1. In a statistics course, a linear regression equation was computed to predict the final-exam score from the score on the first test. The equation was $\hat{y} = 10 + 0.9x$ where $y$ is the final exam score and $x$ is the score on the first test. Carla scored 95 on the first test. What is the predicted value of her score on the final exam? (a) 85.5 (b) 90 (c) 95 (d) 95.5 (e) none of these

2. In the course described in #1, Bill scored a 90 on the first test and a 93 on the final exam. What is the value of his residual? (a) -2.0 (b) 2.0 (c) 3.0 (d) 93 (e) none of these

3. A least-squares regression line for predicting weights of basketball players on the basis of their heights produced the residual plot below. What does the residual plot tell you about the linear model?

(a) A residual plot is not an appropriate means for evaluating a linear model.
(b) The curved pattern in the residual plot suggests that there is no association between the weight and height of basketball players.
(c) The curved pattern in the residual plot suggests that the linear model is not appropriate.
(d) There are not enough data points to draw any conclusions from the residual plot.
(e) The linear model is appropriate, because there are approximately the same numbers of points above and below the horizontal line in the residual plot.
For questions 4-5:
One concern about the depletion of the ozone layer is that the increase in ultraviolet (UV) light will decrease crop yields. An experiment was conducted in a greenhouse where soybean plants were exposed to varying levels of UV, measured in Dobson units. At the end of the experiment the yield (kg) was measured. A regression analysis was performed with the following results:

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>uv</td>
</tr>
</tbody>
</table>

4. The least-squares regression line is the line that
(a) minimizes the sum of the distances between the actual UV values and the predicted UV values.
(b) minimizes the sum of the squared residuals between the actual yield and the predicted yield.
(c) minimizes the sum of the distances between the actual yield and the predicted UV.
(d) minimizes the sum of the squared residuals between the actual UV reading and the predicted UV values.
(e) minimizes the perpendicular distance between the regression line and each data point.

5. Which of the following is correct?
(a) If the UV value increases by 1 Dobson unit, the yield is expected to increase by 0.0463 kg.
(b) If the yield increases by 1 kg, the UV value is expected to decrease by 0.0463 Dobson units.
(c) If the UV value increases by 1 Dobson unit, the yield is expected to decrease by 0.0463 kg.
(d) The predicted yield is 4.3 kg when the UV value is 20 Dobson units.
(e) None of the above is correct.

6. You have data for many families on the parents' income and the years of education their eldest child completes. Your initial examination of the data indicates that children from wealthier families tend to go to school for longer. When you make a scatterplot,
(a) the explanatory variable is parents' income, and you expect to see a negative association.
(b) the explanatory variable is parents' income, and you expect to see a positive association.
(c) the explanatory variable is parents' income, and you expect to see very little association.
(d) the explanatory variable is years of education, and you expect to see a negative association.
(e) the explanatory variable is years of education, and you expect to see a positive association.
7. A certain psychologist counsels people who are getting divorced. A random sample of ten of her patients provided the data in the following scatterplot, where $x =$ number of years of courtship before marriage, and $y =$ number of years of marriage before divorce.

![Scatterplot](image)

**a)** Describe what the scatterplot reveals about the relationship between length of courtship and length of marriage.

*In general, it seems as though the longer the courtship, the longer the couple will be married.*

**b)** Suppose a new point at (4.5, 8), that is, years of courtship = 4.5 and years of marriage = 8, were added to the plot. What effect, if any, will this new point have on the correlation between courtship duration and marriage duration? Explain.

*It wouldn't have too much effect on correlation. The point would add to the linear relationship but not make the correlation much stronger.*

Below is the computer output for the regression of length of marriage versus length of courtship.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.710</td>
<td>1.880</td>
<td>3.04</td>
<td>0.016</td>
</tr>
<tr>
<td>courtship</td>
<td>2.4559</td>
<td>0.6669</td>
<td>3.68</td>
<td>0.006</td>
</tr>
</tbody>
</table>

$S = 2.74982$  $R-$Sq = 62.9%  $R-$Sq(adj) = 58.3%

**c)** What is the least-squares regression equation?

$\text{yrs. of marriage} = 5.710 + 2.4559 \times (\text{years of courtship})$
d) What is the slope? Interpret the slope in the context of this problem.
For each year of courtship, the couple would stay married for 2.4559 years more.

e) What is the coefficient of determination and what does it mean in the context of this problem?
\[ r^2 = 0.029 \]
82.9% of the variation in years of marriage is accounted for by the regression line.

f) Explain what the quantity \( S = 2.74982 \) measures in the context of this problem.
The typical error is 2.74982 years from the regression line.

8. Data is recorded on the lean body mass and resting metabolic rate for 12 women who were subjects in a study of dieting. Lean body mass, given in kilograms, is a person's weight leaving out all fat. Metabolic rate, in calories burned per 24 hours, is the rate at which the body consumes energy. Here are the data.

| (x) Mass: 36.1 54.6 48.5 42.0 50.6 42.0 40.3 33.1 42.4 34.5 51.1 41.2 |
| (y) Rate: 995 1425 1396 1418 1502 1256 1189 913 1124 1052 1347 1204 |

a) Enter the data into your calculator and create a scatterplot.

b) Determine the equation of the least-squares regression line.
\[ \hat{y} = 201.16 + 24.03x \]
\[ \text{Rate} = 201.16 + 24.03(\text{mass}) \]

c) Explain in words what the slope of the regression line tells us.
For every kg of lean body mass, the metabolic rate increases by 24.03 calories/24 hrs.

d) Another woman has a lean body mass of 45 kilograms. What is her predicted metabolic rate?
\[ \hat{y} = 201.16 + 24.03(45) \]
\[ \hat{y} = 1282.51 \]

e) Use your calculator to make a residual plot. Describe what this graph tells you about how well the line fits the data. A linear model would be appropriate because the points show no pattern and are relatively small, except for one possible outlier.
f) Determine the residual for the woman with a body mass of 48.5 kg. Interpret this value in the context of the problem.

\[ \hat{y} = 201.16 + 24.03 \times (48.5) \]
\[ \hat{y} = 13606.615 \text{ (predicted y)} \]

\[ \text{residual} = 13910 - 13606.615 \]

\[ 13910 \text{ (observed y)} \]

\[ \checkmark \]

\[ \text{our line underpredicted} \]

\[ \text{The metabolic rate by 29.385 cal} \]

\[ g) \text{ For the regression you performed earlier,} \quad r^2 = 0.768 \text{ and} \quad s = 95.08. \text{ Explain what each of} \]
\[ \text{these values means in this setting.} \]

* 76.8% of the variation in the metabolic rate is accounted for by the regression line

* The typical error is 95.08 calories from the regression line.