For the Regression, 6740702.01 / 9 is equal to 748966.89. For the Residual, 1240707.79 / 385 equals 3222.6176. These are computed so you can compute the F ratio, dividing the Mean Square Regression by the Mean Square Residual to test the significance of the predictors in the model.

k. The F Value is the Mean Square Regression (748966.89) divided by the Mean Square Residual (3222.61761), yielding F=232.41. The p value associated with this F value is very small (0.0000). These values are used to answer the question "Do the independent variables reliably predict the dependent variable?". The p value is compared to your alpha level (typically 0.05) and, if smaller, you can conclude "Yes, the independent variables reliably predict the dependent variable". You could say that the group of variables **ell**, **meals**, **yr_rnd**, **mobility**, **acs_k3**, **acs_46**, **full**, **emer** and **enroll** can be used to reliably predict **api00** (the dependent variable). If the p value were greater than 0.05, you would say that the group of independent variables do not show a significant relationship with the dependent variable, or that the group of independent variables do not reliably predict the dependent variable. Note that this is an overall significance test assessing whether the group of independent variables when used together reliably predict the dependent variable, and does not address the ability of any of the particular independent variables to predict the dependent variables. The ability of each individual independent variable to predict the dependent variable is addressed in the table below where each of the individual variables are listed.

l. This shows the model number (in this case we ran only one model, so it is model #1). These are the predictors in the regression equation.

m. These are the values for the regression equation for predicting the dependent variable from the independent variable. The regression equation is presented in many different ways, for example...

```
Ypredicted = b0 + b1*x1 + b2*x2 + b3*x3 . . .
```

The column of parameter estimates provides the values for b0, b1, b2, b3, b4, b5, b6, b7, b8 and b9 for this equation. Expressed in terms of the variables used in this example, the regression equation is

```
api00Predicted = 758.942 -.860*ell -2.948*meals -19.889*yr_rnd -1.301*mobility
                 +1.319*acs_k3 +2.032*acs_46 +.610*full -.707*emer -.012*enroll
```

   These estimates tell you about the relationship between the independent variables and the dependent variable. These estimates tell the amount of increase in api00 that would be predicted by a 1 unit increase in the predictor.   Note: For the independent variables which are not significant, the coefficients are not significantly different from 0, which should be taken into account when interpreting the coefficients. (See the columns with the t value and p value about testing whether the coefficients are significant.)

   ell – The coefficient (parameter estimate) is -.86. So, for every unit increase in ell, a .86 unit decrease in api00 is predicted. Or, for every increase of one percentage point of api00, ell is predicted to be lower by .86. This is significantly different from 0.

   meals – For every unit increase in meals, there is a 2.95 unit decrease in the predicted api00.

   yr_rnd – For every unit increase of yr_rnd, the predicted value of api00 would be 19.89 units lower.

   mobility – For every unit increase in mobility, api00 is predicted to be 1.30 units lower.

   acs_k3 – For every unit increase in acs_k3, api00 is predicted to be 1.32 units higher.

   acs_46 – For every unit increase in acs_46, api00 is predicted to be 2.03 units higher.

   full – For every unit increase in mobility, api00 is predicted to be .61 unit higher.

   emer – For every unit increase in mobility, api00 is predicted to be .71 unit lower.

   enroll – For every unit increase in mobility, api00 is predicted to be .01 unit lower.

n. These are the standard errors associated with the coefficients. The standard error is used for testing whether the coefficient is significantly different from 0 by dividing the coefficient by the standard error to obtain a t value (see the column with t values and p values). The standard errors can also be used to form a confidence interval for the parameter.

o. These are the values for a regression equation if all of the variables are standardized to have a mean of zero and a standard deviation of one. Because the standardized variables are all expressed in the same units, the magnitudes of the standardized coefficients indicate which variables have the greatest effects on the predicted value. This is not necessarily true of the unstandardized coefficients. Because the magnitudes of the unstandardized coefficients can largely depend on the units of the variables, the effects of the variable on the prediction can be difficult to gauge. While the standardized coefficients may vary significantly from the unstandardized coefficients in magnitude, the sign (positive or negative) of the coefficients is unchanged.

p. These columns provide the t value and 2 tailed p value used in testing the null hypothesis that the coefficient is 0. If you use a 2 tailed test, then you would compare each p value to your preselected value of alpha. Coefficients having p values less than alpha are significant. For example, if you chose alpha to be 0.05, coefficients having a p value of 0.05 or less would be statistically significant (i.e., you can reject the null hypothesis and say that the coefficient is significantly different from 0). If you use a 1 tailed test (i.e. you predict that the parameter will go in a particular direction), then you can divide the p value by 2 before comparing it to your preselected alpha level. For example, if you used a one tailed test for **yr_rnd** and you predicted the coefficient to be negative, then you could take 0.032 and divide it by 2 yielding 0.016 and that would be the value you would compare to your preselected alpha level. In the examples below, we will use 2 tailed tests with an alpha of 0.05.

The coefficient for meals is significantly different from 0 using alpha of 0.05 because its p value of 0.000 is smaller than 0.05.

The coefficient for yr_rnd (-19.89) is significantly different from 0 because its p value is smaller than 0.05.

The coefficient for mobility is significantly different from 0 using alpha of 0.05 because its p value of 0.003 is smaller than 0.05.

The coefficient for acs_k3 is not significantly different from 0 using alpha of 0.05 because its p value of .559 is greater than 0.05.

The coefficient for acs_46 is significantly different from 0 using alpha of 0.05 because its p value of 0.011 is smaller than 0.05.

The coefficient for full is not significantly different from 0 using alpha of 0.05 because its p value of .201 is greater than 0.05.

The coefficient for emer is not significantly different from 0 using alpha of 0.05 because its p value of .244 is greater than 0.05.

The coefficient for enroll is not significantly different from 0 using alpha of 0.05 because its p value of .469 is greater than 0.05.

The constant is significantly different from 0 at the 0.05 alpha level. However, having a significant intercept is seldom interesting.