Instructions: Complete each of the following six questions, and please explain and justify all appropriate details in your solutions in order to obtain maximal credit for your answers.

1. (2 pts) A sociologist was interested in whether people’s attitudes about raising taxes to support the building of a new school were related to whether they had children in school. One hundred voters with children in the ninth grade were asked whether they supported the idea of higher taxes for the purposes of building a new high school. Five years later they were asked the same question. Explain why this would be classified as a longitudinal study.

2. (6 pts) In an effort to estimate the percentage of US voters that approved of G.W. Bush’s handling of the spy-plane incident, a pollster surveyed 1900 such voters and found that 78 percent of the 1900 approved of G.W. Bush’s actions.
   (a) Describe the population of interest. Is the size of the population given, if so what is it?
   
   (b) Describe the parameter of interest. Is its value known, if so what is it?
   
   (c) Describe the statistic of interest. Is its value known, if so what is it?

3. (4 points) (True or False)
   (a) _____ Given a set of \( n \) ranked data values, the position of the median is the \( \frac{n}{2} \)th place.
   
   (b) _____ The standard deviation of data closely bunched around its mean is smaller than the standard deviation of data that is much more dispersed.
   
   (c) _____ The median for a finite collection of data with an odd number of data is always one of the data values.
   
   (d) _____ The mean is not a resistant measure because it is influenced by extreme values.
4. The following represent scores of a group of 16 students on History and English tests.

Scores on English Test
86  86  87  87  88  89  90  91  92  93  94  95  95  96  99 100

Scores on History test
71  73  76  77  78  85  87  88  90  91  92  93  93  97  98 100

(a) (4 pts) Construct a side by side stem and leaf plot splitting stems 7,8,9 into two parts with leaves 0–4 on one part and 5–9 on the other part?

(b) (2 pts) Find the percentile rank of the score of 89 on the English Test.

c) (2 pts) Find the first quartile $Q_1$ for the History test scores.

d) (2 pts) Which test scores appear to have a larger variance? Don’t compute the variances, but explain why you think this is the case.

e) (2 pts) Given that $\sum x = 1468$ and $\sum x^2 = 134992$ for the English Test scores, find the sample standard deviation for the English test scores.
5. Consider the following data of 27 numbers.

   40  42  58  60  61  62  68  70  71  72  73  75  77  79
   80  82  83  83  83  84  84  89  92  92  93  100 109

(a) (2 pts) Find the median of the data.

(b) (6 pts) Given that $Q_1 = 68$, and $Q_3 = 84$ find the IQR and construct both a boxplot and a modified boxplot for the data.

(c) (4 pts) Construct a frequency histogram for the data where the first class has limits 40–54, be sure to list the class boundaries.
A farmer is interested in the relationship between the amount of fertilizer in hundreds of pounds (x) and the number of soybeans produced in bushels per acre (y). The farmer obtained the following data.

<table>
<thead>
<tr>
<th>Amount of Fertilizer (x)</th>
<th>1.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.0</th>
<th>3.4</th>
<th>4.0</th>
<th>4.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bushels per Acre (y)</td>
<td>25</td>
<td>32</td>
<td>35</td>
<td>32</td>
<td>35</td>
<td>39</td>
<td>41</td>
</tr>
</tbody>
</table>

For this data: \( \sum x = 21.4, \sum x^2 = 73.06, \sum y = 239, \sum y^2 = 8325, \sum xy = 765.5. \)

(a) (4 pts) Find the equation of the least squares regression line.

(b) (2 pts) Use the regression line to predict how many pounds of fertilizer should be used to produce 40 bushels of soybeans per acre.

(c) (1 pt) From just looking at the data, do you expect a positive or negative correlation coefficient? Explain.

(d) (3 pts) Find the correlation coefficient for this data and explain what it means concerning the goodness of the linear fit.
Formulas and Scratch Paper

\[ s^2 = \frac{\sum(x - \bar{x})^2}{n - 1} = \frac{\sum x^2 - \left(\frac{\sum x}{n}\right)^2}{n - 1} = \frac{n(\sum x^2) - (\sum x)^2}{n(n - 1)} \]

\[ \sigma^2 = \frac{\sum(x - \mu)^2}{N} = \frac{\sum x^2 - \left(\frac{\sum x}{N}\right)^2}{N} = \frac{N(\sum x^2) - (\sum x)^2}{N^2} \]

\[ y = mx + b \text{ where } m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \text{ and } b = \bar{y} - m\bar{x}. \]

\[ r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \]