

## AECO320 Economic Statistics Project

Using the **college.csv data**, please perform statistical analysis to provide the answers for each questions. Each of you will choose and/or create independent variables based on the college.csv dataset. You can use R studio program, any statistical software, or excel functions to perform the project. Some excel functions are available in the class slides.

We are interested in effects of some variables (of your choices of minimum three variables) on graduation rate. You will interpret the results, discuss significance and direction, and recommend to college admission offices how they can improve graduation rates. Please select at least three variables you are interested based on college.csv to be discussed.

Please describe **in detail** to get the full credits for each questions. Final answers without proper interpretations will lead to lower grades. This project is worth total of 30% of your final grade. Each question is worth some points (1%-10% each or total of 30%) as indicated.

Note: our dependent or LHS variable is graduation rate variable. **You need to choose at least three variables for independent variables**. If you like, you can create a new variable using the existing variables. For example, we have column H and I, number of full time and part time undergraduate students. Using these two, you can create fraction of full time students. E.g.  $\text{Fraction of full time} = \frac{\# \text{ of full time}}{\# \text{ of full time} + \# \text{ of part time}}$ . Please explore the data. More creative ones with intuitions are better. But it is not necessary to create the new variables.

Please attach all results and code in ONE word file or pdf. Attach R or any other program codes if you used any at the end. If you used excel functions, please describe what you did at the end of the project. For figures or regression results, please attach in each corresponding questions. Please TYPE all answers and submit everything in one document (in word or pdf). **The due date is May 7, 2021.**

1. Provide summary statistics and discuss each of variables of your choice in very detail. For example, if mean graduation rate is 0.8, then this means 80% of students graduate. (3%) (Chapter 3 and 5)
2. Provide correlation tables and discuss each of variables of your choice (you need to choose at least three variables and show the associations with the dependent variable in detail). Discuss the signs and magnitudes. Also talk about intuitions. Does the sign make sense? (2%) (Chapter 11)
3. Using scatter plots to show the association between your dependent variable and the choices of your independent variables (choose at least three variables and show the associations with the dependent variable in detail). Please discuss directions, shape, etc to validate your reasoning. Discuss the intuitions. Do the signs, directions, and shapes make sense? (4%) (Chapter 11)

4. Please perform any statistical analysis to find the association between the response variable and independent variable (e.g. this involves regressions). I.e., show regression of graduation rate on each variables of your interest. Choose at least three and interpret the coefficients and significance. (3%) (Chapter 12, 8, 9, and 10)
5. Please interpret each of your estimates of independent variables from regression results in question 4 as detail as possible. (2%) (Chapter 12)
6. What is the R-square? How to interpret it. Can you say based on R-square that your model is good? (1%) (Chapter 11 and 12)
7. What are the significances for each independent variables (Hint: use t-stat). Please discuss the null hypothesis, p value, and whether you reject or fail to reject it. What does the t stat and p-value tell the associations? (5%) (Chapter 12, 8, 9, and 10)
8. What is your overall recommendation using your results if you were a consultant who can advise to a school official? Please discuss as detail as possible. This part alone is worth **10% of your grade. I expect at least one or two pages of draft.** For example, you can say, based on our results, you should do xxx and yy. Tell as an advisor and consultant who have statistics background. You should write to a general population who did not have any statistics background. Please spend some time on this. If necessary, you might cite some published works (not necessary but plus).