Barreto
BUS 305: Regression with Microdata
26 February 2025 Name: _____

Exam 1

10 questions; each worth 10 points. 6 minutes per question on average. Watch the time.

Open the Excel workbook called Exam1S25.xlsx and Save As to your folder in the I drive.

Do NOT open any other workbooks or files or look in your book. This is a CLOSED book exam. Do NOT communicate with anyone during the exam. Any violation of these rules is cheating and will result in an immediate F.

In the *FakeData* sheet, I created a new hypothetical variable called *Drive*. It is a measure of innate intelligence, motivation, and social skills. The numbers for this Drive variable, however, are completely different than one we used in class. I used the formula =NORMALRANDOM(70, 2) and then did another little tweak to create this variable. It is a FAKE variable.

1) The average of Drive is in cell R2. Please compute the SD of Drive in cell R4. Did you use the population or sample SD? Please explain why you chose the one you did below and why it doesn't really matter which one you use in this case.

2) Your friend watches you compute the SD and notices that there is no SE function in Excel. They are confused and ask you why there is no SE function in Excel. What do you say?

3) Create a scatter plot of INCWAGE as a function of Drive. Fit a Trendline to the scatter and report the equation and R^2 on the chart. Explain what the R^2 tells us and whether it is good or bad.

4) It is hard to see the fitted line, but it is there, close to the *x* axis. Explain why this is the regression line. In your answer, explain why it is not higher up, with an intercept around 400,000.

5) Use LINEST to regress INCWAGE on EducYears. Repeat for ln INCWAGE on EducYears. How are the slope coefficients of these two models interpreted differently? In your answer, please refer to specific numbers.

6) You are worried about omitted variable bias so you regress INCWAGE on Educ Years and Drive. How much omitted variable bias is there? Show your work and explain what is going on in this case.

7) A) Use Excel's PivotTable to compute the average INCWAGE for the three categories of education (High School, College, and Grad School).

   B) If you had to use this these education categories in a regression, what would you do? You do not have to do it, just explain what you would do to incorporate these three categories in a regression.

Proceed to the *DGP* sheet. I'm interested in the sampling distribution of max streak (consecutive same number) in tossing a die 10 times. I implemented the DGP in Excel TWO different ways. The first way uses an IF statement and RAND() to produce a die roll. The second way uses the SAMPLE function.

8) Run a sim of cell F13 (you decide the number of reps). Interpret the sim results. What do the numbers mean? In your answer, refer to specific numbers

9) Compare the Approximate SE from your sim to the Exact SE in cell A13 of the *DGP* sheet. They are very far apart. Does this mean the sim is broken somehow? Please explain what is going on here.

10) Run a single sim (your choice of reps) to compare the two different ways of throwing a die 10 times. Does the sim provide evidence that they are the same? Explain.

Save your workbook now.

Save your workbook now, close Excel, and turn in your exam. You're all done!