Instruction

- Thursday Dec 3rd to Wednesday Dec 9th. This is Final Exam and the due date is Wednesday. Late submissions (even for 1 seconds) receive 0.

- Your submission must include a R-script that is error-less, with your name on top of the script.

- All questions must have numbers in the script. EXAMPLE: “# Questions 1 Part a:”. Note that without the correct R-script I will not grade the submission.

- Please respond to each and every single question on the PAPER as well with numbers and sub-numbers. Otherwise I would consider 0 for your answer if you do not mention what question you are responding to exactly, even if you have the correct answer. For example if you are responding to question 1 part a, in your response you should start with 1.a. ...Again if you are using RStudio I want the response on both the script as well as on paper :).

- Important: if you work with classmates you should submit 1 submission and both your final exam scores will be multiplied by 0.9. If you collaborate and do not inform me your score will be 0. Forget about the group chats :). The data description is attached along with data.

Question 1 : (15/100)

- Question 1 part a (5/100): Please refer to the data description, and define variables ”alcohol”, ”birthweight”, ”smoker”, ”unmarried”, ”drinks”, ”nprevist”, and ”tripre3”; 2 meaning what they are and what is their unit of measurement.

- Question 1 part b (5/100): Provides summary of the variables mentioned above: min, max, mean, and variance of each variable.

- Question 1 part c (5/100): How many observations do we have? What is the unit of observations (is it individuals, households, states,...)? How many dummy variables do we have?

Question 2 : (20/100)

- Question 2 part a (5/100): Run a regression of “birthweight” over “age”.

- Question 2 part b (5/100): What is the model and the predicted equations?

- Question 2 part c (5/100): What is the slope and what is the interpretation of this number?

- Question 2 part d (5/100): Is the slope significant? write down the null and alternative of this hypothesis testing, write down the t-stat/z-score formula and calculate it and provide your conclusion with the 2 approaches, t-stat/z-score and p-value approaches.

Question 3 : (35/100)

- Question 3 part a (5/100): Run a regression of “birthweight” over “age”, “educ”, “unmarried”, “smoker”, “alcohol”.

- Question 3 part b (5/100): What is the model?

- Question 3 part c (5/100): What is the predicted equation?

- Question 3 part d (5/100): Among all the coefficients, which ones are significant and which ones are not. Why?

- Question 3 part e (10/100): In this model, run a ”F” test checking for simultaneous significance of mother’s unhealthy behaviors (smoker and alcohol variables) on the birthweight of the infants.

- Question 3 part e (5/100): In this model (part 3.a) what is the predicted value of an infant's birth weight for a mother who is 27 years old, who has 13 years of education, who is married, who is not a smoker and does not drink alcohol.

Question 4 : (10/100)

- Question 4 part a (5/100): Create log of birthweight and run a single log-linear model of log(“birthweight”) over age.

- Question 4 part b (5/100): What is the predicted slope and what is the interpretation of the slope? (Note it is log-linear so the interpretation is not regular).

Question 5 : (15/100)

- Question 5 part a (5/100): Provide a scatterplot of “birthweight” over “nprevist”.

- Question 5 part b (5/100): Create quadratic and cubic terms for “nprevist”, then run a regression of “birthweight” over “nprevist”, “nprevist2”, “nprevist3”, “unmarried”, and “smoker”.

- Question 5 part c (10/100): Is this a good model or do we have to go with a quadratic or linear relationship between birthweight and “nprevist”? Why?