

EDF 5401 HW 1
Due 05/ 31/ 10

Please type your answers as much as possible.

Total 100 points

(Each question is worth 7 points, except the question 10 which is worth 9 points.)

In this assignment, use an SPSS data file, *hw_1_data_corporation.sav*. It can be found under the Assignments folder where you have downloaded this homework 1 file. Use *job* (job proficiency score) as outcome or dependent variable and *apt* (aptitude test score) as independent variable. You should employ the following statistical procedures:

Analyze, Descriptive Statistics, Descriptives... which provides descriptive statistics for each variable;

Graphs, Scatter... which displays graphically the relationship between the variables;

Analyze, Correlate, Bivariate... which gives the Pearson correlation coefficient; and

Analyze, Regression, Linear... which is the regression procedure.

1. How would you characterize the relationship between *job* and *apt* based on the scatterplot? Does the correlation you have obtained support your description? Are there any points that seem like outliers?
2. Write the theoretical (population) model for the regression equation. (Be careful: which is the dependent variable?) State what each of the variables and parameters represent.
3. Write the "fitted model." You'll want to run SPSS to get these results.
4. Give the predicted value of Y for a subject with the mean value of X. Show your work. Also, explain why calculation is not needed (Hint: use Least squares solution). Comment on the predicted value.
5. Calculate the predicted value of Y for a subject with $X = 200$. What is the problem with computing and interpreting this value?
6. Add the estimated regression line to the X-Y scatterplot printed in your computer output. Check to see that the slope of your line is close to what the output suggests it should be.
7. Run the regression again and save residuals from the model, as well as the *dfbetas*. Be sure to label them so you don't get confused! Which case has the largest absolute residual value? Which case has the largest absolute **standardized** residual? Which case has the largest impact on the slope (largest absolute *dfbeta* for b_1)? Mark the cases on your plot and comment on why the case(s) seem to have these large values.

8. Next plot the residuals (or standardized residuals) and comment on the plots. (You can accomplish this task using either the saved residuals or the “Plots” button in the regression windows.) Do the residuals appear consistent with the assumptions about residuals required for the regression to be valid?
9. Do any points appear to be outliers in the context of this model? What evidence supports your choice(s)?
10. Remove any point(s) that may be outliers and recompute the regression analysis. Does the fitted model seem to be similar to or different from the model in item 3? Add (by hand) this regression line to the plot you made for item 6. Comment on any differences you see between this model and the one from items 3 and 6.
11. Find in the output from the second model
 - a. the estimated regression slope, b_1
 - b. its estimated standard error, S_{b_1}

Then use the results from (a) and (b) to conduct the test of $H_0: \beta_1 = 0$ at $\alpha = .05$. What is your p value? What is your conclusion about *apt* and *job*? Find the t test on your regression output, square this t and compare it to the F on the regression output. Comment.

12. Find a 95% confidence interval for the population regression slope. What does this interval suggest about the relationship between *apt* and *job*? Compare your decision to those given by the tests in question 11.
13. Write the equation for the standardized regression model. As always, be sure to identify the components of the equation.
14. Finally – does having higher *aptitude test score* help improve *job proficiency score*?