

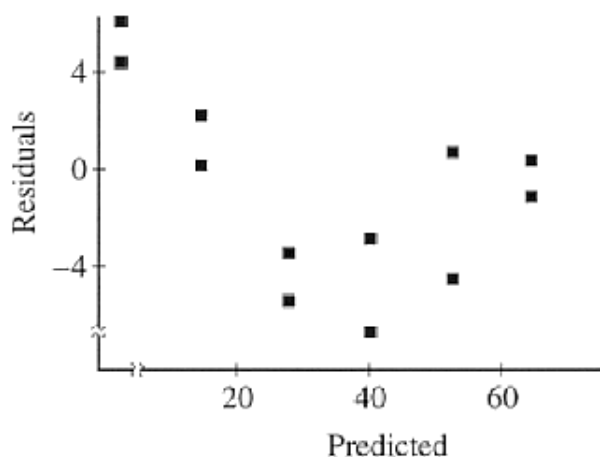
4. In a study of the application of a certain type of weed killer, 14 fields containing large numbers of weeds were treated. The weed killer was prepared at seven different strengths by adding 1, 1.5, 2, 2.5, 3, 3.5, or 4 teaspoons to a gallon of water. Two randomly selected fields were treated with each strength of weed killer. After a few days, the percentage of weeds killed on each field was measured. The computer output obtained from fitting a least squares regression line to the data is shown below. A plot of the residuals is provided as well.

Dependent variable is: percent killed

R squared = 97.2% R squared (adjusted) = 96.9%

$s = 4.505$ with $14 - 2 = 12$ degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	8330.16	1	8330.16	410
Residual	243.589	12	20.2990	
Variable	Coefficient	s.e. of Coeff	t-ratio	Prob
Constant	-20.5893	3.242	-6.35	≤ 0.0001
No. Teaspoons	24.3929	1.204	20.3	≤ 0.0001



- (a) What is the equation of the least squares regression line given by this analysis? Define any variables used in this equation.
- (b) If someone uses this equation to predict the percentage of weeds killed when 2.6 teaspoons of weed killer are used, which of the following would you expect?
- ☐ The prediction will be too large.
 - ☐ The prediction will be too small.
 - ☐ A prediction cannot be made based on the information given on the computer output.
- Explain your reasoning.

GO ON TO THE NEXT PAGE

5. A large university provides housing for 10 percent of its graduate students to live on campus. The university's housing office thinks that the percentage of graduate students looking for housing on campus may be more than 10 percent. The housing office decides to survey a random sample of graduate students, and 62 of the 481 respondents say that they are looking for housing on campus.
- (a) On the basis of the survey data, would you recommend that the housing office consider increasing the amount of housing on campus available to graduate students? Give appropriate evidence to support your recommendation.
 - (b) In addition to the 481 graduate students who responded to the survey, there were 19 who did not respond. If these 19 had responded, is it possible that your recommendation would have changed? Explain.
-

**GO ON TO THE NEXT PAGE**

Free-Response Scoring Guidelines: Question 4**4 Complete Response**

- (a) Correctly gives equation of the regression line as $\hat{y} = -20.5893 + 24.3929x$ (Could use y .) and defines both variables: $x = \#$ of teaspoons of weed killer; $\hat{y} = \%$ killed
 OR
 percent killed $= -20.5893 + 24.3929(\# \text{ of teaspoons of weed killer})$
- (b) Substitutes $x = 2.6$ into the regression equation to get a predicted value of 42.83224, and notes that the residuals around the predicted value of 42.8 (or the middle of the predicted values) are negative. Concludes that since the residual for this prediction is negative, the prediction is expected to be too large.
 OR
 Notes that $x = 2.6$ is about in the middle of the explanatory values and, hence, the predicted percent killed will be close to the middle of the predicted values. Concludes that since the residual for this prediction is negative, the prediction is expected to be too large.
 OR
 Notes that $x = 2.6$ is in the middle of the explanatory values and that the residuals as a function of the explanatory values must exhibit the same pattern of positive and negative residuals. Since the residuals in the middle of the explanatory values are negative, the predicted value is expected to be too large.
- Arithmetic errors in (b) that give reasonable predictions (i.e. predictions between 20 and 60) should not be penalized.

3 Substantial Response

Gives a correct answer for either parts (a) or (b) and a partially correct answer to the other part.

Partially correct answers include but are not limited to:

- (a) Gives the correct equation but fails to define both variables.
- (a) Switches the values for slope and y -intercept in the equation but defines both variables.
- (a) Defines both variables but gives only one correct coefficient in the correct place of the linear equation.
- (b) Correctly explains why the residual at $x = 2.6$ is negative, but incorrectly interprets this negative residual to mean that the predicted value will be too small.
- (b) States that the residual at $x = 2.6$ is negative and thus the predicted value will be too large but fails to specify where the predicted residual for $x = 2.6$ is relative to the other residuals on the residual plot.
- (b) Uses the correct model and gets an incorrect prediction that is not reasonable, but reasons correctly using this prediction. (Unreasonable predictions are below 20 or greater than 60.)
- (b) Gives incorrect residual but interprets it correctly.

2 Developing Response

Gives a correct answer to one of (a) or (b) but not a partially correct answer to the other
 OR

gives a partially correct response to both (a) and (b).

The following responses receive scores of zero for (b):

- Saying that the line cannot be used for prediction or saying that a prediction cannot be based on the computer output because the appropriate model should be quadratic or because the residuals have a non-random pattern. (An r^2 of 97.2% indicates that the fit is good; the residual plot merely reveals that the fit could be improved.)
- Computing a prediction (usually 42.8%) from the equation in (a) and saying that the prediction is “too large” without giving any reasoning from the prediction value to the conclusion of “too large.”

1 Minimal Response

Gives a partially correct answer to either (a) or (b), but not both.
