1.6 Stealers, study components. In a study of the relationship between socio-economic class and unethical behavior, 129 University of California undergraduates at Berkeley were asked to identify themselves as having low or high social-class by comparing themselves to others with the most (least) money, most (least) education, and most (least) respected jobs. They were also presented with a jar of individually wrapped candies and informed that the candies were for children in a nearby laboratory, but that they could take some if they wanted. After completing some unrelated tasks, participants reported the number of candies they had taken. It was found that those who were identified as upper-class took more candy than others. Identify:

(a) the cases,
(b) the variables and their types, and
(c) the main research question

in this study.

<table>
<thead>
<tr>
<th>variable</th>
<th>description/type</th>
</tr>
</thead>
<tbody>
<tr>
<td>social class</td>
<td>ordinal</td>
</tr>
<tr>
<td>money</td>
<td>ordinal</td>
</tr>
<tr>
<td>education</td>
<td>ordinal</td>
</tr>
<tr>
<td>respected job</td>
<td>ordinal</td>
</tr>
<tr>
<td>candies</td>
<td>discrete</td>
</tr>
</tbody>
</table>

c) Does an individual's socio-economic status determine how ethical their behavior is?

1.14 Cats on YouTube. Suppose you want to estimate the percentage of videos on YouTube that are cat videos. It is impossible for you to watch all videos on YouTube so you use a random video picker to select 1000 videos for you. You find that 2% of these videos are cat videos. Determine which of the following is an observation, a variable, a sample statistic, or a population parameter.

(a) Percentage of all videos on YouTube that are cat videos.
(b) 2%.
(c) A video in your sample.
(d) Whether or not a video is a cat video.

a) population parameter
b) sample statistic
c) observation
d) variable
1.38 Mammal life spans. Data were collected on life spans (in years) and gestation lengths (in days) for 62 mammals. A scatterplot of life span versus length of gestation is shown below.

(a) What type of an association is apparent between life span and length of gestation?

(b) What type of an association would you expect to see if the axes of the plot were reversed, i.e., if we plotted length of gestation versus life span?

(c) Are life span and length of gestation independent? Explain your reasoning.

a) positive association (linear)

b) still positive association (and still linear w/ diff slope and intercept)

c) No as the life span seems to depend on the length of the gestation period with a longer gestation period corresponding to a higher life span

1.50 Mix-and-match. Describe the distribution in the histograms below and match them to the box plots.

(a) unimodal and symmetrical so we expect a symmetrical boxplot with maybe a few outliers - (2)

b) uniform with no modes so we expect a symmetrical boxplot with no outliers - (3)

c) unimodal skewed to the right so we can expect a lot of outliers at the higher end of the boxplot - (1)
1.58 Exam scores. The average on a history exam (scored out of 100 points) was 85, with a standard deviation of 15. Is the distribution of the scores on this exam symmetric? If not, what shape would you expect this distribution to have? Explain your reasoning.

Most likely, it isn’t symmetric but rather skewed to the left. In a symmetric distribution, the mean ± std dev defines the range containing about 66% of the values. However, 85 ± 15 = 100 and we can’t have grades greater than 100 suggesting outliers in the lower (left) end of the histogram are leading to a higher std. dev.

1.66 Views on immigration. 910 randomly sampled registered voters from Tampa, FL were asked if they thought workers who have illegally entered the US should be (i) allowed to keep their jobs and apply for US citizenship, (ii) allowed to keep their jobs as temporary guest workers but not allowed to apply for US citizenship, or (iii) lose their jobs and have to leave the country. The results of the survey by political ideology are shown below.\[70\]

<table>
<thead>
<tr>
<th>Response</th>
<th>Conservative</th>
<th>Moderate</th>
<th>Liberal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Apply for citizenship</td>
<td>57</td>
<td>120</td>
<td>101</td>
<td>278</td>
</tr>
<tr>
<td>(ii) Guest worker</td>
<td>121</td>
<td>113</td>
<td>28</td>
<td>262</td>
</tr>
<tr>
<td>(iii) Leave the country</td>
<td>179</td>
<td>126</td>
<td>45</td>
<td>350</td>
</tr>
<tr>
<td>(iv) Not sure</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>372</td>
<td>363</td>
<td>175</td>
<td>910</td>
</tr>
</tbody>
</table>

(a) What percent of these Tampa, FL voters identify themselves as conservatives?

(b) What percent of these Tampa, FL voters are in favor of the citizenship option?

(c) What percent of these Tampa, FL voters identify themselves as conservatives and are in favor of the citizenship option?

(d) What percent of these Tampa, FL voters who identify themselves as conservatives are also in favor of the citizenship option? What percent of moderates share this view? What percent of liberals share this view?

(e) Do political ideology and views on immigration appear to be independent? Explain your reasoning.

\[
a) \frac{total \, conservative}{total} = \frac{372}{910} = .41 \Rightarrow 41\%
\]

\[
b) \frac{total \, citizenship}{total} = \frac{278}{910} = .31 \Rightarrow 31\%
\]

\[
c) \frac{Citizenship \, and \, conservative}{total} = \frac{57}{910} = .06 \Rightarrow 6\%\text{ of }cons
\]

\[
d) \frac{Citizenship \, and \, conservative}{total \, conservative} = \frac{57}{372} = .15 \Rightarrow 15\% \text{ of } cons
\]

\[
\text{Citizenship and moderate} = \frac{120}{363} = .33 \Rightarrow 33\% \text{ of } moder
\]

\[
\text{Citizenship and liberal} = \frac{101}{175} = .58 \Rightarrow 58\% \text{ of } lib
\]

(e) If they were independent then we would expect the percentages in d) to all be nearly the same. So they are not independent.