

# Final Exam; Stata Portion

## EC031-S26

For this part of the exam, you will be asked to answer a series of questions using Stata.

To complete this portion of the exam, you will submit a do-file with you code. **Note: You are free to look at any of your problem sets to help you answer these questions. This exam is open-note, but not open-Google, open-AI, or open-friends. You may pull up old problem set work on your computer, you may use the STATA help menu, and/or you may prepare a “cheat sheet” of notes.**

- Your program should do all of the following tasks. Skip parts that you are unable to do.
- Be sure to answer the few short questions about your results.

For this exam, you will be using data from 157 law schools in the United States. Below is a table of the variables in the data set.

Variable	Description
rank	Law school ranking (1 = highest)
salary	Median starting salary of graduates
cost	Annual tuition and fees (USD)
LSAT	Median LSAT score of entering class
GPA	Median undergraduate GPA of entering class
libvol	Number of volumes in law library
faculty	Number of faculty members
age	Age of the law school (years)
clsize	Size of entering class
north	Located in the North (1 = yes, 0 = no)
south	Located in the South (1 = yes, 0 = no)
east	Located in the East (1 = yes, 0 = no)
west	Located in the West (1 = yes, 0 = no)
studfac	Student-faculty ratio
top10	Ranked in the top 10 (1 = yes, 0 = no)
r11_25	Ranked 11–25 (1 = yes, 0 = no)
r26_40	Ranked 26–40 (1 = yes, 0 = no)

Variable	Description
r41_60	Ranked 41–60 (1 = yes, 0 = no)

First load the dataset into Stata with the following command (Note that this line of code should be all on line, but I have broken it up for readability):

```
import delimited
"https://raw.githubusercontent.com/amichuda/ARE106data/refs/heads/master/
lawsch85.csv"
```

1. What is the mean age of a law school in the dataset?
2. What is the correlation between the age of a law school and its rank? What does that mean?
3. Use the `generate` command to create a new variable, `age_50`, that is equal to 1 if the age of a law school is more than 50 years old.
4. What is the fraction of law schools that are more than 100 years old in the east of the country. Compare that against those in the west of the country.

For questions 5 and 6, run a regression of salary on age, gpa, lsat, cost, and top10

5. Is `cost` statistically significant at the 5% level? Is it economically significant?
6. What is the p-value on `top10`? Why do you think that is?
7. Run a regression to answer the following question: By how many percent does the median starting salary of graduates increase if the law school is in the top 10?
8. For the regression from 7, what is the interpretation of `top10`?
9. What is the  $R^2$  of the regression from 7? What does that mean?
10. Rerun the regression from 8 and 9 adding location variables such that the base location is the South of the country. What is the interpretation on the coefficient of `north`?