Chapter 1: Data Analysis

Objectives: Students will:
- Identify the individuals and variables in a set of data
- Classify variables as categorical or quantitative
- Make, interpret and compare distributions using bar graphs for categorical data; boxplots, dotplots, stemplot, and histograms of quantitative data
- Identify what makes some graphs of categorical data misleading
- Calculate marginal and joint relative frequencies from a two-way table
- Calculate conditional relative frequencies from a two-way table
- Describe the nature of the association between two categorical
- Identify the shape of a distribution from a graph (roughly symmetric or skewed – right/left)
- Describe the overall pattern (shape, center, and variability) of a distribution and identify any major departures from the pattern (outliers)
- Calculate and interpret measures of center (mean, median, mode)
- Calculate and interpret measures of variability (IQR, standard deviation, range) for a distribution of quantitative data
- Explain how outliers and skewness affect measures of center and variability
- Identify outliers using the 1.5 x IQR rule

AP Outline Fit:

I. Exploring Data: Describing patterns and departures from patterns.
   A. Constructing and interpreting graphical displays of distributions of univariate data (dotplot, stemplot, histogram, cumulative frequency plot covered in chapter 2)
      1. Center and spread
      2. Clusters and gaps
      3. Outliers and other unusual features
      4. Shape
   B. Summarizing distributions of univariate data
      1. Measuring center: median, mean
      2. Measuring spread: range, interquartile range, standard deviation
      3. Measuring position: quartiles, percentiles, standardized scores (z-scores covered in chapter 2)
      4. Using boxplots
      5. The effect of changing units on summary measures (covered in chapter 2)
   C. Comparing distributions of univariate data (dotplots, back-to-back stemplots, parallel boxplots)
      1. Comparing center and spread: within group, between group variation
      2. Comparing clusters and gaps
      3. Comparing outliers and other unusual features
      4. Comparing shapes
   E. Exploring categorical data
      1. Frequency tables and bar charts
      2. Marginal and joint frequencies for two-way tables (covered in chapter 5)
      3. Conditional relative frequencies and association

Calculator Functions in this Chapter:

1. Using Lists for data
2. Graphing
   a. Scatter plots
   b. Histograms
   c. Box plots
3. Summary Statistics
   a. Mean
   b. Standard Deviation
   c. Five-Number Summary
Chapter 1: Data Analysis

What You Will Learn:

A. Displaying Distribution
   1. Make a stemplot of the distribution of a quantitative variable.
   2. Trim the numbers or split stems as needed to make an effective stemplot
   3. Make a histogram of the distribution of a quantitative variable
   4. Construct and interpret an ogive of a set of quantitative data

B. Inspecting Distributions (Quantitative Variables)
   1. Look for the overall pattern and for major deviations from the pattern
   2. Assess from a dotplot, stemplot, or histogram whether the shape of a distribution is roughly symmetric, distinctly skewed, or neither.
   3. Assess whether the distribution has one or more major modes
   4. Describe the overall pattern by giving numerical measures of center and spread in addition to a verbal description of shape
   5. Decide which measures of center and spread are more appropriate: the mean and standard deviation (especially for symmetric distributions) or the five-number summary (especially for skewed distributions)
   6. Recognize outliers

C. Measuring Center
   1. Find the mean, x-bar, of a set of observations
   2. Find the median M of a set of observations
   3. Understand that the median is more resistant (less affected by extreme observations) than the mean.
   4. Recognize that skewness in a distribution moves the mean away from the median toward the long fall (tail)

D. Measuring Spread
   1. Find the quartiles Q₁ and Q₃ for a set of data
   2. Give the five-number summary and draw a boxplot, assess center, spread, symmetry, and skewness from a boxplot.
   3. Determine outliers
   4. Using a calculator or software, find the standard deviation, s, for a set of observations
   5. Know the basic properties of s:
      a) s ≥ 0 always;
      b) s = 0 only when all observations are identical; s increases as the spread increases;
      c) s has the same units as the original measurements;
      d) s is increased by outliers or skewness

E. Comparing Distributions
   1. Use side-by-side bar graphs to compare distributions of categorical data
   2. Make back-to-back stemplots or side-by-side Boxplots to compare distributions of quantitative variables
   3. Write narrative comparisons of the shape, center, spread, and outliers for two or more quantitative distributions

F. Introduction to Statistical Software
   1. StatCrunch is a web-based product
   2. Works with Chromebooks
Chapter 1: Data Analysis

Section 1.0: Statistics: The Science and Art of Data

Objectives: Students will be able to:
- Identify the individuals and variables in a set of data
- Classify variables as categorical or quantitative

Vocabulary:
- Individual – an object described in a set of data; individuals can be people, animals or things
- Variable – an attribute that take on different values for different individuals
- Categorical variable – assigns labels that place each individual into a particular group, called a category
- Quantitative variable – takes number values that are quantities; counts or measurements
- Distribution – tells us what values the variable takes on and how often it takes these values
- Inference – using a sample of data to infer (to draw conclusions) about a larger group of data
- Statistics – the science and art of collecting, analyzing, and drawing conclusions from data

Key Concept:

Data Exploration:
- a) Begin by examining each variable by itself
- b) Study the relationship between the variables
- c) Use plots or graphs
- d) Add numerical summaries

Activity:

Hiring Discrimination on pg 6
- Read the scenario and answer the following question:
  Reverse discrimination: Yes or No

Break into pairs and follow the instructions in the activity
- One person work with the cards
- One person acts as a recorder of the data and fills in the following table:

<table>
<thead>
<tr>
<th>Captains</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 4</th>
<th>Trial 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Did your opinion change after the class’s data was displayed?

Examine the homework from last night:
- What variables on the information sheet were categorical? Quantitative?
- What variables on Mr. Starnes’s survey were categorical? Quantitative?

Summary:
- A data set contains information on a number of individuals
- Information is often values for one or more variables
- Variables can be categorical or quantitative
- Distribution of a variable describes what values it can take on and how often

Homework: pg 7-8, problems 1, 3, 5, 7, 8
Chapter 1: Data Analysis

Section 1.1: Analyzing Categorical Data

Objectives: Students will be able to:
- Make and interpret bar graphs for categorical data
- Identify what makes some graphs of categorical data misleading
- Calculate marginal and joint relative frequencies from a two-way table
- Calculate conditional relative frequencies from a two-way table
- Use bar graphs to compare distributions of categorical data
- Describe the nature of the association between two categorical

Vocabulary:
Association – two variables are associated if knowing the value of one variable helps us predict the value of the other
Bar graph – shows each category as a bar. The heights of the bars show the category frequencies or relative frequencies
Conditional relative frequency – gives the percentage or proportion of individuals that have a specific value for one categorical variable among individuals who share the same value of another categorical variable (the condition)
Frequency table – the number of individuals having each value
Joint relative frequency – gives the percent of proportion of individuals that have a specific value for one categorical variable and a specific value for another categorical variable
Marginal relative frequency – the percent or proportion of individuals that have a specific value for one categorical variable
Pie chart – shows each category as a slice of the “pie.” The areas of the slices are proportional to the category frequencies or relative frequencies
Relative frequency table – shows proportion or percent of individuals having each value
Segmented bar graph – displays distribution of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category
Side-by-side bar graph – displays distribution of a categorical variable for each value of another categorical variable
Two-way table – a table of counts that summarizes data on the relationship between two categorical variables for some group of individuals

Key Concepts:

Categorical Charts

Note: association does not imply causation!!!
Chapter 1: Data Analysis

Ex. 1  Construct a bar graph and a pie chart from the following data:

<table>
<thead>
<tr>
<th>Format</th>
<th>Nr of Stations</th>
<th>Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult contemporary</td>
<td>1,556</td>
<td>11.2</td>
</tr>
<tr>
<td>Adult standards</td>
<td>1,196</td>
<td>8.6</td>
</tr>
<tr>
<td>Contemporary Hits</td>
<td>569</td>
<td>4.1</td>
</tr>
<tr>
<td>Country</td>
<td>2,066</td>
<td>14.9</td>
</tr>
<tr>
<td>News/Talk/Info</td>
<td>2,179</td>
<td>15.7</td>
</tr>
<tr>
<td>Oldies</td>
<td>1,060</td>
<td>7.7</td>
</tr>
<tr>
<td>Religious</td>
<td>2,014</td>
<td>14.6</td>
</tr>
<tr>
<td>Rock</td>
<td>869</td>
<td>6.3</td>
</tr>
<tr>
<td>Spanish Language</td>
<td>750</td>
<td>5.4</td>
</tr>
<tr>
<td>Other formats</td>
<td>1,579</td>
<td>11.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>13,838</strong></td>
<td><strong>99.9</strong></td>
</tr>
</tbody>
</table>

Ex. 2  Below are the results of a survey of young adults by gender:

<table>
<thead>
<tr>
<th>Young adults by gender and chance of getting rich</th>
<th>Female</th>
<th>Male</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost no chance</td>
<td>96</td>
<td>98</td>
<td>194</td>
</tr>
<tr>
<td>Some chance, but probably not</td>
<td>426</td>
<td>286</td>
<td>712</td>
</tr>
<tr>
<td>A 50-50 chance</td>
<td>696</td>
<td>720</td>
<td>1416</td>
</tr>
<tr>
<td>A good chance</td>
<td>663</td>
<td>758</td>
<td>1421</td>
</tr>
<tr>
<td>Almost certain</td>
<td>486</td>
<td>597</td>
<td>1083</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2367</td>
<td>2459</td>
<td>4826</td>
</tr>
</tbody>
</table>

a) What are the variables described by this two-way table?

b) How many young adults were surveyed?

Ex. 3  Why can’t we get totals in the super-powers data?

**Homework:**  pg 24-30; probs 11, 16, 20, 25, 30, 38
Chapter 1: Data Analysis

Section 1.2: Displaying Quantitative Data with Graphs

Objectives: Students will be able to:
- Make and interpret dotplots, stemplot, and histograms of quantitative data
- Identify the shape of a distribution from a graph (roughly symmetric or skewed – right/left)
- Describe the overall pattern (shape, center, and variability) of a distribution and identify any major departures from the pattern (outliers)
- Compare distributions of quantitative data using dotplots, stemplot, and histograms

Vocabulary:
- Back-to-back stemplot – two distributions plotted with a common stem
- Bimodal – a graph with two peaks (modes)
- Dotplot – shows each data point as a dot above its location on a number line
- Histogram – shows each interval of values as a bar. Heights of the bars show the frequencies or relative frequencies of values in each interval
- Modes – major peaks in a distribution
- Multimodal – a graph which has more than two peaks (modes)
- Outlier – an observation that falls outside the pattern
- Skewed left – if the left side of the graph is longer than the right (forms a tail to the left)
- Skewed right – if the right side of the graph is longer than the left (forms a tail to the right)
- Stemplot – show each data value separated into two parts: a stem, which consists of all but the final digit and the leaf, the final digit
- Symmetric – if right side of the graph is an approximate mirror images of the left side
- Unimodal – a graph with a single peak (mode)

Key Concepts:

When describing a distribution, remember Statistical Opinions Can Vary! (Shape, Outliers, Center, Variability)

Histogram notes:
1. The choice of intervals in a histogram can affect the appearance (shape) of a distribution
2. Don’t confuse histograms and bar charts
3. Use percent or proportions instead of counts on the vertical axis when comparing distributions with different number of observations
4. Just because a graph looks nice doesn’t make it a meaningful display of data

**Dot plots:**
- Maintains the raw data, while histograms do not maintain the raw data
- Best used when the data sets are small

**Stem and Leaf plots:**
- Maintains the raw data, while histograms do not maintain the raw data
- Best used when the data sets are small
Chapter 1: Data Analysis

**Histograms and Bar Graphs:** Bar graphs have bars touching; histograms don’t

The number of classes, \( k \), to be constructed can be roughly approximated by

\[ k = \sqrt{\text{number of observations}} \]

To determine the width of a class use

\[ w = \frac{\text{maximum} - \text{minimum}}{k} \]

and always round up to the same decimal units as the original data.

**Other Charts:**

- **Relative Frequency Chart**
- **Pareto Chart**
- **Cumulative Frequency Chart**

Plot in the upper right corner is a Pareto chart. It is the same as the relative frequency chart; except the categories are in relative frequency order (from largest to smallest) from left to right. This graph came from the Total Quality Management (TQM) era in the middle to late 1980’s. The bottom chart is also known as an ogive (a confusing chart for students!).

**Cautions:**
- Label all axes and title all graphs
- Histogram rectangles touch each other; rectangles in bar graphs do not touch.
- Can’t have class widths that overlap
- Raw data can be retrieved from the stem-and-leaf plot; but a frequency distribution of histogram of continuous data summarizes the raw data
- Only quantitative data can be described as skewed left, skewed right or symmetric (uniform or bell-shaped)

**Frequency Distributions**

- Uniform
- Mound-like (Bell-Shaped)
- Bi-Modal Skewed Right (− tail)
- Skewed Left (− tail)
Chapter 1: Data Analysis

With the following data
a) Construct a stem graph (in example 1 do a back-to-back [comparative] stem plot)
   b) Construct a histogram

Ex. 1 The ages (measured by last birthday) of the employees of Dewey, Cheatum and Howe are listed below.
Office A: 22 31 21 49 26 42 42 30 28 31 39 39
Office B: 20 37 32 36 35 33 45 47 49 38 28 48

Ex. 2 Below are times obtained from a mail-order company's shipping records concerning time from receipt of order to delivery (in days) for items from their catalogue?

<table>
<thead>
<tr>
<th>3</th>
<th>7</th>
<th>10</th>
<th>5</th>
<th>14</th>
<th>12</th>
<th>6</th>
<th>2</th>
<th>9</th>
<th>22</th>
<th>25</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>7</td>
<td>12</td>
<td>10</td>
<td>22</td>
<td>23</td>
<td>14</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>27</td>
<td>31</td>
<td>13</td>
<td>21</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>10</td>
<td>19</td>
<td>12</td>
<td>11</td>
<td>8</td>
</tr>
</tbody>
</table>

Homework: pg 47-54: probs 45, 51, 62, 66, 79
Chapter 1: Data Analysis

Section 1.3: Describing Quantitative Data with Numbers

Objectives: Students will:
- Calculate and interpret measures of center (mean, median, mode)
- Calculate and interpret measures of variability (IQR, standard deviation, range) for a distribution of quantitative data
- Explain how outliers and skewness affect measures of center and variability
- Identify outliers using the 1.5 x IQR rule
- Make and interpret boxplots of quantitative data
- Use boxplots and numerical summaries to compare distributions of quantitative variables

Vocabulary:
- Boxplot – graphs the five number summary and any outliers
- Degrees of freedom – the number of independent pieces of information that are included in your measurement
- Five-number summary – the minimum, Q1, Median, Q3, maximum
- Interquartile range – the range of the middle 50% of the data; (IQR) = Q3 – Q1
- Mean – the average value (balance point); x-bar
- Median – the middle value (in an ordered list); M
- Mode – the most frequent data value
- Outlier – a data value that lies outside the interval [Q1 – 1.5 x IQR, Q3 + 1.5 x IQR]
- Quartile – multiples of 25th percentile (Q1 = 25th; Q2 = 50th or median; Q3 = 75th)
- Range – difference between the largest and smallest observations
- Resistant – a measure (statistic or parameter) that is not sensitive to the influence of extreme observations
- Standard Deviation – the square root of the variance
- Variance – the average of the squares of the deviations from the mean

Key Concepts:

<table>
<thead>
<tr>
<th>Measure of Central Tendency</th>
<th>Computation</th>
<th>Interpretation</th>
<th>When to use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \mu = \frac{\sum x_i}{N} )</td>
<td>Center of gravity</td>
<td>Data are quantitative and frequency distribution is roughly symmetric</td>
</tr>
<tr>
<td></td>
<td>( \bar{x} = \frac{\sum x_i}{n} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>Arrange data in ascending order and divide the data set into half</td>
<td>Divides into bottom 50% and top 50%</td>
<td>Data are quantitative and frequency distribution is skewed</td>
</tr>
<tr>
<td>Mode</td>
<td>Tally data to determine most frequent observation</td>
<td>Most frequent observation</td>
<td>Data are qualitative or the most frequent observation is the desired measure of central tendency</td>
</tr>
</tbody>
</table>

Center: The mean and the median are the most common measures of center
- If a distribution is perfectly symmetric, the mean and the median are the same
- The mean is not resistant to outliers
- The mode, the data value that occurs the most often, is a common measure of center for categorical data
- Use the mean on symmetric data and the median on skewed data or data with outliers

Spread: Standard deviation is the most common measure of spread. Range and IQR are also measures of spread.
Chapter 1: Data Analysis

Distribution Shape Based on Boxplots:

a. If the median is near the center of the box and each horizontal line is of approximately equal length, then the distribution is roughly symmetric
b. If the median is to the left of the center of the box or the right line is substantially longer than the left line, then the distribution is skewed right
c. If the median is to the right of the center of the box or the left line is substantially longer than the right line, then the distribution is skewed left

Remember identifying a distribution from boxplots or histograms is subjective!

Why Use a Boxplot?
A boxplot provides an alternative to a histogram, a dotplot, and a stem-and-leaf plot. Among the advantages of a boxplot over a histogram are ease of construction and convenient handling of outliers. In addition, the construction of a boxplot does not involve subjective judgments, as does a histogram. That is, two individuals will construct the same boxplot for a given set of data - which is not necessarily true of a histogram, because the number of classes and the class endpoints must be chosen. On the other hand, the boxplot lacks the details the histogram provides.

Dotplots and stemplots retain the identity of the individual observations; a boxplot does not. Many sets of data are more suitable for display as boxplots than as a stemplot. A boxplot as well as a stemplot are useful for making side-by-side comparisons.

Five-number summary

<table>
<thead>
<tr>
<th>Min</th>
<th>Q1</th>
<th>M</th>
<th>Q3</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>smallest value</td>
<td>First, Second and Third Quartiles (Second Quartile is the Median, M)</td>
<td>largest value</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Boxplot

Lower Fence

Smallest Data Value > Lower Fence (Min unless min is an outlier)

Upper Fence

Largest Data Value < Upper Fence (Max unless max is an outlier)

Outlier

Median Notes:
If the number of observations, n, is odd, then the median is the center observation in the ordered list
If the number of observations is even, then the median is the average of the two center observations

Mean Notes:
The mean is pulled toward the tail of the distribution, away from the median, when the distribution is skewed
Outliers can also pull the mean toward them
Chapter 1: Data Analysis

Example 1: Which of the following are resistant measures of central tendency:
Mean,
Median or
Mode?
Mean, Range
Median or Variance
Mode? Standard Deviation
IQR

Example 2: Given the following set of data:
70, 56, 48, 48, 53, 52, 66, 48, 36, 49, 28, 35, 58, 62, 45, 60, 38, 73, 45, 51,
56, 51, 46, 39, 56, 32, 44, 60, 51, 44, 63, 50, 46, 69, 53, 70, 33, 54, 55, 52
What is the mean?
What is the median?
What is the mode?
What is the shape of the distribution?
What is the range?
What is the variance?
What is the standard deviation?
What is the IQR?
What is the Q1?
What is the Q3?
What is the IQR?
What is the upper fence?
What is the lower fence?
Are there any outliers?

Example 3: Given the following types of data and sample sizes, list the measure of central tendency you would use and explain why?

<table>
<thead>
<tr>
<th>Sample of 50</th>
<th>Sample of 200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hair color</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>Parent’s Income</td>
<td></td>
</tr>
<tr>
<td>Number of Siblings</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>Does sample size affect your decision?</td>
<td></td>
</tr>
</tbody>
</table>
Example 4: Consumer Reports did a study of ice cream bars (sigh, only vanilla flavored) in their August 1989 issue. Twenty-seven bars having a taste-test rating of at least “fair” were listed, and calories per bar was included. Calories vary quite a bit partly because bars are not of uniform size. Just how many calories should an ice cream bar contain?

<table>
<thead>
<tr>
<th>Calories</th>
<th>342</th>
<th>377</th>
<th>319</th>
<th>353</th>
<th>295</th>
<th>234</th>
<th>294</th>
<th>286</th>
<th>377</th>
<th>182</th>
<th>310</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>439</td>
<td>111</td>
<td>201</td>
<td>182</td>
<td>197</td>
<td>209</td>
<td>147</td>
<td>190</td>
<td>151</td>
<td>131</td>
<td>151</td>
</tr>
</tbody>
</table>

a) Construct a boxplot of the data
b) Determine if there are any outliers.

Example 5: The weights of 20 randomly selected juniors at MSHS are recorded below:

<table>
<thead>
<tr>
<th>Weight</th>
<th>121</th>
<th>126</th>
<th>130</th>
<th>132</th>
<th>134</th>
<th>137</th>
<th>141</th>
<th>144</th>
<th>148</th>
<th>205</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>125</td>
<td>128</td>
<td>131</td>
<td>133</td>
<td>135</td>
<td>139</td>
<td>141</td>
<td>147</td>
<td>153</td>
<td>213</td>
</tr>
</tbody>
</table>

a) Construct a boxplot of the data
b) Determine if there are any outliers.

Example 6: Using the data from example #5
a) Change the weight from pounds to kilograms and add 2 kg (special uniform)
b) Get summary statistics and compare with example 5
c) Draw a box plot

Homework: pg 75-80; probs 88, 90, 91, 112, 117
Chapter 1: Data Analysis

Chapter 1 Review

Objectives: Students will be able to:
- Summarize the chapter
- Define the vocabulary used
- Know and be able to discuss all sectional and chapter knowledge objectives
- Be able to do all sectional and chapter construction objectives
- Successfully answer any of the review exercises

Vocabulary: None new

Summary:

- Always include labels, scales, legend (if needed) and title on any graph you are asked to make
- Supporting graphs (that you make as supporting evidence) don’t have to have as much detail
- Numerical summaries do not fully describe the shape of a distribution; Always plot your data!
- Median and IQR are resistant; mean and standard deviations are not
- Boxplots do not tell you all about the distribution (especially about the specific type)
- IQR is a number (the length of the box in the boxplot)
- **Describe all distributions** by their Shape, Outliers, Center and Spread (SOCS)
  - Shape – symmetric (uniform or mound-shape) or skewed (left or right); uni-, bi- or multi-modal
  - Outliers – or any unusual data values or patterns
  - Center – mean (symmetric) or median(skewed) (or mode for categorical data)
  - Spread – standard deviation or IQR (or range for categorical data)
- Remember Statistical Opinions Can Vary (Shape, Outliers, Center, Variability)
- Always use comparative language (much greater, less, about the same) when comparing distributions SOCS

Homework: pg 83-85; prob R1.3, 9, 10
Chapter 1: Data Analysis

1. The upper or third quartile for grades on the first calculus test was 85%. Your friend, who has not taken statistics, scored 90% on the test. Explain to your friend how her grade compares to others in her class.

2. Suppose you have test scores of 72%, 91%, 86%, and 95% in your chemistry class. What score do you need to make on the next test in order to have an 85% average?

3. In the computational formula for standard deviation, you sometimes use \( n \) and sometimes use \( (n - 1) \). Under what circumstances should you use \( n \)?

4. Fill in the following blanks:
   a. We studied two measures of central tendency, mean and median. Which of these is the more resistant measure? _______________ Explain why this measure is more resistant.
   
   b. We studied three measures of spread: standard deviation, interquartile range, and range. Which of these is the most resistant measure? _______________

5. In an experiment designed to determine the effect of a drug on reaction time, a subject is asked to press a button whenever a light flashes. The reaction times (in milliseconds) for ten trials are:
   96 101 112 138 93 99 107 93 95 100
   a. Make a stem and leaf plot to display this information. Be sure to include unit information (a legend).
   
   b. What information about the distribution does the stem and leaf plot provide? Be thorough in your response.

6. Data were collected on a sample of Deerfield Academy students. Several of the variables are listed below. Next to each variable, put all of the following words that correctly describe the variable:
   
   Categorical  quantitative  discrete  continuous
   (a) Advisor ______________________________
   (b) Height ______________________________
   (c) Number of courses student is taking this term ______________________________

7. A teacher returned the first test to the five students in a small class. She reported that the median score was 85 and the mean score was 84. The student with the lowest score (62) realized that the teacher had incorrectly calculated her grade and that the correct grade was 72. Assuming that this is still be the lowest score for the seminar students, when the teacher recomputed the summary statistics, the median will equal ____________ and the mean will equal ______________.
Chapter 1: Data Analysis

8. The histogram below displays weight increases (in pounds) for a sample of pigs fed a certain diet. Assume that bars include right endpoints.
   How many pigs were in this sample? __________
   Estimate the median weight increase for the pigs in this sample. __________
   What proportion of these pigs had a weight increase exceeding 20 pounds? __________
   Briefly (but completely) describe the shape of this distribution

9. As I drove through Connecticut several weeks ago, I obtained a sample of prices for a gallon of unleaded gasoline at service stations I passed. Four of these are provided here: $3.09, $3.15, $3.19, $3.29. Use the definition and show work below to find the mean and standard deviation of these prices. Round answers to the nearest cent.
   Mean
   Standard deviation

10. The *Los Angeles Times* reported interest rates for savings accounts at a sample of California banks. Summary statistics are provided below:
    Minimum = 3.15%  Q1 = 3.25%  Median = 3.31%  Q3 = 3.33%  Maximum = 4.35%
    Determine whether the data set has any outliers (check for extremely low and high values). Show work and provide an explanation to support your answer.
This is an anonymous survey for our AP Statistics class. Do not put your name on it or show it to anyone else! I will not share information in any way that reveals who you are.

Amount of sleep you got last night: ____________ hours  Gender: ________

# of siblings (brothers/sisters) you have: __________  Hair color: ________

Height: __________ (in inches)  Cumulative GPA: ________

SAT Math score: __________  SAT CR score: ________

# of AP classes you are taking this year: __________  Birth date: ___________________

Favorite fast-food restaurant: ________________  Left- or right-handed: __________

Ounces of soda consumed yesterday: __________  Pulse rate: __________

Days since you last consumed food from a fast-food restaurant: __________

Do you believe that fast-food restaurants should be held responsible for obesity resulting from eating too much fast-food? Circle: Yes or No

Time spent yesterday: on a computer __________
on the Internet __________
watching TV __________

Thinking back to the most recent time you went to the bathroom; did you wash your hands immediately afterward? Circle: Yes or No

Which of the following best describes what percent of the time you wash your hands immediately after using the bathroom?

100%  75%–99%  50%–74%  25%–49%  0%–24%

Choose one of the following numbers. Then circle your choice. 1  2  3  4

Randomly choose one of the letters S or Q. Circle your choice.

Imagine a sequence of 10 tosses of a fair penny. Write down a sequence of outcomes that you believe is likely to occur. Use H for heads and T for tails.

|  |  |  |  |  |

Guess your teacher’s age: _______