1. The correlation between the heights of fathers and the heights of their (fully grown) sons is \( r = 0.52 \). This value was based on both variables being measured in inches. If fathers' heights were measured in feet (one foot equals 12 inches), and sons' heights were measured in furlongs (one furlong equals 7920 inches), the correlation between heights of fathers and heights of sons would be

(a) much smaller than 0.52
(b) slightly smaller than 0.52
(c) unchanged: equal to 0.52
(d) slightly larger than 0.52
(e) much larger than 0.52

2. All but one of the following statements contains an error. Which statement could be correct?

(a) There is a correlation of 0.54 between the position a football player plays and his weight.
(b) We found a correlation of \( r = -0.63 \) between gender and political party preference.
(c) The correlation between the distance travelled by a hiker and the time spent hiking is \( r = 0.9 \) meters per second.
(d) We found a high correlation between the height and age of children: \( r = 1.12 \).
(e) The correlation between mid-August soil moisture and the per-acre yield of tomatoes is \( r = 0.53 \).

3. There is an approximate linear relationship between the height of females and their age (from 5 to 18 years) described by predicted height = 50.3 + 6.01(age) where height is measured in centimeters and age in years. Which of the following is not correct?

(a) The estimated slope is 6.01, which implies that female children between the ages of 5 and 18 increase in height by about 6 cm for each year they grow older.
(b) The estimated height of a female child who is 10 years old is about 110 cm.
(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 average height of female children when they are 5 years old is about 50% of the average height when they are 18 years old.
(e) My niece is about 8 years old and is about 115 cm tall. She is taller than average for girls her age.

4. 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.
5. Which statements below about least-squares regression are correct?
I. Switching the explanatory and response variables will not change the least-squares regression line.
II. The slope of the line is very sensitive to outliers with large residuals.
III. A value of $r^2$ close to 1 does not guarantee that the relationship between the variables is linear.

(a) Only I is correct.
(b) Only II is correct.
(c) Only III is correct.
(d) Both II and III are correct.
(e) All three statements—I, II, and III—are correct.

6. Other things being equal, larger automobile engines consume more fuel. You are planning an experiment to study the effect of engine size (in liters) on the gas mileage (in miles per gallon) of sport utility vehicles. In this study,

(a) gas mileage is a response variable, and you expect to find a negative association.
(b) gas mileage is a response variable, and you expect to find a positive association.
(c) gas mileage is an explanatory variable, and you expect to find a strong negative association.
(d) gas mileage is an explanatory variable, and you expect to find a strong positive association.
(e) gas mileage is an explanatory variable, and you expect to find very little association.

7. 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.

8. Suppose we fit a least-squares regression line to a set of data. What is true if a plot of the residuals shows a curved pattern?

(a) A straight line is not a good model for the data.
(b) The correlation must be 0.
(c) The correlation must be positive.
(d) Outliers must be present.
(e) The regression line might or might not be a good model for the data, depending on the extent of the curve.

9. A regression of the amount of calories in a serving of breakfast cereal vs. the amount of fat gave the following results: Predicted Calories = 97.1053 + 9.6525(Fat). Which of the following is false?

(a) It is estimated that for every additional gram of fat in cereal, the number of calories increases by about 10.
(b) It is estimated that in cereals with no fat, the total amount of calories is about 97.
(c) If a cereal has 2 g of fat, then it is estimated that the total number of calories is about 116.
(d) The correlation between amount of fat and calories is positive.
(e) If one cereal has 140 calories and 5 g of fat. Its residual is about 5 calories.
10. 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.

11. One weekend, a statistician notices that some of the cars in his neighborhood are very clean and others are quite dirty. He decides to explore this phenomenon, and asks 15 of his neighbors how many times they wash their cars each year and how much they paid in car repair costs last year. His results are in the table below:

| \( x = \text{number of car washes per year} \) & Mean & Standard deviation |
|--------------------------------|------|------------------|
| \( y = \text{repairs costs for last year} \) & $955.30 & $323.50 |

The correlation for these two variables is: \( r = -0.71 \)

(a) Find the equation of the least-squares regression line (with \( y \) as the response variable).
(b) What percentage of the variation in repair costs can be explained by the number of times per year a car is washed?

12. How are traffic delays related to the number of cars on the road? Statisticians gathered 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). 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>-3.629</td>
<td>7.367</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>

\( s = 3699.57 \) \( \text{R-Sq} = 48.6\% \) \( \text{R-Sq(adj)} = 42.8\% \)

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

13. Because elderly people may have difficulty standing to have their heights measured, a study looked at predicting overall height from height to the knee. Here are data (in centimeters) for five elderly men:

<table>
<thead>
<tr>
<th>Knee Height, cm.</th>
<th>57.7</th>
<th>47.4</th>
<th>43.5</th>
<th>44.8</th>
<th>55.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height, cm</td>
<td>192</td>
<td>153</td>
<td>146</td>
<td>163</td>
<td>169</td>
</tr>
</tbody>
</table>

(a) Which variable is explanatory and which is response in this situation?
(b) Construct a scatterplot on your calculator and draw a rough sketch of your calculator’s display.
(c) Describe the form, direction, and strength of the relationship that you see.