9.2-9.3 Does Seat Location Matter? Part 2

Do students who sit in the front rows do better than students who sit farther away? A teacher took a random sample of 30 students from her classes and found these results.

Row	Front of Class					
1	76	77	94	99	88	90
2	83	85	74	79	77	79
3	90	88	68	78	83	79
4	94	72	101	70	63	76
5	76	65	67	96	79	96

Below is the computer output of the results.

Predictor	Coef	SE Coef	T	P
Constant	85.95	4.418	19.454	<0.001
Row	-1.517	1.332	-1.139	0.265

$s = 10.318$

1. What is the equation of the Least Squares Regression Line (LSRL)?

2. What is the value of the slope? Circle it in the computer output and write it below.

3. What is the standard error of the slope? Put a box around it in the computer output.
    a. Interpret the standard error of the slope in context.

4. If another teacher were to rerun this experiment with 30 similar students, do you think the slope of their LSRL would be the same? Similar? Why?

9.2-9.3 Confidence Interval for Slope

We are going to construct a 95% confidence interval for the slope of the population regression line.

Choose

State the name of the inference procedure:

Identify the parameter and statistic:

Parameter: ____________________

Statistic: ____________________

Calculate

General Formula:

Specific Formula:

Work:

Conclude

Conditions

Linear: *The **scatterplot** needs to show a linear relationship.*

Sketch it below.

Independent: *You know this one!*

Normal: *A **dotplot** of the residuals cannot show strong skew or outliers.*

Make one using technology and sketch it below.

Equal SD: *The **residual plot** is homoskedastic.*

Sketch it below.

Random: *You know this one too!*

THE INVESTIGATIVE TASK

Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

An engineer is developing a polymer material and is concerned that the mean density, d, of the material is not sufficiently close to the desired target value of 1.37 kilograms per milliliter (kg/mL). Nine different samples of the material were prepared. The volumes, in milliliters, for the samples were 10, 20, 30, 40, 50, 60, 70, 80 and 90. The engineer carefully measured the mass (in kilograms) of each sample. For the range of volumes, the true density, d, (in kg/mL) of the material can be estimated by the slope of the least-squares regression line fit to the resulting data. (Recall that density is defined as mass divided by volume.) Computer output for the regression analysis is shown below.

Mass = (Intercept) + d*Volume + Error

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	0.9759	0.49276	1.98
Volume	1.3812	0.00876	157.73

s = 0.6783

a. Since the process the engineer used to measure the mass of samples does not always provide the true density value, the regression model shown above contains a random error term. To use the t-distribution to perform a test of hypotheses or construct a confidence interval for the slope of the least-squares regression line, the random errors must conform to some model assumptions. These include the assumption that the random errors are independent of each other. In the context of this experiment, this means that the error that the engineer makes in measuring the mass of one sample has no influence on the error made in determining the density of any other sample.

   Describe two other assumptions about the distribution of the random errors that are needed to use the t-distribution to perform a test of hypotheses or construct a confidence interval for the slope of the least squares regression line.

b. Assuming that all of the assumptions that you considered in part (a) are satisfied, construct a 95 percent confidence interval for d, the true density of the material. With respect to the target density of 1.37 kg/mL, what conclusion can be reached?

c. In context, what does the intercept in the least squares regression line that is associated with the computer output represent?

d. Use your answer from part (c) to explain why it might be reasonable to set the intercept equal to 0 and consider the resulting alternative model

$$Mass = d * Volume + Error$$

as a model for the true density of the material.

The engineer fit a regression model with an intercept of 0 to the data from the nine samples and obtained the following results for the least squares estimate of the slope.

Mass = d*Volume + Error

Coefficients:

	Estimate	Std. Error	t value
Volume	1.3966	0.00469	297.49

s = 0.7925

The engineer then wanted to use the results to construct a 95 percent confidence interval for the slope, but could not decide if it should be constructed as

$1.3966 \pm (2.306)(0.00469)$ or as $1.3966 \pm (2.262)(0.00469)$.

The first expression uses the critical value 2.306, the 97.5 percentile of a t-distribution with 8 degrees of freedom, and the second expression uses the critical value 2.262, the 97.5 percentile of a t-distribution with 9 degrees of freedom. The engineer decided to perform a simulation study to determine the appropriate formula to use to construct a 95 percent confidence interval for the slope.

e. To perform the simulation study, the engineer will simulate samples of mass observations using the model

$$Mass = 1.3966 * Volume + Error$$

for the nine volumes of material (10, 20, 30, 40, 50, 60, 70, 80, and 90) used in the study. The standard deviation of the random errors is assumed to be 0.7925. Explain how the engineer can simulate a sample of observations for the nine amounts using a computer program that can generate random samples from a standard normal distribution.

f. Given the method of simulating samples described in part (e), explain how the engineer can determine which of the two expressions provides an appropriate method of constructing a 95 percent confidence interval for the slope of a least squares regression line, using a model with an intercept of 0.

Answer Key

Sample Correct Response

a. One assumption is that the random errors at each volume have the same variance. The level of variation in the mass measurements about the true mass does not vary with the volume. A second assumption is that the random errors at each volume are normally distributed.

b. A 95 percent confidence interval for the slope is 1.3812 ± (2.365)(0.00876) or (1.3605, 1.4019) kg/mL. Since a 95 percent confidence interval has a probability of 0.95 of containing the true value, and the interval provided by the data includes 1.37, this result is consistent with a material with a true density of 1.37 kg/mL.

c. The intercept represents the mean mass in kilograms for a sample with a volume of 0.

d. Since a sample with no volume should have no mass, it would be very reasonable to consider a regression model with an intercept of 0.

e. The basic steps for simulating a data set with nine observations are:
    - Use a calculator or computer to generate a random observation z from a standard normal distribution.
    - For a specific volume, compute a simulated mass measurement as $y = 1.3966 \times volume + 0.7925 \times z$.
    - Repeat each step in the previous two bulleted statements for each of the nine amounts—10, 20, 30, 40, 50,60, 70, 80, and 90—to obtain a sample of nine mass measurements.

f. The engineer would need to simulate a large number of data sets, say 10,000, using the procedure described in part (e). For each simulated data set, the engineer would need to fit the least-squares regression line with an intercept of 0 to obtain the estimated slope and its standard error and evaluate both of the proposed formulas for constructing a 95 percent confidence interval for the slope. For the appropriate method, about 95 percent of the simulated confidence intervals should contain 1.3966, the slope of the “true” line used to simulate the data sets.

Scoring:

There are six parts for this question. Parts (a) and (b) are scored together as a single score—essentially correct, partially correct, or incorrect. Parts (c) and (d) are scored together as a single score—also as essentially correct, partially correct, or incorrect. Part (e) is scored as essentially correct, partially correct, or incorrect. Part (f) is scored as essentially correct, partially correct, or incorrect. Each essentially correct response in each part is worth 1 point, each partially correct response is worth ½ point, and each incorrect response is worth 0 points.

Parts (a) and (b):

- are essentially correct if the student presents two correct assumptions about the distribution of the random errors in part (a) and constructs a correct confidence interval in part (b) with a corresponding correct conclusion in context.

- are partially correct if the student correctly responds to part (a) in its entirety but only partially correctly to part (b), OR correctly responds to part (b) in its entirety but only partially correctly to part (a), OR responds partially correctly to each of the two parts. In order to earn credit for a partially correct response in either part, the part (a) response must have one correct assumption stated, and the part (b) response must have a correct confidence interval with an incorrect or incomplete conclusion OR a confidence interval that contains only a computational error with a corresponding conclusion that is correct based on that interval.

- are incorrect if the student's response does not meet at least the criteria described in the partially correct category.

Parts (c) and (d):

- are essentially correct if the student presents a correct interpretation for the intercept in part (c) and a correct explanation in part (d) that supports the use of an intercept of 0 for the alternative model.

- are partially correct if the student answers either part (c) or part (d) correctly.

- are incorrect if the student's response does not meet at least the criteria described in the partially correct category.

Part (e)

- is essentially correct if the student presents a correct method for simulating a data set with nine observations.

- is partially correct if the student presents an incomplete method for simulating the data set. The method must convey some information beyond the general level (i.e., use a random number table or use the random number generator on the calculator).

- is incorrect if the student's response does not meet at least the criteria described in the partially correct category.

Part (f)

- is essentially correct if the student fully explains the process for determining which of the two provided expressions will result in a confidence interval for the slope of a least-squares regression line for the model with an intercept of 0.

- is partially correct if the student utilizes the information in part (e) to fit the least-squares regression line with an intercept of 0 for the purpose of obtaining an estimated slope and standard error but does not complete the process OR makes a mistake in carrying out this process.

- is incorrect if the student's response does not meet at least the criteria described in the partially correct category.