

Individual Assignment #1: Regression Basics

Due: 1/15/2024

You are working for a tutoring company looking to target its potential clients better. The company has partnered with a large, prestigious university in the South. Many first-year students take the introductory economics sequence but stop at the intermediate level after doing poorly in intermediate microeconomics. As part of the partnership, your company will help the university identify students at risk for poor academic performance in their intermediate economics class in exchange for the university subsidizing discounts for these students to sign up for your tutoring services.

The university has provided you with data on past students in the intermediate micro class in the file *econgrades.csv*, including the following information: grades in intermediate micro, intro micro, and intro macro; class size; freshman GPA; standardized test scores; gender; and some information about high school courses. A complete list of variables and their definitions is included at the end of this assignment description.

Your manager has asked you to build a linear regression model with grade in econ 231 as the dependent variable. Your model should use the following explanatory variables:

- grade in econ 100
- grade in econ 101
- class size
- freshman GPA
- math SAT score
- verbal SAT score
- whether the student completed 2 or more years of advanced math in high school
- gender

Hint: Before running your regressions, you will need to create some new variables:

- A dummy variable for male (or female) students
- A single dummy variable capturing whether the student completed 2 *or more* (that is, either 2 or 3) years of advanced math in high school

Use your model to answer the following questions.

1. According to your regression results:
 - a. What is the implied effect of scoring one additional point on the math SAT on Econ 231 grade?
 - b. Is this relationship statistically significant at the 5% confidence level?
 - c. What is the implied effect of scoring 10 additional points on the math SAT on Econ 231 grade? Briefly discuss the economic significance of the effect.

2. The university shared a new dataset on a group of rising second-year students who have enrolled in econ 231 for the upcoming quarter. Briefly explain how you will use your results to identify students at risk for poor academic performance in this class.
3. The economics department is concerned that female students only enroll in econ 231 if they are highly confident that they will perform well and that this is driving a gender gap in students choosing to major in economics. One potential sign of this would be if the female students who enroll in econ 231 perform systematically better than the male students.
 - a. After controlling for all the other explanatory variables in the regression, how much higher or lower is female students' average grade than male students' average grade? Is this difference statistically significant?
 - b. If you reversed the baseline category for the gender variable in your regression (e.g., if Male was your baseline category before, you now make Female the baseline category), by how much would the regression intercept term increase or decrease? (Note: Try answering this question without running an additional regression first. You can confirm with a regression later.)
4. Before partnering with your company, the university ran its own internal analysis to understand the drivers of intermediate microeconomics grades. The analyst segmented the data into eight groups of students by race (black/white/other/unknown) and first-generation status (yes/no). For precisely one of these categories, first-generation college students of unknown race, the coefficient on the verbal SAT score was positive with a p-value of 0.03. Explain briefly to the administrator why the conclusion that better performance on the verbal SAT is predictive of a better grade may not be valid despite the small p-value. How would you improve the analysis?

Variables in the Data

Variable name	Variable description
grade	Grade earned in Economics 231, A = 4, A- = 3.7, etc.
ge100	Grade in Economics 100 (intro macro)
ge101	Grade in Economics 101 (intro micro)
size	Class size
fgap	Freshman GPA
gender	Gender
msat	Score on Math section of the SAT
vsat	Score on Verbal section of the SAT
am1	Dummy variable = 1 if student completed 1 year of advanced math in high school
am2	Dummy variable = 1 if student completed 2 years of advanced math in high school
am3	Dummy variable = 1 if student completed 3 years of advanced math in high school