Homework 1

Identify the population, the sample, and the variable of interest.

2. During the first day of classes, a professor collects the following information about his students:
name, age, gender, number of siblings, favorite color, current GPA, Zip Code of home address, and use of computers in class (Yes or No).
For each of those variables, state whether it is categorical (qualitative) or numerical (quantitative).

   Age: ____________________________

   Gender: _________________________

   Number of siblings: ______________

   Favorite color: __________________

   Current GPA: ____________________

   Zip Code: _______________________

   Use of computers: ________________

3. A statistics teacher wants to know how his students feel about an introductory statistics course.
He decides to administer a survey to a random sample of students taking the course.
He has several sampling designs to choose from.
(a) Name the sampling design in each case.

   - There are four ranks of students taking the class: freshman, sophomore, junior and senior.
     Randomly select 15 students from each class rank.

   - Randomly select a class rank (freshman, sophomore, junior, or senior) and every student in that class rank.

   - Each student has a nine-digit student number. Randomly choose 60 numbers.

   - Using the class roster, select every second student from the list, using a random start.
3. (b) Explain why the second plan suggested above, sampling all students from one class rank, might not be the best sampling plan.

3. (c) Name and describe the kind of selection bias that might be present if instead of randomly selecting students, the teacher simply:
- asks students to volunteer for the survey.
- gives the survey during class one day.

4. Students in an Intro Stats course were asked to describe their political stance as Liberal, Moderate, or Conservative. Here are the results:

<table>
<thead>
<tr>
<th></th>
<th>Liberal</th>
<th>Moderate</th>
<th>Conservative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>35</td>
<td>36</td>
<td>6</td>
<td>77</td>
</tr>
<tr>
<td>Male</td>
<td>50</td>
<td>44</td>
<td>21</td>
<td>115</td>
</tr>
<tr>
<td>Total</td>
<td>85</td>
<td>80</td>
<td>27</td>
<td>192</td>
</tr>
</tbody>
</table>

(a) What percent of the class is male?

(b) What percent of the class considers themselves to be Conservative?
(c) What percent of the males in the class consider themselves to be Conservative?

(d) What percent of all students in the class are males who consider themselves to be Conservative?

(e) Find the conditional distributions (percentages) of political views for the females.

(f) Find the conditional distributions (percentages) of political views for the males.

(g) Do the variables Politics and Gender appear to be independent? Explain.

5. Consider the following stem and leaf plot of data on the weights (in lbs) of 32 female students.

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>11</td>
<td>000</td>
</tr>
<tr>
<td>12</td>
<td>0457</td>
</tr>
<tr>
<td>13</td>
<td>0002455667889</td>
</tr>
<tr>
<td>14</td>
<td>024555</td>
</tr>
<tr>
<td>15</td>
<td>05</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
</tr>
</tbody>
</table>
```

Give the five-number summary of this data set.
6. In the Super Bowl, by how many points does the winning team outscore the losers? Here are the winning margins for the last 10 Super Bowl games: 3, 11, 12, 3, 4, 14, 6, 4, 13, 35.

(a) Give the five-number summary of this data set.

(b) Find the range and the inter quartile range.

(c) Draw the boxplot of this data set using the scale below.

(d) Are there any outliers in this data set? If so, which values?

(e) Find the mode (or modes) of this data set.

(f) Compute the mean of this data set \textit{by hand}, showing all your calculations.
6. (g) Compute the variance and the standard deviation of this data set by hand, showing all your calculations.

7. Below is a boxplot of the seal strengths of 100 chips bags.

(a) Approximately, what percentage of bags had seal strengths below 350?

(b) Approximately, what percentage of bags had seal strengths between 250 and 325?

(c) Circle the correct answer choice:
The distribution of seal strengths is skewed to the left / symmetric about its mean / skewed to the right.

(d) Identify the outlier by circling in on the boxplot and write down its approximate value.

(e) Find the approximate range of this data set.
8. For each of the following histograms, state whether it is unimodal or bimodal. If it is unimodal, also state whether it is symmetric, left-skewed, or right-skewed.

(a)

(b)

(c)

(d)
9. Refer back to Problem 6 (Super Bowl winning margins). Compute the z-score of each value in the data set.

10. A high school senior uses the Internet to get information on February temperatures in the town where he'll be going to college. He finds a Web site with some statistics, but they are given in degrees Celsius. The conversion formula is \[ ^\circ F = \frac{9}{5} ^\circ C + 32. \]
    Determine the Fahrenheit equivalents for the summary information below. Show your work.

    Maximum temperature = 11 °C = \[ \text{°F} \]

    Range = 33 °C = \[ \text{°F} \]

    Mean = 1 °C = \[ \text{°F} \]

    Standard deviation = 7 °C = \[ \text{°F} \]

    Median = 2 °C = \[ \text{°F} \]

    IQR = 16 °C = \[ \text{°F} \]
11. What percent of a standard Normal model is found in each region? Draw a picture first.

(a) $z > -2.05$

(b) $z < -0.33$

(c) $1.2 < z < 1.8$

(d) $|z| < 1.28$
12. In a standard Normal model, what value(s) of $z$ cut(s) off the region described? Remember to draw a picture first.

(a) the lowest 12%

(b) the highest 30%

(c) the highest 7%

(d) the middle 50%
13. Draw a picture for each of the following questions. Based on the Normal model $N(100,16)$ describing IQ scores, what percent of people’s IQs would you expect to be

(a) over 80?

(b) under 90?

(c) between 112 and 132?