Assignment II: Marking Guide

In this assignment you will have the opportunity to explore the relationship between two variables. The questions below draw upon the dataset World (World.sav). You can download the data from Avenue (in the folder "Datasets"). The dataset was taken from: https://edge.sagepub.com/pollock/student-resources/an-spss-companion-to-political-analysis-5th-edition/datasets.

Is there a relationship between a state's overall level of economic freedom, and its level of economic development? The dataset World.sav includes variables measuring states' overall level of economic freedom (Free_Overall - a composite index with scores ranging from 0 (completely unfree) to 100 (completely free) see http://www.heritage.org/index/about if you are interested in a fuller explanation of the operational definition), as well as states' GDP/capita (Gdppcap08). Economic freedom will be treated as our independent variable, and GDP/capita will be treated as our dependent variable.

i). Transform the variable "Gdppcap08" so that it is scaled according to $1000s rather than $1s. Label the new variable Gdppcap1000 (Transform – Compute Variable – under "Target Variable" type Gdppcap1000 (the name of our new variable) – under "Numeric Expression" type Gdppcap08/1000 (our old variable with the scores divided by 1000)). This new variable simply tells us the GDP/capita scores in units of $1000. So, the mean of Gdppcap1000 is 13.59 – i.e. $13 590 – the same as the mean for Gdppcap08). This transformation will make it easier to interpret the regression [2 marks].

The descriptive statistics for the original and transformed variable are as follows:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
GDP/capita WB Gdppcap1000	152 152	188 .19	85868 85.87	13594.30 13.5943	15451.794 15.45179
Valid N (listwise)	152				

ii). Generate a scatterplot (be certain to add the least squares regression line ("Fit Line at Total")). Does it look like there is a relationship between Free_Overall and Gdppcap1000 [2 marks]?
As the slope of the trend of the regression line is upward sloping, there seems to be a relationship between the two variables.
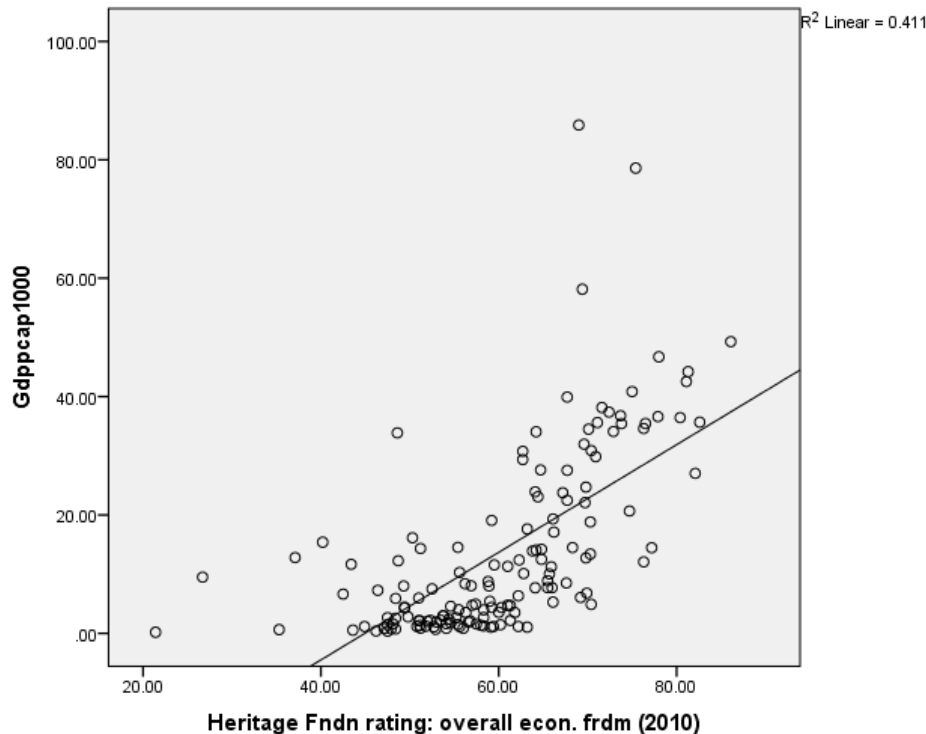
In what direction [2 marks]?

As the slope of the trend of the regression line is upward sloping, there seems to be a positive relationship between the two variables.

Does the relationship look linear [2 marks]?

Yes, as the regression line is a upward sloping and the trend does not seem to be non-linear.

Are there any possible outliers – if so, identify the country(ies) and describe what effect the outlier(s) have on the regression line (use your judgement here – there are tests we can use to identify outliers but that is not required for this exercise) [4 marks]?

R^2 Linear = 0.411

There seems to be three outliers in the scatter plot in the upper right corner for Luxembourg, Qatar and Norway. The relationship between the two variables for these three countries seems to be more positive than the rest of the data affecting the regression line and resulting in more positive slope.

Describe the relationship between the variables using the information presented on the linear regression line (i.e. the information contained in the box "y=…..") – use both the language of statistics, as well as jargon-free English (imagine that you are relaying the results of your study to CBC News) [10 marks].

The relationship between the two variables is positive as the regression line is an upward sloping line which suggests that increase on overall level of economic index results in higher GDP per capita in thousands and decrease on overall level of economic index results in lower GDP per capita in thousands. The value of R-square as displayed on the scatter plot is 0.411 which means that approximately 41% of the variation in GDP per capita can be explained by Free_overall variable. The equation for the model is as follows:

$Y = -40.85 + 0.909*x$

The above equation shows that the value of constant is negative which means that if the value of overall level of economic index is zero, the computed value of GDP per capita in thousands will come to be negative at -40.85 which is redundant here as the GDP cannot be negative. The coefficient for the x variable (i.e. overall level of economic index) is .909 which suggests that there is a positive impact of the overall level of economic index on the GDP per capita in thousands i.e increase on overall level of economic index results in higher GDP per capita in thousands and decrease on overall level of economic index results in lower GDP per capita in thousands. Furthermore, it also tells that on an average, increase in overall level of economic index by 1 score results in increase in GDP per capita by 909 dollars and on an average, decrease in overall level of economic index by 1 score results in decrease in GDP per capita by 909 dollars

iii). Run a linear regression of Free_Overall and Gdppcap1000. Clearly describe and interpret your output [10 marks].

The output from the regression model shows that the value of F-stat is very large and its p-value is close to zero which means that the independent variable included in the model has significant impact on the dependent variable. Furthermore, the value of R-square is equal to 0.411 which means that approximately 41% of the variation in GDP per capita can be explained by overall level of economic index variable. Thus, the goodness of fit of the model is quiet low and there is need to include more significant variables in the model which explains more variation in the GDP per capita.

The regression equation for the model is as follows:

GDP per capita in thousands = -40.85 + 0.909* overall level of economic index

The above equation shows that the value of intercept is negative at -40.85. This means that if the value of overall level of economic index is zero, the computed value of GDP per capita in thousands will come to be negative at -40.85 which is redundant here as the GDP cannot be negative. The coefficient for the overall level of economic index is .909 and since the value of the coefficient is positive, it suggests that there is a positive impact of the overall level of economic index on the GDP per capita in thousands i.e increase on overall level of economic index results in higher GDP per capita in thousands and decrease on overall level of economic index results in lower GDP per capita in thousands. Furthermore, the coefficient of overall level of economic index shows that on an average, increase in overall level of economic index of a country by 1 score results in increase in GDP per capita by 909 dollars and on the other hand, on an average, decrease in overall level of economic index by 1 score for a country results in decrease in GDP per capita by 909 dollars.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.641[a]	.411	.407	11.94066

a. Predictors: (Constant), Heritage Fndn rating: overall econ. frdm (2010)
b. Dependent Variable: Gdppcap1000

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	14716.238	1	14716.238	103.214	.000[b]
	Residual	21101.752	148	142.579		
	Total	35817.989	149			

a. Dependent Variable: Gdppcap1000
b. Predictors: (Constant), Heritage Fndn rating: overall econ. frdm (2010)

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		

1	(Constant)	-40.850	5.461			-7.481	.000
	Heritage Fndn rating: overall econ. frdm (2010)	.909	.089	.641		10.159	.000

a. Dependent Variable: Gdppcap1000

iv). Select two new variables from World.sav (select variables that are interval-ratio, or ordinal (with at least 5 possible scores)) [5 marks for selecting two reasonable variables]. Follow the same procedure as in the first example (i.e. repeat steps i-iii as necessary).
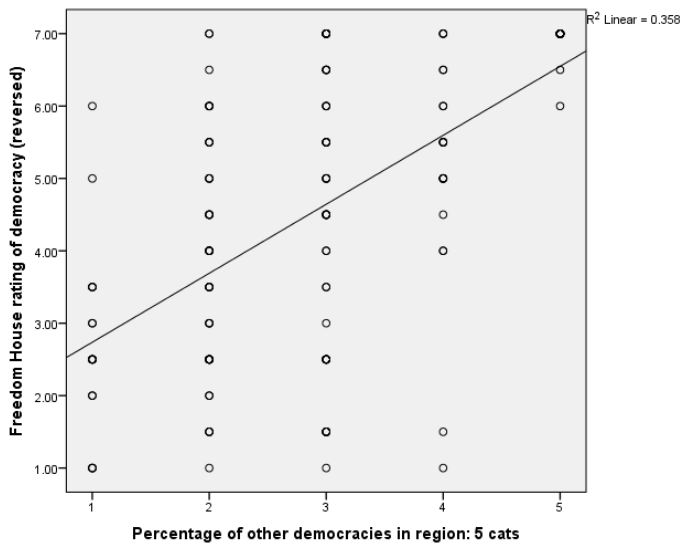
The two variables selected are percentage of other democracies in region and Freedom house rating of democracy (reversed). The two variables have been selected as the two variables are ordinal variables where the value of the former ranges from 1-5 and the latter ranges from 1 to 7. The analysis will use percentage of other democracies in region as the independent variable while Freedom house rating of democracy (reversed) as the dependent variable.

i). Transform the variable [2 marks].

There is no need to transform the two variables as the two variables are easily interpretable with values of percentage of other democracies in region ranges from 1-5 and the values of Freedom house rating of democracy (reversed) ranges from 1 to 7.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Percentage of other democracies in region: 5 cats	153	1	5	2.92	1.222
Freedom House rating of democracy (reversed)	153	1.00	7.00	4.5686	1.94502

ii). Generate a scatterplot (be certain to add the least squares regression line ("Fit Line at Total")). Does it look like there is a relationship between Free_Overall and Gdppcap1000 [2 marks]?

As the slope of the trend of the regression line is upward sloping, there seems to be a relationship between the two variables.

In what direction [2 marks]?
As the slope of the trend of the regression line is upward sloping, there seems to be a positive relationship between the two variables.

Does the relationship look linear [2 marks]?
It seems difficult to comment on the linearity of the relationship.

Are there any possible outliers – if so, identify the country(ies) and describe what effect the outlier(s) have on the regression line (use your judgement here – there are tests we can use to identify outliers but that is not required for this exercise) [4 marks]?
There does not seem to be any outliers in the scatter plot as the range of values for both the dependent and independent variable is quiet limited.

Describe the relationship between the variables using the information presented on the linear regression line (i.e. the information contained in the box "y=…..") – use both the language of statistics, as well as jargon-free English (imagine that you are relaying the results of your study to CBC News) [10 marks].

The relationship between the two variables is positive as the regression line is an upward sloping line which suggests that increase in percentage of other democracies in region results in

higher values of Freedom house rating of democracy (reversed) and increase in percentage of other democracies in region results in higher values of Freedom house rating of democracy (reversed). The value of R-square as displayed on the scatter plot is 0.358 which means that approximately 36% of the variation in Freedom house rating of democracy (reversed) can be explained by percentage of other democracies in region variable. The equation for the model is as follows:

$Y = 1.785 + 0.953*x$

The above equation shows that the value of constant is 1.785 which suggests that if the value of percentage of other democracies in region is zero, the value of Freedom house rating of democracy (reversed) will come to be 1.785. The coefficient for the percentage of other democracies variable is .953 which suggests that there is a positive impact of the percentage of other democracies on the Freedom house rating of democracy (reversed) i.e increase in percentage of other democracies results in higher Freedom house rating of democracy and decrease in percentage of other democracies results in lower Freedom house rating of democracy. Furthermore, it also tells that on an average, increase in percentage of other democracies by 1 scale results in increase in Freedom house rating of democracy scale by 0.953 and on an average, decrease in percentage of other democracies by 1 scale results in decrease in Freedom house rating of democracy scale by 0.953.

iii). Run a linear regression of Free_Overall and Gdppcap1000. Clearly describe and interpret your output [10 marks].

The output from the regression model shows that the value of F-stat is very large and its p-value is close to zero which means that the independent variable included in the model has significant impact on the dependent variable. Furthermore, the value of R-square is equal to 0.358 which means that approximately 36% of the variation in Freedom house rating of democracy scale can be explained by percentage of other democracies. Thus, the goodness of fit of the model is quiet low and there is need to include more significant variables in the model which explains more variation in the Freedom house rating of democracy.

The regression equation for the model is as follows:

Freedom house rating of democracy = 1.785 + 0.953* Percentage of other democracies

The above equation shows that the value of intercept is 1.785. This suggests that if the value of percentage of other democracies in region is zero, the value of Freedom house rating of democracy (reversed) will come to be 1.785. The coefficient for the percentage of other democracies variable is .953 and since the value of the coefficient is positive, it suggests that there is a positive impact of the percentage of other democracies on the Freedom house rating of democracy (reversed) i.e increase in percentage of other democracies results in higher Freedom house rating of democracy and decrease in percentage of other democracies results in lower Freedom house rating of democracy. Furthermore, it also tells that on an average, increase in percentage of other democracies by 1 scale results in increase in Freedom house rating of democracy scale by 0.953 and on an average, decrease in percentage of other democracies by 1 scale results in decrease in Freedom house rating of democracy scale by 0.953.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.599[a]	.358	.354	1.56310

a. Predictors: (Constant), Percentage of other democracies in region: 5 cats
b. Dependent Variable: Freedom House rating of democracy (reversed)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	206.096	1	206.096	84.352	.000[b]
	Residual	368.934	151	2.443		
	Total	575.029	152			

a. Dependent Variable: Freedom House rating of democracy (reversed)
b. Predictors: (Constant), Percentage of other democracies in region: 5 cats

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.785	.328		5.437	.000

Percentage of other democracies in region: 5 cats	.953	.104	.599		9.184	.000

a. Dependent Variable: Freedom House rating of democracy (reversed)