

---

# Homework 5

## Multiple Linear Regression

---

### Purpose

Build and evaluate multiple regression models, interpret coefficients while holding other predictors constant, check model assumptions, and avoid treating automated model output as a conclusion.

### Part A: Tree Volume

Use `Trees.csv`. Model tree volume as a function of girth and height.

1. Inspect variable names, data types, missing values, and numerical summaries. Create plots that show the response against each predictor.
2. Fit the multiple regression model and write the fitted equation.
3. Interpret the coefficient of girth and the coefficient of height in context.
4. Report the overall F test, individual coefficient tests, adjusted  $R^2$ , and residual standard error. Explain what each contributes to the evaluation.
5. Check linearity, constant variance, residual normality, multicollinearity, and influential observations.
6. Use the model to make one prediction within the observed range. Report both a confidence interval for the mean response and a prediction interval for a new tree. Explain the difference.

### Part B: Ames Housing Prices

Use `ames.csv`. The response is `SalePrice`.

1. Select a manageable set of candidate predictors using subject-matter reasoning and data quality. State which variables you selected and why. Do not put every column into a model without inspection.
2. Prepare the variables for analysis. Document how you handled missing values and categorical predictors.
3. Fit an initial multiple regression model and assess its assumptions and influential observations.
4. Examine variance inflation factors. Identify any multicollinearity concern and explain how you addressed it.
5. Develop a final model that is useful and interpretable. Compare it with the initial model using appropriate evidence.
6. Interpret at least three important coefficients in context.

7. Evaluate predictive performance using a train/test split or cross-validation. Report the procedure and metric.
8. Write a concise conclusion explaining what the model does well, where it is limited, and how it could support a housing or hospitality-related decision.

## Submission Checklist

Submit through WebCampus (Canvas) by the deadline posted there. Submit both of the following files:

- a rendered PDF report; and
- the complete `.Rmd` source file used to produce the report.

The PDF must include your name and assignment title, answer every question in the order presented, show readable and labeled tables and figures, and explain results in complete sentences. The `.Rmd` file must contain all code needed to reproduce the report. Both files must run from the same data supplied for the assignment. Do not submit unexplained software output. You are responsible for checking all results and explaining every choice you made.

## A Note About AI

AI can be used as a resource, but it cannot do the assignment for you. I do not want to read AI slop. Do not paste pages of generated writing, code, or interpretations into your report and expect me to sort through it. In the real world, you cannot send AI garbage to a boss, client, or customer. You should not submit it in this class either.

If you use AI, use it to help you learn. Check what it gives you. Make sure the code works. Internalize the material, interpret it, and then explain it in your own words and voice. You should understand every part of your submission and be able to defend the choices you made. Include a short note at the end of the report that names the tool and tells me how you used it. If the work reads like unreviewed AI output, it will receive a 0%, and you will have the opportunity to redo it under the policy in the syllabus.

## How Your Work Will Be Evaluated

Credit is based on the statistical reasoning, correct use of R, appropriate checks of assumptions, quality of visual and numerical evidence, and accuracy of the interpretation. A correct number without an explanation is not a complete answer.