# DS201/DSL253: Statistical Programming

# Assignment 09

24.04.2025

**Instructions for Submission:** You can submit your solution as a Jupyter Notebook/Matlab file with comments and discussions on the results obtained in each step.

1. Follow Standard Report Format: Include sections like Introduction, Data, Methodology, Results, Discussion, and Conclusion.
2. File Naming Convention: Adhere to the specified naming convention for each file you submit (e.g., RollNumber FirstName Asg1).
3. Refrain from using zip files. If necessary, submit multiple files.
4. Include comments in the code explaining the logic and any assumptions made.
5. Include References: Cite any external sources or references used in your assignment.
6. Code Quality: Ensure your code follows best practices, is well-organized, and avoid plagiarism as a plagiarism check will be conducted.
7. Be aware that late submissions are not permitted; ensure timely submission.
8. Coding can be done in any language.

1. A data analyst is investigating how different car features influence fuel efficiency measured in miles per gallon (MPG). The dataset is given below :

| Vehicle | Engine Size (L) | Weight (kg) | Horsepower | MPG |
|---|---|---|---|---|
| 1 | 1.6 | 1200 | 110 | 34 |
| 2 | 2.0 | 1300 | 130 | 30 |
| 3 | 2.4 | 1500 | 150 | 27 |
| 4 | 1.8 | 1250 | 115 | 32 |
| 5 | 2.2 | 1400 | 140 | 28 |
| 6 | 3.0 | 1600 | 180 | 22 |
| 7 | 2.0 | 1350 | 135 | 29 |
| 8 | 1.5 | 1100 | 105 | 36 |
| 9 | 2.5 | 1550 | 160 | 25 |
| 10 | 3.2 | 1650 | 190 | 20 |
| 11 | 1.4 | 1050 | 100 | 38 |
| 12 | 2.1 | 1380 | 138 | 28 |
| 13 | 3.5 | 1700 | 200 | 18 |
| 14 | 1.6 | 1150 | 108 | 35 |
| 15 | 2.3 | 1450 | 145 | 26 |
| 16 | 2.8 | 1580 | 170 | 23 |
| 17 | 2.6 | 1520 | 155 | 24 |
| 18 | 1.3 | 1020 | 98 | 39 |
| 19 | 3.1 | 1620 | 185 | 21 |
| 20 | 1.7 | 1180 | 112 | 33 |

Using the data provided:

(a) Fit a multiple linear regression model to predict MPG using:

- Engine Size
- Weight
- Horsepower

(b) Write the corresponding regression equation

(c) Report the following from the regression output:

- Coefficients and intercept
- p-values for each predictor
- R-squared value
- Also plot residual values
- Conduct a hypothesis test for each predictor to determine whether it has a statistically significant effect on MPG.

(d) Show the results for different levels of significance.

(e) Identify which predictors are statistically significant and interpret the regression results.

2. A study was conducted to examine how the height of a child is influenced by the heights of their parents. Data were collected from 10 families, and the heights (in inches) of the father, mother, and son were recorded.

   The data are presented in the table below:

| **Father's Height (in)** | 60 | 62 | 64 | 65 | 66 | 67 | 68 | 70 | 72 | 74 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Mother's Height (in)** | 61 | 63 | 63 | 64 | 65 | 66 | 66 | 67 | 68 | 69 |
| **Son's Height (in)** | 63.6 | 65.2 | 66.0 | 65.5 | 66.9 | 67.1 | 67.4 | 68.3 | 70.1 | 70.0 |

(a) Fit a multiple linear regression model to predict the son's height using the heights of the father and mother.

(b) Interpret the regression coefficients.

(c) Using multiple linear regression, determine whether the data supports the idea that children of unusually short or tall parents tend to be closer to the average height—that is, test for regression toward the mean by examining if the regression coefficients for father's and mother's heights are each significantly less than 1.

(d) Also plot residual values

(e) Comment on the implications of your results.