DATA SCIENCE FOR ECONOMISTS

ECON 220 LAB

Jafet Baca-Obando

Week 12, Analysis of Variance – 11/14/2025

Outline

What is ANOVA?

- Stands for **Analysis of Variance**
- Used to compare **means(averages) across more than two independent groups**.

- More interesting questions:
  - Example: Do average earnings differ across education levels?

The hypothesis for ANOVA

Null hypothesis (H_0)

- Coincides with our assertion.

- States that the means of **ALL** groups are equal:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

Alternative hypothesis (H_a)

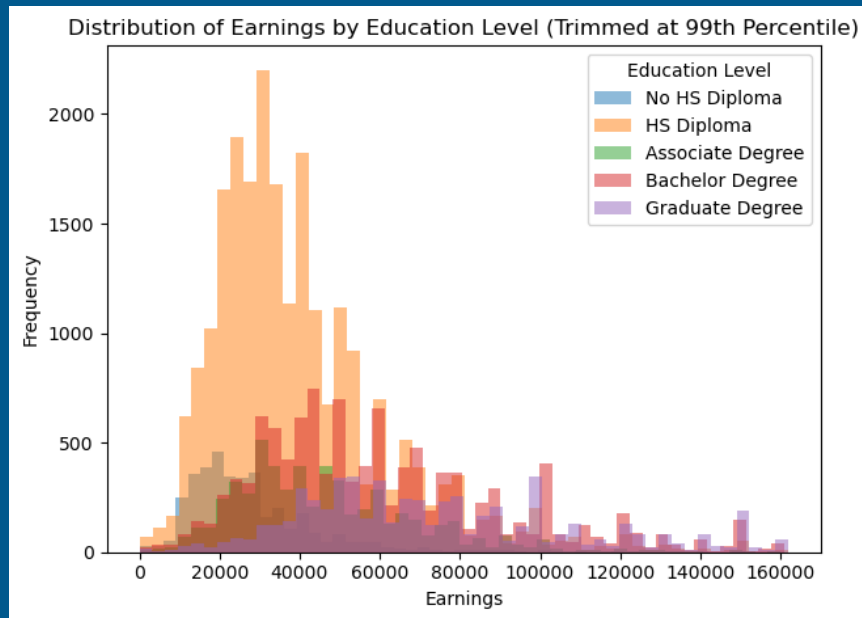- States that AT LEAST one group mean is different from the others:

$$H_a: \text{not all group means are equal}$$

Example: Earnings by education level

Distribution of Earnings by Education Level (Trimmed at 99th Percentile)

F-statistic: 3223.8
p**-value: 0.000**
Conclusion: Reject H_0, there are statistically significant differences in mean earnings across education levels

To-do list

- **Complete Data Exercise 8**

  - Upload Jupyter notebook (.ipynb file) and HTML file on **November 16**

- **Complete Data Exercise 9**

  - Upload Jupyter notebook (.ipynb file) and HTML file on **November 23**