*Multiple Regression Assignment*

*Data posted in the StatCrunch group*

*The following questions use the HAPPINESS data set. The* World Database of Happiness *is an online registry of scientific research on the subjective appreciation of life. It is available at* worlddatabaseofhappiness.eur.nl *and is directed by Dr. Ruut Veenhoven, Erasmus University, Rotterdam. One inventory presents the “average happiness” score for various nations between 2007 and 2008. This average is based on individual responses from numerous general population surveys to a general life satisfaction (well-being) question. Scores ranged between 0 (dissatisfied) to 10 (satisfied). The NationMaster Web site,* www.nationmaster.com, *contains a collection of statistics associated with various nations. For this data set, the factors considered are the GINI Index: measures the degree of inequality in the distribution of income (higher score* = *greater inequality); the degree of corruption in government (higher score* = *less corruption); average life expectancy; and the degree of democracy (higher score* = *more political liberties).*

1. [**Predicting a nation’s “average happiness” score.**](javascript:top.OpenSupp('exercise','11',31)) Consider the five statistics for each nation: LSI, the average life-satisfaction score; GINI, the GINI index; CORRUPT, the degree of corruption in government; LIFE, the average life expectancy; and DEMOCRACY, a measure of civil and political liberties.

1. **(a)** Use boxplots to compare the distribution of each variable.
2. **(b)** Using a correlation matrix describe the relationship between each pair of variables.
3. **(c)** Construct four scatterplots, one for each four explanatory variables with the response variable, y, (LSI).

2. [**Building a multiple linear regression model.**](javascript:top.OpenSupp('exercise','11',32)) Let’s now build a model to predict the life-satisfaction score, LSI.

1. **(a)** Consider a simple linear regression using GINI as the explanatory variable. Run the regression and summarize the results. Check the assumptions – only for this model.
2. **(b)** Now consider a model using GINI and LIFE. Run the multiple regression and summarize the results.
3. **(c)** Now consider a model using GINI, LIFE, and DEMOCRACY. Run the multiple regression and summarize the results.
4. **(d)** Now consider a model using all four explanatory variables. Again summarize the results.

3. [**Selecting from among several models.**](javascript:top.OpenSupp('exercise','11',33)) Refer to the results from the previous exercise.

**(a)** Complete the table below giving the estimated regression coefficients, standard errors, *t* statistics, and *P*-values for each of your models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Estimate** | **Standard errors** | **t-stat** | **p-value** | **r^2 adjusted** |
| **Model1** |  |  |  |  |  |
| Intercept |  |  |  |  |  |
| GINI |  |  |  |  |  |
| **Model2** |  |  |  |  |  |
| Intercept |  |  |  |  |  |
| GINI |  |  |  |  |  |
| Life |  |  |  |  |  |
| **Model 3** |  |  |  |  |  |
| Intercept |  |  |  |  |  |
| GINI |  |  |  |  |  |
| Life |  |  |  |  |  |
| Democracy |  |  |  |  |  |
| **Model 4** |  |  |  |  |  |
| Intercept |  |  |  |  |  |
| GINI |  |  |  |  |  |
| Life |  |  |  |  |  |
| Democracy |  |  |  |  |  |
| Corrupt |  |  |  |  |  |

**(b)** Describe how the coefficients and *P*-values change for the four models.

**4.** Based on the table of coefficients, suggest another model. Run that model, summarize the results, and compare it with the other ones. Which model would you choose to explain LSI? Explain.