Tim Xeriland
Unit 1

A. Descriptive, Graphical

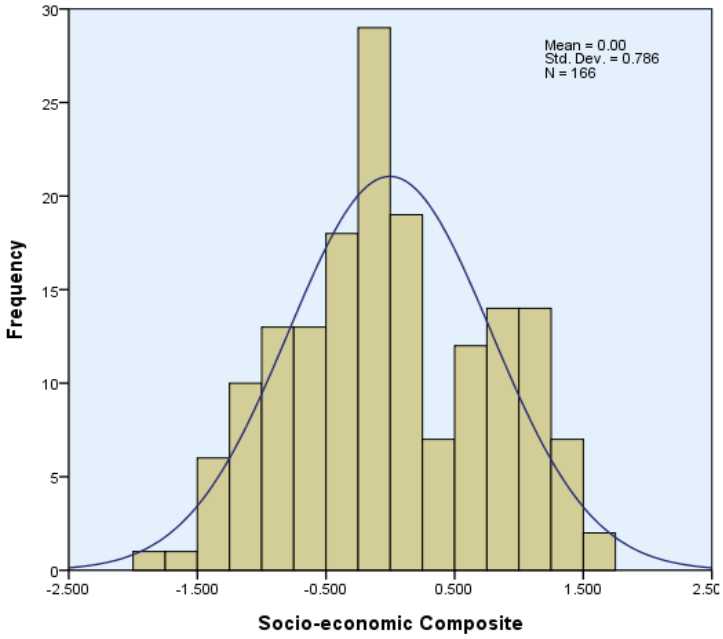
1. Examine the distribution of each variable separately by considering the descriptive statistics (mean, standard deviation, minimum, and maximum) and a histogram. Describe the shape of each distribution. Do the values of the variable seem sensible (for example, it would not be sensible to have a negative value for someone's height)?

Social Economic Status (SES):
The following descriptive statics table shows the basic descriptive statistics for SES. The mean of SES is very close to zero, and the minimum value is slightly more extreme than the maximum value. Skewness value close to zero indicates that the shape of the histogram will be bell shaped and the kurtosis is indicated that the slight left tail shape. The following histogram supports the above comment.

Descriptive Statistics

		Statistic	Std. Error
Socio-economic Composite	N	166	
	Range	3.500	
	Minimum	-1.862	
	Maximum	1.638	
	Mean	-.00342	.061038
	Std. Deviation	.786423	
	Variance	.618	
	Skewness	.073	.188
	Kurtosis	-.758	.375
Valid N (listwise)	N	166	

Histogram

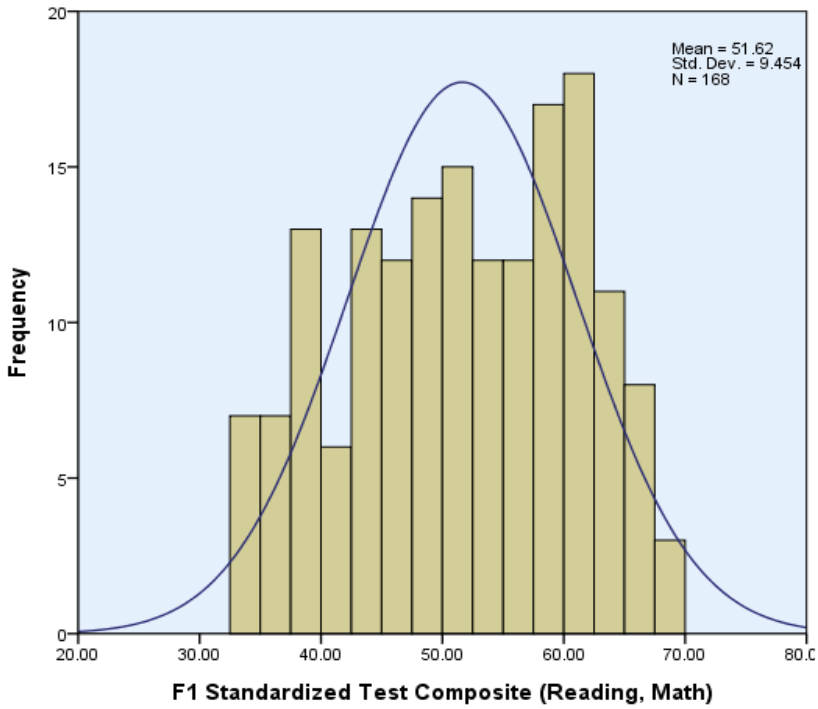


Student achievement in reading and mathematics (f1txcomp):
Student achievement in reading and mathematics scores are positive and mean are quite modest and this indicate that score is sensible as the negative average scores would be insensible. The skewness and kurtosis values indicate that the shape of the distribution is not bell and slightly negatively skewed. The histogram shows that the data has discontinuity in normal shape. It also shows that the standard deviation is large may cause a great variation in data.

Descriptive Statistics

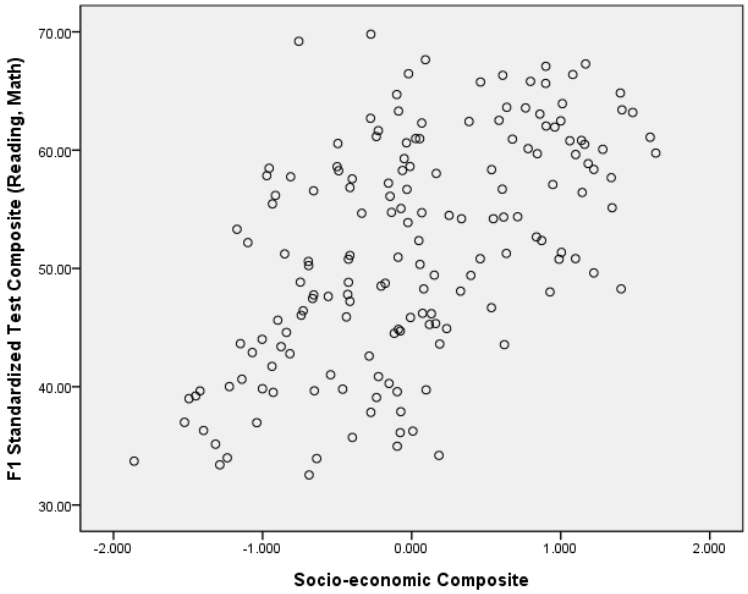
		Statistic	Std. Error
F1 Standardized Test Composite (Reading, Math)	N	168	
	Range	37.26	
	Minimum	32.54	
	Maximum	69.80	
	Mean	51.6236	.72937
	Std. Deviation	9.45366	
	Variance	89.372	
	Skewness	-.170	.187
	Kurtosis	-.996	.373
Valid N (listwise)	N	168	

Histogram



2. Generate a scatter plot representing the relationship between social economic status (ses) and student achievement (f1txcomp). What kind of a relationship do you see? (Consider the following aspects of a relationship: linear vs. non-linear, positive vs. negative, strength of relationship). Be sure the graph is included in your answer.

The scatter plot shown bellow shows an indication that there might be a linear relationship between the variables included here. The majority of the points tend to fall around an imaginary line. The relationship between **ses** and student achievement can be considered strong. It is surely a positive relationship because high scores on the X-axis are associated with high scores on the Y-axis. The coefficient of correlation is 0.559 (table 2) indicates that the strength of the correlation is moderate. The scatter plot also supports this conclusion because the points are neither too close nor too scattered.



Correlations

		Socio-economic Composite	F1 Standardized Test Composite (Reading, Math)
Socio-economic Composite	Pearson Correlation	1	.559**
	Sig. (2-tailed)		.000
	N	166	166
F1 Standardized Test Composite (Reading, Math)	Pearson Correlation	.559**	1
	Sig. (2-tailed)	.000	
	N	166	168

**. Correlation is significant at the 0.01 level (2-tailed).

3. Do you think it is sensible to use regression to estimate the nature of the relationship between the two variables? Explain your answer.

Since the correlation is highly significant (P value < 0.0001) (Table 1), the regression analysis is sensible to estimate the nature of the relationship between the two variables, because one of the assumptions for using a regression is that both variables should have a correlation and linear relationship.

B. Population Model and Statistical Estimation

1. Write down the population model for the regression. Define each term in the model. Identify the population parameters to be estimated, and tell us in words what they represent about the relationship between social economic status (**ses**) and student achievement (**f1txcomp**).

   Consider the population regression model:

$$y = \alpha + \beta x + \varepsilon$$

Where: y is the dependent variable or the outcome variable

x is the independent variable

α is the intercept term

β is the regression coefficient

ε is the error term of the model.

If we consider y = F1 Standardized Test Composite (Reading, Math) (**f1txcomp**) and x = then the parameter α Socio-economic Composite (**SES**) indicates the minimum value of **SES** when there is no achievement of student. The parameter β is the regression coefficient indicating the average change in y for one unit change in x.

2. Use SPSS to obtain estimates of the parameters, and include a table that includes the values of the coefficients, standard errors, t-statistic, and p-value.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	51.708	.611		84.682	.000
	Socio-economic Composite	6.725	.779	.559	8.635	.000

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

3. Provide an interpretation for each parameter estimate. That is, tell us what each parameter estimate means in terms of the relationship between social economic status (ses) and student achievement (f1txcomp), and use the actual values of the estimates in your interpretation.

The estimated value of α is 51.708 indicates that students score in F1 Standardized Test Composite (Reading, Math) when Socio-economic Composite (SES) is zero. The above table indicates that the constant is highly significant.

While the estimated value of β is 6.725 indicates that the average increment in F1 Standardized Test Composite (Reading, Math) (f1txcomp) when there is one unit increment in Socio-economic Composite (SES). The regression coefficient is also significant.

The standard error of the **f1txcomp** coefficient is 0.779. A 95% confidence interval for the regression coefficient for **f1txcomp** is constructed as $(6.725 \pm t * 0.779)$, where t is the appropriate percentile of the t distribution with degrees of freedom equal to 164.

C. Hypothesis Testing

1. Pose the null and alternative hypothesis regarding the slope in the model. Specify these hypotheses in words first and then by using symbols.

Null hypothesis (H_0) : Socio-economic Composite (SES) is insignificant.

Alternative hypothesis (H_1) : Socio-economic Composite (SES) is significant.

Symbolically:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

2. What test statistic is appropriate for testing your null hypothesis?

t-test is the appropriate for testing the null hypothesis. A t-test for testing significance of regression coefficient is used to test the significance of regression coefficients in linear and multiple regression setups.

3. What is the probability of obtaining the slope estimate you've obtained (or one even more extreme) under the null hypothesis?

Here we have to calculate $P(|t| \geq 6.725)$ under null hypothesis, i.e. p-value of estimated coefficient. The SPSS table shows that the p-value is less than 0.0001 based on the t-statistic and the degrees of freedom, and thus we cannot accept the null hypothesis. The P-value is the probability that a t-statistic having 164 degrees of freedom is more extreme than 6.725. Since this is a two-tailed test, "more extreme" means greater than 6.725 or less than -6.725.

4. Make a decision about the null hypothesis (e.g., do you reject or retain the null hypothesis?). What does this tell you about the relationship between social economic status (ses) and student achievement (f1txcomp)?

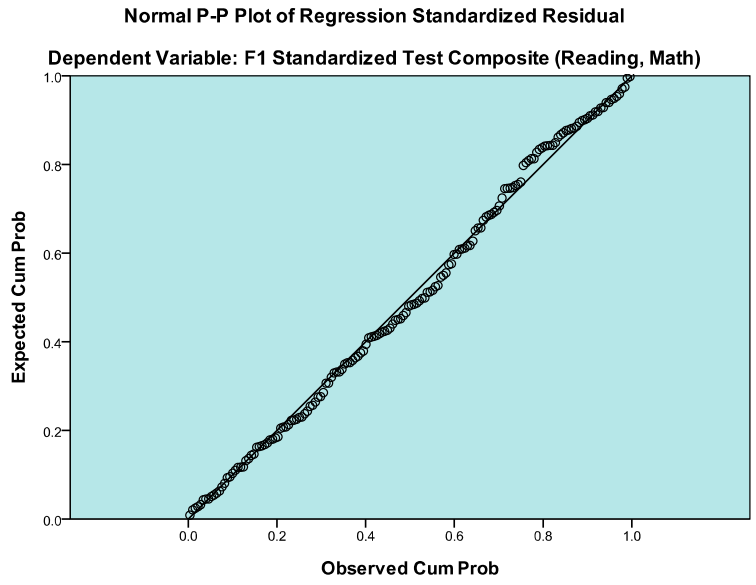
Since the probability of rejecting null hypothesis is less than 0.05 (p-value < 0.0001), we may reject the null hypothesis. So f1txcomp has a significant influence on Socio-economic Composite (SES).

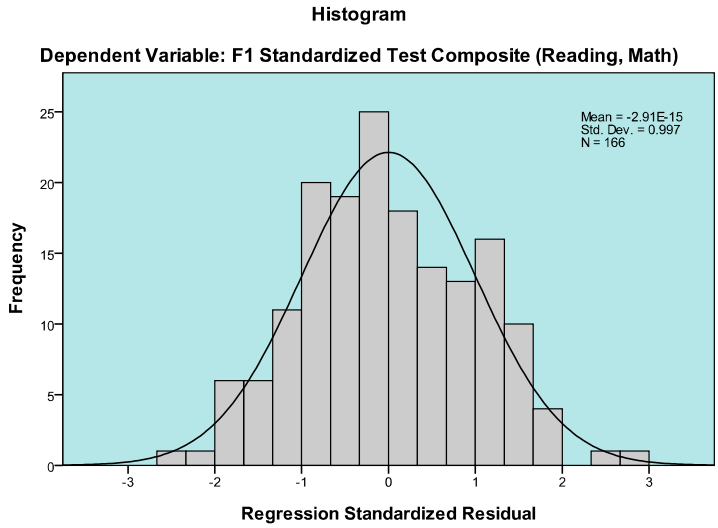
5. Evaluate the assumption of normally distributed errors by visually inspecting a histogram of the residuals. Include relevant SPSS output. What is your assessment of this assumption?

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	39.1863	62.7229	51.6848	5.28848	166
Residual	-18.74839	22.58284	.00000	7.84325	166
Std. Predicted Value	-2.363	2.087	.000	1.000	166
Std. Residual	-2.383	2.871	.000	.997	166

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)





The statistics table and the histogram show that residuals are normally distributed. Our frequency is symmetric about zero, and fits the normal curve reasonably. Another indicator is the P-P plot of regression standardized residuals shows points follow the straight diagonal line, which indicates that assumption of residual normality is not violated.

Part II

A. Multiple Regression: When we interpret our model we also assume that the variables are measured accurately and that the model has been specified correctly. In the case of model specification, there may be other variables which are related to the predictor social economic status (ses) and to student achievement (f1txcomp).

1. Identify at least one other variable in the data set that might have collinearity with social economic status (ses). Explain your reasoning.

We must build a correlation table of SES will all other variables from SPSS. The significant correlations are highlighted in the following table.

Correlations Variables=Socio-economic Composite

dimension0	Socio-economic Composite	Pearson Correlation	1
		Sig. (2-tailed)	
		N	166
	Sex of Respondant	Pearson Correlation	- 0.073
		Sig. (2-tailed)	0.347
		N	166
	Race of Respondant	Pearson Correlation	0.027
		Sig. (2-tailed)	0.730
		N	165
	BY Locus of control 2 measure	Pearson Correlation	0.274**
		Sig. (2-tailed)	0.000
		N	162
	BY Self-concept 2 measure	Pearson Correlation	0.190*
		Sig. (2-tailed)	0.016
		N	162
	BY Hoodlumism 1 Measure	Pearson Correlation	- 0.046
		Sig. (2-tailed)	0.560
		N	161
	F1 Locus of control 2 measure	Pearson Correlation	0.190*
		Sig. (2-tailed)	0.014
		N	165
	F1 Self-concept 2 measure	Pearson Correlation	0.096
		Sig. (2-tailed)	0.218
		N	165

	F1 Hoodlumism 1 Measure	Pearson Correlation	-0.183*
		Sig. (2-tailed)	0.019
		N	162
	F1 Substance Abuse Measure	Pearson Correlation	- 0.125
		Sig. (2-tailed)	0.152
		N	133
	F1 Parent Control 2 Measure	Pearson Correlation	- 0.030
		Sig. (2-tailed)	0.720
		N	149
	F1 Student Morale Composite	Pearson Correlation	0.205**
		Sig. (2-tailed)	0.010
		N	158
	BY Reading Standardized Score	Pearson Correlation	0.448**
		Sig. (2-tailed)	0.000
		N	161
	BY Math Standardized Score	Pearson Correlation	0.579**
		Sig. (2-tailed)	0.000
		N	160
	BY Science Standardized Score	Pearson Correlation	0.309**
		Sig. (2-tailed)	0.000
		N	160
	BY History/Cit/Geog Standardized Score	Pearson Correlation	0.481**
		Sig. (2-tailed)	0.000
		N	160
	BY Standardized Test Composite (Reading, Math)	Pearson Correlation	0.557**
		Sig. (2-tailed)	0.000
		N	161
	F1 Reading Standardized Score	Pearson Correlation	0.436**
		Sig. (2-tailed)	0.000
		N	166
	F1 Math Standardized Score	Pearson Correlation	0.572**
		Sig. (2-tailed)	0.000
		N	165

	F1 Science Standardized Score	Pearson Correlation	0.461**
		Sig. (2-tailed)	0.000
		N	162
	F1 History/Cit/Geog Standardized Score	Pearson Correlation	0.510**
		Sig. (2-tailed)	0.000
		N	161
	F1 Standardized Test Composite (Reading, Math)	Pearson Correlation	0.559**
		Sig. (2-tailed)	0.000
		N	166

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

The highlighted variables are significantly correlated with Socio-economic Composite (SES). One of these variables can be used as an additional predictor of **f1txcomp** which will have collinearity with SES . By considering the value of correlation, we select F1 History/Cit/Geog Standardized Score as additional predictor to model.

Pearson correlation r between SES and BY History/Cit/Geog Standardized Score (**bytxhstd**) 0.481 this means that SES has a low positive correlation with F1 History/Cit/Geog Standardized Score. This indicates that the multicollinearity should be quite low.

2. Write a population model that includes social economic status (ses) and the predictor you listed in question (1) to predict student achievement. Be sure to define each term in the model.

Consider the population regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Where, y is the dependent variable.

x_1 and x_2 are the independent variables

α is the intercept term

β_1 and β_2 are the regression coefficients corresponding to the variables x_1 and x_2

ε is the error term of the model.

If we consider y = F1 Standardized Test Composite (Reading, Math) (f1txcomp) and x_1 = Socio-economic Composite (SES), x_2 = BY History/Cit/Geog Standardized Score (**bytxhstd**) then the parameter α indicates the minimum value of **f1txcomp** when there is no values for SES present. The parameter β_1 is the regression coefficient indicating the average change in y for one unit change in x_1 and β_2 is the regression coefficient indicating the average change

in y for one unit change in x_2.

3. Use SPSS to obtain estimates of the parameters you specified in question (2) and report them in a table that includes the values of the coefficients, standard errors, t-statistics, and p-values.

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF
(Constant)	16.850	3.080		5.471	.000	10.767	22.934		
Socio-economic Composite	2.842	.694	.231	4.094	.000	1.471	4.214	.769	1.301
BY History/Cit/Geog Standardized Score	.669	.059	.646	11.436	.000	.553	.785	.769	1.301

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
(Constant)	16.850	3.080		5.471	.000	10.767	22.934					
Socio-economic Composite	2.842	.694	.231	4.094	.000	1.471	4.214	.542	.311	.203	.769	1.301
BY History/Cit/Geog Standardized Score	.669	.059	.646	11.436	.000	.553	.785	.757	.674	.566	.769	1.301

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

The above table shows the parameter coefficients values for constant term and for independent variables. Here, $\alpha = 16.850$, $\beta_1 = 2.842$ with $\text{se} = 0.694$ and $\beta_2 = 0.669$ with $\text{se} = 0.059$. Both coefficients are significant while p less than 0.001

4. State the null and alternative hypotheses in words and symbols for the "slope" associated with social economic status for the model you defined in question (2).

The null hypothesis (H_0) states that the partial slope for SES is equal to zero, in other words SES is unable to predict Student Achievement.

The alternative hypothesis (H_1) states that the partial slope for SES is not equal to zero, in other words SES has a predictive power in the population.

Symbolically:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

5. Do you reject the null hypothesis? Why?

From the result table, $t(\beta_1) = 4.094$ with a $p < 0.05$, which means that t is statistically significant. We may reject the null hypothesis.

6. Explain conceptually, in terms of overlapping variances, how multiple regression controls for a covariate such as the variable that you chose when estimating the effect of social economic status on student achievement.

Multiple regression controls for a covariate because it finds the correlation between independent and dependent variable and between the independent variables themselves. By using the Pearson coefficient from each of the separate linear relationships and then factoring in the areas that overlap the covariate can be taken into account. Multiple regression, unlike simple linear regressions, looks at the effect of changing one independent variable while the other independent variables stay constant.

It is important to continually check for collinerarity between the independent variables. Examining a scatter plot of correlations can detect collineraity. Another option is to calculate the variation in the dependent variable and the independent variable (to control for a covariate either a partial or semi-partial correlation can be used).

For example, using the data included above we calculate the shared variance and isolate the effect of a particular variable. partial correlation between X and Y given a set of n controlling variables

- Using partial correlation: $r^2_{YX_1.X_2} = (0.331)^2$ this method describe the relationship between two variables while taking away the effects of another variable, For example, SES explains 10.90% of variation in Student Achievement not explained by other explanatory variable.
- Using semi-partial correlation: $r^2_{YX_1.X_2} = (0.203)^2$ while partial correlation computes the correlation between X and Y while holding Z constant for both. For semi-partial, however, Z is held constant for just X or just Y. SES accounts for 4.12% of the variation in student achievement once the overlapping between other explanatory variable and Student Achievement is removed. Semi-partial correlation is lower in quantity than the partial correlation.

7. Has the standard error for the slope associated with social economic status changed from the model you estimate in Part I B.2 to the model you estimated in Part II A.3? Explain any differences.

The equation for se is-

$$\sqrt{\frac{\frac{\sum(y_i - \hat{y}_i)^2}{n-p-1}}{\sum(x_i - \bar{x})^2(1 - R^2_{1.2})}}$$

The value of standard error of the coefficient of SES is changed from 0.779 to 0.694. The SE has decreased slightly as more independent variables are added the numerator in the equation shown above becomes less. Additionally as there is a low correlation between the old

predictor (SES) and the newly introduced one (F1 History/Cit/Geog Standardized Score), SE has no reason to increase.

8. Has the p-value for the test of the slope of social economic status changed from the model you estimate in Part I B.3 to the model you estimated in Part II A.3? Explain any differences.

P value for SES in part I was approximately zero
P value for SES in part II is also approximately zero.
No change and no effect on the statistical significance of the SES partial slopes.

9. Report the VIF for socioeconomic status as well as the variable you chose. Would you consider these VIFs to be 'high'?

The Variance Inflation Factor VIF: 1.301
We know, $VIF = \frac{1}{1-r_{x_1x_2}^2}$,
Because the denominator of the equation is one minus the correlation when the correlation approaches zero the VIF is approaches one. In this case the VIF is 1.301 which is low and suggests multicollinearity is not a factor in this case.

B. Pick an additional independent variable to add to the multiple regression model you specified in Part II, A.3.

1. Using SPSS, compute the partial and semi-partial correlations for each of the three independent variables.

Consider the population regression model:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

Where, y is the dependent variable.

x_1 , x_2 and x_3 are the independent variables

α is the intercept term

β_1 , β_2 and β_3 are the regression coefficients corresponding to the variables x_1 , x_2 and x_3

ε is the error term of the model.

If we consider y = F1 Standardized Test Composite (Reading, Math) (f1txcomp) and x_1 = Socio-economic Composite (SES), x_2 = BY History/Cit/Geog Standardized Score (**bytxhstd**) and x_3 = sex then the parameter α indicates the minimum value of **f1txcomp** when there is no other value is present. The parameter β_1 is the regression coefficient indicating the average change in y for one unit change in x_1 and β_2 is the regression coefficient indicating the average change in y for one unit change in x_2 and β_3 is the regression coefficient indicating the average change in y for one unit change in x_3.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	14.838	3.508		4.230	.000					
	Socio-economic Composite	2.861	.693	.233	4.125	.000	.542	.314	.204	.768	1.302
	BY History/Cit/Geog Standardized Score	.673	.059	.650	11.502	.000	.757	.677	.569	.766	1.306
	Sex of Respondant	1.135	.951	.059	1.194	.234	-.006	.095	.059	.993	1.007

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

Socio-economic Composite (**SES**)

- Partial correlation: $r^2_{YX_1.X_2X_3} = (0.314)^2$ represents the portion of variance in Y that is not associated with any other predictors but with the variance in x_1. i.e, SES explains 9.86% of variation in Student Achievement not explained by other explanatory variables.
- Semi-Partial correlation: $r^2_{Y(X_1.X_2X_3)} = (0.204)^2$ represents the portion of variance in once we have accounted for the other independent variable. i.e, SES accounts for 4.12% of the variation in student achievement once the overlapping between explanatory variable is removed. Semi-partial correlation is lower in quantity than the partial correlation.

BY History/Cit/Geog Standardized Score (**bytxhstd**)

- Partial correlation: $r^2_{YX_2.X_1X_3} = (0.677)^2$ represents the portion of variance in Y that is not associated with any other predictors but with the variance in x_1. i.e, **bytxhstd** explains 45.83% of variation in Student Achievement not explained by other explanatory variables.
- Semi-Partial correlation: $r^2_{Y(X_2.X_1X_3)} = (0.569)^2$ represents the portion of variance in once we have accounted for the other independent variable. i.e, **bytxhstd** accounts for 32.37% of the variation in student achievement once the overlapping between

explanatory variable is removed. Semi-partial correlation is lower in quantity than the partial correlation.

Sex of Respondent

- Partial correlation: $r^2{}_{YX_3.X_2X_1} = (0.095)^2$ represents the portion of variance in Y that is not associated with any other predictors but with the variance in x_1. i.e, sex explains 0.90% of variation in Student Achievement not explained by other explanatory variables.
- Semi-Partial correlation: $r^2{}_{Y(X_3.X_2X_1)} = (0.059)^2$ represents the portion of variance in once we have accounted for the other independent variable. i.e, sex accounts for 0.35% of the variation in student achievement once the overlapping between explanatory variable is removed. Semi-partial correlation is lower in quantity than the partial correlation.

2. Provide an interpretation in words (and using the values reported by SPSS) of the squared partial correlation of each of the independent variables.

The squared correlation $(0.314)^2$=0.0986 or 9.86% represents the portion of variance in Y that is not associated with any other predictors but with the variance in x_1. i.e, SES explains 9.86% of variation in Student Achievement not explained by other explanatory variables.

$$r^2{}_{Y(X_1.X_2X_3)} = \left(\frac{t_{x_1}^2}{Residual\ DF}\right) * (1 - r^2{}_{Y.X_1X_2X_3})$$

Therefore it is also the amount of that R^2 would be reduced if X_i were not included in the regression model.

a. R^2 would be reduced by 0.35% if **sex** is dropped.
b. R^2 would be reduced by 4.12% if **ses** is dropped.
c. R^2 would be reduced by 32.37% if **bytxhstd** is dropped.

3. What does the squared semi-partial correlation suggested about the addition of the third independent variable that you added to the regression model?

It suggests that we are adding no more explaining power to our model by adding a second and third predictor. As we added sex, we see that this explains just 0.35% of the variation in student achievement once we account for the other two predictors.

4. Perform a statistical test of the differences between multiple Rs for this model and the model without the third independent variable. Specify the null and alternative hypotheses, compute the test statistic, and determine whether you reject the null based on an alpha of 0.05.

Model 1: With sex variable

Model Summary

Model		R	R Square	Adjusted R Square	Std. Error of the Estimate
dimension0	1	.786[a]	.618	.611	5.91168

a. Predictors: (Constant), Sex of Respondant, Socio-economic Composite, BY History/Cit/Geog Standardized Score

b. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

Model 2: Without sex variable

Model Summary

Model		R	R Square	Adjusted R Square	Std. Error of the Estimate
dimension0	2	.784[a]	.615	.610	5.91967

a. Predictors: (Constant), BY History/Cit/Geog Standardized Score, Socio-economic Composite

b. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

Model 1: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Adjusted ${R_1}^2 = 0.661$

Model 2: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

Adjusted ${R_1}^2 = 0.610$

Here- $y =$ f1t1comp

$x_1 = SES$

$x_2 =$ bytxhstd

$x_3 = SEX$

Test Statistics: F test

Null hypothesis ($\boldsymbol{H_0}$): adding the extra predictor (sex) gives no explanatory value to the model

Alternative hypothesis($\boldsymbol{H_1}$): addition of the extra predictor (sex) explains the variation caused by existing predictors (SES and bytxhstd)

$$F = \frac{\left[\frac{R_1^2 - R_2^2}{p_1 - p_2}\right]}{\left[\frac{1 - R_1^2}{n - p_1 - 1}\right]}$$

Here, $p_1 = 3$; $p_2 = 2$; n=160

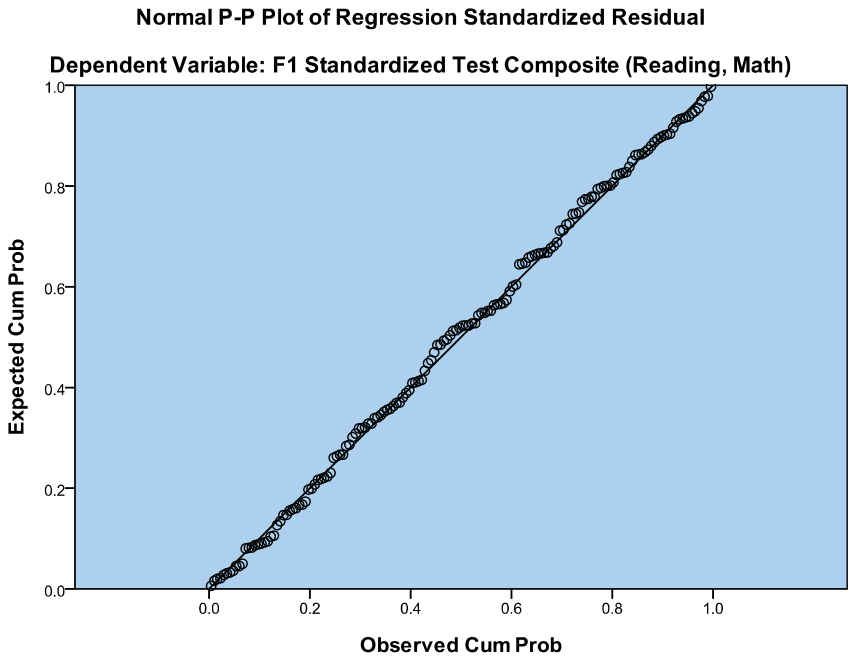
$df_1 = p_1 - p_2 = 1$

$$df_1 = n - p_1 - 1 = 156$$

F = 0.0512 which is less than the calculated value of F1,156,0.05=2.73, indicates that we fail to reject the null hypothesis. This means sex may have explained more variance in the outcome (student achievement).

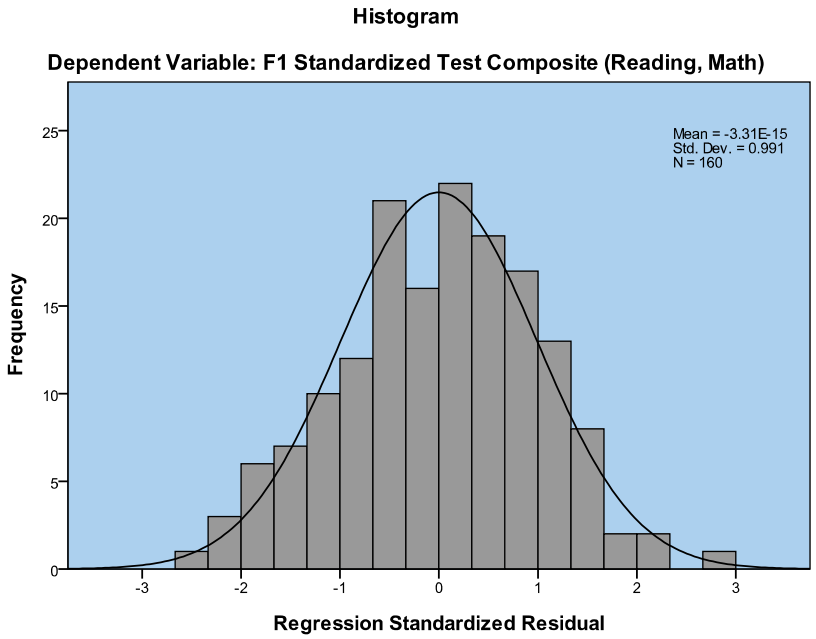
C. Using the model you specified in Part II, B., you will examine the tenability of the assumptions of multiple regression and evaluate the model for misfit.

1. Generate a P-P plot of the standardized residuals. Comment on what this plot suggests about the normality of the residuals.



As the graph shows, the majority of the points stay firm to the diagonal line. This means that the normality of residuals was not violated.

2. Examine a histogram of the standardized residuals. Comment on what this plot suggests about the normality of the residuals.



This histogram of residuals has a standard deviation close to one. This is representative of a normal curve because it is somewhat symmetric about zero and gives evidence that the residuals are normally distributed.

3. Generate a plot of Cook's D and a plot of leverage. What do these plots suggest about the presence of any outliers?

Residuals Statistics[a]

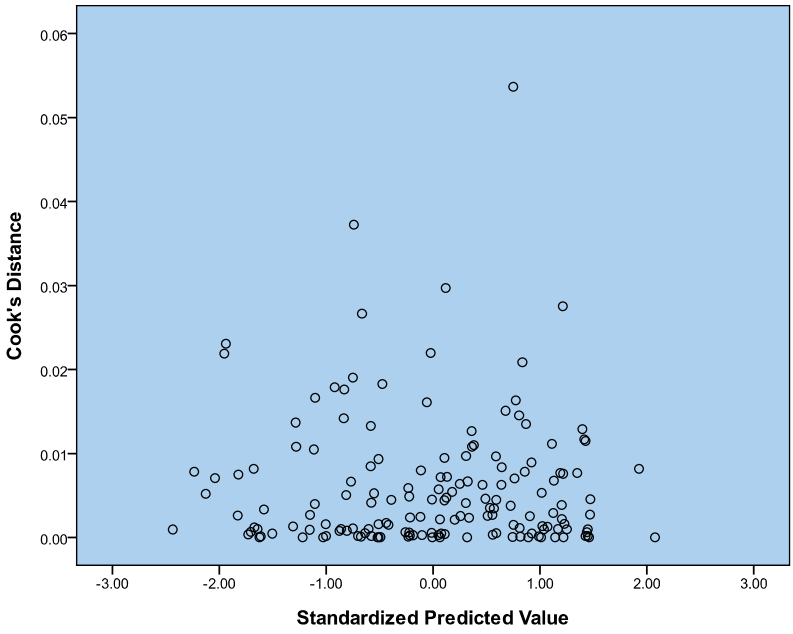
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	33.6358	67.2408	51.7738	7.45213	160
Std. Predicted Value	-2.434	2.076	.000	1.000	160
Standard Error of Predicted Value	.613	1.502	.917	.182	160
Adjusted Predicted Value	33.5522	67.2389	51.7725	7.46170	160
Residual	-14.81624	16.44655	.00000	5.85564	160
Std. Residual	-2.506	2.782	.000	.991	160
Stud. Residual	-2.529	2.808	.000	1.002	160
Deleted Residual	-15.09142	16.75723	.00130	5.99462	160
Stud. Deleted Residual	-2.575	2.873	.000	1.008	160
Mahal. Distance	.718	9.276	2.981	1.627	160
Cook's Distance	.000	.054	.006	.008	160

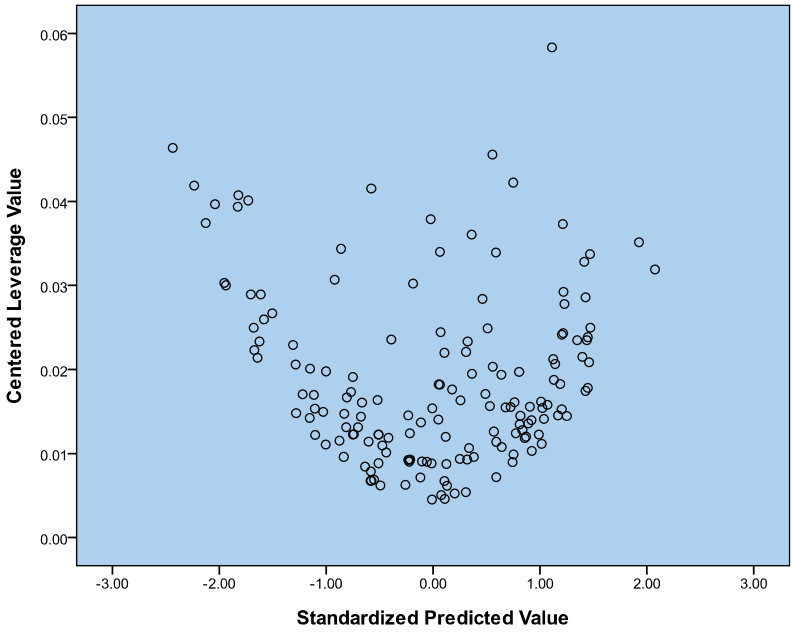
Centered Leverage Value	.005	.058	.019	.010	160

a. Dependent Variable: F1 Standardized Test Composite (Reading, Math)

With this Leverage Plot, if h>8/160=0.05, the point is worth examining. There appears to be only one point above that line.

Using the graph we do not see any outliers corresponding to the scale of the y axis (because no value of cook's D is greater than 1).





4. Given what you found in Part II, C.1 – C.3, what do you think about the usefulness of the regression model?

The regression model is useful with SES and Science Score to predict student achievement. However, calculating the partial correlation has shown that sex has a low predication value. The regression model proved useful because there were no major outliers, residuals appeared to be normally distributed, and collineraity among independent variables was low. There is no need to transform our model or add important independent variables.