2.22

a. The scatter shown below plots the five measurements along the horizontal axis. Since there are two “1”s, the corresponding dots are placed one above the others. The approximate center of the data appears to be around 2.

![Figure 2.18](image)

b. The mean is the sum of the measurements divided by the number of measurements, or

\[
\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{0 + 5 + 1 + 1 + 3}{5} = \frac{10}{5} = 2
\]

To calculate the median, the observations are first ranked from smallest to largest: 0, 1, 1, 3, 5. Then, since \( n = 5 \) is odd, the median is the \( (n+1)/2 = 3^{rd} \) measurement, or \( median = 1 \). The mode is the measurement occurring most frequently, or \( mode = 1 \).

c. The three measures in the part b, are located on the scatterplot. Since the median and mode are to the left of the mean, we conclude that the measurements are skewed to the right.

2.26

a. There appear to be two companies with unusually high earnings per share ($1.28 and $1.64). Hence, the data may be skewed to the right.

b. \( Mean: \ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{11.41}{20} = 0.5705 \)

Median: To find the median, the observations are ranked from smallest to largest:

\[0.01, 0.21, 0.29, 0.29, 0.29, 0.32, 0.33, 0.33, 0.43, 0.44, \]
\[0.54, 0.56, 0.56, 0.62, 0.72, \]
Since \( n = 20 \) is even, the median is the average of the 10\(^{th}\) and 11\(^{th}\) observations, or \[ \text{median} = \frac{0.44 + 0.54}{2} = 0.49 \]

**Mode:**
The mode is the observation that occurs most frequently, or \( \text{mode} = 0.29 \)

c. The relative frequency histogram generated is shown Figure 2.20, with the mean, median, and mode located along the horizontal axis. Note that the mean is larger than the median, indicating that the data is skewed to the right.

\[
\begin{align*}
\sum_{i=1}^{n} x_i &= 17 \\
n &= 8 \\
\bar{x} &= \frac{\sum_{i=1}^{n} x_i}{n} = \frac{17}{8} = 2.125
\end{align*}
\]

b. Create a table of differences from the sample mean \((x_i - \bar{x})\) and their squares \((x_i - \bar{x})^2\)

\[
\begin{array}{ccc}
x_i & (x_i - \bar{x}) & (x_i - \bar{x})^2 \\
4 & 1.875 & 3.515625 \\
1 & -1.125 & 1.265625 \\
3 & 0.875 & 0.765725 \\
1 & -1.125 & 1.265625 \\
3 & 0.875 & 0.765725 \\
1 & -1.125 & 1.265625 \\
2 & -0.125 & 0.015625 \\
2 & -0.125 & 0.015625 \\
\text{Total} & 0 & 8.875
\end{array}
\]
Then, 
\[ s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{8.875}{7} = 1.268 \implies s = \sqrt{1.268} = 1.126 \]

c. Using the computational formula, we have:

\[
\begin{align*}
\sum_{i=1}^{n} x_i^2 &= \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} - \frac{45 - 289}{8} = 1.268 \implies s = \sqrt{1.268} = 1.126
\end{align*}
\]

The results of part a and b are identical.

2.41

(i) mean = 7.729
median = 7
mode = 8

(ii) We can expect it to be nearly symmetric

(a) We choose to use twelve classes of length 1.0. The tally and relative frequency histogram follow.

<table>
<thead>
<tr>
<th>Class i</th>
<th>Class Boundaries</th>
<th>Tally</th>
<th>( f_i )</th>
<th>Relative Frequency, ( f_i/n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.5 - 2.5</td>
<td>I</td>
<td>1</td>
<td>1.0/70.0</td>
</tr>
<tr>
<td>2</td>
<td>2.5 - 3.5</td>
<td>I</td>
<td>1</td>
<td>1.0/70.0</td>
</tr>
<tr>
<td>3</td>
<td>3.5 - 4.5</td>
<td>III</td>
<td>3</td>
<td>3.0/70.0</td>
</tr>
<tr>
<td>4</td>
<td>4.5 - 5.5</td>
<td>III</td>
<td>5</td>
<td>5.0/70.0</td>
</tr>
<tr>
<td>5</td>
<td>5.5 - 6.5</td>
<td>III</td>
<td>5</td>
<td>5.0/70.0</td>
</tr>
<tr>
<td>6</td>
<td>6.5 - 7.5</td>
<td>12</td>
<td>12.0/70.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7.5 - 8.5</td>
<td>18</td>
<td>18.0/70.0</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8.5 - 9.5</td>
<td>15</td>
<td>15.0/70.0</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9.5 - 10.5</td>
<td>6</td>
<td>6.0/70.0</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10.5 - 11.5</td>
<td>3</td>
<td>3.0/70.0</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11.5 - 12.5</td>
<td>0</td>
<td>0/70.0</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>12.5 - 13.5</td>
<td>1</td>
<td>1.0/70.0</td>
<td></td>
</tr>
</tbody>
</table>
(b) The sample mean is \( \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{541}{70} = 7.729 \)

\[ \text{x} \]

Using the definition of a percentile, 90% of the students taking the test scored lower than you: only 10% scored higher.

The ordered data are 12, 18, 22, 23, 24, 25, 25, 26, 26, 27, 28

(i) The \textit{minimum} is 12, and the \textit{maximum} is 28.
(ii) The \textit{median} is the 6th observation = 25.

Note that there are two alternative ways to determine the 1st quartile \( Q_1 \) and the 3rd quartile \( Q_3 \), which give approximately the same answers. Any of them will be considered correct.

\textit{Way 1 (from the lecture notes of Prof. Wang):}

The position of \( Q_1 \) is \( 0.25(n + 1) = 0.25(12) = 3 \). That is, \( Q_1 = 22 \)

The position of \( Q_3 \) is \( 0.75(n + 1) = 0.75(12) = 9 \). That is, \( Q_3 = 26 \)

\textit{Way 2:}

The first quartile is the median of the lower half of the observations, i.e., \( Q_1 = 22 \)

The third quartile is the median of the upper half of the observations, i.e., \( Q_3 = 26 \)

The boxplot is given in the figure below:
To check for outliers, we compute (using the results from the Way 1 computations):

\[ Q_1 - 1.5 \times IQR = 22 - 1.5(4) = 16 \quad \text{and} \quad Q_3 + 1.5 \times IQR = 26 + 1.5(4) = 32 \]

Clearly, there is one observation below 16 and it is 12. There are no observations exceeding 32. Therefore, the only outlier is 12.

3.1
This experiment involves tossing a single die and observing the outcome. The sample space for this experiment consists of the following outcomes:

- E1: observe a 1
- E2: observe a 2
- E3: observe a 3
- E4: observe a 4
- E5: observe a 5
- E6: observe a 6

We define the events a through F in the following way:

- A = (E2)
- B = (E1, E3, E5)
- C = (E1, E2, E3)
- D = contains no simple events
- E = (E1, E2, E3, E5)
- F = (E2)

To find the probability of an event, we sum the probabilities assigned to the outcomes in that event. Since the outcomes \( E_i, \ i=1,2,\ldots,6 \) are equally likely, \( P(E_i) = 1/6 \), so that

\[
\begin{align*}
P(A) &= 1/6 \\
P(B) &= 3/6 = 1/2 \\
P(C) &= 3/6 = 1/2 \\
P(D) &= 0 \\
P(E) &= 4/6 = 2/3 \\
P(F) &= P(E2) = 1/6
\end{align*}
\]
The assignment of probabilities is valid in this experiment, since each probability lies between 0 and 1 and since the sum of the four probabilities equal 1.

\[ 3.3 \]
In order to have a valid assignment of probabilities, probabilities must some to 1: \( \sum P(E_i) = 1 \)
Since \( P(E_1) + P(E_2) + P(E_3) + P(E_4) = 4(.15) = .6 \), we must have \( P(E_5) = 1 -.6 = .4 \)

\[ 3.7 \]
(a) The tree diagram is similar to Fig 3.1 in the text

(b) It is required that \( \sum P(E_i) = 1 \). Hence, \( P(E_2) = 1 -.01 -.09 -.81 = .09 \)

(c) The company will hit on at least one of the two drillings if it hits on the first, the second, or both. The associated simple events are \( E_1, E_2, \) and \( E_3 \) and
\[ P[\text{hits on at least one}] = P(E_1) + P(E_2) + P(E_3) = .19 \]

\[ 3.10 \]
(a) Experiment: A taster tastes and ranks three varieties of tea, A,B, and C, according to preference.

(b) Simple events in \( S \) are in triplets form:
\[
\begin{align*}
E_1 &= (A,B,C) & E_2 &= (A,C,B) & E_3 &= (B,A,C) \\
E_4 &= (B,C,A) & E_5 &= (C,B,A) & E_6 &= (C,A,B)
\end{align*}
\]

Here the tea in the 1\(^{st}\) position is the most desirable, the tea in the 2\(^{nd}\) position is the next most desirable, and the tea in the 3\(^{rd}\) is the least desirable.
(c) Define the events 
\[ D = \text{variety A is ranked 1}^{\text{st}} \]
\[ F = \text{variety A is ranked 3}^{\text{rd}} \]

Then, \[ P(D) = P(E_1) + P(E_2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \]

The probability that A is least desirable is \[ P(F) = P(E_4) + P(E_5) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \]