
Homework for Lecture 16: Line of Best Fit

The following homework assignment should be completed as an RMarkdown document. Only show your working code in the code chunks, and place comments/explanations in the “white space” to help me evaluate your work. Please submit both the knitted RMarkdown document (.docx) and the .Rmd file. You may discuss the assignment with your classmates, but your write-up and work should be done independently of your classmates. The use of AI is considered a violation of academic policy for this assignment. When in doubt, ask me what is and is not allowed.

Answer the following questions. When you run code, make sure you explain what you are doing and what the output is telling us. This should be done in the written ‘white’ section of the RMarkdown document. The ‘grey’ section is where your code will go, and when you knit the document, the output will be present.

1. Conceptual Questions

Answer the following questions in complete sentences.

- (a) What is the purpose of a line of best fit?
- (b) What is the difference between an independent and a dependent variable?
- (c) What does the slope of a line of best fit tell us?
- (d) What does the intercept of a line of best fit tell us?

2. Getting to Know the Dataset

Use the `trees` dataset to answer the following questions.

- (a) Display the first few rows of the dataset.
- (b) Use `str()` to inspect the structure of the dataset.
- (c) What variables are included in the dataset?
- (d) In this homework, we are using `Girth` to predict `Volume`. Which variable is the independent variable, and which variable is the dependent variable?

3. Visualizing the Relationship

Create a scatterplot with `Girth` on the x-axis and `Volume` on the y-axis.

- (a) Add an appropriate title and axis labels.
- (b) Does the relationship appear to be positive, negative, or roughly flat?
- (c) Does the relationship appear approximately linear?
- (d) In 2–3 sentences, explain what the scatterplot tells us.

**** MORE PROBLEMS ON PAGE 2 ****

4. Creating the Linear Model

Use the `lm()` function to create a linear model predicting **Volume** from **Girth**.

- (a) Save the model as `tree_model`.
- (b) Display the model output.
- (c) Write the equation of the line of best fit using the slope and intercept from R.

5. Interpreting the Model

Use your model from Question 4 to answer the following questions.

- (a) Interpret the slope in context.
- (b) Interpret the intercept in context.
- (c) Does the intercept have a meaningful real-world interpretation in this problem? Explain why or why not.

6. Adding the Line of Best Fit

Recreate your scatterplot from Question 3 and add the line of best fit using `abline()`.

- (a) Make the line a different color from the points.
- (b) Does the line seem to fit the data reasonably well?
- (c) Are there any points that appear far away from the line?
- (d) In 2–3 sentences, explain what the line helps us see.

7. Making Predictions

Use your line of best fit to make predictions.

- (a) Predict the volume of a tree with a girth of 10 inches.
- (b) Predict the volume of a tree with a girth of 15 inches.
- (c) Predict the volume of a tree with a girth of 20 inches.
- (d) Show your calculations or R code.
- (e) Explain how the predicted volume changes as girth increases.

8. Thinking About Extrapolation

Use R to find the minimum and maximum values of **Girth** in the **trees** dataset.

- (a) What is the smallest girth in the dataset?
- (b) What is the largest girth in the dataset?
- (c) Would it be reasonable to use this model to predict the volume of a tree with a girth of 35 inches? Explain why or why not.