

SPSS Annotated Output Regression Analysis

This page shows an example regression analysis with footnotes explaining the output. These data ([hsb2](#)) were collected on 200 high schools students and are scores on various tests, including science, math, reading and social studies (**socst**). The variable **female** is a dichotomous variable coded 1 if the student was female and 0 if male.

In the syntax below, the **get file** command is used to load the data into SPSS. In quotes, you need to specify where the data file is located on your computer. In the **regression** command, the **statistics** subcommand must come before the **dependent** subcommand. You list the independent variables after the equals sign on the **method** subcommand. The **statistics** subcommand is not needed to run the regression, but on it we can specify options that we would like to have included in the output.

Please note that SPSS sometimes includes footnotes as part of the output. We have left those intact and have started ours with the next letter of the alphabet.

```
get file "c:\hsb2.sav".
```

```
regression
  /statistics coeff outs r anova ci
  /dependent science
  /method = enter math female socst read.
```

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	reading score, female, social studies score, math score ^a		Enter

a. All requested variables entered.

b. Dependent Variable: science score

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.699 ^a	.489	.479	7.14817

a. Predictors: (Constant), reading score, female, social studies score, math score

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	9543.721	4	2385.930	46.695	.000 ^a
	Residual	9963.779	195	51.096		
	Total	19507.500	199			

a. Predictors: (Constant), reading score, female, social studies score, math score

b. Dependent Variable: science score

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	12.325	3.194		3.859	.000	6.027	18.624
	math score	.389	.074	.368	5.252	.000	.243	.535
	female	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score	.050	.062	.054	.801	.424	-.073	.173
	reading score	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

Variables in the model**Variables Entered/Removed^b**

Model ^c	Variables Entered ^d	Variables Removed ^e	Method ^f
1	reading score, female, social studies score, math score ^a		Enter

a. All requested variables entered.

b. Dependent Variable: science score

c. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

d. **Variables Entered** - SPSS allows you to enter variables into a regression in blocks, and it allows stepwise regression. Hence, you need to know which variables were entered into the current regression. If you did not block your independent variables or use stepwise regression, this column should list all of the independent variables that you specified.

e. **Variables Removed** - This column listed the variables that were removed from the current regression. Usually, this column will be empty unless you did a stepwise regression.

f. **Method** - This column tells you the method that SPSS used to run the regression. "Enter" means that each independent variable was entered in usual fashion. If you did a stepwise regression, the entry in this column would tell you that.

Overall Model Fit

Model Summary

Model ^b	R ^c	R Square ^d	Adjusted R Square ^e	Std. Error of the Estimate ^f
1	.699 ^a	.489	.479	7.14817

a. Predictors: (Constant), reading score, female, social studies score, math score

b. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

c. **R** - R is the square root of R-Squared and is the correlation between the observed and predicted values of dependent variable.

d. **R-Square** - This is the proportion of variance in the dependent variable (**science**) which can be explained by the independent variables (**math, female, socst and read**). 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** - This is an adjustment of the R-squared that penalizes the addition of extraneous predictors to the model. Adjusted R-squared is computed using the formula $1 - ((1 - Rsq)(N - 1) / (N - k - 1))$ where k is the number of predictors.

f. **Std. Error of the Estimate** - This is also referred to as the root mean squared error. It is the standard deviation of the error term and the square root of the Mean Square for the Residuals in the ANOVA table (see below).

Anova Table**ANOVA^b**

Model ^c		Sum of Squares ^e	df ^f	Mean Square ^g	F ^h	Sig. ^h
1	Regression ^d	9543.721	4	2385.930	46.695	.000 ^a
	Residual ^d	9963.779	195	51.096		
	Total ^d	19507.500	199			

a. Predictors: (Constant), reading score, female, social studies score, math score

b. Dependent Variable: science score

c. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

d. **Regression, Residual, Total** - Looking at the breakdown of variance in the outcome variable, these are the categories we will examine: Regression, Residual, and Total. The Total variance is partitioned into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Error).

e. **Sum of Squares** - These are the Sum of Squares associated with the three sources of variance, Total, Model and Residual. 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).

f. **df** - These are the degrees of freedom associated with the sources of variance. The total variance has N-1 degrees of freedom. The Regression degrees of freedom corresponds to the number of coefficients estimated minus 1. Including the intercept, there are 5 coefficients, so the model has $5 - 1 = 4$ degrees of freedom. The Error degrees of freedom is the DF total minus the DF model, $199 - 4 = 195$.

g. **Mean Square** - These are the Mean Squares, the Sum of Squares divided by their respective DF.

h. **F and Sig.** - This is the F-statistic the p-value associated with it. The F-statistic is the Mean Square (Regression) divided by the Mean Square (Residual): $2385.93/51.096 = 46.695$. The p-value is compared to some alpha level in testing the null hypothesis that all of the model coefficients are 0.

Parameter Estimates

Coefficients^a

Model ^b		Unstandardized Coefficients		Standardized Coefficients	t ^g	Sig. ^g	95% Confidence Interval for B	
		B ^d	Std. Error ^e	Beta ^f			Lower Bound ^h	Upper Bound ^h
1	(Constant) ^c	12.325	3.194		3.859	.000	6.027	18.624
	math score ^c	.389	.074	.368	5.252	.000	.243	.535
	female ^c	-2.010	1.023	-.101	-1.965	.051	-4.027	.007
	social studies score ^c	.050	.062	.054	.801	.424	-.073	.173
	reading score ^c	.335	.073	.347	4.607	.000	.192	.479

a. Dependent Variable: science score

b. **Model** - SPSS allows you to specify multiple models in a single **regression** command. This tells you the number of the model being reported.

c. This column shows the predictor variables (**constant**, **math**, **female**, **socst**, **read**). The first variable (**constant**) represents the constant, also referred to in textbooks as the Y intercept, the height of the regression line when it crosses the Y axis. In other words, this is the predicted value of **science** when all other variables are 0.

d. **B** - 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:

$$Y_{\text{predicted}} = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4$$

The column of estimates provides the values for b_0 , b_1 , b_2 , b_3 and b_4 for this equation.

math - The coefficient for **math** is .389. So for every unit increase in **math**, a 0.39 unit increase in **science** is predicted, holding all other variables constant.

female - For every unit increase in **female**, we expect a -2.010 unit decrease in the **science** score, holding all other variables constant. Since **female** is coded 0/1 (0=male, 1=female) the interpretation is more simply: for females, the predicted science score would be 2 points lower than for males.

socst - The coefficient for **socst** is .050. So for every unit increase in **socst**, we expect an approximately .05 point increase in the science score, holding all other variables constant.

read - The coefficient for **read** is .335. So for every unit increase in **read**, we expect a .34 point increase in the science score.

e. **Std. Error** - These are the standard errors associated with the coefficients.

f. **Beta** - These are the standardized coefficients. These are the coefficients that you would obtain if you standardized all of the variables in the regression, including the dependent and all of the independent variables, and ran the regression. By standardizing the variables before running the regression, you have put all of the variables on the same scale, and you can compare the magnitude of the coefficients to see which one has more of an effect. You will also notice that the larger betas are associated with the larger t-values and lower p-values.

g. **t** and **Sig.** - These are the t-statistics and their associated 2-tailed p-values used in testing whether a given coefficient is significantly different from zero. Using an alpha of 0.05:

The coefficient for **math** is significantly different from 0 because its p-value is 0.000, which is smaller than 0.05.

The coefficient for **soest** (.0498443) is not statistically significantly different from 0 because its p-value is definitely larger than 0.05.

The coefficient for **read** (.3352998) is statistically significant because its p-value of 0.000 is less than .05.

The intercept is significantly different from 0 at the 0.05 alpha level.

h. **95% Confidence Limit for B Lower Bound and Upper Bound** - These are the 95% confidence intervals for the coefficients. The confidence intervals are related to the p-values such that the coefficient will not be statistically significant if the confidence interval includes 0. These confidence intervals can help you to put the estimate from the coefficient into perspective by seeing how much the value could vary.

[How to cite this page](#)

[Report an error on this page](#)

UCLA Researchers are invited to our [Statistical Consulting Services](#)

We recommend others to our list of [Other Resources for Statistical Computing Help](#)

These pages are [Copyrighted \(c\) by UCLA Academic Technology Services](#)

The content of this web site should not be construed as an endorsement of any particular web site, book, or software product by the University of California.
