Calculators and one 8.5" by 11" sheet of handwritten notes allowed. Show all work and answers clearly in the space provided. Questions 1-6 are worth 15 points each; question 7 is 10 points. There are 100 points possible.

1. The values below are snow depths (in centimeters) measured at different locations.

\[19 \ 18 \ 12 \ 25 \ 22 \ 8 \ 8 \ 16 \ 8 \ 12 \ 18 \ 19 \ 22 \ 25\]

a. Find the mean and median.

\[\bar{x} = \frac{\sum x}{n} = \frac{128}{8} = 16\]

\[\text{median} = \frac{16 + 18}{2} = 17\]

b. Calculate the standard deviation using the formula \(s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}\). Show work.

\[
\begin{array}{c|cc}
X & X - \bar{x} & (X - \bar{x})^2 \\
8 & 8 - 16 = -8 & 64 \\
8 & -8 & 64 \\
12 & -4 & 16 \\
14 & 0 & 0 \\
18 & 2 & 4 \\
19 & 3 & 9 \\
22 & 6 & 36 \\
25 & 9 & 81 \\
\end{array}
\]

\[s = \sqrt{\frac{274}{8-1}} = \sqrt{\frac{274}{7}} = 6.26\]

c. Calculate \(Q_1\)

\[\text{position} = \frac{1}{4} (n+1) = \frac{1}{4} (9) = 2.25\]

between 8 and 12

\[Q_1 = 8 + 0.25(12 - 8) = 8 + 1 = 9\]
2. An experiment is conducted to determine which of two diets is more effective. The subjects are randomly assigned to Diet 1 or Diet 2. After six months on the assigned diet, the weight loss of each subject is measured. The boxplots below summarize the resulting data.

![Boxplots](image)

a. Use the boxplots to estimate the median weight loss for Diet 1 and Diet 2.
   \[ m_1 = Q_2, \quad m_2 = 12 \]

b. Which diet group had the largest interquartile range (IQR)? Estimate the IQR for this group. Explain what this IQR means in simple English (pretend you are explaining it to someone who has never taken statistics).
   - Diet 2 has largest IQR.
   - IQR = \( Q_3 - Q_1 \approx 17 - 5 = 12 \)
   - The IQR is the spread of the middle 50% of the data.

c. Calculate the upper fence for the diet group from part (b). What is the upper fence used for?
   \[ Q_3 + 1.5 \times \text{IQR} = 17 + 1.5(12) = 35 \]
   - Upper fence is used to identify outliers that are large.

d. Do the boxplots indicate symmetric or skewed data?
   - Since the median is not in the middle of the box, data are skewed.
3. The histogram and summary statistics below are for a data set consisting of the wealth in billions of dollars of 233 billionaires.

<table>
<thead>
<tr>
<th>Column</th>
<th>n</th>
<th>Mean</th>
<th>Variance</th>
<th>Std. Dev.</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>wealth</td>
<td>233</td>
<td>2.681545</td>
<td>11.014701</td>
<td>3.3188403</td>
<td>37</td>
<td>1.3</td>
<td>3</td>
</tr>
</tbody>
</table>

![Histogram of wealth data](image)

a. Is there skewness in this data set? If so, toward what values?
   - Skewed toward higher values (right)

b. The empirical rule states that approximately 68% of data fall within one standard deviation of the mean.

i. Calculate mean +/- one standard deviation for this data. (OK to round to the nearest tenths.)
   \[ 2.7 \pm 3.3 \]  
   \[ (-0.6, 6.0) \]

ii. Approximately what percent of the data points fall in this interval? (Hint: The height of the bar at 37 represents one person.)
   \[ \frac{125 + 70 + 25}{233} = 0.42 \]  
   \[ 42\% \]

iii. Is your answer to (ii) consistent with the empirical rule? If not, does this mean the empirical rule is not valid? Explain.
   - No. Empirical Rule says about 68% of the data are within one standard dev. of the mean, we have about 42% within the standard deviation.

c. This data set contains two outliers – one person at 24 and one at 37 billion dollars. Suppose these two individuals had a fit of generosity and each gave away all but 15 billion dollars of their wealth. Would this change the standard deviation? If so, higher or lower?
   - If the data would be less spread out about the mean so the standard deviation would be lower.
4. There are three finalists in a beauty pageant: Flora, Sabrina and Janet. Assume the contestants are equally beautiful and talented so the judges randomly choose who will win first, second and third place.

a. List the sample space for this experiment. (Hint: Tree diagram)

Calculate the probability that:

b. First prize goes to Flora, second to Sabrina and third to Janet.

\[ \frac{8}{216} = \frac{1}{27} \]

c. Flora wins first prize.

\[ \frac{8}{216} \]

d. Flora wins second or third prize.

\[ 1 - \frac{1}{3} = \frac{2}{3} \]

e. Janet wins third prize.

\[ \frac{2}{6} = \frac{1}{3} \]

5. Workers at a business are classified according to religion and gender. The results are shown in the table below.

<table>
<thead>
<tr>
<th>Religion</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Catholic</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>Jewish</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

A worker is randomly selected from this business. What is the probability the worker is

a. Catholic

\[ \frac{75}{215} = \frac{15}{43} \]

b. Catholic, given the worker is a female

\[ \frac{30}{88} \]

c. Jewish and Male

\[ \frac{15}{215} \]

d. Jewish or Male

\[ \frac{127 + 30}{215} = \frac{157}{215} \]
e. Are the events Jewish and Male mutually exclusive? Why or why not?

No. Like are 45 subjects that are Jewish and Male. So these events can occur at the same time.

f. Male, given the worker is Jewish

45/75

6. Consider the bivariate data in the table below for this problem. X = price of a widget in dollars, Y = number of widgets sold (in thousands).

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

\[ \Sigma xy = 25 \]

a. Calculate the correlation coefficient for this data. Hint: The standard deviations of x and y are 1.29 and 1.41, respectively.

\[ s_{xy} = \frac{\Sigma xy - \frac{1}{n} \Sigma x \Sigma y}{n-1} = \frac{25 - \frac{1}{4} (10)(12)}{4-1} = \frac{25-30}{3} = -\frac{5}{3} \]

\[ r = \frac{s_{xy}}{s_x s_y} = \frac{-5/3}{1.29(1.41)} = -0.92 \]

b. Calculate the best fit regression line.

\[ b = \frac{s_{xy}}{s_x^2} = \frac{-5/3}{1.29^2} = -1.07 \]

\[ a = \bar{y} - b \bar{x} = (\frac{12}{4}) - (-1.07)(\frac{12}{4}) = 3 + 2.5 = 5.5 \]

\[ y = 5.5 - x \]

2. c. Use the line in part (b) to predict the number of widgets sold when the price is 2.50.

\[ y = 5.5 - 2.50 = 3 \text{ thousand} \]

2. d. Interpret the slope calculated in part (b). Give a specific interpretation involving price of widgets and sales.

For every dollar increase in widget price, sales decline by about 1000 units.
a. Match the correlation coefficient with the appropriate plot. Note there are more correlation coefficients than plots so some will not be used: -1.00, -0.70, 0, 0.90, 1.00, 1.10.

![Plots with correlation coefficients](image)

b. Which of the above correlation coefficients would you prefer if you were interested in accurate prediction of y for a given value of x?

Plot 4