PSTAT 5E
Quiz Homework 2
Solutions

2.25
\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{54.5}{7} = 7.786 \]
a. The ranked observations are: 7.4, 7.5, 7.7, 7.8, 7.9, 8.0, 8.2. Since the sample size \( n = 7 \) is odd, the median is the \((n+1)/2 = 4^{th}\) ranked observation, or \(\text{median} = 7.8\).
b. The consumer would more likely be interested in the company with the highest or maximum satisfaction rating.

2.28
The Wall Street Journal is probably referring to the average number of cubes used per glass measured for some population that they have chosen.

2.30
\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{91}{10} = 9.1 \]
a. To find the median, the observations are ranked from smallest to largest: 5, 6, 8, 9, 9, 10, 10, 10, 11, 13. Since \( n = 10 \) is even, the median is the average of the \( n / 2 = 5^{th}\) and \((n / 2) + 1 = 6^{th}\) ranked observations, or \(\text{median} = (9+10) / 2 = 9.5\).

b. Since the mean and the median are almost the same, the data is nearly symmetric. Therefore, either the mean or the median provides a good measure of the center of the distribution.

2.34
\[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12}{6} = 2 \]
a. The following table will be helpful in calculating the standard deviation \(s\):

<table>
<thead>
<tr>
<th>(x_i)</th>
<th>(x_i - \bar{x})</th>
<th>((x_i - \bar{x})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
\[
\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{10}{4} = 2.5 \Rightarrow s = \sqrt{2.5} = 1.581
\]

or, using the computational formula, we have:

\[
\frac{\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2}{n - 1} = \frac{30 - \frac{100}{5}}{4} = 2.5 \Rightarrow s = \sqrt{2.5} = 1.581
\]

\[2.57\]

Using the definition of percentiles given in the text, we can conclude that:

$20,000$ is the 60\textsuperscript{th} percentile, $30,000$ is the 78\textsuperscript{th} percentile, and $40,000$ is the 87\textsuperscript{th} percentile.

\[2.61\]

The ordered data are given in the text.

(i) The \textbf{minimum} of the whole data set is 9.2, and the \textbf{maximum} of the whole data set is 132.8.

(ii) The \textbf{median} is the average of the 25\textsuperscript{th} and the 26\textsuperscript{th} observations (its position is \(0.5(n + 1) = 25.5\)):

\[
\text{median} = \frac{14 + 16.2}{2} = 15.1.
\]

Note that there are two alternative ways to determine the 1\textsuperscript{st} quartile \(Q_1\) and the 3\textsuperscript{rd} 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 \(0.25(n + 1) = 0.25(51) = 12.75\). That is, we need to interpolate between the 12\textsuperscript{th} and the 13\textsuperscript{th} observations in the following way: \(Q_1 = 11.8 + 0.75(11.8 - 11.8) = 11.8\)

The position of \(Q_3\) is \(0.75(n + 1) = 0.75(51) = 38.25\). That is, we need to interpolate between the 38\textsuperscript{th} and the 39\textsuperscript{th} observations in the following way: \(Q_3 = 29.9 + 0.25(30.2 - 29.9) = 29.975 \approx 29.98\)

\textit{Way 2:}

The first quartile is the median of the lower half of the observations, i.e., the average of the 13\textsuperscript{th} and the 14\textsuperscript{th} observations: \(Q_1 = \frac{11.8 + 11.8}{2} = 11.8\)
The third quartile is the median of the upper half of the observations, i.e., the average of the 37th and 38th observations: $Q_3 = \frac{29.9 + 30.2}{2} = 30.05$

The boxplot, using the results from the Way 1 computations, is given in the figure below:

![Boxplot Image]

To check for outliers, we compute (using the results from the Way 1 computations):

$Q_1 - 1.5 \text{IQR} = 11.8 - 1.5(18.18) = -15.47$ and $Q_3 + 1.5 \text{IQR} = 29.98 + 1.5(18.18) = 57.25$

Clearly, there are no observations below -15.47. There are, however, there are six observations above 57.25. Those are the outliers: 57.4, 62.2, 65.1, 100.8, 103.5, and 132.8.

Note that 57.4, 62.2 and 65.1 are very close to the upper fence, so although we classify them as outliers, they won’t influence (skew) our sample statistics (like sample mean and sample standard deviation) as much as 100.8, 103.5, and 132.8 would.

2.64

a. “Ethnic origin” is a qualitative variable since a quality (ethnic origin) is measured.
b. “Score” is a quantitative variable since a numerical quantity (0-100) is measured.
c. “Type of establishment” is a qualitative variable since a category (Carl’s Jr., McDonalds or Burger King) is measured.
d. “Mercury concentration” is a quantitative variable since a numerical quantity is measured.
2.65
a. The number of new clients acquired by a law firm in a month is a discrete variable since it can take only the values 0, 1, 2…

b. The shelf life of a particular drug is a continuous variable, since it can be any of the infinite number of positive real values.

c. The weight of a railway carload of wheat is a continuous variable.

d. The velocity of a pitched baseball is a continuous variable.

e. This is a discrete variable since it can take only the values 0, 1, 2…

2.75
The answer can vary depending on the choice of class boundaries. As an example, we calculate the range to be $R = 3.7 - 1.8 = 1.9$. We could use ten classes of length .2, or we could use seven classes of length .3. Using the latter, we obtain the tabulation.

<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>1.75 - 2.05</td>
<td>111</td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>2.05 - 2.35</td>
<td>1111</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>2.35 - 2.65</td>
<td>1111111</td>
<td>7</td>
<td>0.28</td>
</tr>
<tr>
<td>4</td>
<td>2.65 - 2.95</td>
<td>1111</td>
<td>4</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>2.95 - 3.25</td>
<td>111</td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>6</td>
<td>3.25 - 3.55</td>
<td>111</td>
<td>3</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>3.55 - 3.85</td>
<td>1</td>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Here, the lowest point of division is chosen just the smallest observed measurement. The relative frequency histogram is shown in Figure 2.31.

3.4
It is given that $P(E_1) = P(E_2) = .15$ and $P(E_3) = .4$. Since the probabilities of the individual outcomes in an experiment sum to 1 ($\sum P(E_i) = 1$), we have:

(i) $P(E_4) + P(E_5) = 1 - .15 - .15 - .4 = .30$

(ii) $P(E_4) = 2 \times P(E_5)$
We have two equations with two unknown that can be solved simultaneously for $P(E_4)$ and $P(E_5)$. Substituting equation (ii) into equation (i), we have

$$2*P(E_5) + P(E_5) = .3$$
$$3*P(E_5) = .3, \text{ so that } P(E_5) = .1$$

Then, from (i), $P(E_4) + .1 = .3$ and $P(E_4) = .2$

3.6

a. There are thirty-eight simple events, each corresponding to a single outcome of the wheel’s spin. The thirty-eight simple events are indicated below.

- $E_1$: Observation a 1  …  $E_{37}$: Observation a 37
- $E_2$: Observation a 2  …  $E_{38}$: Observation a 38

b. Since any pocket is just as likely as any other, $P(E_i) = 1/38$

c. The event $A$ contains two simple events, $E_{37}$ and $E_{38}$. Then,

$$P(A) = P(E_{37}) + P(E_{38}) = 2 / 38 = 1 / 19.$$ 

d. Define event $B$ as the event that you win on a single spin. Since you have bet on the number 1 through 18, event $B$ contains eighteen simple events, $E_1$, $E_2$, ..., $E_{18}$. Then,

$$P(B) = P(E_1) + P(E_2) + ... + P(E_{18}) = 18 / 38 = 9 / 19$$ 

3.8

The two-way table given in the exercise shows the four possible simple events and their associated probabilities. For clarity, define the four simple events as follows:

- $E_1$: adult is male and lives in the suburbs
- $E_2$: adult is male and lives in the city
- $E_3$: adult is female and lives in the suburbs
- $E_4$: adult is female and lives in the city

a. $P$ [customer resides in suburbs] = $P(E_1) + P(E_3) = .17 + .67 = .84$

b. $P$ [customer is female and lives in the city] = $P(E_4) = .12$

c. $P$ [customer is male] = $P(E_1) + P(E_2) = .17 + .04 = .21$

3.9

Define following events

- $G$: stock market gains 100 points or more
- $L$: stock market loses 100 points or more
- $N$: stock market changes less than 10 points

The experiment consists of pairs of forecasts, each of which will be one of the three events defined above. A tree diagram will assist in finding these simple events.
b. The simple events in A are GG, GN, NG, GL, and LG.
c. The simple events in B are GG, LL, and NN.
d. Since each analyst is as likely to select any one of the three choices as any other, 
   \[ P(E) = \frac{1}{9} \text{ and } P(A) = \frac{5}{9}. \]
e. \[ P(B) = \frac{3}{9} = \frac{1}{3}. \]