CHAPTER 1
What Is Statistics?

1.1 a The population of interest is the population of measurements consisting of the number of a particular item that will be demanded by customers in a particular month, measured for all months, in the past and in the future.
b The manager can possibly obtain monthly demand data for past months, but it is impossible to obtain the future data.
c Although it is not possible to itemize the entire population, the manager may be able to use information about past demand for a single item to draw conclusions about possible demand for that item in the future.

1.2 a The population of interest is the population of measurements consisting of the appraisals of the land by all experienced appraisers who might be asked to appraise the land.
b Although the population might be quite large, it is possible to get appraisals from all existing appraisers, and to enumerate the population.
c Yes, the populations are different. Buyers might tend to underestimate the appraisal, sellers might overestimate, and the public at large might be inexperienced at appraisals (hence they would either underestimate or overestimate). Experienced appraisers should give more unbiased and accurate appraisals.
d More than one appraisal should be used in order to get a more accurate estimate of the "value of the land" as measured by the population of appraisals.

1.3 a The population consists of measurements on the variable of interest (top speed, range before recharging, time to accelerate from 0 to 60) for all Tropicas that have been manufactured to date, or that will be manufactured in the future.
b The figures given in the article were based on a sample taken from the population, since the population does not exist in fact, but is instead hypothetical (not all Tropicas have yet been manufactured).

1.4 a The population of interest consists of a series of measurements (yes or no) in answer to the question "Would you be interested in buying the Tropical?" taken on all residents (of driving age) in the state of California.
b The sample could be obtained using a list of registered drivers, perhaps supplied by the Department of Motor Vehicles. Once the drivers have been selected, their answers (yes or no) would be obtained by telephone or mail interviews.
c No. The residents of Los Angeles County, who reside in an urban area, would not necessarily be representative of all drivers in the state. Los Angeles County residents, many of whom are commuters, will be more interested in speed and range before recharging than would rural drivers in some of the eastern or central counties in California.

1.5 a The population of interest consists of a series of measurements (damaged or not damaged) taken for each pine tree in Yosemite National Park at a particular point in time (either 1985 or 1990).
b The percentages are probably estimated.
c One possible method is to divide the park into a number of grids and to sample trees randomly within each grid. If the trees are numbered by the park rangers using some method of categorization, this numbering system might be employed.

1
1.6 a The percentage return of mutual stock funds is a function of the economic conditions during the past year. The characterization of this population will vary, depending on the year under consideration by the student. The population does exist and can be enumerated for a given time period.

b The percentage return of mutual stock funds over the next twelve months is conceptual and cannot be enumerated.

1.7 a The population consists of shopper opinions (prefer or do not prefer background music) for all shoppers who might patronize the supermarket, now or in the future.

b It would not be possible to enumerate the entire population, since the identities of the future shoppers would not be known to the sampler.

c No, but it will serve as an estimate of the population percentage.

1.8 a The population consists of the valuation criteria of stocks of 8000 public companies as listed in Barron's magazine in November, 1990. The sample of valuation criteria for thirty-four stocks is drawn based on various valuation criteria under which a stock is considered "cheap.

b The sample is not random, but is specifically chosen to include the stocks which are believed to be "cheap."
The line graph charts the line corresponding to the consumer price index for total services over the three years of interest. The vertical axis measures the size of the index.
b) The figure below shows the two consumer price indexes (each studied differently), each measured for the three years of interest. The vertical axis still measures the size of the index.

![Graph showing consumer price indexes](image)

**Figure 2.5**

Using the line graphs in Figures 2.4 and 2.5, we see that the "Total Services" index and the "Physicians Fees" index behaved similarly over the period, both increasing dramatically, but at the same rate. The "Hospital Room Charges" index also increased dramatically over the period, but the rate of increase was greater from 1982 to 1992 than the rate of increase for the other two indexes.

2.6 This is similar to Exercise 2.3. The pie chart and bar graph are shown below.

![Pie chart and bar graph](image)

**Figure 2.6**

Since there are only three categories for this qualitative variable, either method of presentation is probably equally effective. The bar graph is easier to construct.

**Figure 2.7**
This is similar to Exercise 2.9. The range of the data is 32 - 0 = 32. We choose to use eleven class intervals of length 3 (32/11 = 2.9, which, when rounded to the next largest integer, is 3). The subintervals 0.5 to 2.5, 2.5 to 5.5, 5.5 to 8.5, and so on, are convenient and the tally is shown below.

<table>
<thead>
<tr>
<th>Class</th>
<th>Boundaries</th>
<th>Tally</th>
<th>( f_i )</th>
<th>Relative Frequency, ( f_i/n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.5 - 2.5</td>
<td>I I I I I I I I I I</td>
<td>14</td>
<td>14/50</td>
</tr>
<tr>
<td>2</td>
<td>2.5 - 5.5</td>
<td>I I I I I I I I I</td>
<td>10</td>
<td>10/50</td>
</tr>
<tr>
<td>3</td>
<td>5.5 - 8.5</td>
<td>I I I I I I I I I</td>
<td>9</td>
<td>9/50</td>
</tr>
<tr>
<td>4</td>
<td>8.5 - 11.5</td>
<td>I I I I I I I I I</td>
<td>4</td>
<td>4/50</td>
</tr>
<tr>
<td>5</td>
<td>11.5 - 14.5</td>
<td>I I I I I I I I I</td>
<td>4</td>
<td>4/50</td>
</tr>
<tr>
<td>6</td>
<td>14.5 - 17.5</td>
<td>I I I I I I I I I</td>
<td>4</td>
<td>4/50</td>
</tr>
<tr>
<td>7</td>
<td>17.5 - 20.5</td>
<td>I I I I I I I I I</td>
<td>4</td>
<td>4/50</td>
</tr>
<tr>
<td>8</td>
<td>20.5 - 23.5</td>
<td>I I I I I I I I I</td>
<td>2</td>
<td>2/50</td>
</tr>
<tr>
<td>9</td>
<td>23.5 - 26.5</td>
<td>I I I I I I I I I</td>
<td>2</td>
<td>2/50</td>
</tr>
<tr>
<td>10</td>
<td>26.5 - 29.5</td>
<td>I I I I I I I I I</td>
<td>0</td>
<td>0/50</td>
</tr>
<tr>
<td>11</td>
<td>29.5 - 32.5</td>
<td>I I I I I I I I I</td>
<td>1</td>
<td>1/50</td>
</tr>
</tbody>
</table>

\( ^a \) The relative frequency histogram is shown in Figure 2.14.
Looking at the data, we see that thirty-six commercial banks and/or lending institutions granted ten or fewer loans. Therefore, the fraction is $36/50 = 0.72$. 
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.

\[ \bar{x} = \frac{\sum x_i}{n} = \frac{11.41}{20} = .5705 \]

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

\[ .10 .21 .29 .29 .29 .32 .33 .43 .44 .54 .56 .62 .62 .72 .72 .84 .89 1.28 1.64 \]

Since \( n = 20 \) is even, the median is the average of the \( \frac{n}{2} = 10 \)th and \( \frac{n}{2} + 1 = 11 \)th ranked observations, or

\[ m = (.44 + .54)/2 = .49 \]

The mode is the observation that occurs most frequently, or mode = .29.

c The relative frequency histogram generated using Minitab is shown in 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.
a \[ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{787}{10} = 78.7 \]

b \[ z = \frac{\sum_{i=1}^{n} z_i}{n} = \frac{2731}{10} = 273.1 \]

e The average overall value rating and the average estimated street price are not very useful, since a consumer would generally pick his preferred brand, using a combination of high value rating and/or low estimated street price. Once this choice is made, the consumer could then use the averages to compare his choice to the overall market.
a. The distribution of weekly wages is probably symmetric about the mean, with some higher and some lower wages.

b. According to the Empirical Rule, approximately 95% of the measurements will fall in the interval \( \mu \pm 2\sigma = 364 \pm 52 \) or 312 to 416. Hence, approximately 5% of the measurements will fall outside this interval. Because of the symmetry of a mound-shaped distribution, we can infer that half of 5%, or 2.5%, will be less than 312.

c. According to the Empirical Rule, almost all of the measurements will fall in the interval \( \mu \pm 3\sigma = 364 \pm 78 \) or 286 to 442. Hence, almost no contractors pay their workers less than 286. You could justly be accused of underpaying your workers.