Part 1: Multiple Choice. Circle the letter corresponding to the best answer.

Questions 1 and 2 refer to the following information:

For children between the ages of 18 months and 29 months, there is an approximately linear relationship between height and age. The relationship can be represented by $\hat{y} = 64.93 + 0.63x$, where $y$ represents height (in centimeters) and $x$ represents age (in months).

1. Joseph is 22.5 months old. What is his predicted height?
   (a) 50.80  (b) 64.96  (c) 65.96  (d) 79.11  (e) 87.40

2. Loretta is 20 months old and is 80 centimeters tall. What is her residual?
   (a) $-2.47$  (b) $2.47$  (c) $-12.60$  (d) $12.60$  (e) $77.53$

3. 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.

4. A community college announces that the correlation between college entrance exam grades and scholastic achievement was found to be $-1.08$. On the basis of this you would tell the college that
   (a) the entrance exam is a good predictor of success.
   (b) the exam is a poor predictor of success.
   (c) students who do best on this exam will be poor students.
   (d) students at this school are underachieving.
   (e) the college should hire a new statistician.

5. An agricultural economist says that the correlation between corn prices and soybean prices is $r = 0.7$. This means that
   (a) when corn prices are above average, soybean prices also tend to be above average.
   (b) there is almost no relation between corn prices and soybean prices.
   (c) when corn prices are above average, soybean prices tend to be below average.
   (d) when soybean prices go up by 1 dollar, corn prices go up by 70 cents.
   (e) the economist is confused, because correlation makes no sense in this situation.
6. Which of the following are true statements?

   I. Correlation requires that there are clearly-identified explanatory and response variables.
   II. Scatterplots require that both variables be quantitative.
   III. Every least-squares regression line passes through \((\bar{x}, \bar{y})\).

   (a) I and II only
   (b) I and III only
   (c) II and III only
   (d) I, II, and III
   (e) None of the above

7. There is an approximate linear relationship between the height of females and their age (from 5 to 18 years) described by the equation \(\text{Height} = 50.3 + 6.1\, \text{(Age)}\), where height is measured in centimeters and age in years. Which one of the following statements must be true?

   (a) The estimated slope is 6.1, which implies that female children between the ages of 5 and 18 increase in height by about 6.1 cm for each year they grow older.
   (b) The correlation between Height and Age is negative.
   (c) The estimated intercept is 50.3 cm. We can conclude from this that the typical height of female children at birth is 50.3 cm.
   (d) The heights of 68% of female children who are 10 years old are within 6.1 cm. of 110.4 cm.
   (e) My niece is 8 years old and is 115 cm tall. She is shorter than average for girls her age.

8. You are interested in predicting the cost of heating houses on the basis of how many rooms the house has. A scatterplot of 25 houses reveals a strong linear relationship between these variables, so you calculate a least-squares regression line. “Least-squares” refers to

   (a) Minimizing the sum of the squares of the 25 houses’ heating costs.
   (b) Minimizing the sum of the squares of the number of rooms in each of the 25 houses.
   (c) Minimizing the sum of the products of each house’s actual heating costs and the predicted heating cost based on the regression equation.
   (d) Minimizing the sum of the squares of the difference between each house’s heating costs and number of rooms.
   (e) Minimizing the sum of the squares of the residuals.
9. A study of the fuel economy for various automobiles plotted the fuel consumption (in liters of gasoline used per 100 kilometers traveled) vs. speed (in kilometers per hour). A least-squares line was fit to the data. Here is the residual plot from this least-squares fit.

What does the residual plot tell you about the linear model?
(a) The residual plot confirms the linearity of the fuel economy data.
(b) The residual plot does not confirm nor rule out the linearity of the data.
(c) The residual plot suggests that the model may be linear, but more data points are needed to confirm this.
(d) The residual plot clearly indicates that the data isn’t linear.
(e) A residual plot is not an appropriate means for evaluating a linear model.

10. Leonardo da Vinci, the renowned painter, speculated that an ideal human would have an armspan (distance from the outstretched fingertip of the left hand to the outstretched fingertip of the right hand) that was equal to his height. Is it possible to predict armspan from height? The following computer regression printout shows the results of a least-squares regression of armspan on height, both in inches, for a sample of 18 high school students.

<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>11.5474</td>
<td>5.6</td>
<td>2.06</td>
<td>0.0558</td>
</tr>
<tr>
<td>Height</td>
<td>0.84024</td>
<td>0.08091</td>
<td>10.4</td>
<td>0.000</td>
</tr>
<tr>
<td>R-sq = 87.1%</td>
<td>R-sq(adj.) = 86.3%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The students’ armspans ranged from 62 to 76 inches.

Which of the following statements is true?
(a) If one of the students in the sample had a height of 70 inches and an armspan of 68 inches, then the residual for this student would be about –2.36 inches.
(b) The correlation between height and armspan is .871.
(c) Contrary to da Vinci’s speculation, the regression model suggests that, for these students at least, height is about 84% of armspan.
(d) For every one-inch increase in armspan, the regression model predicts about a 0.84-inch increase in height.
(e) For a student 66 inches tall, this model would predict an armspan of about 68 inches.
Part 2: Free Response

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

11. How are traffic delays related to the number of cars on the road? Below is data on the total number of hours of delay per year at 10 major highway intersections in the western United States versus traffic volume (measured by average number of vehicles per day that pass through the intersection).

(a) Describe what the scatterplot reveals about the relationship between traffic delays and number of cars on the road.

(b) Suppose another data point at (200000, 24000), that is 200,000 vehicles per day and 24,000 hours of delay per year, were added to the plot. What effect, if any, will this new point have on the correlation between these two variables? Explain.
Below is computer output for the regression of hours of delay versus number of vehicle per day.

<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>-3629</td>
<td>7367</td>
<td>-0.49</td>
<td>0.634</td>
</tr>
<tr>
<td>vehicles per day</td>
<td>0.07822</td>
<td>0.02684</td>
<td>2.91</td>
<td>0.017</td>
</tr>
</tbody>
</table>

S = 3899.57  R-Sq = 48.6%  R-Sq(adj) = 42.8%

(c) What is the slope of the regression line? Interpret the slope in the context of this problem.

(d) Explain what the quantity S = 3899.57 measures in the context of this problem.

(e) Below is the same scatterplot, but with the six intersections in California plotted as circles and the four in other western states plotted as squares.

Comment on how the relationship between average number of vehicles per day and hours of delay per year differs between the California intersections and the intersections in other western states.
12. An ecologist studying breeding habits of the common crossbill in different years finds that there is a linear relationship between the number of breeding pairs of crossbills and the abundance of the spruce cones. Below are statistics on eight years of measurements, where \( x \) = average number of cones per tree and \( y \) = number of breeding pairs of crossbills in a certain forest.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x ) = mean number of cones/tree</td>
<td>23.0</td>
<td>16.2</td>
</tr>
<tr>
<td>( y ) = number of crossbill pairs</td>
<td>18.0</td>
<td>15.1</td>
</tr>
</tbody>
</table>

The correlation between \( x \) and \( y \) is \( r = 0.968 \).

(a) Find the equation of the least-squares regression line (with \( y \) as the response variable).

(b) What percentage of the variation in numbers of breeding pairs of crossbills can be accounted for by this regression?

(c) Based on these data, can we conclude that the abundance of spruce cones is responsible for the number of breeding pairs of crossbills? Explain.