

Multiple Regression

Multiple regression is an extension of simple (bi-variate) regression. The goal of multiple regression is to enable a researcher to assess the relationship between a dependent (predicted) variable and several independent (predictor) variables. The end result of multiple regression is the development of a regression equation (line of best fit) between the dependent variable and several independent variables.

There are several types of multiple regression analyses (e.g. standard, hierarchical, setwise, stepwise) only two of which will be presented here (standard and stepwise). Which type of analysis is conducted depends on the question of interest to the researcher.

Suppose, for example, a college admissions officer was interested in using verbal SAT scores and high school grade point averages (as independent or predictor variables) to predict college grade point averages (as a dependent or predicted variable).

Standard multiple regression would be used to address a couple of questions: a) what is the size of the overall relationship between college GPA (the predicted variable) and the independent (predictor) variables of verbal SAT scores and high school GPA?; and b) how much does **each** independent (predictor) variable uniquely contributed to that relationship? In **standard** multiple regression **all** predictor variables are entered into the regression equation at once.

Stepwise multiple regression would be used to answer a different question. The focus of **stepwise** regression would be the question of what the best combination of independent (predictor) variables would be to predict the dependent (predicted) variable, e.g. college GPA. In **stepwise** regression **not all** independent (predictor) variables, e.g. high school GPA and verbal SAT scores, may end up in the equation.

In a **stepwise** regression, predictor variables are entered into the regression equation **one at a time** based upon statistical criteria. At each step in the analysis the predictor variable that contributes the most to the prediction equation in terms of increasing the multiple correlation, R , is entered first. This process is continued only if additional variables add anything statistically to the regression equation. When no additional predictor variables add anything statistically meaningful to the regression equation, the analysis stops. Thus, **not all** predictor variables may enter the equation in **stepwise** regression. Listed below are the verbal SAT scores, college GPAs, and high school GPAs collected on 11 students by an admissions officer.

Student	Verbal SAT	College GPA	High School GPA
Jane	760	3.95	98
Bob	720	3.68	95
Rich	710	3.66	94
Laura	700	3.20	92
Karen	650	3.10	90
Randy	580	2.90	88
Jim	570	2.70	85
Paul	520	2.70	82
Glen	520	2.50	80
Bill	500	2.30	78
Mary	490	2.00	70

1. Logon to system
2. **Click Start > Programs > SPSS for Windows > SPSS 10.1 for Windows.** At this point a window will appear asking you what you would like to do. Click on the circle next to Type in Data (2nd option in list) and then click **OK** at the bottom of the window.
3. A Data Editor will appear. Look in the lower left corner of the screen. You should see a **Data View** tab and to the right of it a **Variable View** tab. The **Variable View** tab will be used first for the Data **Definition** Phase of creating a data file. The **Data View** tab will be used to actually enter the raw numbers listed above. (See pages 1-3 for a more detailed explanation of creating data files.)

DATA DEFINITION PHASE

- Click on the **Variable View** tab in the lower left corner. A new screen will appear with the following words at the top of each column.

Name Type Width Decimals **Label** **Values** Missing Columns Align Measure

- Click on the white cell in **Row 1** under the word **Name** and type in the word **verbsat** (for Verbal SAT score).
- Click on the white cell in **Row 1** under the word **Label** and type in **Verbal SAT**. (Doing this will provide you with a more expansive label in the results output).
- Click on the white cell in **Row 2** under the word **Name** and type in the word **colgpa** (for College GPA).
- Click on the white cell in **Row 2** under the word **Label** and type in **College GPA**. (Doing this will provide you with a more expansive label in the results output).
- Click on the white cell in **Row 3** under the word **Name** and type in **hsgpa** (for high school GPA).
- Click on the white cell in **Row 3** under the word **Label** and type in **high school gpa** (Doing this will provide you with a more expansive label in the results output).

DATA ENTRY PHASE

- Click on the **Data View** tab in the lower left corner. The data **view** screen will now appear with Column 1 named **verbsat** (for the Verbal Sat variable) and Column 2 named **colgpa** (for the College GPA variable) and Column 3 named **hsgpa** (for the High School GPA variable)..
- Enter data the data for the 11 students (Jane through Mary) as follows> Click on the top left cell under the first column **verbsat** and enter:

760 tab 3.95 tab 98 enter
720 tab 3.68 tab 95 enter

Then mouse to second row to enter the data for the second case.
Then mouse to the third row to enter the data for the third case etc. for the remaining cases.

710 tab 3.66 tab 94 enter
700 tab 3.20 tab 92 enter
650 tab 3.10 tab 90 enter
580 tab 2.90 tab 88 enter
570 tab 2.70 tab 85 enter
520 tab 2.70 tab 82 enter
520 tab 2.50 tab 80 enter
500 tab 2.30 tab 78 enter
490 tab 2.00 tab 70 enter

The data may also be entered down one column at a time, entering all the verbsat data, then moving on to column 2 and entering the data for the college gpa, and then on to column 3 and entering the data for the high school gpa.

Data Analysis

- Click on **Analyze** at top of the screen then
 - Click on **Regression** then
 - Click on **Linear**
- Highlight **colgpa** by clicking on it and then
 - Click on **arrow >** to transfer this name to the **Dependent** Box
- Highlight **verbsat** by clicking on it and then
 - Click on **arrow >** to transfer this name to the **Independent(s)** Box
- Highlight **hsgpa** by clicking on it and then

- a. Click on **arrow >** to transfer this name to the **Independent (s)** Box
5. Click on **Down arrow** adjacent to the **Method** Box and then
 - a. Click on **Enter**
6. Click on **OK**
7. Your results will appear in a Window. Scroll up using the slide bar on the right to the top of the output. The **key results** of this analysis are presented below. **There are other important tables which may appear on your screen that are NOT reproduced below.**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.981 ^a	.963	.954	.13190

a. Predictors: (Constant), High School GPA, Verbal SAT

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.633	2	1.816	104.395	.000 ^a
	Residual	.139	8	.017		
	Total	3.772	10			

a. Predictors: (Constant), High School GPA, Verbal SAT

b. Dependent Variable: College GPA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.245	.622		-3.611	.007
	Verbal SAT	2.531E-03	.001	.411	1.989	.082
	High School GPA	4.241E-02	.015	.584	2.824	.022

a. Dependent Variable: College GPA

8. **Interpretation and APA writing template for the Standard Multiple Regression Results Above:**

A standard multiple regression analysis was conducted to evaluate how well high school grade point average and verbal SAT scores predicted college GPA. The linear combination of high school GPA and verbal SAT scores was significantly related to college GPA, $F((2,8) = 104.395, p < .001$. The multiple correlation coefficient was .98, indicating that approximately 96% of the variance of the college GPA can be accounted for by the linear combination of high school GPA and verbal SAT scores. The regression equation for predicting the college GPA was:

$$\text{Predicted College GPA} = .042406 \times \text{high school GPA} + .002531 \times \text{Verbal SAT Score} - 2.244615$$

Data Analysis for a Stepwise Regression

1. Click on **Analyze** at top of the screen then
 - a. Click on **Regression** then
 - b. Click on **Linear**
2. Highlight **colgpa** by clicking on it and then
 - a. Click on **arrow >** to transfer this name to the **Dependent** Box
3. Highlight **verbsat** by clicking on it and then
 - a. Click on **arrow >** to transfer this name to the **Independent(s)** Box
4. Highlight **hsgpa** by clicking on it and then
 - a. Click on **arrow >** to transfer this name to the **Independent (s)** Box
5. Click on **Down arrow** adjacent to the **Method** Box and then
 - a. Click on **Stepwise**
6. Click on **OK**
7. Your results will appear in a Window. Scroll up using the slide bar on the right to the top of the output. The **key** results of this analysis are presented below. **There are other important tables which may appear on your screen that are NOT reproduced below.**

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	high school gpa	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: College GPA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.972 ^a	.945	.939	.1520

a. Predictors: (Constant), high school gpa

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.564	1	3.564	154.212	.000 ^a
	Residual	.208	9	2.311E-02		
	Total	3.772	10			

a. Predictors: (Constant), high school gpa

b. Dependent Variable: College GPA

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-3.139	.494		-6.352	.000
	High School GPA	7.061E-02	.006	.972	12.418	.000

a. Dependent Variable: College GPA

Excluded Variables^b

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Verbal SAT	.411 ^a	1.989	.082	.575	.108

a. Predictors in the Model: (Constant), High School GPA

b. Dependent Variable: College GPA

8. **Interpretation and APA writing template for the Stepwise Multiple Regression Results Above:**

A stepwise multiple regression was conducted to evaluate whether **both** high school grade point average and verbal SAT scores were necessary to predict college GPA. At step 1 of the analysis high school GPA entered into the regression equation and was significantly related to college GPA $F(1,9) = 154.21, p < .001$. The multiple correlation coefficient was .97, indicating approximately 94.5% of the variance of the college GPA could be accounted for by high school GPA scores. Verbal SAT scores did not enter into the equation at step 2 of the analysis ($t = 1.989, p > .05$). Thus the regression equation for predicting college GPA was:

$$\text{Predicted College GPA} = .0706 \times \text{high school GPA} - 3.139$$