

Topics:

- Describing Scatterplots and Bivariate Regression Analysis
- Lessons Covered: 39 - 45
- Textbook Chapter (Optional): 12 and 13.1

Grading:

- Points are listed next to each question and should total 25 points.
- Late assignments will not be graded.

Deadlines:

- Final Submission: Monday, March 4th

Instructions:

- Download or view questions below.
- Clearly label and type answers, *without* question prompts, in word, google docs, PDF, or other word processing software.
- Insert diagrams or plots as a picture in an appropriate location.
- Math Formulas need to be typed with Math Type, LaTeX, or clearly using key board symbols such as +, -, *, /, sqrt() and ^
- Submit assignment Verify the correct document has been uploaded. If not, resubmit. You can submit up to three times.

Allowances:

- Any resources listed or posted in our class.
- You are encouraged to discuss the problems with other students, the instructor and TAs, however, all work must be your own words. Duplicate wording will be considered plagiarism.
- Outside Resources need to be cited. Websites such as Chegg, Coursehero, Koofers are discouraged, but if used need to be cited and within the boundaries of academic honesty.

The dataset *ExamDataW19.csv* represents the midterm and final exam grades for students in the ST314 online and campus courses for the last three consecutive terms. Use this data to complete a regression analysis in R and answer the following questions.

Part 1. (5 points) Describing the relationship between your two variables

- a. **(3 points)** Graphically: Make a scatterplot of the relationship between ST314 student midterm grades and final exam grades. Paste your plot. Describe in context the relationship from the scatterplot. Include strength, direction, form and outliers (if any).
- b. **(2 points)** Numerically: Calculate the correlation coefficient r . Describe in context the strength of the relationship based on your value.

Part 2. (6 points) Calculate the Least Square Regression Line (Model) and Check Conditions for Inference

- a. **(2 points)** Using R, calculate the least squares regression line that predicts final exam scores from midterm exam scores for ST314 students. Paste the R output for the model summary. State the least squares regression line (model).
- b. **(4 points)** Plot the residuals for the model. Include a reference line at 0. Paste your plot. Check the linearity, normality and constant variation conditions using the residual plot. State why each condition is met or why it is not met.

Part 3. (7 points) Is your model a good fit? Use your R output from the model in Part 2a. From the output, is there statistical evidence midterm exam score is a significant predictor of final exam score? Use a significance level of 0.05.

- a. **(1 point)** State the null and alternative hypothesis for the individual t test on the slope.
- b. **(1 point)** State the test statistic, degrees of freedom and p-value from the output.
- c. **(2 points)** Make a conclusion. Include context, a statement in terms of the alternative and whether to reject the null based on the level of significance.
- d. **(3 points)** Calculate the 95% confidence interval for the slope. Interpret the point and interval estimate for β_1 .

Part 4. (7 point) Prediction: A common goal in regression is to use the estimated model for prediction.

- a. **(2 point)** Using your model from part 2a, provide a final exam score prediction for students that earned a midterm exam score of 200. Show work by hand! You may validate with code provided in R.
- b. **(1 point)** The following intervals are the 95% confidence and prediction intervals for a midterm value of 200. Which is the confidence interval? Which is the prediction interval? How do you know?

(153.4, 246.5)

(198.3, 201.5)

- c. **(2 point)** Write an appropriate interpretation of both the intervals. Include context.
- d. **(1 point)** Based on your midterm exam score, calculate your predicted final exam score based on the model.
- e. **(1 point)** Do you think the predicted value is a reasonable score to assume for your final exam grade? Why or why not? Consider the data and the strength of the model. Consider other factors that might influence your grade.