
Homework 3

Correlation and Simple Linear Regression

Purpose

Use correlation and simple linear regression to describe relationships, assess model fit, make predictions, and explain the limits of the evidence. Complete the full analysis for both datasets.

Dataset A: E-commerce Sales and Advertising Cost

Use `eCommerce Sales vs Ad Costs.csv`. Treat advertising cost as the predictor and e-commerce sales as the response.

Dataset B: Automobile Price and Age

Use `Auto Price vs Age.csv`. Treat age as the predictor and price as the response.

Tasks for Each Dataset

1. Create a scatterplot. Describe the direction, form, strength, and any unusual observations before fitting a model.
2. Calculate Pearson's correlation coefficient. Interpret its sign and magnitude in context.
3. Fit a simple linear regression model and write the fitted equation using the variable names and estimated coefficients.
4. Interpret the slope and intercept. If the intercept has no useful meaning in context, say so.
5. Report and interpret R^2 . Verify numerically that $R^2 = r^2$ for this simple linear regression with an intercept.
6. Add the fitted regression line to the scatterplot. Also create a version showing the confidence band for the mean response.
7. Examine residuals versus fitted values, a normal Q-Q plot, and observations with unusual influence. Discuss whether linearity, constant variance, normality, and independence appear reasonable.
8. Give one example of a prediction that would be reasonable with this model and one that would be an unsafe extrapolation.
9. Write a short conclusion for a manager. Distinguish association from causation.

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.