

**EDF 5401 HW 2**  
**Due 6/ 21/ 10**

**Please type your answers as much as possible.**

**Total 100 points**

(Each question is worth 10 points, except question 5 which is worth 20 points.)

In this assignment, use an SPSS data file, *hw\_2\_data\_school.sav*. It can be found under the assignments folder in our course website. Use *f1txcomp* (Standardized test composite (reading & math) score) as outcome or dependent variable and *tchcomm* (Teacher Community (High values=lots o' community)), *bytxcomp* (Base year student achievement), *f1ses* (Socio-Economic status composite), as independent variables.

1. Begin with a model that includes *f1txcomp* as the outcome and has two predictors – *tchcomm* (Teacher Community (High values=lots o' community)), *bytxcomp* (Base year student achievement). First write the population regression model for this analysis. Write this model at the level of the individual student (include subscript *i*.) and as always be sure to identify all parameters and components in the model. Also specify any assumptions for the errors.

2. Obtain descriptive statistics and three separate histograms of the dependent variable (*f1txcomp*) and the predictors *tchcomm* and *bytxcomp*. Describe the distributions. Do they seem normal? Skewed? Unimodal or bimodal?

3. Next obtain scatter plots for the four variables. You may want to make a scatterplot matrix for this item. How would you characterize the relationships of the predictor variables to the outcome in terms of the strength, direction and shape of the relationship?

4. Now consider the correlations among the four variables listed above. What do you conclude about the relationships of the 3 Xs to the outcome? Also comment about the relations among the predictors. Which Xs are interrelated the highest? Is there any evidence of multicollinearity in the *r* value?

5. Use SPSS to obtain estimates of the parameters you specified in item 2.

- a. Write the estimated model.
- b. State the null hypothesis in words and symbols for the slope associated with *tchcomm* in your model.
- c. What is the *p*-value for this slope? What does the *p*-value tell you? Do you reject the hypothesis you stated? Why?
- d. What statistical assumptions do we make regarding the error term in order to interpret the *p*-value?
- e. Have the assumptions for this model been satisfied? Provide evidence where possible to support your answer.

6. Now add *flses* to the equation. Run this model with 3 predictors and write the new estimated model.

7. Reconsider the role of *tchcomm* in this model. Is it a significant predictor? Compare the value of the slope for *tchcomm* in this model to the value from the model above (with only *tchcomm* and *bytxcomp*). Has the slope changed? How about its standard error? What do you think has happened in this model? Also, describe how you check for multicollinearity.

8. Do you think this model with three predictors is properly specified (parsimonious)? Explain your answer. If your answer is “no”, be sure to describe a model you think may be more parsimonious. Compare how much variance is explained by these two models.

9. Select one variable that you feel is most important and discuss the results you have found in item 8. Why is this variable most important (what is your evidence for its importance)? Is the predictor significant? Are any other variables similar in their levels of importance?