Part I. Multiple Choice, True/False, Short Answer. (3 points each)

How many movies does an adult American watch per year? An opinion poll calls 2000 randomly chosen residential telephone numbers, then asks to speak with an adult member of the household. The interviewer asks, "How many movies have you watched in a movie theatre in the past 12 months?". In all 1131 people respond to this question. The results are in the following table.

<table>
<thead>
<tr>
<th>Number of movies seen in the last 12 months</th>
<th>0</th>
<th>1 to 3</th>
<th>4 to 8</th>
<th>9 or more</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>385</td>
<td>463</td>
<td>181</td>
<td>102</td>
</tr>
</tbody>
</table>

1. The type of data being asked for in this opinion poll is
   a) quantitative   
   b) explanatory    
   c) categorical    
   d) response       
   e) none of these

2. Pick one student at random from this population. What is the probability that the chosen student's SAT score is within ±10 points of the mean score 1075?
   a) 0.0399   
   b) 0.3829   
   c) 0.6171   
   d) 0.9601   
   e) none of these

3. Jane makes a SAT score of \( x = 1350 \). What is Jane's z-score?

   Answer: \( z = \frac{x - \mu}{\sigma} = \frac{1350 - 1075}{200} = 1.375 \)

4. John took the SAT, and his z-score was \( z = 1.2 \). What was John's SAT score?
   a) \( x = 835 \)   
   b) \( x = 1490 \)   
   c) \( x = 1075 \)   
   d) \( x = 1315 \)   
   e) none of these

5. What SAT score would a person need in order to be among the top 5% of all SAT test-takers for this year?
   a) \( x = 746 \)   
   b) \( x = 1228 \)   
   c) \( x = 1404 \)   
   d) \( x = 1535 \)   
   e) none of these

6. Circle all of the items below that are measures of spread of a distribution.
   a) mean   
   b) standard deviation   
   c) median   
   d) IQR   
   e) none of these

7. Circle all of the items below that are measures of center of a distribution.
   a) mean   
   b) standard deviation   
   c) median   
   d) IQR   
   e) none of these

8. Circle all of the items below that are resistant to outliers.
   a) mean   
   b) standard deviation   
   c) median   
   d) IQR   
   e) none of these
9. The mean time spent waiting in line at the local Walmart is somewhat higher than the median time spent waiting in line. This means that the distribution of survival times is most likely
   a) left-skewed  b) right-skewed  c) symmetric  d) unimodal  e) none of these

10. If a distribution is strongly skewed, then the most appropriate numerical summary of the distribution is
   a) the two-number summary  b) the five-number summary  c) the correlation coefficient
      d) the $\chi^2$-statistic  e) none of these

11. **True** or **False** (circle one) If people with high cholesterol tend to be overweight, we would expect the correlation between cholesterol and weight to be negative.

12. **True** or **False** (circle one) If two variables are strongly correlated (i.e. $r^2 \approx 1$), then we can be confident that the explanatory variable is causing the changes in the response variable.

13. **True** or **False** (circle one) If two variables have a linear relationship and are strongly correlated (i.e. $r^2 \approx 1$), then a scatterplot of the two variables will be close to a straight line.

Use the following scenario for problems 14 through 16. For a biology project, you measure the weight in grams and the tail length in millimeters (mm) of a group of mice. You then use this data to find the least-squares regression line for predicting tail length from weight: $\hat{y} = 20 + 3x$

14. What tail length would you predict for a mouse that weighs 50 grams?
   a) 20 mm  b) 10 mm  c) 150 mm  d) 170 mm  e) none of these

15. How much on average does tail length increase for each additional gram of weight?
   a) 3 mm  b) 20 mm  c) 23 mm  d) 1/3 mm  e) none of these

16. The correlation coefficient between weight and tail length is $r = 0.80$. What fraction of the variation in tail length is explained by the least squares regression onto weight?
   a) 0.80  b) 0.64  c) 0.20  d) 0.36  e) none of these

Use the following scenario for problems 17 through 19. Below is a boxplot of test scores on a standardized skills test from a random sample of 36 employees from a large company.

17. What is the median of the distribution?
   Answer: 79

18. What is the interquartile range (IQR) of the distribution?
   Answer: $85 - 54 = 31$

19. Based on this boxplot, which of the following would best describe the distribution of test scores?
   a) left-skewed  b) right-skewed  c) symmetric  d) normal  e) none of these
Use the following scenario for problems 20 and 21. At right you see the histograms of four random samples, A, B, C, and D. They are drawn using the same scales on both the horizontal and vertical axis, so they can be directly compared against each other.

20. Which of the four samples (A, B, C, or D) has the largest sample mean?

Answer: D

21. Which of the four samples (A, B, C, or D) has the largest sample standard deviation?

Answer: A

22. Below are three scatterplots, A, B, C. Rank them in order of least correlated to most correlated (i.e. from lowest correlation coefficient to highest).

Answer: C, A, B

Use the following scenario for questions 23 through 26. Below is a histogram of light bulb lifetimes from a simple random sample of 30 bulbs taken from an assembly line. Note that the vertical axis is given in counts (i.e. number of bulbs). Assume that no data point fell exactly on a class (i.e. bin) boundary.
23. Circle all words below that describe the shape of the distribution of light bulb lifetimes:
unimodal  bimodal  uniform  symmetric  left-skewed  right-skewed.

24. What percent of light bulbs in the sample lasted less than 1100 hours?

Answer: \( \frac{25}{30} = 0.8333 \) \( \approx 83.33\% \)

25. Which of the following classes contains the median light bulb lifetime?
   a) 1000 to 1100  b) 1100 to 1200  c) 1200 to 1300  d) 1300 to 1400  e) none of these

26. Use the histogram to estimate the sample mean lifetime.

Answer: 1220

---

Part II. Analysis and Written Response: Show your work, and all calculator commands and inputs. Where appropriate, use the “plus four” method for confidence intervals.

3. (20 points) Is there a correlation between test anxiety and performance on final exams? The Psychology Department did a study on a sample of 9 students, using a test that they know measures anxiety: the higher the score on this test, the more anxious the student is. They gave this anxiety test to each of the 9 students in the sample, and then they gave the students the final exam. Scores for each student are shown in the scatterplot below: the explanatory variable is the student’s score on the anxiety test, and the response variable is the student’s final exam score. The sample statistics of the data are given in the accompanying table. Find the equation of the least-squares regression line for this data, and plot the line accurately on this scatterplot.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std. dev.</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>anxiety level</td>
<td>16.22</td>
<td>6.24</td>
<td>r = -0.8323</td>
</tr>
<tr>
<td>exam score</td>
<td>53.67</td>
<td>10.33</td>
<td></td>
</tr>
</tbody>
</table>

The least-squares regression line is:

\[
\hat{y} = 76.12 - 1.3776x
\]