Assignment 1 - MDA 9159

Instructor: Dr. Guowen Huang

Fall 2025

- **This assignment is due Friday, Oct 3, 2025, at 11:55 pm.**
- You must write your answers and R code using Rmarkdown (template provided) and generate a single PDF file. **Submissions not generated by Rmarkdown will not be graded and receive zero marks.**
- Submissions must be done via Gradescope. You must carefully assign questions to their corresponding pages. Submissions without questions assigned to pages will **not** be graded. **Questions with no pages assigned to them will receive zero marks.**
- Always show all your work and add comments to your code explaining what you are doing.
- **Students are encouraged to create their own work and articulate their answers and explanations in their own words. It's essential to understand that scores are not based on the language used, as long as the core concepts and meanings are clearly conveyed. Therefore, using AI tools (like ChatGPT) to generate responses and directly copying them is not permitted. The focus of the assignment is to demonstrate your understanding, not that of the AI. This approach will also serve as valuable preparation for the upcoming closed-book midterm test.**

Dataset: Salary Survey (P130.txt)

Variables: Salary (S), Experience (X, years), Education (E: 1=HS, 2=College, 3=Advanced), Management (M: 0=Other, 1=Manager)

Source: BrightSpace: P130.txt

Q1. Summaries and Plots

(a) Compute the mean and SD of salary.

(b) Compute mean salary by education level.

(c) Draw a scatterplot of salary vs. experience, colored by education, faceted by management. Interpret the plot.

Q2. Simple Linear Regression (SLR)

Fit a regression of salary on experience: $S \sim X$.

(a) Report the fitted regression line.

(b) Interpret the slope.

(c) What percent of salary variation is explained by experience?

Q3. Multiple Regression (MLR) with a Categorical Predictor

Fit the model $S \sim X + \text{Education}$ (treat Education as a factor).

(a) Write the regression equation using dummy variables (reference = **High School**).

(b) Interpret the coefficient for **College**.

(c) Conduct an overall F-test for whether **Education** matters (i.e., all education-level effects = 0).

Q4. Confidence & Prediction Intervals

At **10 years of experience**, predict mean salary and compute:

(a) a 95% **confidence interval** for the mean salary,

(b) a 95% **prediction interval** for a single individual.

Q5. Conceptual

(a) Why do we drop one indicator when coding education with dummies?

(b) Explain (in words) why a prediction interval is wider than a confidence interval.