

Regression Inference (sec. 2.1-2.5).

Regression slope and intercept. Estimation, testing, and prediction

1. (2.10) For each of the following questions, explain whether a confidence interval for a mean response or a prediction interval for a new observation is appropriate.
 - (a) What will be the humidity level in this greenhouse tomorrow when we set the temperature level at 31°C ?
 - (b) How much do families whose disposable income is \$23,500 spend, on the average, for meals away from home?
 - (c) How many kilowatt-hours of electricity will be consumed next month by commercial and industrial users in the Twin Cities service area, given that the index of business activity for the area remains at its present level?
2. (2.11) A person asks if there is a difference between the “mean response at $X = x$ ” and the “mean of m new observations at $X = x$ ”. Reply.
3. (Continued from HW-2) The time it takes to transmit a file always depends on the file size. Suppose you transmitted 30 files, with the average size of 126 Kbytes and the standard deviation of 35 Kbytes. The average transmittance time was 0.04 seconds with the standard deviation of 0.01 seconds. The correlation coefficient between the time and the size was 0.86. In the previous homework, we fit a regression model that predicted the time it will take to transmit a 400 Kbyte file. According to this model, the standard deviation of responses is estimated by $s = s_y \sqrt{\frac{n-1}{n-2}(1 - r^2)} = 0.0052$.
 - (a) Construct a 95% confidence interval for the regression slope.
 - (b) Based on this interval, is the slope significant at the 5% level?
 - (c) State the null and alternative hypotheses in (b). Calculate the test statistic and the p-value.
 - (d) When you answered questions (b) and (c), it was correct to conduct a two-sided a two-sided test. However, in this given example, why does it make more sense to consider a one-sided, right-tail alternative?
4. (Continued from HW-2) At a gas station, 180 drivers were asked to record the mileage of their cars and the number of miles per gallon. The results are summarized in the table.

	Sample mean	Standard deviation
Mileage	24,598	14,634
Miles per gallon	23.8	3.4

The sample correlation coefficient is $r = -0.17$. In the previous homework, we fit a regression model that described how the number of miles per gallon depends on the mileage. According to this model, the standard deviation of responses is estimated by $s = s_y \sqrt{\frac{n-1}{n-2}(1 - r^2)} = 3.36$.

- (a) You purchase a used car with 35,000 miles on it. Predict the number of miles per gallon. Give a 95% prediction interval for your car and a 95% confidence interval for the average number of miles per gallon of all cars with such a mileage.
- (b) Do the given data present a significant evidence that cars with higher mileage are less economic? Formulate appropriate null hypothesis and alternative and conduct the test.

5. **(2.52+, Stat-615 only)** In this problem, we develop statistical methods for the regression *intercept*. You can review our inference about the regression slope and follow similar steps.

- (a) Derive the expression for the variance $\text{Var}(b_0)$.
- (b) Derive the t-statistic for testing $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$. Suppose that the standard regression assumptions are satisfied.
- (c) Test significance of the intercept for the Miles Per Gallon data in the previous exercise.
- (d) Write the general expression for a $(1 - \alpha)100\%$ confidence interval for the population intercept β_0 .
- (e) Construct a 95% confidence interval for the intercept for the Miles Per Gallon data in the previous exercise.

Hint: In (a), notice again that $b_0 = \sum a_i Y_i$ is a linear combination of responses Y_1, \dots, Y_n . In (b,d), use our general methods for t-tests and confidence intervals, applying them to the regression intercept β_0 and its estimator b_0 .

6. Computer project **(2.4, 2.13ab, 2.67)**.

Grade point average (this data set was already used in Homework-2).

- (a) Obtain a 99% confidence interval for β_1 . Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?
- (b) Test whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of 0.01. State the alternatives, decision rule, and conclusion.
- (c) What is the P-value of your test in part (b)? How does it support the conclusion reached in part (b)?
- (d) Obtain a 95 percent confidence interval for the mean freshman GPA for students whose ACT test score is 28. Interpret your confidence interval.
- (e) Mary Jones obtained a score of 28 on the ACT. Predict her freshman GPA using a 95 percent prediction interval. Interpret your prediction interval.
- (f) On the same graph, plot
 - the data
 - the least squares regression line for ACT scores
 - the 95 percent confidence band for the true regression line for ACT scores between 20 and 30.

Does the confidence band suggest that the true regression relation has been precisely estimated? Discuss.