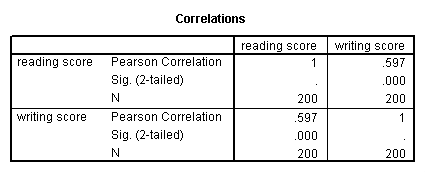
**Correlation**

A correlation is useful when you want to see the relationship between two (or more) normally distributed interval variables.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/spss/whatstat/#hsb) we can run a correlation between two continuous variables, **read** and **write**.

**correlations**

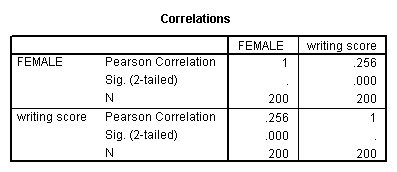
**/variables = read write.**



In the second example, we will run a correlation between a dichotomous variable, **female**, and a continuous variable, **write**. Although it is assumed that the variables are interval and normally distributed, we can include dummy variables when performing correlations.

**correlations**

**/variables = female write.**



In the first example above, we see that the correlation between **read** and **write** is 0.597.  By squaring the correlation and then multiplying by 100, you can determine what percentage of the variability is shared.  Let’s round 0.597 to be 0.6, which when squared would be .36, multiplied by 100 would be 36%.  Hence **read** shares about 36% of its variability with **write**.  In the output for the second example, we can see the correlation between **write** and **female** is 0.256.  Squaring this number yields .065536, meaning that **female** shares approximately 6.5% of its variability with **write**.

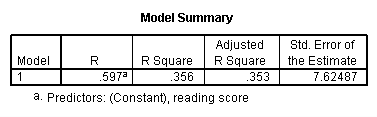
## Simple linear regression

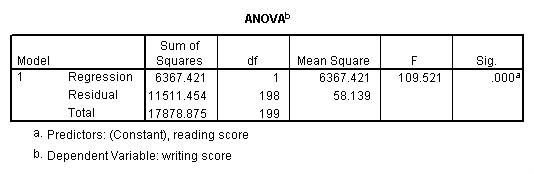
Simple linear regression allows us to look at the linear relationship between one normally distributed interval predictor and one normally distributed interval outcome variable.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/spss/whatstat/#hsb), say we wish to look at the relationship between writing scores (**write**) and reading scores (**read**); in other words, predicting **write** from **read**.

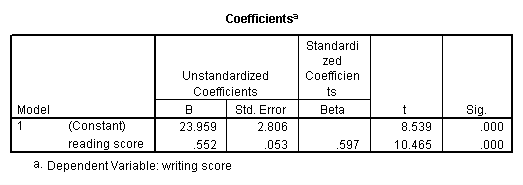
**regression variables = write read**

**/dependent = write**

**/method = enter.**





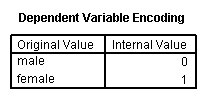


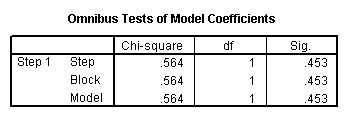
We see that the relationship between **write** and **read** is positive (.552) and based on the t-value (10.47) and p-value (0.000), we would conclude this relationship is statistically significant.  Hence, we would say there is a statistically significant positive linear relationship between reading and writing.

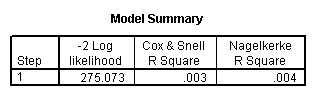
## Simple logistic regression

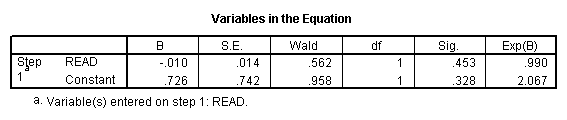
Logistic regression assumes that the outcome variable is binary (i.e., coded as 0 and 1).  We have only one variable in the [hsb2 data file](https://stats.idre.ucla.edu/spss/whatstat/#hsb) that is coded 0 and 1, and that is **female**.  We understand that **female** is a silly outcome variable (it would make more sense to use it as a predictor variable), but we can use **female** as the outcome variable to illustrate how the code for this command is structured and how to interpret the output.  The first variable listed after the **logistic** command is the outcome (or dependent) variable, and all of the rest of the variables are predictor (or independent) variables.  In our example, **female** will be the outcome variable, and **read** will be the predictor variable.  As with OLS regression, the predictor variables must be either dichotomous or continuous; they cannot be categorical.

**logistic regression female with read.**









The results indicate that reading score (**read**) is not a statistically significant predictor of gender (i.e., being female), Wald = .562, p = 0.453. Likewise, the test of the overall model is not statistically significant, LR chi-squared – 0.56, p = 0.453.

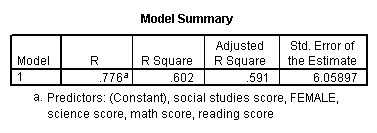
## Multiple regression

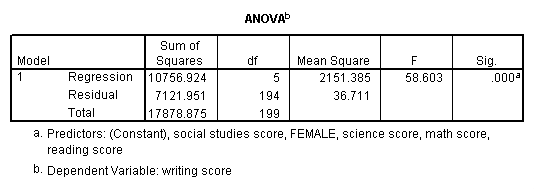
Multiple regression is very similar to simple regression, except that in multiple regression you have more than one predictor variable in the equation.  For example, using the [hsb2 data file](https://stats.idre.ucla.edu/spss/whatstat/#hsb) we will predict writing score from gender (**female**), reading, math, science and social studies (**socst**) scores.

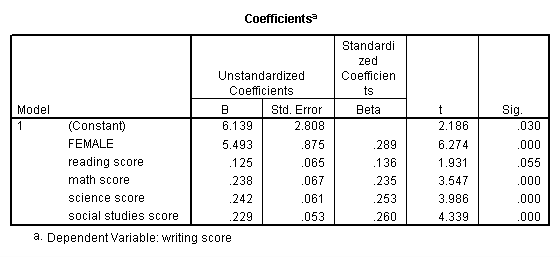
**regression variable = write female read math science socst**

**/dependent = write**

**/method = enter.**





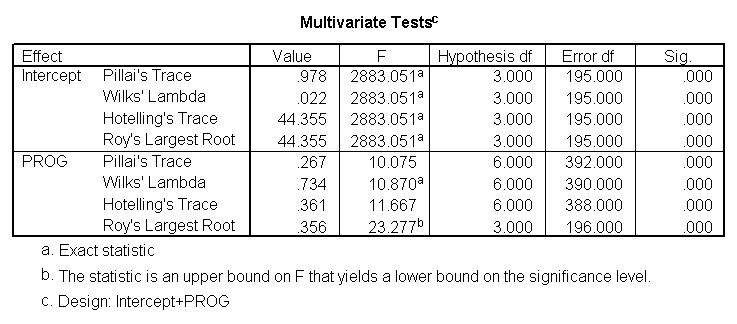


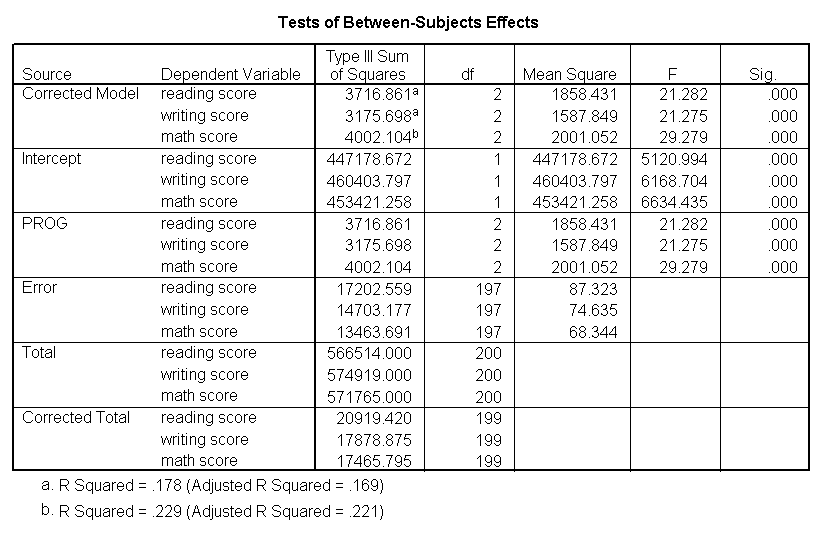
The results indicate that the overall model is statistically significant (F = 58.60, p = 0.000).  Furthermore, all of the predictor variables are statistically significant except for **read**.

## One-way MANOVA

MANOVA (multivariate analysis of variance) is like ANOVA, except that there are two or more dependent variables. In a one-way MANOVA, there is one categorical independent variable and two or more dependent variables. For example, using the [hsb2 data file](https://stats.idre.ucla.edu/spss/whatstat/#hsb), say we wish to examine the differences in **read**, **write** and **math** broken down by program type (**prog**).

**glm read write math by prog.**



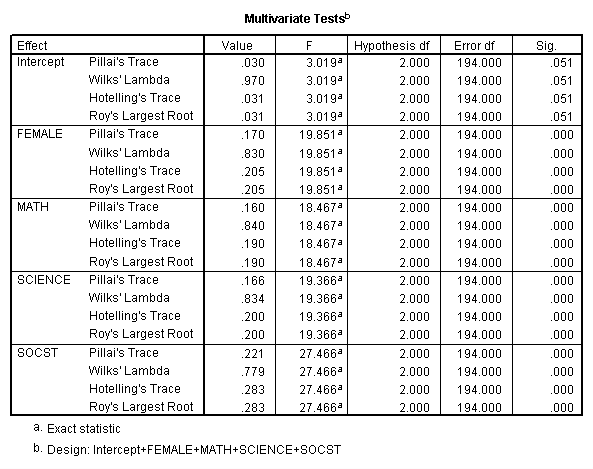


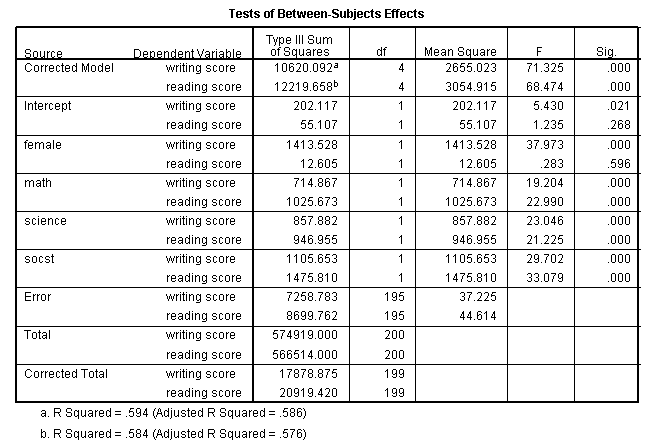
The students in the different programs differ in their joint distribution of **read**, **write** and **math**.

## Multivariate multiple regression

Multivariate multiple regression is used when you have two or more dependent variables that are to be predicted from two or more independent variables.  In our example, we will predict **write** and **read** from **female**, **math**, **science** and social studies (**socst**) scores.

**glm write read with female math science socst.**





These results show that all of  the variables in the model have a statistically significant relationship with the joint distribution of **write**and**read**.

## Factor analysis

Factor analysis is a form of exploratory multivariate analysis that is used to either reduce the number of variables in a model or to detect relationships among variables.  All variables involved in the factor analysis need to be interval and are assumed to be normally distributed.  The goal of the analysis is to try to identify factors which underlie the variables.  There may be fewer factors than variables, but there may not be more factors than variables.  For our example, let’s suppose that we think that there are some common factors underlying the various test scores.  We will include subcommands for varimax rotation and a plot of the eigenvalues.  We will use a principal components extraction and will retain two factors. (Using these options will make our results compatible with those from SAS and Stata and are not necessarily the options that you will want to use.)

**factor**

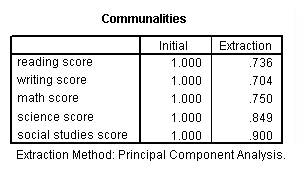
**/variables read write math science socst**

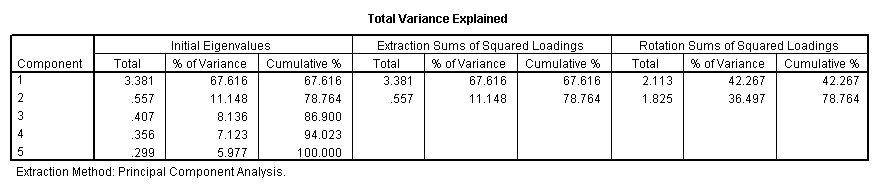
**/criteria factors(2)**

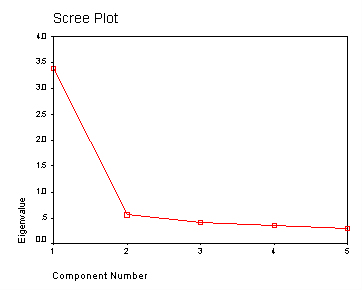
**/extraction pc**

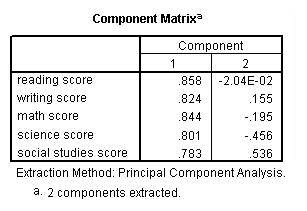
**/rotation varimax**

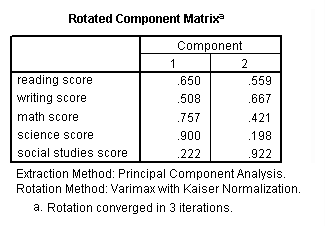
**/plot eigen.**

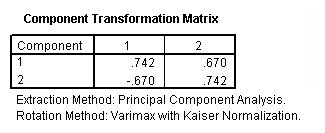












Communality (which is the opposite of uniqueness) is the proportion of variance of the variable (i.e., **read**) that is accounted for by all of the factors taken together, and a very low communality can indicate that a variable may not belong with any of the factors.  The scree plot may be useful in determining how many factors to retain.  From the component matrix table, we can see that all five of the test scores load onto the first factor, while all five tend to load not so heavily on the second factor.  The purpose of rotating the factors is to get the variables to load either very high or very low on each factor.  In this example, because all of the variables loaded onto factor 1 and not on factor 2, the rotation did not aid in the interpretation. Instead, it made the results even more difficult to interpret.