1. Give a brief definition of the terms descriptive statistics and inferential statistics. (10 points)

2. Give a brief definition of the terms population and sample. (10 points)

3. Based on a study of 2121 children between the ages of one and four, researchers at the Medical College of Wisconsin concluded that there was an association between iron deficiency and the length of time that a child is bottle-fed (Milwaukee Journal Sentinel, November 26, 2005). Describe the sample and the population in this study. (10 points)

4. For following numerical attributes, state whether each is discrete or continuous. (8 points – 2 each)
   (a) The number of insufficient-funds checks received by a grocery store in a given month.
   (b) The amount by which 1-pound package of ground beef decreases in weight (because of moisture loss) before purchase.
   (c) The number of New York Yankees during a given year that will not play for the Yankees the following year.
   (d) The amount of Mountain Dew actually in a given 2-liter bottle.

5. Classify each of the following attributes as either categorical, discrete numerical, or continuous numerical. (6 points – 2 each)
   (a) Zip Code
   (b) Price of a text book.
   (c) State of birth in which an individual was born.

6. The article “So Close, Yet So Far: Predictors of Attrition in College Seniors” (Journal of College Student Development [1998]: 343 – 348) examined the reasons that college seniors leave their college programs before graduating. Forty-two college seniors at a large public university who dropped out before graduation were interviewed and asked the main reason for discontinuing enrollment at the university. (12 points)

<table>
<thead>
<tr>
<th>Reason for Leaving</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Academic Problems</td>
<td>7</td>
</tr>
<tr>
<td>(b) Poor Advising/Teaching</td>
<td>3</td>
</tr>
<tr>
<td>(c) Needed a Break</td>
<td>2</td>
</tr>
<tr>
<td>(d) Economic Reasons</td>
<td>11</td>
</tr>
<tr>
<td>(e) Family Responsibilities</td>
<td>4</td>
</tr>
<tr>
<td>(f) Switching Schools</td>
<td>9</td>
</tr>
<tr>
<td>(g) Personal Problems</td>
<td>3</td>
</tr>
<tr>
<td>(h) Other</td>
<td>3</td>
</tr>
</tbody>
</table>

   Build a bar chart for the frequency of each response, and also build a relative frequency table for this data.

7. The numbers below show the percentage of full time freshman in fall 1993 who had earned a college degree by August 1999 in nineteen Division I schools in Texas. Build a dot plot to represent this data. (10 points)

8. Give a brief definition of the terms selection bias, measurement bias, and non-response bias. (9 points)

9. According to a report on CNN (September 17, 1998) titled “Majority of U.S. Teens Are Not Sexually Active,” 52% of surveyed teenagers had never had sexual intercourse. A very large random sample of 16252 high school students was the source of this information. If the population of interest was all teenagers in the U.S., what type of sampling bias does this scenario have? Explain. (10 points)

10. “Attending Church Found Factor in Longer Life” was the title of an article that appeared in USA Today (August 9, 1999). Based on a nationally representative survey of 3617 Americans, the authors of the article concluded that “attending (church) services extends the life span about as much as moderate exercise.” Discuss the validity of this conclusion. Be specific by discussing the type of study that is being done here and what type(s) of conclusions can be drawn from such studies. (10 points)
Math 156–Sat: HW #1 - Solutions

1. Descriptive Statistics = methods for organizing and summarizing data
   Inferential Statistics = generalizing information from a sample to the population from which it was selected.

2. Population = the entire collection of individuals or objects about which information is desired
   Sample = any subset of the population

3. Population = all children between the ages of one and four that are bottle-fed
   Sample = the 2121 children that were part of the study

4. (a) discrete
   (b) continuous
   (c) discrete
   (d) continuous

5. (a) categorical
   (b) discrete numerical (only allowed 2 decimal places)
   (c) categorical

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>7</td>
<td>0.1667</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>0.0714</td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td>0.0476</td>
</tr>
<tr>
<td>d</td>
<td>11</td>
<td>0.2619</td>
</tr>
<tr>
<td>e</td>
<td>4</td>
<td>0.0952</td>
</tr>
<tr>
<td>f</td>
<td>9</td>
<td>0.2143</td>
</tr>
<tr>
<td>g</td>
<td>3</td>
<td>0.0714</td>
</tr>
<tr>
<td>h</td>
<td>3</td>
<td>0.0714</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>About 1</td>
</tr>
</tbody>
</table>

7. 

![Frequency Chart]

![Reason for Leaving Bar Chart]
8. *Selection Bias* = the method used to select a sample systematically excludes part of the population from being selected

*Measurement Bias* = the method of observation tends to produce values that systematically differ from the population in some way (i.e. an improperly calibrated scale)

*Nonresponse Bias* = responses are not actually obtained for all members included in the sample (i.e. a nonresponse is automatically assumed to be “no”)

9. The largest problem is that there is gross selection bias. The population of interest is *all* U.S. teens, but only high school students were surveyed. So teenagers not yet in high school, already out of high school, or just not going to high school for some reason have been systematically excluded.

10. This is the description of an observational study, and while correlation between going to church and living longer may be established, a cause and effect relationship between the two cannot be established. The authors are claiming a cause and effect relationship based on the study, and this is not valid.