Please communicate as clearly as possible with your answers.

1. A company wanted to determine the health care costs of its employees. A sample of 25 employees were interviewed and their medical expenses for the previous year were determined. Later the company discovered that the highest medical expense in the sample was mistakenly recorded as 10 times the actual amount. However, after correcting the error, the correct amount was still greater than or equal to any other medical expense in the sample. Tell how the following summary statistics would have been affected:

   a) Mean

   b) Median

   c) IQR

   d) Range

   e) Variance

2. Camila and Mari have cars in desperate need of repair. On any given day, the probability that Camila’s car will break down is 0.5 and the probability that Mari’s car will break down is 0.4. Assume that their cars’ needs for repair are independent of each other. What is the probability that at least one of the cars will break down?

3. Given a set of ordered pairs \((x, y)\) so that \(s_x = 1.6, s_y = 0.75, r = 0.55\). What is the slope of the least squares regression line for this data?
4. Data was collected in 2008 on the salaries of the 88 quarterbacks on the National Football League rosters. The mean and the median are $3,475,000 and $1,695,000, but I’m not sure which is which. Tell which one you think the mean is, and why.

5. A university is proposing a new procedure for professors to gain tenure. To gauge sentiment about the proposal, the university intends to randomly sample five full professors, five assistant professors, five associate professors, five adjunct professors, and five visiting professors. This is an example of what?

6. Following are the SAT math scores for 20 students applying to Princeton University: 730, 720, 610, 740, 680, 740, 740, 690, 770, 580, 760, 780, 520, 730, 760, 670, 770, 790, 680, and 750. Tell if the distribution of scores is symmetric, skewed to the left, skewed to the right, approximately uniform or approximately normal.

7. Two students went to the local mall to conduct a survey. They wanted to know how the local population felt about boys coloring their hair. Both students had neat haircuts, but one had dyed hair and the other did not. Discuss the types of bias that might occur in this study.
8. A sample of 56 college students was asked to report the number of hours devoted to study during a typical week. The stem-and-leaf plot of the results is shown below. Describe the data and draw a boxplot.

```
0 | 3
0 | 555667777778999
1 | 000000000112222333344444
1 | 555668
2 | 0
2 | 56
3 | 00
3 | 5
4 | 0
```

9. A random sample size 10 was taken from a population. The sample has a variance of zero. Which of the following statements must be true?
   I. The population also has a variance of zero.
   II. The sample mean is equal to the sample median.
   III. The ten data points in the sample are equal in numerical value.

10. Which of the following distributions are more likely to be skewed to the right than skewed to the left? Which are more likely to be skewed to the left than skewed to the right?
   A. Household incomes
   B. Home prices
   C. Ages of teenage drivers
11. Here are the test scores of 32 students.

Draw a boxplot and histogram that illustrate these data. Identify any outliers.

32  37  41  44  46  48  53  55
56  57  59  63  65  66  68  69
70  71  74  74  75  77  78  79
80  82  83  86  89  92  95  99

12. Last weekend police ticketed 18 men whose mean speed was 72 miles per hour, and 30 women going an average of 64 mph. Overall, what was the mean speed of all the people ticketed?

13. Here are some summary statistics for the recent English exam: lowest score = 34, mean score = 74, median score = 88.2, range = 76. Is the distribution symmetric, skewed to the left, or skewed to the right? Explain.
14. A study was conducted to compare the average time spent in the lab each week versus course grade for computer students. The results are recorded in the table below.

<table>
<thead>
<tr>
<th>Number of hours spent in lab</th>
<th>Grade (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>96</td>
</tr>
<tr>
<td>11</td>
<td>51</td>
</tr>
<tr>
<td>16</td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td>58</td>
</tr>
<tr>
<td>7</td>
<td>89</td>
</tr>
<tr>
<td>15</td>
<td>81</td>
</tr>
<tr>
<td>16</td>
<td>46</td>
</tr>
<tr>
<td>10</td>
<td>51</td>
</tr>
</tbody>
</table>

Find the correlation between the two variables, and write the linear regression line that predicts the grade from the number of hours spent in the lab.

15. Here are some scores from a recent Mathematics exam: 95.5, 65.9, 93.2, 80.6, 56.8, 50, 86.4, 54.5, 40.9, 77.3, 79.5, 10, 65.9, 70.5, 15, 77.3, 81.8, 12, 50, 79.5, 60.2. Which is a better summary of the scores, the mean or the median? Explain

16. The volumes of soda in quart soda bottles can be described by a Normal model with a mean of 32.3 oz and a standard deviation of 1.2 oz. What percentage of bottles can we expect to have a volume less than 32 oz?
17. Fill in the missing information:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$s_x$</th>
<th>$y$</th>
<th>$s_y$</th>
<th>$r$</th>
<th>$\hat{y} = b_0 + b_1 x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>2</td>
<td>21</td>
<td>3</td>
<td>0.5</td>
<td>$\hat{y} =$ ?</td>
</tr>
</tbody>
</table>

18. Tell what the following residual plots would tell you about your linear model:

![Residual plots](image)

19. Suppose the correlation between SAT Verbal scores and Math scores is 0.57 and that these scores are normally distributed. If a student's Verbal score places her at the 90th percentile, at what percentile would you predict her Math score to be?

20. A regression analysis of students' college grade point averages (GPAs) and their high school GPAs found $r^2 = 0.311$. Which of the following statements is/are true?

I. High school GPA accounts for 31.1\% of college GPA.

II. 31.1\% of college GPAs can be correctly predicted with this model.

III. 31.1\% of the variance in college GPA can be accounted for by the model.
21. A philosophy professor has found a correlation of 0.80 between the number of hours students study for his exams and their exam performance. During the time he collected the data, students studied an average of 10 hours with a standard deviation of 2.5 hours, and scored an average of 80 points with a standard deviation of 7.5 points. Write the least squares regression line that estimates the number of points a student will score on the next exam from the number of hours the student studies.

22. The average number of calories in a 1.5 ounce chocolate bar is 225. Suppose that the distribution of calories is approximately Normal, with \( \sigma = 10 \). Find the percentage of chocolate bars that have …
   a. Between 200 and 220 calories
   
   b. Less than 200 calories.

23. The average salary for a Queens College full professor is $85,900 (from the Chronicle of Higher Education). If the average salaries are normally distributed with a standard deviation of $11,000, what percentage of the full professors at Queens College have salaries that are…
   a. More than $90,000?
   
   b. More than $75,000?

24. The average length of a hospital stay is 5.9 days. If we assume a normal distribution and a standard deviation of 1.7 days, …
   a. 15% of the hospital stays are less than how many days?
   
   b. 20% of the hospital stays are longer than how many days?
25. The average waiting time to be seated for dinner at a popular restaurant is 23.5 minutes, with a standard deviation of 3.6 minutes. The waiting times are normally distributed.

   a. What percentage of the customers have to wait longer than 25 minutes?

   b. What percentage of the customers wait between 15 and 22 minutes?

   c. 98% of the customers waited longer than I did. How long did I wait?

   d. What is the IQR of this data?

26. A new medication has been developed to treat sleep-onset insomnia (difficulty in falling asleep). Researchers want to compare this drug to a drug that has been used in the past by comparing the length of time it takes subjects to fall asleep. How (sample, observational study, experiment) would you obtain this information? Describe your process in detail.

27. What is “control” in the context of an experiment?

28. How can I tell if a linear model is appropriate for my data? How can I tell the strength of the association?
29. A manufacturer of boots plans to conduct an experiment to compare a new method of waterproofing to the current method. The appearance of the boots is not changed by either method. The company recruits 100 volunteers in Seattle, where it rains frequently, to wear the boots as they normally would for 6 months. At the end of the 6 months, the boots will be returned to the company to be evaluated for water damage.

   a. Describe a design for this experiment that uses the 100 volunteers. Include a few sentences on how it would be implemented.


30. A least squares regression line is fitted to the weights (in pounds) versus age (in months) of many young children. The equation of the line is \( \hat{y} = 16.6 + .65t \), where \( \hat{y} \) is the predicted weight and \( t \) is the age of the child. A 20-month-old child in this group has an actual weight of 25 pounds. What is the residual for this child?

31. Why would someone ever use an observational study instead of an experiment to get data, if it’s impossible to infer a causal relationship from an observational study? Give an example.
32. Commercial airlines need to know the operating cost per hour of flight for each plane in their fleet. In a study of the relationship between operating cost per hour and number of passenger seats, investigators computed the regression of operating cost per hour on the number of passenger seats. The 12 sample aircraft used in the study included planes with as few as 216 passenger seats and planes with as many as 410 passenger seats. Operating cost per hour ranged between $3,600 and $7,800. Some computer output from regression analysis of these data is shown below.

**SCATTERPLOT WITH LEAST SQUARES REGRESSION LINE**

- **Operating Cost per Hour** vs **Number of Passenger Seats**
- Points on the scatterplot represent individual aircraft data points.
- The line of best fit (least squares regression line) is superimposed on the scatterplot.
- The correlation coefficient (R-Sq) is 57.0%.
- The adjusted R-Sq (adj) is 52.7%.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>StDev</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1136</td>
<td>1226</td>
<td>0.93</td>
<td>0.376</td>
</tr>
<tr>
<td>Seats</td>
<td>14.673</td>
<td>4.027</td>
<td>3.64</td>
<td>0.005</td>
</tr>
<tr>
<td>S = 845.3</td>
<td>R-Sq = 57.0%</td>
<td>R-Sq (adj) = 52.7%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. What is the equation of the least squares regression line that describes the relationship between operating cost per hour and number of passenger seats in the plane? Define any variables used in this equation.

b. What is the correlation coefficient?

c. Interpret the slope in this particular context.
d. Interpret $r^2$ in this particular context.

e. Suppose that you want to describe the relationship between operating cost per hour and number of passenger seats in the plane for planes only in the range of 250 to 350 seats. Does this line shown in the scatterplot still provide the best description of the relationship for data in this range? Why or why not?

33. A researcher interested in eye color versus success in a math program collected the following data from a random sample of 2000 high school students.

<table>
<thead>
<tr>
<th></th>
<th>brown</th>
<th>blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>190</td>
<td>10</td>
</tr>
<tr>
<td>pass</td>
<td>1710</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>1900</td>
<td>100</td>
</tr>
</tbody>
</table>

a. What is the probability that a student from this group fails the math program?

b. What is the probability that a student from this group fails the math program given that he/she has blue eyes?

c. Are blue eyes and failing the math program independent or dependent?

34. Of the 60 obese teenagers in a recent study, 15 had type II diabetes, 20 had high blood pressure, and 10 had both high blood pressure and type II diabetes. Suppose one of these 60 obese teenagers is randomly selected.

a. Given that the teenager has type II diabetes, what is the probability that he or she also has high blood pressure?

b. If the obese teenager does NOT have high blood pressure, what is the probability that he or she also does not have type II diabetes?
35. The probability that Michael misses a free throw shot is .1. If he goes to the line to shoot three free throws (due to a foul on a three-point shot),

   a. What is the probability that Michael misses all three shots? What assumptions did you make in order to calculate this probability?

   b. What is the probability that Michael makes at least one of the three shots?

   c. What is the probability that Michael makes the first shot but not the second or third?

36. Of the 10,000 freshman at the University of Texas, 7000 must take English, 6000 must take History, and 5000 must take both. Suppose that a student is randomly selected.

   a. What is the probability that the selected student must take English?

   b. What is the probability that the selected student must take both English and History?

   c. Suppose you learn that the selected student must take English, what is the probability that this student must take both English and History?

   d. Are the outcomes must take English and must take History independent? Explain.

   e. Answer the question posed in part d if only 4200 of the students must take both English and History.

37. Two shipping services offer overnight delivery of parcels, and both promise delivery before 10 A.M.. A mail-order catalog company ships 30% of its overnight packages using shipping service 1 and 70% using service 2. Service 1 fails to meet the 10 A.M. delivery promise 10% of the time, whereas service 2 fails to deliver by 10 A.M. 8% of the time. Suppose that you made a purchase from this company and were expecting your package by 10 A.M., but it is late. What is the probability that it was shipped by service 1?