

2024 AP DAILY: PRACTICE SESSIONS

AP Statistics Session 8 – FRQ (Part B, Question 6: Investigative Task)

The instructor of a training course for a certain coding language hopes that a student's score on a pretest will predict the student's score on the certification final taken at the end of the course. The final scores indicate proficiency with the coding language. A random sample of students who completed the pretest and the training course was selected, and the final scores of the students were recorded. The following tables show summary statistics and computer output for a regression analysis of the sample.

	Mean	Standard Deviation
Pretest	54.833	10.408
Final	78.867	6.946

Predictor	Coef	SE	T	P
Constant	49.2804	6.9199	7.12	0.0000
Pretest	0.5396	0.1242	4.35	0.0015
$S = 4.2682$		$R\text{-sq} = 65.38\%$		$R\text{-sq}(\text{adj}) = 61.92\%$

- Identify the equation of the least-squares regression line for predicting final score from pretest score. Identify the variables that appear in the equation in context.
- Consider the information presented in the tables and your answer from part (a).
 - Determine the value of the predicted final score for students with a pretest score of 65.
 - Interpret $S = 4.2682$ in the context of the problem.

For any particular pretest score, two point estimates can be calculated from the given simple linear regression analysis.

Point Estimate I: An estimate of the mean final score among all students with a particular pretest score who complete the course.

Point Estimate II: A prediction for the final score of one student with a particular pretest score who completes the course.

The values of both point estimates are equal to the predicted value of the least-squares regression equation for the particular pretest score, such as the estimate from part (b-i) for a pretest score of 65. Although the values of the two point estimates are the same, the sampling variability is not the same.

- c. Which of the two point estimates has the larger sampling variability? Explain your reasoning.
- d. The value calculated in part (b-i) is an estimate of the mean final score for all students with a pretest score of 65 who completed the course. If a list of the pretest scores and final scores for all students who completed the course is obtained, explain how this information could be used to investigate the sampling distribution of the estimated mean final scores for all students with a pretest score of 65.