

## Simple Regression

For a simple (bivariate) linear regression problem, a researcher collects data for each participant in the study on an independent or predictor variable (X), and a dependent or criterion variable (Y). Simple linear regression evaluates how well a single independent (predictor) variable predicts a participant's response on a dependent (predicted) variable. If a linear relationship exists between two variables, regression methods utilize the existence of that relationship to predict the values of one variable (the predicted variable) from those of another (the predictor variable).

Suppose, for example, an admission's officer at a university is concerned about the methods that are being used to select students for entry into the university. In the past the university had been using student scores on the verbal section of the SAT as a predictor of likelihood of success while in college. The admissions officer was wondering whether this was a wise practice. The question of interest was whether success in college as measured by a student's college grade point average (GPA) could be predicted from knowledge of that student's score on the verbal section of the SAT. The admissions officer then took a random test sample of students who had recently graduated from the university and looked up their final college GPA as well as their entrance scores on the verbal section of the SAT which s/he had collected 4 years earlier. The admissions officer obtained the following information:

Student	Verbal SAT Score	Final College GPA
Jane	760	3.95
Bob	720	3.68
Rich	710	3.66
Laura	700	3.20
Karen	650	3.10
Randy	580	2.90
Jim	570	2.70
Paul	520	2.70
Glen	520	2.50
Bill	500	2.30
Mary	490	2.00

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 (2<sup>nd</sup> 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

4. 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

5. Click on the white cell in **Row 1** under the word **Name** and type in the word **verbsat** (for Verbal SAT score).
6. 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).
7. Click on the white cell in **Row 2** under the word **Name** and type in the word **colgpa** (for College GPA).

8. 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).

## DATA ENTRY PHASE

9. 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).
10. Enter data for cases 1 through 11 (Jane through Mary) as follows> Click on the top left cell under the first column **verbsat** and enter:

760 tab 3.95 enter Then mouse to the second row to enter the data for the second case.

720 tab 3.68 enter. Then mouse to the third row to enter the data for the third case etc. for the remaining cases

710 tab 3.66 enter

700 tab 3.20 enter

650 tab 3.10 enter

580 tab 2.90 enter

570 tab 2.70 enter

520 tab 2.70 enter

520 tab 2.50 enter

500 tab 2.30 enter

490 tab 2.00 enter

The data may also be entered down one column at a time, entering all the verbal sat data first, then moving on to column 2 and entering the data for college gpa.

## Data Analysis

1. Click on **Analyze** at top of 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** Box
4. Underneath the Independent(s) box adjacent to the word **Method** should be the word **Enter**. If not, then click on down arrow and select the method of Enter.
5. Click **Statistics** button toward the bottom of Linear Regression Screen then
  - a. Click on White box next to **Model Fit** to place a check mark in it then
  - b. Click on White box next to **Descriptives** to place a check mark in it
  - c. Click on the **Continue** tab then
6. Click **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 results of this analysis are presented below.
8. Note that at step 5 above, adjacent to the **Statistics** button is a **Plot** button. You have an option at that point to select among many plot figures.

**Descriptive Statistics**

	Mean	Std. Deviation	N
College GPA	2.9718	.61415	11
Verbal SAT	610.9091	99.74513	11

**Correlations**

		College GPA	Verbal SAT
Pearson Correlation	College GPA	1.000	.962
	Verbal SAT	.962	1.000
Sig. (1-tailed)	College GPA	.	.000
	Verbal SAT	.000	.
N	College GPA	11	11
	Verbal SAT	11	11

**Variables Entered/Removed<sup>b</sup>**

Model	Variables Entered	Variables Removed	Method
1	Verbal SAT <sup>a</sup>	.	Enter

a. All requested variables entered.

b. Dependent Variable: College GPA

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.962 <sup>a</sup>	.926	.918	.17572

a. Predictors: (Constant), Verbal SAT

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.494	1	3.494	113.147	.000 <sup>a</sup>
	Residual	.278	9	.031		
	Total	3.772	10			

a. Predictors: (Constant), Verbal SAT

b. Dependent Variable: College GPA

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.648	.344		-1.883	.092
	Verbal SAT	5.926E-03	.001	.962	10.637	.000

a. Dependent Variable: College GPA

9. For the problem above the null and alternative hypotheses are spelled out below:

$H_{\text{null}}$ : There is no linear relationship (or predictability) between verbal SAT score and college GPA

$H_{\text{alt}}$ : There is a positive linear relationship between verbal SAT score and college GPA

10. **Interpretation and APA writing template for Results Above:**

A linear regression analysis was conducted on the bivariate data set to evaluate the prediction of college grade point average (GPA) from a student's score on the verbal component of the college SAT. The results indicate that there is a linear relationship between the two variables. As a student's score on the verbal section of the SAT increased, their overall GPA at the end of four years of college also tended to increase. The regression equation for predicting the college four year GPA is:

$$\text{Predicted Four Year College GPA} = .005926 \text{ Verbal SAT Score} - .648$$

As expected students who scored well on the verbal component of the SAT tended to have higher GPAs at the end of four years of college. The correlation between Verbal SAT score and college GPA was  $r(9) = +.962$ ,  $p < .05$ . Approximately 93% of the variance in the GPA scores was accounted for by its linear relationship with score on the verbal component of the SAT.