Chapter 5 – Analyzing Univariate Data

Introduction

Now that we have discussed some methods for collecting data, we can look at what to do with those findings. Whether you have collected categorical or numerical data, you will want to choose an appropriate type of graphical display so that you can visualize the data. Charts and graphs of various types, when created carefully, can provide important information about a data set. You will also need to analyze the data with numerical and summary statistics. Once you have constructed a graphical display and have calculated numerical statistics, it will be necessary to describe your findings verbally. Statisticians can then make appropriate conclusions and comparisons based on the data and statistics while avoiding opinion and judgment statements. This chapter will focus on some of the more common visual presentations of data, numerical analyses of data, and verbal descriptions of data.

5.1 Categorical Data

Learning Objectives

- Organize categorical data in tables
- Construct bar graphs and pie charts by hand and with technology
- Describe, summarize, and compare categorical data

Each student in the class should complete the following survey. The data collected will be used in your homework problems. Notice that the variables in each question are categorical.

1. What is your gender? Choose one
   - Female
   - Male

2. What is your favorite season? Choose one
   - Winter
   - Spring
   - Summer
   - Fall

3. Which of these is your favorite type of food? Choose one
   - Italian
   - Asian
   - Mexican
   - American

4. What type of pet(s) do you have? Choose all that apply
   - Dog
   - Cat
   - Fish
   - Reptile
   - Rodent
   - Other
   - None

https://bit.ly/probstatsSection5-1
(3 video links included)
Frequency Tables and Bar Graphs

When analyzing categorical data (also called qualitative data), bar graphs are commonly used. A bar graph is a graph in which each bar shows how frequently a given category occurs. It is usually helpful to organize the data in a frequency table, a table that shows the number of occurrences for each category, before constructing the bar graph. The bars can go either horizontally or vertically, should be of consistent width, and need to be equally spaced apart. The categories are separate and can be put in any order along the axis. It is common to put them in alphabetical order, but not required. As with all the graphs you will construct, be sure to use a consistent scale, include a title, labels for axes, numbers to mark axes as necessary, and a key whenever needed.

Example 1

A bar graph could show the number of different types of pets for a group of students. The number and types of pets owned by a class of 33 geometry students are shown to the right.

<table>
<thead>
<tr>
<th>Type of Pet</th>
<th>#of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td>14</td>
</tr>
<tr>
<td>Cat</td>
<td>8</td>
</tr>
<tr>
<td>Fish</td>
<td>3</td>
</tr>
<tr>
<td>Reptile</td>
<td>2</td>
</tr>
<tr>
<td>Rodent</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
</tr>
<tr>
<td>None</td>
<td>7</td>
</tr>
</tbody>
</table>

a) What could cause the numbers to add up to more than 33?
b) Construct a bar graph to display this data set.
c) Describe what the graph shows.

Solution

a) They add up to more than 33 because some students likely own more than one type of pet and are being counted in more than one category.
b) Here is a bar graph that was created using Microsoft Excel.
c) For this class, the most common pet is a dog. Fourteen students, or 42% of the class, own a dog. Having a cat or no pet at all are the next most common results. Five students own some type of rodent, two have reptiles for pets, and three have fish. There are also two students who own some other type of pet.
Example 2

A great deal of electronic equipment ends up in landfills as people update their computers, TVs, cell phones, etc. This is a concern because the chemicals from batteries and other electronics add toxins to the environment. This electronic waste has been studied in an effort to decrease the amount of pollution and hazardous waste. The following frequency table shows the amount of tonnage of the most common types of electronic equipment discarded in the United States in 2005. Construct a bar graph and comment on what it shows.

<table>
<thead>
<tr>
<th>Electronic Equipment</th>
<th>Thousands of Tons Discarded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cathode Ray Tube (CRT) TV’s</td>
<td>7591.1</td>
</tr>
<tr>
<td>CRT Monitors</td>
<td>389.8</td>
</tr>
<tr>
<td>Printers, Keyboards, Mice</td>
<td>324.9</td>
</tr>
<tr>
<td>Desktop Computers</td>
<td>259.5</td>
</tr>
<tr>
<td>Laptop Computers</td>
<td>30.8</td>
</tr>
<tr>
<td>Projection TV’s</td>
<td>132.8</td>
</tr>
<tr>
<td>Cell Phones</td>
<td>11.7</td>
</tr>
<tr>
<td>LCD Monitors</td>
<td>4.9</td>
</tr>
</tbody>
</table>

Electronics Discarded in the US (2005)
Source: National Geographic, January 2008. Volume 213 No.1, pg 73

Solution

The type of electronic equipment is a categorical variable, and therefore, this data can easily be represented using the bar graph below.

According to this 2005 data, the most commonly disposed of electronic equipment was CRT TV’s, by more than 19 times than that of the next most common type of electronic discard.
Pie Charts

Pie charts (or circle graphs) are used extensively in statistics. These graphs are used to display categorical data and appear often in newspapers and magazines. A pie chart shows each category (sectors) as a part of the whole (circle). The relationships between the parts, and to the whole, are visible in a pie chart by comparing the sizes of the sectors (slices).

Constructing a pie chart uses the fact that the whole of anything is equal to 100%. All of the sectors equal the whole circle. Remember from geometry that the central angles of a circle total 360°. In regards to pie charts, 360° = 100% of the circle. The sections should have different colors or patterns to enable an observer to clearly see the difference in size of each section.

Pie charts are an appropriate choice when you are working with categorical data that can be viewed as covering 100% of all results. It is not an appropriate choice when you aren’t working with 100% of the data, when choices may include overlaps, or results come from different categories. For example, when we asked every student in this class to list the pets they currently have, we found some students who had more than one pet. A pie chart would not be an appropriate way to display the data in this case. The sectors in a circle graph do not allow for overlaps such as this. Another time when pie charts are not appropriate is when the choices do not cover all possibilities. For example, the electronic waste example above does not include every possibility, so the categories would not add to 100%. In such cases, a bar graph would be a more appropriate choice because it allows for overlaps and does not need to cover exactly 100% of the choices.

Example 3: How to Construct a Pie Chart

The Red Cross Blood Donor Clinic had a very successful morning collecting blood donations. Within three hours twenty-five people had made donations. The types of blood donated are shown in table 5.2 below.

<table>
<thead>
<tr>
<th>Blood Type</th>
<th>A</th>
<th>B</th>
<th>O</th>
<th>AB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Donors</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

a) Construct a pie chart to represent the data.

b) Comment on what the graph shows.

Solution

a) **Step 1:** Determine the total number of donors. $7 + 5 + 9 + 4 = 25$

**Step 2:** Express each donor number as a percent of the whole by using the formula

$$\text{Percent} = \frac{f}{n} \cdot 100\%$$ where $f$ is the frequency and $n$ is the total number.

$$\frac{7}{25} \cdot 100\% = 28\% \quad \frac{5}{25} \cdot 100\% = 20\% \quad \frac{9}{25} \cdot 100\% = 36\% \quad \frac{4}{25} \cdot 100\% = 28\%$$
Step 3: Express each donor number as the number of degrees of a circle that it represents by using the formula \[ \text{Degree} = \frac{f}{n} \cdot 360° \] where \( f \) is the frequency and \( n \) is the total number.

\[
\frac{7}{25} \cdot 360° = 100.8° \\
\frac{5}{25} \cdot 360° = 72° \\
\frac{9}{25} \cdot 360° = 129.6° \\
\frac{4}{25} \cdot 360° = 57.6°
\]

Step 4: Using a protractor or technology, draw the central angles for each section of the circle.

Step 5: Write the label and correct percentage inside or next to the section. Color each section a different color. Be sure to include a title, and a key if needed.

In order to create a pie graph by using the circle, it is necessary to use the percent of a section to compute the correct degree measure for the central angle. The blood type graph labels each section with context and percent, not the degrees. This is because degrees would not be meaningful to an observer trying to interpret the graph. If the sections are not labeled directly as in this example, it becomes necessary to include a key so that the observers will know what each section represents.

b) From the graph, you can see that more donations were of Type O (36%) than any other type. The least amount of blood collected was of Type AB (16%).

Graphs on Computer Software

The above pie chart could be created by using a protractor and graphing each section of the circle according to the number of degrees needed for each section. However, bar graphs and pie charts are most frequently made with computer software programs such as Excel or Google Docs. You will be asked to create bar graphs and pie charts both by hand and by using computer software programs. Always remember to include titles, labels, and keys as needed. Be sure to ‘fix’ the graph generated by the software program so that it looks the way you want it to look and shows clearly whatever it is you are trying to convey.

Example 4

Comment on what the graph shows:
Solution

Several people were asked to choose their favorite fruits from a list of six options. Apples were the favorite choice with 35% of the participants choosing them. The second favorite fruit was cherries at 25%, followed by grapes with 20%. Ten percent of the people said that dates were their favorite fruit. However, only 7% chose bananas from the choices provided and the remaining 3% liked some fruit other than those listed.

Pictographs

Another type of graph that is sometimes used to display categorical data is a pictograph. A **pictograph** is basically a bar graph with pictures instead of bars. A problem with pictures in graphs is that the area that they take up can mislead the observer. The width and height both increase as the picture gets larger. Pictographs are often used in advertisements and magazines. They can be a fun way to make the graphs more interesting in appearance. However, pictographs can be misleading and can be distracting, so they are generally avoided in serious statistical representations.

Example 5

The following graph compares the number of wins for high school football teams during the 2010 seasons. Explain why the pictograph is misleading.

Solution

The pictures increased in both height and width. So when something should be doubled, it actually looks four times as big. For example, when comparing the number of wins between Eisenhower and Adams the graph should show 4 times as many wins. However, in this pictograph it looks as though Adams had 16 times as many wins (4 times as wide X 4 times as tall).
Problem Set 5.1

Exercises

1) Many students at SRHS were given a questionnaire regarding their interests outside of school. The results of one of the questions, ‘What is your favorite After-School Activity?’, are shown in the table below. Each student chose exactly one of the choices in the table.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Play Sports</td>
<td>45</td>
</tr>
<tr>
<td>Talk on Phone</td>
<td>53</td>
</tr>
<tr>
<td>Visit With Friends</td>
<td>99</td>
</tr>
<tr>
<td>Earn Money</td>
<td>44</td>
</tr>
<tr>
<td>Chat Online</td>
<td>66</td>
</tr>
<tr>
<td>School Clubs</td>
<td>22</td>
</tr>
<tr>
<td>Watch TV</td>
<td>37</td>
</tr>
</tbody>
</table>

Source: http://www.mathgoodies.com

a) Create a bar graph for this data.
b) Would a pie chart also be appropriate for this example?
c) Calculate the percent of total for each category and the central angle for each category.
d) Create a pie chart for this data.

2) Based on what you can see in the graph, write a brief description of what it is showing. This should be at least three sentences and be written in context.

3) Use the Type of Pet data collected from your class to complete each problem.
   a) Construct a frequency table to show the Type of Pet data from your class.
   b) Create a bar graph that shows the types of pets the students in our class have. This may be done by hand or with technology.
   c) Write a brief description of what your graph shows.

4) Use the Favorite Season data collected from your class to complete each problem.
   a) Construct a frequency table to show the Favorite Season data from your class.
   b) Create a pie chart that shows the favorite season of the year for the students in your class. This may be done by hand or with technology.
   c) Write a brief description of what your graph shows.

5) Look at the school lunch graph that was created by some students:

![School Lunch Graph]

   a) In what way is this graphical representation misleading? Explain.
   b) Create a better graphical representation for this same data.

6) Use the Favorite Food and Gender data from your class to complete each problem.
   a) Construct a frequency table to show the Favorite Food data separately for males and females from your class.
   b) Create two pie charts that compare the favorite food types for the boys and girls in our class. The charts should ‘match’ as much as possible. In other words, they should be the same size, use the same colors, use the same fonts, etc. This may be done by hand or with technology.
   c) Write a brief description comparing the male and female choices for favorite food. Look for similarities and differences.
Review Exercises

7) The following table has statistics for the Minnesota Wild hockey team for the 2015-2016 season for a selection of players. Thirteen variables are listed across the top of the page.
   a) Identify the individuals.
   b) Identify what each variable represents, for example, GP = games played. You may need to do some research or ask classmates.
   c) Classify each variable as numerical or categorical.

<table>
<thead>
<tr>
<th>Forwards &amp; Defensemen</th>
<th>GP</th>
<th>G</th>
<th>A</th>
<th>P</th>
<th>+/-</th>
<th>PIM</th>
<th>PP</th>
<th>SH</th>
<th>GW</th>
<th>S</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 C MIKKO KORVO</td>
<td>82</td>
<td>17</td>
<td>39</td>
<td>56</td>
<td>6</td>
<td>40</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>141</td>
<td>12.10</td>
</tr>
<tr>
<td>20 D RYAN SUTER</td>
<td>82</td>
<td>8</td>
<td>43</td>
<td>51</td>
<td>10</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>188</td>
<td>4.30</td>
</tr>
<tr>
<td>3 C CHARLIE COYLE</td>
<td>82</td>
<td>21</td>
<td>21</td>
<td>42</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>140</td>
<td>15.00</td>
</tr>
<tr>
<td>64 C MIKAEL GRANLUND</td>
<td>82</td>
<td>13</td>
<td>31</td>
<td>44</td>
<td>-12</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>150</td>
<td>8.10</td>
</tr>
<tr>
<td>22 R NINO NIEDERREITER</td>
<td>82</td>
<td>20</td>
<td>23</td>
<td>43</td>
<td>9</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>159</td>
<td>12.60</td>
</tr>
<tr>
<td>24 D MATT DUMBE</td>
<td>81</td>
<td>10</td>
<td>16</td>
<td>26</td>
<td>1</td>
<td>38</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>152</td>
<td>5.60</td>
</tr>
<tr>
<td>46 D JARED SPURGEON</td>
<td>71</td>
<td>11</td>
<td>18</td>
<td>29</td>
<td>11</td>
<td>14</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>122</td>
<td>9.00</td>
</tr>
<tr>
<td>56 C ERIK HAUHA</td>
<td>76</td>
<td>14</td>
<td>20</td>
<td>34</td>
<td>21</td>
<td>24</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>99</td>
<td>14.10</td>
</tr>
<tr>
<td>29 R JASON FOMINVILLE</td>
<td>75</td>
<td>11</td>
<td>25</td>
<td>36</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>187</td>
<td>5.90</td>
</tr>
<tr>
<td>26 L THOMAS VANEK</td>
<td>74</td>
<td>18</td>
<td>23</td>
<td>41</td>
<td>-10</td>
<td>22</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>146</td>
<td>12.30</td>
</tr>
<tr>
<td>6 D MARCO SCANDELLA</td>
<td>73</td>
<td>5</td>
<td>16</td>
<td>21</td>
<td>6</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>126</td>
<td>4.00</td>
</tr>
<tr>
<td>16 L JASON ZUCKER</td>
<td>71</td>
<td>13</td>
<td>10</td>
<td>23</td>
<td>-4</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>158</td>
<td>8.20</td>
</tr>
<tr>
<td>11 L ZACH PARISE</td>
<td>70</td>
<td>25</td>
<td>28</td>
<td>53</td>
<td>-3</td>
<td>36</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>234</td>
<td>10.70</td>
</tr>
</tbody>
</table>

8) John forgot to study for his history quiz, so he will guess on each question. The quiz has 5 true-false questions and 5 multiple-choice questions (with 4 choices each). He will guess an answer for each question. In how many possible ways might John answer all of the questions?

9) What is the probability that John will get all of the questions correct?
5.2 Time Plots & Measures of Central Tendency

Learning Objectives

- Construct time plots
- Describe trends in time plots
- Calculate range and measures of central tendency: mean, median, mode
- Understand how a change in the data will effect the statistics

Line Graphs as Time Plots

We are often interested in how something has changed over time. The type of graphical display that shows this the most clearly is the time plot, or line graph. When one of the variables is time, it will almost always be plotted along the horizontal axis as the explanatory variable. A time plot is a continuous graph that allows us to examine if there is some type of trend in how the response variable behaves over a period of time.

Example 1

The total municipal waste generated in the US by year is shown in the data set below.

<table>
<thead>
<tr>
<th>Year</th>
<th>Municipal Waste Generated (Millions of Tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>269</td>
</tr>
<tr>
<td>1991</td>
<td>294</td>
</tr>
<tr>
<td>1992</td>
<td>281</td>
</tr>
<tr>
<td>1993</td>
<td>292</td>
</tr>
<tr>
<td>1994</td>
<td>307</td>
</tr>
<tr>
<td>1995</td>
<td>323</td>
</tr>
<tr>
<td>1996</td>
<td>327</td>
</tr>
<tr>
<td>1997</td>
<td>327</td>
</tr>
<tr>
<td>1998</td>
<td>340</td>
</tr>
</tbody>
</table>

Source: [http://www.zerowasteamerica.org](http://www.zerowasteamerica.org)
Solution

a) In this example, the time (in years) is considered the explanatory variable, and is graphed along the horizontal axis. The amount of municipal waste is the response variable, and is graphed along the vertical axis. Time plots can be drawn by hand most easily using graph paper. They can also be created with computer software programs or graphing calculators. This graph was made using Microsoft Excel.

b) This graph shows that the amount of municipal waste generated in the United States increased at a fairly steady rate during the 1990s. Between 1991 and 1992 there was a decrease of 13 million tons of municipal waste, but every other year during the 1990s had an increase.

c) It should be noted that factors other than the passage of time cause our waste to increase. Population growth, economic conditions, and societal habits and attitudes may also be contributing factors.

Example 2

Here is a line graph that shows how the hourly minimum wage changed from when it was first mandated through 1999.

a) During which decade did the hourly wage increase by the greatest amount?

b) During which decade did it increase the most times?

c) When did it stay constant for the longest?

Solution

a) The greatest increase appears to have happened during the 1990’s, when it went from ≈$3.75 to ≈$5.20.

b) The 1970’s appear to have had 5 or 6 increases in the minimum wage.

c) The longest constant minimum wage was during the 1980’s.
Measures of Central Tendency & Spread

The mean, the median, and the mode are all measures of central tendency. They all show where the center of a set of data “tends” to be. Each one is useful at different times. Any one of these three measures of may be referred to as the center of a set of data.

Mean

The mean, often called the ‘average’ of a numerical set of data, is found by taking the sum of all of the numbers divided by the number of values in the data set. This value is sometimes called the arithmetic mean. Geometrically, the mean is the balance point of a distribution. The mean is a summary statistic that gives you a description of the entire data set and is especially useful with large data sets where you might not have the time to examine every single value. However, the mean can be dramatically impacted by outliers (unusual values), and can end up leaving the observer with the wrong impression of a data set.

Example: Suppose these are the hourly wages for the employees at Burger Boy: $9.25, $9.55, $10.15, $9.40, $9.25, $10.90, $18.75, $10.10. If you calculate the mean wage, you would get $10.92. If someone were to report the average wage at Burger Boy to be $10.92 it would give the impression that this is what the average employee makes. However, this is misleading because all employees other than the manager makes less than this amount. In this particular situation, the mean is misleading. The outlier (the manager’s salary) is causing a significant increase in the mean.

Median

The median is the number in the middle position once the data has been organized from smallest to largest. This is the only number for which there are as many values above it as below it in the set of organized data. The median is sometimes referred to as the equal areas point. The median, for a data set with an odd number of values, is the value that is exactly in the middle of the ordered list. It divides the data into two halves. The median for data set with an even number of values, is the mean of the two values in the middle of the ordered list. The median is a useful measure of center when there are outliers in the data set because the middle number will stay in the middle. The median often gives a good impression of the center because half of the values are above the median and half of the values below the median. It doesn’t matter how big the largest values are or how little the smallest values are.

Example: If you calculate the median salary for the Burger Boy employees you get $9.83. This is a much better description of what the typical employee at Burger Boy gets paid because half the employees make more than this amount and half make less than this amount. The manager’s higher salary does not affect the median.

Mode

The mode of a set of data is simply the number that appears most frequently in the data set. There are no calculations required to find the mode of a data set. You simply need to look for the most common result. Be aware that it is not uncommon for a data set to have no mode, one mode, two modes or even more than two modes. If there is more than one mode, simply list them all. If there is no mode, write ‘no mode’. No matter how many modes, the same set of data will have only one mean and only one median.

https://bit.ly/probstatsSection5-2a

MMM and Range
The mode is a measure of central tendency that is simple to locate but is not used much in practical applications. It is the only one of these three values that can be for either categorical or numerical data. Remember the example regarding pets from section 5.1? The mode was ‘dog’ because that was the most common response.

**Range**

The range of a data set describes how spread out the data is. It is one measure of variability. To calculate the range, subtract the smallest value from the largest value (maximum value – minimum value = range). This value provides information about a data set that we cannot see from only the mean, median, or mode. For example, two students may both have a quiz average of 75%, but one of them may have scores ranging from 70% to 82% while the other may have scores ranging from 24% to 90%. In a case such as this, the mean would make the students appear to be achieving at the same level, when in reality one of them is much more consistent than the other.

**Example 3**

Stephen has been working at Wendy’s for 15 months. The following numbers are the number of hours that Stephen worked at Wendy’s during the past seven months:

24, 24, 31, 50, 53, 66, 78

What is the mean number of hours that Stephen worked per month for the last seven months?

**Solution**

Stephen has worked at Wendy’s for 15 months but note we are only given data for the last seven months. Therefore, this set of data represents a sample of the population. The mean of a sample is denoted by $\bar{x}$ which is called “x bar” and is found using the formula below.

The number of data points for a sample is written as $n$. The formula to the right shows the steps that are involved in calculating the mean for a data sample.

The formula can now be written using symbols.

You can now use the formula to calculate the mean number of hours that Stephen worked.

The mean number of hours that Stephen worked during this time period was 47 hours per month.
Example 4
The ages of several randomly selected customers at a coffee shop were recorded. Calculate the mean, median, mode, and range for this data.

\[23, 21, 29, 24, 31, 21, 27, 23, 24, 32, 33, 19\]

Solution

**mean:**

\[
\frac{23 + 21 + 29 + 24 + 31 + 21 + 27 + 23 + 24 + 32 + 33 + 19}{12} = \frac{307}{12} = 25.58
\]

**median:** Organize the ages in ascending order: 19, 21, 21, 23, 23, 24, 24, 27, 29, 31, 32, 33

Count in to find the middle value. Note that 24 & 24 are both in the middle. The middle value will be halfway between these two values or the average of 24 and 24.

\[
\frac{24 + 24}{2} = 24
\]

**mode:** Look for the values that occur most frequently (21, 23, 24). This data set has three modes.

**range:** Subtract the smallest value from the largest value (max - min = range) 33 - 19 = 14.

**Solution:** Make your conclusion in context.

At this coffee shop, the mean age of the people in our sample was 25.58 years old and the median age was 24 years old. There were three modes for age at 21, 23, and 24 years old and the range for ages was 14 years.

Example 5
Lulu is obsessing over her grade in health class. She just simply cannot get anything lower than an A- or she will cry! She knows that the grade will be based on her average (mean) test grade and that there will be a total of six tests. They have taken five so far, and she has received 85%, 95%, 77%, 89%, and 94% on those five tests. The third test did not go well, and she is getting worried. The cutoff score for an A is 93% and 90% is the cutoff score for an A-. She wants to know what she has to get on the last test.

a) What is the lowest grade Lulu will need to get on the last test in order to get an A in health?

b) What is the lowest grade Lulu will need to get on the last test in order to get an A- in health?

**Solution**

a) Set up an equation thinking about how Lulu would calculate her average test grade if she knew all six scores. Knowing that she wants the final average to equal 93%, she puts an ‘x’ in the place of the last test score, and then does some algebra to solve for x.

\[
\frac{85 + 95 + 77 + 89 + 94 + x}{6} = 93
\]

\[
(85 + 95 + 77 + 89 + 94 + x) = 93 \cdot 6
\]

\[
85 + 95 + 77 + 89 + 94 + x = 558
\]

\[
440 + x = 558
\]

\[
x = 118
\]

Oh no! There is no way she can get 118%. So, there is no possible hope for her to get an A.
b) It is time to try for an A-, but that 118% scared her, so she is going to think of the lowest possible score that will still be an A-. Because her teacher rounds grades, she knows that she can get an A- if her mean score is 89.5%. The algebra for this calculation is shown to the right.

There is hope! As long as she gets a 97% or higher on this last test, she can get an A-. She is going to study like crazy!

\[
\frac{85 + 95 + 77 + 89 + 94 + x}{6} = 89.5 \\
(440 + x) = 89.5 \cdot 6 \\
440 + x = 537 \\
x = 97
\]
Problem Set 5.2

Exercises

1) Determine the mean, median, mode, and range for each of the following sets of values:
   a) 20, 14, 54, 16, 38, 64
   b) 22, 51, 64, 76, 29, 22, 48
   c) 40, 61, 95, 79, 9, 50, 80, 63, 109, 42

2) The mean weight of five men is 167.2 pounds. The weights of four of the men are 158.4 pounds, 162.8 pounds, 165 pounds and 178.2 pounds. What is the weight of the fifth man?

3) The mean height of 12 boys is 5.1 feet. The mean height of 8 girls is 4.8 feet.
   a) What is the total height of the boys?
   b) What is the total height of the girls?
   c) What is the mean height of all 20 boys and girls together?

4) The following data represents the number of mailing advertisements received by ten families during the past month. Make a statement describing the ‘typical’ number of advertisements received by each family during the month. Be sure to include statistics to support your statement.
   43 37 35 30 41 23 33 31 16 21

5) Mica’s chemistry teacher bases grades on the average of each student’s test scores during the trimester. Mica has been kind of slacking this year, but hasn’t been too concerned because he knows that he will at least get the credit (60% = passing). However, his parents just informed him that he will not be allowed to use the car if he has any grades below a C which begin at 73%. Below are Mica’s chemistry test scores for the first eight chapters.
   10, 70, 71, 82, 65, 76, 58, 75
   a) Calculate the mean, median, mode, and range for Mica’s chemistry test scores. What grade will Mica receive in chemistry based on this?
   b) His teacher has decided that each student may retake any one of his or her tests in an effort to improve his or her grade. Mica jumps at this opportunity, studies chapter one for hours and retakes the test. To his, and his mother’s delight, his 10% turns into a 70%!! Woo-hoo! Calculate the mean, median, mode, and range for Mica after this change. Which of these values changed? Which did not? What grade will Mica receive now?
   c) Suppose after Mica turned the 10% into a 70%, he studied only a little bit and earned a 60% on the chapter 9 test and a 76% on the chapter 10 test. What would his final average be in this case?
   d) Suppose instead that after Mica turned the 10% into a 70%, he studied hard and earned an 85% on the chapter 9 test and a 90% on the chapter 10 test. What would his final average be in this case?
6) **Deals on Wheels:** The table below lists the retail price and the dealer’s costs for 10 cars at a local car lot this past year.

<table>
<thead>
<tr>
<th>Car Model</th>
<th>Retail Price</th>
<th>Dealer’s Cost</th>
<th>Amount of Mark-Up</th>
<th>Percent of Mark-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nissan Sentra</td>
<td>$24,500</td>
<td>$18,750</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ford Fusion</td>
<td>$26,450</td>
<td>$21,300</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyundai Elantra</td>
<td>$22,660</td>
<td>$19,900</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chevrolet Malibu</td>
<td>$25,200</td>
<td>$22,100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pontiac Sunfire</td>
<td>$16,725</td>
<td>$14,225</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mazda 5</td>
<td>$27,600</td>
<td>$22,150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyota Corolla</td>
<td>$14,280</td>
<td>$13,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honda Accord</td>
<td>$28,500</td>
<td>$25,370</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volkswagen Jetta</td>
<td>$29,700</td>
<td>$27,350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subaru Outback</td>
<td>$32,450</td>
<td>$28,775</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) Calculate the **amount** each car was marked up.

b) Calculate the **percent** that each car was marked up and report answers rounded to the nearest tenth of a percent.

c) Calculate the mean, median, mode and range for the **percent of mark-up** column.

d) Do the “amount of mark-up column” and the “percent of mark-up column” put the cars in the same order for profit? Explain or give an example.

7) Write a brief description of what the line graph for platinum prices shows. Be sure that you do this in context using complete sentences and that you include at least three observations.

**Line Graph: Platinum Prices, 1960 to 2005**

The line graph shows the price of platinum per ounce in US dollars between 1960 and 2005.

**Source:** http://www.admc.hct.ac.ae
8) According to the U.S. Census Bureau, “household median income” is defined as “the amount which divides the income distribution into two equal groups, half having income above that amount, and half having income below that amount.” The table shows the median U.S. household incomes every 3 years from 1975 until 2008, according to the U.S. Census Bureau.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>$11,800</td>
<td>$15,064</td>
<td>$19,074</td>
<td>$22,415</td>
<td>$26,061</td>
<td>$29,943</td>
<td>$31,241</td>
<td>$35,492</td>
<td>$40,696</td>
<td>$42,409</td>
<td>$46,326</td>
<td>$50,303</td>
</tr>
</tbody>
</table>

a) Construct a time plot for the median household data. You may do this by hand, on graph paper, or by using technology.

b) Write a brief description of what the line plot shows. This should be done using complete sentences in context and it should include at least three distinct observations.

Review Exercises

For each of the following problems, decide whether you will use a combination, a permutation, or the fundamental counting principle. Then, set up and solve the problem.

9) A camp counselor is in charge of 10 campers. The kids will be going horseback riding today. There are 5 horses, so they will go in two shifts. In how many ways can the camp counselor assign campers to the specific horses for the first shift?

10) In how many ways can the camp counselor select four of the ten campers to attend the afternoon archery class?

11) How many different three-topping pizzas are possible if there are 12 toppings from which to select?

12) Luigi has 3 pairs of shoes, 7 pairs of jeans, and 8 shirts that he likes to wear that happen to be clean. He is going to put together an outfit for his hot date tonight. If he will choose one of each item, how many different outfits are possible?

13) Eleven skiers are to be in a race. Prizes will be awarded for 1st, 2nd, and 3rd place. Assuming no ties, in how many ways can the prizes be awarded?
5.3 Numerical Data: Dot Plots & Stem Plots

Learning Objectives

- Construct dot plots, stem plots and split-stem plots
- Calculate numerical statistics for quantitative data
- Identify potential outliers in a distribution
- Describe distributions in context – including shape, outliers, center, and spread

Dot Plots

One convenient way to organize numerical data is a dot plot. A dot plot is a simple display that places a dot (or X, or another symbol) above an axis for each datum value (datum is the singular of data). The axis should cover the entire range of the data including numbers that will have no data marked above them. This will visually show outliers or gaps in the data set. There is a dot for each value, so values that occur more than once will be shown by stacked dots. Dot plots are especially useful when you are working with a small set of data across a reasonably small range of values. This type of graph gives the observer a clear view of the shape, mode, and range of the set of data. Outliers are also often easy to spot. Finally, since the numbers are already in order, locating the median is also a simple process.

Ages of all of the Sales People at Stinky’s Car Dealership.

Describing a Numerical Distribution

Once you have constructed a graphical representation of a data set, you should try to describe what the graph shows. There are several characteristics that should be mentioned when describing a numerical distribution and your description needs to explain what this specific data represents. Describe the shape of the graph, whether or not there are any outliers present in the data, the location of the center of the data, and how spread out the data is. All of this should be done in the specific context of the individuals and variable being studied. We will use an acronym to help you remember what to include in your descriptions (S.O.C.C.S.) - shape, outliers, context, center and spread. An explanation of each of these characteristics follows.
Shape

Once a graphical display is constructed, we can describe the distribution. When describing the distribution, we should be sure to address its shape. Although many graphs will not have a clear or exact shape, we can usually identify the shape as symmetrical or skewed. A symmetrical distribution will have a middle through which we can draw an imaginary line. The portions of the graphs on the two sides of this line should be fairly equal mirror images of one another. If you were to fold along the imaginary center line, the two sides would almost match up. Many symmetrical distributions are bell shaped; they will be tall in the middle with the two sides thinning out as you move away from the middle. The sides are referred to as tails. A skewed distribution is one in which the bulk of the data is concentrated on one end with the other side having less data and a longer tail. The direction of the longer tail is the direction of the skew. Skewed right data sets will have a longer tail to the right while skewed left data sets will have a longer tail to the left. Other shapes that you might see are uniform distributions which have nearly consistent heights all the way across the data set and bimodal distributions which have two peaks in the distribution.

Outliers

We should be sure to mention any outliers, gaps, groupings, or other unusual features of a distribution. An outlier is a value that does not fit with the rest of the data. Some distributions will have several outliers, while others will not have any. We should always look for outliers because they can affect many of our statistics. Also, sometimes an outlier is actually an error that needs to be corrected. If you have ever ‘bombed’ one test in a class, you probably discovered that it had a big impact on your overall average in that class. This is because the mean is impacted by outliers and will be pulled toward outliers. This is another reason why we should be sure to look at the data and not just at the statistics about the data. When an outlier occurs in the data set and we do not realize it, we can be misled by the mean to believe that the numbers are higher or lower than they really are.
Context

Do not forget that the graph, the numbers and the descriptions are all about something. There is a context. All of the elements of the distribution should be described in the specific context of the situation in question.

Center

The center of a distribution needs to be included in the verbal analysis as well. People often wonder what the ‘average’ is. The measure for center can be reported as the median, mean, or mode. Even better, give more than one of these in your description. Remember, an outlier will impact the mean but it will not impact the median. For example, while the median of a data set will stay in the center even when the largest value increases tremendously, the mean will change, sometimes significantly.

Spread

In our description of a data set, we should also mention the spread. The spread is a measure of variability and can be reported as the range of values of the data set. When analyzing a distribution, we often don’t want to simply report the range (saying that the range is equal to some number is not always enough information). It can be much more informative to say that the data ranges from ____ to _____ (minimum value to maximum value). For example, suppose the TV news reports that the temperature in St. Paul had a range of 20° during a given week. This could mean very different temperatures depending upon the time of year. It would be more informative to give specific information such as the temperature in St. Paul ranged from 68° to 88° last week.

S.O.C.C.S.

When you describe the distribution of a numerical variable, there are several key pieces of information to include. This text will use the acronym S.O.C.C.S (Shape, Outliers, Context, Center, Spread) to help us remember what characteristics to include in our descriptions.

Example 1

An anthropology instructor at the community college is interested in analyzing the age distribution of her students. The students in her Anthropology 102 class are: 21, 23, 25, 26, 25, 24, 26, 19, 18, 19, 26, 28, 24, 22, 24, 19, 23, 24, 24, 21, 23, and 28 years old. Organize the data in a dot plot. Calculate the mean, median, mode, and range for the distribution. Describe the distribution. Be sure to include the shape, outliers, center, context, and spread.

Solution

- Construct a dot plot:

```
  X
  X
  X
  X X X X X X X
  X X X X X X X
  18 19 20 21 22 23 24 25 26 27 28
```

Ages of Students in Anthropology 102
Solution (continued)

- **mean:**
  \[
  \frac{18 + 19 + 19 + 19 + 21 + 21 + 22 + 23 + 23 + 24 + 24 + 24 + 24 + 24 + 25 + 26 + 26 + 26 + 28 + 28}{22}
  \]
  \[= \bar{x} = 23.27 \text{ years old}\]

- **median:** With the numbers listed in order, count to locate the middle number. It is between 24 and 24 so calculate the mean of these two numbers. \((24+24)/2=24\) The median = 24 years old.

- **mode:** The most frequent age is 24. The mode is 24 years old.

- **range:** The minimum age is 18 and the maximum is 28 so the range is 28 - 18 = 10 years and the ages range from 18 to 28 years old.

- **describe:** Address the shape, outliers, center, context, and spread of the distribution.

  The distribution of student ages in this Anthropology 102 class is fairly symmetrical with no clear outliers. Student ages range from 18 to 28 years old. The median and mode for age are both 24 years old and the mean is 23.27 years. Thus, the typical student in this class is 23-24 years of age.

**Stem Plots**

In statistics, data is represented in tables, charts or graphs. One disadvantage of representing data in these ways is that sometimes the specific data values are often not retained. Using a stem plot is one way to ensure that the data values are kept intact. A **stem plot** is a method of organizing the data that includes sorting the data and graphing it at the same time. This type of graph uses the stem as the leading part of the data value and the leaf as the remaining part of the value. The result is a graph that displays the sorted data in groups or classes. A stem plot is used with numerical data when it will be helpful to see the actual values organized in order.

To construct a stem plot you must first determine the range of your distribution. Build the stems so that they cover the entire range. Include every stem even if it will have no values after it. This will allow us to see the true shape of the distribution including outliers, whether it is skewed, and if there are any gaps. We then place all of the “leaves” after the appropriate stems. Place the numbers in ascending order and include all values. In other words, repeats will show up more than once. Some people like to put the numbers in order before they construct the stem plot, some like to try to put them in order as they make the plot, and others like to make a rough draft first without regard to order and then make a final copy with the numbers in the correct order. Any of these methods will result in a correct stem plot if completed carefully.
Example 2

A researcher was studying the growth of a certain plant. She planted 25 seeds and kept watering, sunlight, and temperature as consistent as possible. The following numbers represent the growth (in centimeters) of the plants after 28 days.

\begin{itemize}
  \item[a)] Construct a stem plot
  \item[b)] Describe the distribution.
\end{itemize}

**Solution**

\begin{itemize}
  \item[a)] **Construct a stem plot:** Notice that the stem plot has the numbers in ascending order and includes a key and title.
  \item[b)] **Describe the distribution:** Be sure to address shape, outliers, center, context, & spread.
\end{itemize}

The distribution of growth at 28 days ranged from 10 to 61 centimeters for these plants with the majority of plants growing to at least 30cm. The median height was 41cm after 28 days. The shape is bimodal and there is a gap in the distribution because there are no plants in the 20-29 cm class. There are some possible low outliers, but no high outliers for plant growth.

Example 3

Sometimes a stem plot ends up looking too crowded. When the data is concentrated in a few rows, or classes, it can be difficult to determine what the shape is or whether there are any outliers in the data. In the stem plot that follows, the ages of a group of people was concentrated in the 30’s and 40’s as shown in the plot on left. However, the statistician looking at this was not satisfied with the crowded appearance, so she decided to ‘split’ the stems. The resulting graph on the right, called a **split-stem plot**, shows very different results. Describe the distribution based on the split-stem plot.

**Solution**

To split the stems, each stem was written twice. The top one is for the first half of the leaves in that class, and the second one is for the leaves in the second half of that class. For example the first stem of 4 gets 40 to 44, and the second 4 gets 45 to 49. When splitting stems into two separate groupings, the number 5 is the cutoff for moving into the second grouping, just like we normally round numbers.

The split-stem plot shows that the distribution of ages in this example is bimodal and also roughly symmetrical. It also shows that the ages of 20 and 22 appear to be low outliers. None of this was visible in the original stem plot. Both plots show that the ages range from 20 to 54 years, with a median age of 41 years old, a mean of 41.3 years old, and a mode of 47 years old.
Problem Set 5.3

Exercises

1) The following is data representing the percentage of paper packaging manufactured from recycled materials for a select group of countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>% of Paper Packaging Recycled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estonia</td>
<td>34</td>
</tr>
<tr>
<td>New Zealand</td>
<td>40</td>
</tr>
<tr>
<td>Poland</td>
<td>40</td>
</tr>
<tr>
<td>Cyprus</td>
<td>42</td>
</tr>
<tr>
<td>Portugal</td>
<td>56</td>
</tr>
<tr>
<td>United States</td>
<td>59</td>
</tr>
<tr>
<td>Italy</td>
<td>62</td>
</tr>
<tr>
<td>Spain</td>
<td>63</td>
</tr>
<tr>
<td>Australia</td>
<td>66</td>
</tr>
<tr>
<td>Greece</td>
<td>70</td>
</tr>
<tr>
<td>Finland</td>
<td>70</td>
</tr>
<tr>
<td>Ireland</td>
<td>70</td>
</tr>
<tr>
<td>Netherlands</td>
<td>70</td>
</tr>
<tr>
<td>Sweden</td>
<td>70</td>
</tr>
<tr>
<td>France</td>
<td>76</td>
</tr>
<tr>
<td>Germany</td>
<td>83</td>
</tr>
<tr>
<td>Austria</td>
<td>83</td>
</tr>
<tr>
<td>Belgium</td>
<td>83</td>
</tr>
<tr>
<td>Japan</td>
<td>98</td>
</tr>
</tbody>
</table>


The dot plot for this data is shown below.

a) Calculate the mean, median, mode, and range for this set of data
b) Describe the distribution in context. Remember your S.O.C.C.S!
2) At the local veterinarian school, the number of animals treated each day over a period of 20 days was recorded.
   a) Construct a stem plot for the data.
   b) Describe the distribution thoroughly.
      Remember your S.O.C.C.S!

28 34 23 35 16
17 47 05 60 26
39 35 47 35 38
35 55 47 54 48

3) The following table reports the percent of students who took the SAT for the 20 U.S. States with the highest participation rates for the 2004 SAT test.

<table>
<thead>
<tr>
<th>STATE</th>
<th>SAT Participation Rate 2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>87%</td>
</tr>
<tr>
<td>Connecticut</td>
<td>85%</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>85%</td>
</tr>
<tr>
<td>New Jersey</td>
<td>83%</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>80%</td>
</tr>
<tr>
<td>D.C.</td>
<td>77%</td>
</tr>
<tr>
<td>Maine</td>
<td>76%</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>74%</td>
</tr>
<tr>
<td>Delaware</td>
<td>73%</td>
</tr>
<tr>
<td>Georgia</td>
<td>73%</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>72%</td>
</tr>
<tr>
<td>Virginia</td>
<td>71%</td>
</tr>
<tr>
<td>North Carolina</td>
<td>70%</td>
</tr>
<tr>
<td>Maryland</td>
<td>68%</td>
</tr>
<tr>
<td>Florida</td>
<td>67%</td>
</tr>
<tr>
<td>Vermont</td>
<td>66%</td>
</tr>
<tr>
<td>Indiana</td>
<td>64%</td>
</tr>
<tr>
<td>South Carolina</td>
<td>62%</td>
</tr>
<tr>
<td>Hawaii</td>
<td>60%</td>
</tr>
<tr>
<td>Oregon</td>
<td>56%</td>
</tr>
</tbody>
</table>

   a) Create a split-stem plot for the data.
   b) Find the median for this data set.
   c) If we included the data from the other 30 states, would our mean and median be higher or lower? Explain.
   d) Describe the distribution thoroughly. Remember to use S.O.C.C.S. Specifically identify states as needed.

4) This stem plot is one that looks too crowded.
   a) Create a split-stem plot for this example.
   b) Name at least two things that are visible in the second plot that were not apparent in the first plot.
   c) Invent a scenario that this data could represent.

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>4789</td>
</tr>
<tr>
<td>7</td>
<td>022222233344445566667</td>
</tr>
<tr>
<td>8</td>
<td>0001112222</td>
</tr>
</tbody>
</table>
5) Several game critics rated the **Wow So Fit** game, on a scale of 1 to 100 with 100 being the highest rating. The results are presented in the stem plot to the right.

   a) Find the three measures of central tendency for the game rating data (mean, median, and mode).

   b) Which of these three measures of central tendency gives the best impression of the ‘average’ (typical) rating for this game? Explain.

<table>
<thead>
<tr>
<th>Key: 6</th>
<th>7 = 67</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>2 3 5 7</td>
</tr>
<tr>
<td>6</td>
<td>2 2 4 6 7 7 9</td>
</tr>
<tr>
<td>7</td>
<td>0 0 2 5 6 7 7 9 9</td>
</tr>
<tr>
<td>8</td>
<td>1 1 2 2 3 5 6 6 7 9</td>
</tr>
<tr>
<td>9</td>
<td>0 2 2 2 6 7</td>
</tr>
</tbody>
</table>

6) These dot plots do not have any numbers or context. For each of the following dot plots:

   a) Identify the shape of each distribution and whether or not there appear to be any outliers.

   b) For each plot, determine whether the mean or median would be greater, or if they would be similar.

   c) Suggest a possible variable that might have such a distribution. (In other words, invent a context that fits the graph.)
The table below displays statistics for 23 Minnesota Wild hockey players for the 2015-2016 regular season. We will use this data from the players in problems 7 and 8.

<table>
<thead>
<tr>
<th>Forwards &amp; Defensemen</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>64</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>55</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

Source: [http://wild.nhl.com/club/stats](http://wild.nhl.com/club/stats)

7) Analyze the variable “GP”; which stands for games played.
   a) Create a stem plot for the number of games played by these Wild players.
   b) Calculate the mean, median, mode, range for the number of games played by these Wild players.
   c) Describe the distribution of the number of games played by these players. Remember your S.O.C.C.S!

8) Now, you will examine the +/- statistic data.
   a) Find out what +/- stands for?
   b) Construct a dot plot to show the +/- data.
   c) Describe the distribution.
Review Exercises

9) A random poll was conducted in Springfield to determine what percent of people enjoy watching the television show *The Simpsons*. Of the 1245 people surveyed, 1002 said that they do enjoy watching *The Simpsons*. Identify each of the following:

a) population of interest
b) parameter of interest
c) sample
d) statistic
e) estimated margin of error
f) estimated 95% confidence interval
g) confidence statement
5.4 Numerical Data: Histograms

Learning Objectives

- Construct histograms
- Describe distributions including shape, outliers, center, context, and spread.

Histograms

When it is not necessary to show every value the way a stem plot would do, a histogram is a useful graph. Histograms organize numerical data into ranges, but do not show the actual values. The histogram is a summary graph showing how many of the data points fall within various ranges. Even though a histogram looks similar to a bar graph, it is not the same. Histograms are for numerical data sets and each ‘bar’ covers a range of values. Each of these ‘bars’ is called a class or bin. Histograms are a great way to see the shape of a distribution and can be used even when working with a large set of data.

The bin width is the most important decision that needs to be made when constructing a histogram. The bins need to be of consistent width so that they cover the same range. A well-built histogram will not have fewer than 5 and not more than 15 bins. Find the range and divide by 10. This will give you an idea of how wide to make your bins. From there it becomes a judgment call as to what is a reasonable bin width. For example, it really does not make any sense to count by 11.24 just because that is what the range divided by 10 is equal to. In such a case, it might make more sense to count by 10’s or 12’s depending on the specific data.

Example 1

Suppose that the test scores of 27 students were recorded. The scores were: 8, 12, 17, 22, 24, 28, 31, 37, 39, 40, 42, 43, 47, 48, 51, 57, 58, 59, 60, 65, 65, 74, 75, 84, 88, 91. The lowest score was an 8 and the highest was a 91. Construct a histogram.

Solution

**Plan bin width:** The first step is to look at the range which is 91 - 8 = 83. Divide the range by 10 to get 83/10 = 8.3. It doesn’t make any sense to count by bins of 8.3 points, so we may use 8, or 10, or 12. Next we look at where to start. The first number is 8. It doesn’t make any sense to start counting at 8 either, or to end at 91. We will probably want to start from 0 and end at 100. Counting by 10’s should work nicely.

*Where to begin, and what to count by, are not obvious to a calculator or many computer software programs. The graphing calculator would probably start at 8, and count by 8.3. Leaving you with bins of [8 -16.3); [16.3-24.6); [24.6 -32.9); etc. If you are using technology to create a histogram, you will generally need to ‘fix’ the window so that the bin widths make sense.*
**Mark the horizontal axis:** Mark your scale along the horizontal axis to cover your entire range and to count by the decided upon bin width. Include values where you marked your scale.

**Count the number of values within each bin:** We note that only one value falls between 0 and <10 so we will make the first bin one unit tall. There are two results between 10 and <20 so we make this bin two units tall. Continue counting in this fashion. A frequency table may be helpful here. You need to know how tall to make each bin. You especially need to know how tall to make the tallest of the bins so that your vertical axis will be scaled properly.

**Mark the vertical axis:** Your vertical axis needs to reach the height of the tallest bin. Mark your vertical axis by consistent steps so that it will reach the value needed. Include labels.

*For instance, if you need to get to 2,460; then you should probably count by steps of 250’s or even a larger number.*

**Make your histogram:** Make the bins the correct heights, shade or color them in, add labels including any units, a title, and a key if needed.

**TEST SCORES**

![Histogram Image](https://bit.ly/probstatsSection5-4b)

*Source: [http://www.netmba.com](http://www.netmba.com)*

The bins in this example are [0 to 10); [10 to 20); etc. This means that zero up to, but not including 10 are in the first bin. Note that 9.999 would be in bin #1, but 10 would be in bin #2.

You may be creating your histograms with paper and pencil. Graphing calculators are also a great way to create histograms as well because you have the opportunity to try out different bin widths without needing to erase and start all over. Also, you may want to try to create histograms in Excel or Google Sheets. When you use technology to create your graphs, you should sketch an approximate picture of what you see. Your sketch will look similar to the graph that the technology produces but you will still need to add labels and titles.
Example 2

a) Construct a histogram to look at the distribution of acceptance rates for these U.S. Universities.

b) Describe your findings.

Solution

a) Try this on your calculator: Enter the data in a list and set up a histogram.

Plan bin width: Determine the range (72 - 11 = 61). Divide by 10 (61/10 = 6.1) to get a rough idea of a good bin width. We can try a variety of bin widths of 5, 7.5, 8, or 10, etc. We must start before the minimum of 11 (start at 0 or 10), and pass the maximum of 72 (80).

After trying a few of these options, we decide to use a bin width of 10, starting at 10 and ending at 80. Here is the window that was used on a TI-84 graphing calculator: {Xmin = 10, Xmax = 80, Xscl = 10, Ymin = -2, Ymax = 5, Yscl = 1}

Mark the horizontal axis: Mark your scale along the horizontal axis to cover your entire range and to count by your decided upon bin width. Include values.

Count the number of values within each bin: A frequency table may be helpful here. You need to know how tall to make each bin. You especially need to know how tall to make the tallest of the bins.

Mark the vertical axis: Your vertical axis needs to reach the height of the tallest bin. Mark your vertical axis by consistent steps so that it will reach the number needed. Include values.

Make your histogram: Make the bins the correct heights, shade or color them in, add labels, and include units, a title, and a key if needed.

b) The median and mean are difficult to identify from just a histogram. You will often only be able to estimate them. In this case, we were given all of the original data so we can find the exact values. When possible, identify outliers specifically.

The median acceptance rate for these Universities is 30%. The percent of students, who were accepted to these universities ranged from 11% to 72%. Note that 72% was a high outlier because the next highest rate was 49%. Most of these schools accepted 36% or fewer of those who applied. The distribution is skewed to the right with the high outlier of American University.
Problem Set 5.4

Exercises

1) This graph shows the distribution of salaries (in thousands of dollars) for the employees of a large school district. Answer the questions that follow.

   a) Approximately how many employees make $77,000 or more per year?
   b) What is the bin width here? Be careful.
   c) Without calculating anything, how would you describe the typical salary of an employee of this school district?

2) Jessica is a freshman at the University of Minnesota Duluth. She has been watching her weight because she is afraid of gaining that ‘freshman fifteen’ she keeps hearing about. She has weighed herself every Monday morning since school started. Here is a histogram showing the results in pounds of all of her Monday morning weight checks.

   a) Describe the distribution. Remember your S.O.C.C.S!
   b) What is the range for the bin that has 6 observations?
   c) For her height, Jessica feels that 140 lbs. is her ideal weight. What percent of the time has she been within 5 lbs. of her ideal weight?
3) Pretend you are a journalist.
   a) What do you notice that is wrong with this graph?
   b) Based on only what you can see in the graph and labels, write several sentences that could go with this graph. (Think S.O.C.C.S!) Ignore the mistakes from part (a).

Source: Department of Health

Source: Men and exercise graph: http://www2.le.ac.uk
4) Here again are the statistics from several of the 2015-2016 Minnesota Wild players. We are going to analyze the **Penalties in Minutes (PIM)** data.

<table>
<thead>
<tr>
<th></th>
<th>GP</th>
<th>G</th>
<th>A</th>
<th>P+/-</th>
<th>PIM</th>
<th>PP</th>
<th>SH</th>
<th>GW</th>
<th>S</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIKKO KOIVU</td>
<td>82</td>
<td>17</td>
<td>39</td>
<td>55</td>
<td>6</td>
<td>40</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>141</td>
</tr>
<tr>
<td>RYAN SUTER</td>
<td>82</td>
<td>8</td>
<td>43</td>
<td>51</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>188</td>
</tr>
<tr>
<td>CHARLIE COYLE</td>
<td>82</td>
<td>21</td>
<td>21</td>
<td>42</td>
<td>1</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>MIKAEL GRANLUND</td>
<td>82</td>
<td>13</td>
<td>31</td>
<td>44</td>
<td>-12</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>160</td>
</tr>
<tr>
<td>NINO NIEDERREITER</td>
<td>82</td>
<td>20</td>
<td>23</td>
<td>43</td>
<td>9</td>
<td>35</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>159</td>
</tr>
<tr>
<td>MATT DUMBA</td>
<td>81</td>
<td>10</td>
<td>18</td>
<td>20</td>
<td>1</td>
<td>38</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>152</td>
</tr>
<tr>
<td>JARED SPURGEON</td>
<td>77</td>
<td>11</td>
<td>18</td>
<td>29</td>
<td>11</td>
<td>14</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>122</td>
</tr>
<tr>
<td>JASON POMINVILLE</td>
<td>75</td>
<td>11</td>
<td>25</td>
<td>35</td>
<td>10</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>187</td>
</tr>
<tr>
<td>THOMAS VANEK</td>
<td>74</td>
<td>18</td>
<td>23</td>
<td>41</td>
<td>-10</td>
<td>22</td>
<td>6</td>
<td>0</td>
<td>5</td>
<td>146</td>
</tr>
<tr>
<td>MARCO SCANDINNA</td>
<td>73</td>
<td>5</td>
<td>16</td>
<td>21</td>
<td>5</td>
<td>22</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>125</td>
</tr>
<tr>
<td>JASON ZUCKER</td>
<td>71</td>
<td>13</td>
<td>10</td>
<td>23</td>
<td>-4</td>
<td>20</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>158</td>
</tr>
<tr>
<td>ZACH PARISE</td>
<td>70</td>
<td>25</td>
<td>28</td>
<td>53</td>
<td>-3</td>
<td>36</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>234</td>
</tr>
<tr>
<td>JONAS BRODIN</td>
<td>68</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>-5</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td>CHRIS PORTER</td>
<td>61</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td>-6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>RYAN CARTER</td>
<td>60</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>-3</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>62</td>
</tr>
<tr>
<td>JUSTIN FONTAINE</td>
<td>60</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>3</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>54</td>
</tr>
<tr>
<td>NATE FROSSER</td>
<td>54</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>JARRET STOLL</td>
<td>51</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>-4</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
</tr>
<tr>
<td>MIKE REILLY</td>
<td>29</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>-4</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>JORDAN SCHROEDER</td>
<td>26</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>CHRISTIAN FOLIN</td>
<td>26</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>-1</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>DAVID JONES</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

a) Construct a histogram for PIM (Penalty Minutes) for the Wild players shown above.

b) Describe the distribution. Remember your S.O.C.C.S!

5) Sketch a histogram that fits each of the following scenarios: *(you will have 5 different histograms)*

a) Symmetrical with a few high outliers and a few low outliers.

b) Strongly skewed right with no outliers.

c) Bimodal and symmetrical.

d) Skewed left with a few outliers.

e) Doesn’t fit any of the descriptions we have learned.
6) The table to the right lists the average life expectancy for people in several countries, as of 2010.

<table>
<thead>
<tr>
<th>Country</th>
<th>Life Expectancy (in years) for 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>48</td>
</tr>
<tr>
<td>Australia</td>
<td>82</td>
</tr>
<tr>
<td>Brazil</td>
<td>73</td>
</tr>
<tr>
<td>Canada</td>
<td>81</td>
</tr>
<tr>
<td>China</td>
<td>73</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>79</td>
</tr>
<tr>
<td>Fiji</td>
<td>69</td>
</tr>
<tr>
<td>France</td>
<td>81</td>
</tr>
<tr>
<td>Germany</td>
<td>80</td>
</tr>
<tr>
<td>Guatemala</td>
<td>71</td>
</tr>
<tr>
<td>India</td>
<td>65</td>
</tr>
<tr>
<td>Italy</td>
<td>82</td>
</tr>
<tr>
<td>Japan</td>
<td>83</td>
</tr>
<tr>
<td>Kenya</td>
<td>56</td>
</tr>
<tr>
<td>Madagascar</td>
<td>66</td>
</tr>
<tr>
<td>Mexico</td>
<td>77</td>
</tr>
<tr>
<td>Nigeria</td>
<td>51</td>
</tr>
<tr>
<td>Pakistan</td>
<td>65</td>
</tr>
<tr>
<td>Peru</td>
<td>74</td>
</tr>
<tr>
<td>Poland</td>
<td>76</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>69</td>
</tr>
<tr>
<td>Singapore</td>
<td>82</td>
</tr>
<tr>
<td>South Africa</td>
<td>52</td>
</tr>
<tr>
<td>United States</td>
<td>78</td>
</tr>
<tr>
<td>Vietnam</td>
<td>75</td>
</tr>
</tbody>
</table>

a) Construct a histogram for the distribution of life expectancies for these countries (start at Xmin = 45 and use a bin width of 5).

b) Based on the shape of your graph, do you expect the mean or median to be higher?

c) Calculate the range and the three measures of central tendency for this data set.

d) Which of these three measures of central tendency is most appropriate in this context? Explain.

Review Exercises

7) The local booster club is holding a raffle. There will be one prize of $1000, two prizes of $250, five prizes of $50, and 10 prizes of $25. They are selling 500 tickets at $10 each.

a) Construct a probability model that shows the different prizes and the probabilities of winning those prizes.

b) What is the expected value of a single raffle ticket?

c) Is this raffle considered a “fair game”? Explain why or why not.

8) A fish bowl on a counter contains 4 gold fish, 7 turquoise fish, and 5 pink fish. Simon the cat is playing a game where he closes his eyes, reaches into the bowl, grabs a fish and sees what color the fish is. He then puts the fish back and repeats the process because Simon is sometimes a very kind cat. Find each of the probabilities below.

a) \( P(2 \text{ turquoise fish}) \)

b) \( P(\text{exactly one of the fish is gold}) \)

c) \( P(\text{a pink fish, then a gold fish}) \)

9) If Simon changes the game so that he eats the fish after he takes them out of the bowl, find the following probabilities.

a) \( P(2 \text{ pink fish}) \)

b) \( P(\text{exactly one of the fish is turquoise}) \)

c) \( P(\text{no gold fish}) \)
5.5 Numerical Data: Box Plots & Outliers

Learning Objectives

- Calculate the five number summary for a set of numerical data
- Construct box plots
- Calculate IQR and standard deviation for a set of numerical data
- Determine which numerical summary is more appropriate for a given distribution
- Determine whether or not any values are outliers based on the 1.5*(IQR) criterion
- Describe distributions in context— including shape, outliers, center, and spread

Box Plots

A box plot (also called box-and-whisker plot) is another type of graph used to display data. A box plot divides a set of numerical data into quarters. It shows how the data are dispersed around a median, but does not show specific values in the data. It does not show a distribution in as much detail as does a stem plot or a histogram, but it clearly shows where the data is located. This type of graph is often used when the number of data values is large or when two or more data sets are being compared. The center and spread of the distribution are very obvious from the graph. It is easy to see the range of the values as well as how these values are distributed around the middle value. The smaller the box plot is, the more consistent the data values are with the median of the data. The shape of the box plot will give you a general idea of the shape of the distribution, but a histogram or stem plot will do this more accurately. Any outliers will show up as long ‘whiskers’. The box in the box plot contains the middle 50% of the data, and each ‘whisker’ contains 25% of the data.

The Five-Number Summary

In order to divide into fourths, it is necessary to find five numbers. This list of five values is called the five-number summary. The numbers in the list are {Minimum, Quartile 1, Median, Quartile 3, Maximum}. We have already learned how to find the median of a set of numbers by putting values in order and find the middle value. Clearly, the minimum and maximum are the smallest and largest values. We now will learn how to find the quartiles.

\[ \text{5\# sum} = \{\text{min}, \ Q_1, \ \text{Med}, \ Q_3, \ \text{max}\} \]
Quartiles

The first step is to list all of the values in order from least to greatest. The minimum and maximum are now on the ends of the list and we can count in to find the median. It is a good idea to write down or circle these three values as you find them. Finding the quartiles is just like finding the median except you are only dealing with half of the data set. **Quartile 1** is the ‘median’ of all of the values to the left of the median. **Quartile 3** is the ‘median’ of all of the values to the right of the median. **Do not include the median when finding the Q1 and Q3.**

Constructing a Box Plot

Start by listing the five-number summary in order {Min, Q1, Med, Q3, Max}. The next step is to mark an axis that covers the entire range of the data. Mark the numbers along the axis before you make the box plot, so that the resulting plot shows the shape of the data. The last step is to place a dot above the axis for each of the 5 numbers from the five-number summary, and then to make a ‘box’ through the second and fourth dots, mark a line through the middle dot to show the median, and mark ‘whiskers’ from the box out to the first and fifth dots.

Example 1

You have a summer job working at Paddy’s Pond which is a recreational fishing spot where children can go to catch salmon which have been raised in a nearby fish hatchery and then transferred into the pond. The cost of fishing depends upon the length of the fish caught (0.75 per inch). Your job is to transfer 15 fish into the pond three times a day. But, before the fish are transferred, you must measure the length of each one and record the results. Below are the lengths (in inches) of the first 15 fish you transferred to the pond this morning. Calculate the five number summary, and construct a box plot for the lengths of these fish.

<table>
<thead>
<tr>
<th>Length of Fish (in.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>13 14 6 9 10</td>
</tr>
<tr>
<td>21 17 15 15 7</td>
</tr>
<tr>
<td>10 13 13 8 11</td>
</tr>
</tbody>
</table>

Solution

Since box plots are based on the median and quartiles, the first step is to organize the data in order from smallest to largest.

6, 7, 8, 9, 10, 10, 10, **13**, 13, 13, 14, 15, 15, 17, 21

The minimum is the smallest number (min = 6), and the maximum is the largest number (max = 21). Next, we need to find the median. This has an odd number of values, so the median of all the data is the value in the middle position (Med = 13). There are 7 numbers before and 7 numbers after 13. The next step is the find the median of the first half of the data – the 7 numbers before the median, not including the median. This is called the lower quartile since it
marks the point above the first quarter of the data. On the graphing calculator this value is referred to as Q1.

6, 7, 8, 9, 10, 10, 11

Quartile 1 is the median of the lower half of the data (Q1 = 9).

This step must be repeated for the upper half of the data – the 7 numbers above the median of 13. This is called the upper quartile since it is the point that marks the third quarter of the data. On the graphing calculator this value is referred to as Q3.

13, 13, 14, 15, 15, 17, 21

Quartile 3 is the median of the upper half of the data (Q3 = 15).

Now that the five numbers have all been determined, it is time to construct the actual graph. The graph is drawn above a number line that includes all the values in the data set. Graph paper works very well since the numbers can be placed evenly using the lines of the graph paper. For this example we will need to mark from at least 6 to at least 21. Be sure to mark your axis before you start to construct the box plot. Next, represent the following values by placing dots above their corresponding values on the number line:

Minimum − 6 
Quartile 1 − 9 
Median − 13 
Quartile 3 − 15 
Maximum − 21

The five data values listed above are often called the five number summary for the data set and are necessary to graph every box plot.

Make the ‘box’ part around the Q1 and Q3 values, make ‘whiskers’ out to the min and max values, and make a vertical line to show the location of the median. This will complete the box plot.

The five numbers divide the data into four equal parts. In other words, for this example:

- One-quarter of the data values are located between 6 and 9
- One-quarter of the data values are located between 9 and 13
- One-quarter of the data values are located between 13 and 15
- One-quarter of the data values are located between 15 and 21
More Measures of Spread

Range
We have already learned how to find the range of a set of data. The range represents the entire spread of all of the data.

The formula for calculating the range is \( \text{max} - \text{min} = \text{range} \).

Interquartile Range
The quartiles give us one more measure of spread (variability) called the interquartile range. The interquartile range (IQR) is the range between the lower and upper quartile. To find the IQR, subtract the quartile 1 value from the quartile 3 value (\( Q_3 - Q_1 = \text{IQR} \)). The IQR represents the spread, or range, of the middle 50% of the data. The IQR is a measure of spread that is used when the median is the measure of central tendency.

The formula for calculating the IQR is \( Q_3 - Q_1 = \text{IQR} \).

Note that while the range is impacted by outliers, the IQR is resistant to outliers.

Standard Deviation
Another measure of spread or variability that is used in statistics is called the standard deviation. The standard deviation measures the spread around the mean. This value is more difficult to calculate than range or IQR, but the formula used takes all of the data values in the distribution into account. Standard deviation is the appropriate measure of spread when the mean is the measure of center. However, the standard deviation is easily affected by outliers or skewness because every value is calculated in the formula. The symbol for standard deviation of a sample is \( s \) (on the graphing calculators it is \( S_x \)) and for a population it is \( \sigma \) (sigma).

The standard deviation can be any number zero or greater. It will only be equal to zero if there is no spread (i.e. all values are exactly the same). The more spread out the data is, the larger the standard deviation will be. The standard deviation is most appropriate when you have a very symmetrical, bell-shaped distribution called a normal distribution. We will study this type of distribution in chapter 7.

Which Numerical Summary Should We Use?
We have learned several statistics that are measures of central tendency and several that are measures of spread. How do we know which ones to use? The mean and standard deviation go together while the median will go with the IQR (or range). It is important to remember that the mean and the standard deviation are both affected by outliers and by skewness in a distribution. If either of these issues are present, then the mean and standard deviation are not appropriate. However, it is often interesting to calculate all of the statistics and compare them to one another. The general guidelines are given in the following diagram.
How to Calculate the Standard Deviation by using the Formula

In order to calculate the standard deviation you must have all of the values. Complete the steps below.

1) Calculate the mean of the values.
2) Subtract the mean from each data value. These are the individual deviations.
3) Each of these deviations is squared.
4) All of the squared deviations are added up.
5) The total of the squared deviations is divided by one less than the number of deviations. This is the variance.
6) Take the square root of the variance to get the standard deviation.

The formula for calculating the variance is:

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

The formula for calculating standard deviation is:

\[ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2} \]

As you can probably tell, this formula is very time consuming when you have a large set of data. Also, it is easy to make a mistake in your calculations. We will show this process with a small set of data but generally we will use our calculator to find the standard deviation. See Appendix C for calculator instructions on how to find the standard deviation.
Example 2
There are five teenage girls on Buhl Street that the Millers often use to babysit their three rambunctious sons. The babysitters’ ages are 12, 15, 14, 17, and 19 years old. Find the mean and standard deviation for the ages of the Miller’s babysitters.

Solution

Calculate the mean of the values. \[
\frac{12+15+14+17+19}{5} = 15.4
\]

Subtract the mean from each data value. These are the individual deviations.

Each of these deviations is squared.

All of the squared deviations are then added up.

This total of the squared deviations is divided by one less than the number of deviations. This is the variance.

Take the square root of the variance. This is the standard deviation.

<table>
<thead>
<tr>
<th>Data values</th>
<th>Value – mean = deviation</th>
<th>Deviation squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>(x - ( \bar{x} )) = ( x - \bar{x} )</td>
<td>(( x - \bar{x} ))^2</td>
</tr>
<tr>
<td>12</td>
<td>(12 - 15.4) = -3.4</td>
<td>(-3.4)^2 = 11.56</td>
</tr>
<tr>
<td>15</td>
<td>(15 - 15.4) = -0.4</td>
<td>(-0.4)^2 = 0.16</td>
</tr>
<tr>
<td>14</td>
<td>(14 - 15.4) = -1.4</td>
<td>(-1.4)^2 = 1.96</td>
</tr>
<tr>
<td>17</td>
<td>(17 - 15.4) = 1.6</td>
<td>(1.6)^2 = 2.56</td>
</tr>
<tr>
<td>19</td>
<td>(19 - 15.4) = 3.6</td>
<td>(3.6)^2 = 12.96</td>
</tr>
</tbody>
</table>

Sum of the squared deviations 29.2

\[
\text{Variance} = \frac{\text{sum}}{n-1} = \frac{29.2}{5-1} = 7.3
\]

\[
\text{Standard Deviation} = \sqrt{s^2} = \sqrt{7.3} = 2.7019
\]

The mean age of the Miller family’s babysitters is 15.4 years old and the standard deviation is 2.7019 years.

The standard deviation is tedious to calculate. For any problem where you are asked to calculate the standard deviation, you may wish to use your calculator or a computer to find it.

Example 3
After one month of growing, the heights of 30 parsley seed plants were measured and recorded. The measurements (in inches) are shown in the table below.

<table>
<thead>
<tr>
<th>Table 5.6: Heights of Parsley (in.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

a) Calculate the five-number summary and construct a box plot to represent the data.

b) Describe the distribution.

c) Calculate the mean and standard deviation.

d) Calculate the median, and IQR
Solution

a) Order the values first. The data organized from smallest to largest is shown in the table below. (Note that you could use your calculator to quickly sort these values.)

<table>
<thead>
<tr>
<th>6</th>
<th>8</th>
<th>11</th>
<th>11</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>23</td>
<td>25</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>30</td>
<td>33</td>
<td>34</td>
<td>37</td>
<td>37</td>
</tr>
<tr>
<td>38</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>46</td>
<td>49</td>
</tr>
</tbody>
</table>

Now find the 5-number summary. This time there is an even number of data values so the median will be the mean of the two middle values. Med = \( \frac{26 + 26}{2} = 26 \) Note that we will not use the median when finding the quartiles. The median of the lower half is the number in the 8th position which is 17. The median of the upper half is the number in the 22nd position (or 8th from the top) which is 37. The smallest number is 6 and the largest number is 49.

5-number summary = \( \{6, 17, 26, 37, 49\} \) (All values are measured in inches.)

b) We will remember to reference S.O.C.C.S. to guide us on our description.

The heights of these parsley plants ranged from 6 inches to 49 inches after one month. The distribution is very symmetrical and does not contain any outliers. The median height for these parsley plants was 26 inches tall. The middle 50% of the plants were all between 17 inches and 37 inches tall.

c) The mean and standard deviation can be calculated using technology. The mean is 25.93 inches and the standard deviation is 11.47 inches.

d) The median is part of the five-number summary and is 26 inches. The IQR = Q3 - Q1 = 37 - 17 = 20 inches.
Outliers

We have been noticing some values that appear to be outliers, but have not defined a specific criteria to be considered an outlier. The common outlier test, used to determine whether or not any of the values are outliers utilizes the IQR. This outlier test, often called the \(1.5\times(IQR)\) Criterion, says that any value that is more than one and one-half times the width of the IQR box away from the box is an outlier.

Example 4

Test the sodium in the McDonald’s® sandwiches for outliers. The data can be found in the Section 5.5 Exercises, problem #1. Use the \(1.5\times(IQR)\) Criterion. Show your steps.

Solution

Calculate the five number summary for the Amount of Sodium (in mg):

Five-number summary = \{480, 680, 1030, 1180, 1470\}

Find the IQR: \(IQR = 1180 - 680 = 500\)

Test for low outliers: \(Q1 - 1.5(IQR) = 680 - 1.5(500) = -80\)

Test for high outliers: \(Q3 + 1.5(IQR) = 1180 + 1.5(500) = 1930\)

Check the data to see if we have any outliers:

We certainly have no sandwiches with less than -80 mg sodium so we have no low outliers. We also have no values that are greater than the cut off of 1930 mg so we also have no high outliers.
Problem Set 5.5

Exercises

Here is some nutritional information about a few of the sandwiches on the McDonald’s® menu.

1) Determine the median and the IQR for the following data regarding the McDonald’s® sandwiches:
   a) Calories from fat
   b) Cholesterol

2) Analyze the calories for these McDonald’s® sandwiches.
   a) Calculate the five number summary and construct an accurate box plot for the calories for these sandwiches.
   b) Use the outlier test to determine whether there are any outliers for calories. Test for both high and low outliers. Show your steps.
   c) Describe the distribution in context - Remember your S.O.C.C.S!

3) Analyze the sodium content further.
   a) Construct a box plot for sodium.
   b) Calculate the median and IQR for sodium by hand and compare your results to Example 4.
   c) Calculate the mean and standard deviation for sodium by using a calculator.
   d) It turns out that the Angus Bacon Cheeseburger has 2,070 mg of sodium. Would it be considered an outlier?
The following table shows the potential energy that could be saved by manufacturing each type of material using the maximum percentage of recycled materials, as opposed to using all new materials.

<table>
<thead>
<tr>
<th>Manufactured Material</th>
<th>Energy Saved (millions of BTU’s per ton)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aluminum Cans</td>
<td>206</td>
</tr>
<tr>
<td>Copper Wire</td>
<td>83</td>
</tr>
<tr>
<td>Steel Cans</td>
<td>20</td>
</tr>
<tr>
<td>LDPE Plastics (e.g. trash bags)</td>
<td>56</td>
</tr>
<tr>
<td>PET Plastics (e.g. beverage bottles)</td>
<td>53</td>
</tr>
<tr>
<td>HDPE Plastics (e.g. household cleaner bottles)</td>
<td>51</td>
</tr>
<tr>
<td>Personal Computers</td>
<td>43</td>
</tr>
<tr>
<td>Carpet</td>
<td>106</td>
</tr>
<tr>
<td>Glass</td>
<td>2</td>
</tr>
<tr>
<td>Corrugated Cardboard</td>
<td>15</td>
</tr>
<tr>
<td>Newspaper</td>
<td>16</td>
</tr>
<tr>
<td>Phone Books</td>
<td>11</td>
</tr>
<tr>
<td>Magazines</td>
<td>11</td>
</tr>
<tr>
<td>Office Paper</td>
<td>10</td>
</tr>
</tbody>
</table>


a) Calculate the five number summary and construct an accurate box plot for the Energy Saved data.

b) Use the outlier test to determine whether there are any outliers. Show your steps.

c) Calculate the mean and standard deviation for the Energy Saved data. How do the mean and the median compare?

d) Delete any outliers. Recalculate the five number summary, mean and standard deviation. Which values changed?

The table shows the mean travel time to work (in minutes), for workers age 16+ for 16 cities in Minnesota. This is according to the U.S. Census website.

Source: http://quickfacts.census.gov

a) Construct a box plot for the mean travel time for residents of these Minnesota cities.

b) Make a statement, in context, about what the ‘box’ part of the box plot tells you.

c) Describe the distribution. Remember your S.O.C.C.S! Identify any unusual values specifically.
6) The *Burj Khalifa*, in Dubai, is the world’s tallest building. It is more than twice the height of the Empire State Building in New York. The chart below lists the 20 tallest buildings in the world.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Building &amp; Location</th>
<th>Year Completed</th>
<th>Architectural top (meters)</th>
<th>Architectural top (feet)</th>
<th>Floors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Burj Khalifa, Dubai, United Arab Emirates</td>
<td>2010</td>
<td>828</td>
<td>2,717</td>
<td>163</td>
</tr>
<tr>
<td>2</td>
<td>Abraj Al Bait, Mecca, Saudi Arabia</td>
<td>2011</td>
<td>601</td>
<td>1,972</td>
<td>120</td>
</tr>
<tr>
<td>3</td>
<td>One World Trade Center, New York City, USA</td>
<td>2013</td>
<td>541.3</td>
<td>1,776</td>
<td>104</td>
</tr>
<tr>
<td>4</td>
<td>Taipei 101, Taipei, Taiwan</td>
<td>2004</td>
<td>509</td>
<td>1,670</td>
<td>101</td>
</tr>
<tr>
<td>5</td>
<td>Shanghai World Financial Center, Shanghai, China</td>
<td>2008</td>
<td>492</td>
<td>1,614</td>
<td>101</td>
</tr>
<tr>
<td>6</td>
<td>International Commerce Centre, Hong Kong</td>
<td>2010</td>
<td>484</td>
<td>1,588</td>
<td>118</td>
</tr>
<tr>
<td>7</td>
<td>Petronas Towers, Kuala Lumpur, Malaysia</td>
<td>1998</td>
<td>452</td>
<td>1,483</td>
<td>88</td>
</tr>
<tr>
<td>8</td>
<td>Zifeng Tower, Nanjing, China</td>
<td>2009</td>
<td>450</td>
<td>1,480</td>
<td>89</td>
</tr>
<tr>
<td>9</td>
<td>Willis Tower, Chicago, United States</td>
<td>1974</td>
<td>442</td>
<td>1,450</td>
<td>108</td>
</tr>
<tr>
<td>10</td>
<td>Jin Mao Building, Shanghai, China</td>
<td>1998</td>
<td>421</td>
<td>1,381</td>
<td>88</td>
</tr>
<tr>
<td>11</td>
<td>Two International Finance Centre, Hong Kong</td>
<td>2003</td>
<td>415</td>
<td>1,362</td>
<td>88</td>
</tr>
<tr>
<td>12</td>
<td>CITIC Plaza, Guangzhou, China</td>
<td>1997</td>
<td>391</td>
<td>1,283</td>
<td>80</td>
</tr>
<tr>
<td>13</td>
<td>Shun Hing Square, Shenzhen, China</td>
<td>1996</td>
<td>384</td>
<td>1,260</td>
<td>69</td>
</tr>
<tr>
<td>14</td>
<td>Empire State Building, New York City, United States</td>
<td>1931</td>
<td>381</td>
<td>1,250</td>
<td>102</td>
</tr>
<tr>
<td>15</td>
<td>Central Plaza, Hong Kong</td>
<td>1992</td>
<td>374</td>
<td>1,227</td>
<td>78</td>
</tr>
<tr>
<td>16</td>
<td>Bank of China Tower, Hong Kong</td>
<td>1990</td>
<td>367</td>
<td>1,204</td>
<td>70</td>
</tr>
<tr>
<td>17</td>
<td>Bank of America Tower, New York City, United States</td>
<td>2008</td>
<td>366</td>
<td>1,201</td>
<td>54</td>
</tr>
<tr>
<td>18</td>
<td>Almas Tower, Dubai, United Arab Emirates</td>
<td>2008</td>
<td>350</td>
<td>1,150</td>
<td>74</td>
</tr>
<tr>
<td>19</td>
<td>Emirates Office Tower, Dubai, United Arab Emirates</td>
<td>2000</td>
<td>355</td>
<td>1,165</td>
<td>54</td>
</tr>
<tr>
<td>20</td>
<td>Tuntex Sky Tower, Kaohsiung, Taiwan</td>
<td>1997</td>
<td>348</td>
<td>1,142</td>
<td>85</td>
</tr>
</tbody>
</table>

*Source: https://en.wikipedia.org*

a) Calculate the five number summary for the heights (in feet) of the 20 buildings and construct an accurate box plot.

b) Use the outlier test to determine whether there are any outliers among the heights of these 20 buildings. Test for both high and low outliers. Show your steps. Identify any outliers by name.

c) Describe the shape of the distribution. Remember your S.O.C.C.S!

d) Within what range of heights are the middle 50% of these buildings?

e) Calculate the range and IQR for the number of floors for these 20 buildings.

f) Use the outlier test to determine whether or not there are any outliers for the number of floors. Do your results match your results in part (b)?
7) Several game critics rated the Wow So Fit game, on a scale of 1 to 100 (100 being the highest). The results are presented in this stem plot:

<table>
<thead>
<tr>
<th>Critics Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 8</td>
</tr>
<tr>
<td>5 2 3 5 7</td>
</tr>
<tr>
<td>6 2 2 4 6 7 7 9</td>
</tr>
<tr>
<td>7 0 0 2 5 6 7 7 9 9</td>
</tr>
<tr>
<td>8 1 1 2 2 3 5 6 6 7 9</td>
</tr>
<tr>
<td>9 0 2 2 2 6 7</td>
</tr>
</tbody>
</table>

a) Calculate the five number summary for the Wow So Fit data.

b) Construct a box plot for the data.

c) Describe this distribution. (S.O.C.C.S.)

d) Make a statement, in context, about what the “box” part of the box plot tells us.

Review Exercises

8) Read each of the criticisms below regarding game ratings and determine whether the person making the statement is questioning the validity, the reliability, or the presence of bias in the test. Explain.

a) “The game critics get free copies of the games for their families. So, these ratings are inflated.”

b) “The game critics have no set guidelines on which to use to critique the games. So, these ratings are meaningless.”

c) “The game critics may give different ratings to the same game, when asked at different times. So, these ratings are inconsistent.”

9) Construct a tree diagram that shows all possible outcomes, in regard to sex of the children, of a family with three children.

10) Assuming that \( P(\text{boy}) = P(\text{girl}) = 0.5 \), find the following probabilities for a family with three children.

a) \( P(\text{boy, girl, then boy}) \)

b) \( P(\text{exactly two girls}) \)

c) \( P(\text{at least one boy}) \)
5.6 Numerical Data: Comparing Data Sets

Learning Objectives

- Construct parallel box plots
- Construct back-to-back stem plots
- Compare more than one set of numerical data in context

Parallel Box Plots

Parallel box plots (also called side-by-side box plots) are very useful when two or more numerical data sets need to be compared. The graphs of parallel box plots are plotted, parallel to each other, along the same number axis. This can be done vertically or horizontally and for as many data sets as needed.

Example 1

The figure shows the distributions of the temperatures for three different cities. By graphing the three box plots along the same axis, it becomes very easy to compare the temperatures of the three cities. What are some conclusions that can be drawn about the temperatures in these three cities?

Temperature Range by City

http://www.mathworksheetscenter.com
**Solution**

Here are some conclusions, based on these graphs that might be made. Think S.O.C.C.S! Be sure to compare the distributions to one another and use statistics to support your observations.

- Quartile 1 for City 2 is higher than the quartile 3 in City 1 and the median in City 3. Also, the minimum temperature in City 2 is about the same as the median for the other two cities.
- City 2 is generally warmer than both of the other cities. Cities 1 and 3 have nearly the same median temperature, around 60° to 63° while the median temperature in City 2 is around 82°.
- City 3 has a much larger range in temperatures (35° to 85°), than City 1 (45° to 75°) or City 2 (62° to 95°). The temperature in City 3 varies the most of these three cities.
- The temperature distributions in all three cities are fairly symmetrical and none have any outliers.

**Comparing Numerical Data Sets**

When you are given numerical sets of data for more than one variable and asked to compare them, it will be necessary to construct graphical representations for each data set. In order to compare them to one another the scales must match. When comparing more than one box plot, we construct parallel box plots. When using histograms, we can match the horizontal and vertical scales so that the separate histograms can ‘line up’. Dot plots will work the same way as histograms. Such comparisons are also possible when working with stem plots. Two sets of numerical data can simply share the stems in the middle, with one set’s ‘leaves’ going to the right and the other set’s ‘leaves’ going to the left. On both sides of the plot, the ‘leaves’ will go in numerical order out. Plots like these are called **back-to-back stem plots**.

Once you have constructed any of these types of comparative graphical representations (using the same scale,) you can make observations about how the data sets are the same and how they are different. Just as we have been doing up to this point, those comparisons should be done in context. The observations made might address the shapes of the distributions and whether or not any outliers are present. It is important to compare the centers of the distributions (means, medians, or modes). The spread for each of the distributions should also be addressed (ranges, IQRs, or standard deviations).

**Example 2**

A teacher gave the same physics exam to her two sections of physics. She has been wondering whether the first period and fifth period classes are learning the same amount as one another. She constructed this back-to-back stem plot to compare the test scores for the two different classes.

<table>
<thead>
<tr>
<th>Class A Leaves</th>
<th>Stems</th>
<th>Class B Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

**Source:** [http://www.basic-mathematics.com](http://www.basic-mathematics.com)
Solution

a) The numbers in the stem plots are already in order, so these statistics could be found either by hand or with a graphing calculator.

The five-number summary for Class A is {60, 75, 90.5, 94, 100}.

The five-number summary for Class B is {60, 71, 75.5, 85, 92}.

b) These statistics are most efficiently found using a graphing calculator.

For Class A, the mean is $\bar{x} \approx 85.71$ points and the standard deviation is $S \approx 12.64$ points.

For Class B, the mean is $\bar{x} \approx 76.64$ points and the standard deviation is $S \approx 10.05$ points.

c) Comparison:

Overall, Class A did better on this test than Class B did. Class A’s scores on this test are skewed to the left, but Class B’s scores are somewhat skewed to the right. Neither class has any outliers among the test scores. Class A has a mean score of about 9 points higher (85.7 compared to 76.6) and a median score of 15 points higher (90.5 compared to 75.5). The overall range for the two classes is fairly similar, but the Class A students’ scores were less consistent. The ranges (40 and 32), IQRs (19 and 14), and standard deviations (12.6 and 10.1), all show that Class B’s scores are less spread out than Class A’s scores.

Example 3

An oil company claims that its premium grade gasoline contains an additive that significantly increases gas mileage. They conducted the following experiment in an effort to prove their claim. They selected 15 drivers who all drove the same make, model, and year of car. Starting with an empty gas tank, each car was filled with 45L of one of the two types of gasoline (selected in a random order). The driver was asked to drive until the fuel warning light came on. The number of kilometers was recorded and then the car was filled with the other type of gasoline (whichever they had not yet used). The process was repeated and the number of kilometers was again recorded. The results below show the number of kilometers each car traveled.

<table>
<thead>
<tr>
<th>Regular Gasoline</th>
<th>Premium Gasoline</th>
</tr>
</thead>
<tbody>
<tr>
<td>640 570 640 580 610</td>
<td>659 619 639 629 664</td>
</tr>
<tr>
<td>540 555 588 615 570</td>
<td>635 709 637 633 618</td>
</tr>
<tr>
<td>550 590 585 587 591</td>
<td>694 638 689 589 500</td>
</tr>
</tbody>
</table>

Display each set of data to examine whether or not the claim made by the oil company is true or false.
Solution

Order the data – List the values in order for each set of data.

<table>
<thead>
<tr>
<th>Regular Gasoline</th>
<th>Premium Gasoline</th>
</tr>
</thead>
<tbody>
<tr>
<td>540 550 555 570 570</td>
<td>500 589 618 619 629</td>
</tr>
<tr>
<td>580 585 587 588 590</td>
<td>633 635 637 638 639</td>
</tr>
<tr>
<td>591 610 615 640 660</td>
<td>659 664 689 694 709</td>
</tr>
</tbody>
</table>

5-# summaries – Determine the five number summary for each set of data separately. Be sure to report your five number summary whether you are asked to report it or not.

<table>
<thead>
<tr>
<th>Five-Number Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smallest #</td>
</tr>
<tr>
<td>$Q_1$</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>$Q_3$</td>
</tr>
<tr>
<td>Largest #</td>
</tr>
</tbody>
</table>

Box plots – Mark your axis so that it covers the entire range needed. Here, the smallest value is 500 and the largest value is 709. Graph each box plot along the same axis so that they are parallel to each other. This allows for the two data sets to be easily compared to one another.

Key: blue (top graph) = regular gasoline
gold (bottom graph) = premium gasoline

Conclusions – Make comparisons by looking for similarities and/or differences between the two distributions. Remember your S.O.C.C.S!

Based on this experiment, the number of kilometers that the cars were able to travel on the premium gasoline was generally greater than the number of kilometers that the same cars were able to travel with on regular gasoline. The median number of kilometers for premium gasoline was 637 compared to 587 for regular gas. The first quartile for premium was higher than the third quartile for regular. Also, 25% of those with the premium gasoline went further than all of those using regular gasoline. The distribution for the regular fuel is slightly skewed to the right, and it doesn’t have any outliers. The premium distribution is strongly pulled to the left toward one outlier on the low end (500 km). Based on these results, it appears that the additive in the premium gasoline does improve gas mileage for this make and model of car. Further tests should be done on other types of vehicles.
Example 4

The heights of a group of students are all included in the first histogram. The second histogram only contains the data from the male students and the third is a graph of the heights of only the girls. Explain what these histograms show.

Solution

The range of heights of all students in this group is approximately 20 inches. However, the female heights only have a range of about 11 inches while the male heights have a range of about 13 inches. The female height distribution is the most symmetrical of all three. There is one male whose height is a high outlier and there are no outliers for the females. The median height for the entire group is around 70 inches, for males it is slightly higher around 72 inches, and for females it is around 65 inches. In general, the female students tend to be shorter than the male students.
Problem Set 5.6

Exercises

1) Compare the %Daily Value for Total Fat to the %Daily Value for Saturated Fat for these McDonald’s® sandwiches.

a) Calculate the five-number summary for both %Daily Values.

b) Construct parallel box plots for both.

c) Make at least four observations to compare these two distributions.

<table>
<thead>
<tr>
<th>Nutrition Facts</th>
<th>Serving Size</th>
<th>Calories</th>
<th>Calories from Fat</th>
<th>Total Fat (g)</th>
<th>% Daily Value**</th>
<th>Saturated Fat (g)</th>
<th>% Daily Value**</th>
<th>Trans Fat (g)</th>
<th>Cholesterol (mg)</th>
<th>% Daily Value**</th>
<th>Sodium (mg)</th>
<th>% Daily Value**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Mac</td>
<td>7.4 oz (211 g)</td>
<td>150</td>
<td>240</td>
<td>27</td>
<td>42</td>
<td>10</td>
<td>48</td>
<td>1</td>
<td>85</td>
<td>28</td>
<td>960</td>
<td>40</td>
</tr>
<tr>
<td>Quarter Pounder with Cheese+</td>
<td>7.1 oz (202 g)</td>
<td>120</td>
<td>240</td>
<td>26</td>
<td>41</td>
<td>12</td>
<td>61</td>
<td>1</td>
<td>95</td>
<td>31</td>
<td>1100</td>
<td>46</td>
</tr>
<tr>
<td>Bacon Clubhouse Burger</td>
<td>5.6 oz (270 g)</td>
<td>120</td>
<td>360</td>
<td>62</td>
<td>15</td>
<td>75</td>
<td>1.5</td>
<td>110</td>
<td>38</td>
<td>1470</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Quarter Pounder Bacon &amp; Cheddar Ranch+</td>
<td>5.3 oz (235 g)</td>
<td>110</td>
<td>250</td>
<td>51</td>
<td>10</td>
<td>64</td>
<td>1.5</td>
<td>100</td>
<td>35</td>
<td>1150</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>Quarter Pounder Bacon &amp; Cheese+</td>
<td>5.8 oz (227 g)</td>
<td>100</td>
<td>280</td>
<td>29</td>
<td>45</td>
<td>13</td>
<td>63</td>
<td>1.5</td>
<td>100</td>
<td>24</td>
<td>1440</td>
<td>60</td>
</tr>
<tr>
<td>Quarter Pounder Deluxe+</td>
<td>5.5 oz (244 g)</td>
<td>140</td>
<td>260</td>
<td>27</td>
<td>42</td>
<td>11</td>
<td>54</td>
<td>1.5</td>
<td>96</td>
<td>29</td>
<td>960</td>
<td>40</td>
</tr>
<tr>
<td>Double Quarter Pounder with Cheese+</td>
<td>5.6 oz (233 g)</td>
<td>160</td>
<td>380</td>
<td>63</td>
<td>10</td>
<td>94</td>
<td>2.5</td>
<td>160</td>
<td>63</td>
<td>1200</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Hamburger</td>
<td>3.5 oz (339 g)</td>
<td>240</td>
<td>75</td>
<td>8</td>
<td>12</td>
<td>3</td>
<td>15</td>
<td>1</td>
<td>30</td>
<td>10</td>
<td>480</td>
<td>20</td>
</tr>
<tr>
<td>Cheesburger</td>
<td>4 oz (113 g)</td>
<td>250</td>
<td>100</td>
<td>11</td>
<td>18</td>
<td>5</td>
<td>27</td>
<td>0.5</td>
<td>45</td>
<td>15</td>
<td>650</td>
<td>23</td>
</tr>
<tr>
<td>BBQ Ranch Burger</td>
<td>4 oz (113 g)</td>
<td>240</td>
<td>130</td>
<td>13</td>
<td>22</td>
<td>6</td>
<td>29</td>
<td>0.5</td>
<td>45</td>
<td>16</td>
<td>670</td>
<td>25</td>
</tr>
<tr>
<td>Grilled Onion Cheddar</td>
<td>4 oz (113 g)</td>
<td>240</td>
<td>110</td>
<td>13</td>
<td>20</td>
<td>6</td>
<td>29</td>
<td>0.5</td>
<td>45</td>
<td>16</td>
<td>640</td>
<td>27</td>
</tr>
<tr>
<td>Double Cheesburger</td>
<td>5.7 oz (161 g)</td>
<td>200</td>
<td>190</td>
<td>21</td>
<td>22</td>
<td>10</td>
<td>52</td>
<td>1</td>
<td>90</td>
<td>30</td>
<td>1040</td>
<td>43</td>
</tr>
<tr>
<td>McDouble</td>
<td>5.2 oz (147 g)</td>
<td>180</td>
<td>150</td>
<td>17</td>
<td>26</td>
<td>6</td>
<td>40</td>
<td>1</td>
<td>75</td>
<td>25</td>
<td>840</td>
<td>35</td>
</tr>
<tr>
<td>Jalapeno Double</td>
<td>5 oz (129 g)</td>
<td>200</td>
<td>210</td>
<td>23</td>
<td>36</td>
<td>9</td>
<td>44</td>
<td>1</td>
<td>80</td>
<td>27</td>
<td>1050</td>
<td>43</td>
</tr>
<tr>
<td>Bacon McDouble</td>
<td>5.7 oz (161 g)</td>
<td>240</td>
<td>200</td>
<td>22</td>
<td>34</td>
<td>10</td>
<td>49</td>
<td>1</td>
<td>90</td>
<td>30</td>
<td>1110</td>
<td>46</td>
</tr>
</tbody>
</table>

Source: http://nutrition.mcdonalds.com June 2016
2) The heights of the students in a statistics class were all measured to the nearest inch. The results are presented in this back-to-back stem plot. Notice that it is also a split stem plot. The girls’ heights are ordered out to the right side of the graph and the boys’ heights are ordered out to the left side of the graph.

a) Compute the standard deviation, the range, and the IQR for both girls and boys.

b) Compare the spread for the two groups, based on your answers to (a), in context.

c) Compute the mean, median, and mode for both boys and girls.

d) Compare the center for the two groups, based on your answers to (c), in context.

e) Compare the shape of the distributions, based on the graph, in context.

3) Compare the results of the Probability and Statistics District Common Assessment for two statistics classes.

**CLASS 3:**
45, 37, 14, 42, 24, 33, 41, 16, 39, 24, 38, 35, 35, 32, 51, 46, 30, 42, 25, 37, 37, 19, 26, 23, 28, 38, 16, 35, 21

**CLASS 4:**

a) Construct back-to-back stem plots (use split-stems) for these two classes.

b) Calculate the five number summaries for both classes.

c) Calculate the following statistics for both classes: mean, standard deviation, mode, range, and IQR.

d) Compare and contrast the two distributions. This should be in context and you should make at least four distinct observations.
4) The number of home-runs during a season is one of the statistics recorded about baseball players. The following table has the number of home-runs (over many seasons) for several of the best hitters in baseball. Compare the home-run hitting performance of these exceptional baseball players.

<table>
<thead>
<tr>
<th>Player</th>
<th>Home-Runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babe Ruth</td>
<td>54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22</td>
</tr>
<tr>
<td>Mark McGwire</td>
<td>49, 32, 33, 39, 22, 42, 9, 9, 39, 52, 58, 70, 65, 32, 29</td>
</tr>
<tr>
<td>Barry Bonds</td>
<td>16, 25, 24, 19, 33, 25, 34, 46, 37, 33, 42, 40, 37, 34, 49, 73, 46</td>
</tr>
<tr>
<td>Roger Maris</td>
<td>13, 23, 26, 16, 33, 61, 28, 39, 14, 8</td>
</tr>
</tbody>
</table>

a) Calculate the following statistics for all four players:

\[ \bar{x} = \_, \quad s_x = \_, \quad IQR = \_, \quad 5\text{# summary} = \{\_, \_, \_, \_, \_, \_\} \]

b) Construct Parallel Box Plots for the four players. Be sure to use the same scale for all four graphs and to label each graph.

c) Test for outliers, for all four players, using the 1.5*IQR criterion. Show your work.

d) Compare and contrast the four distributions. This should be in context and you should make at least four distinct observations.

5) The following box plots show the average miles per gallon (city) for various types of vehicles. Comment on what these parallel box plots show. This should be in context and include at least 4 distinct observations. Any dots represent outliers for that data set.

![Box Plot – Average city MPG by Car Type](http://www.fort.usgs.gov)
6) Refer to the four dot plots to answer the questions that follow.

![Graphs I to IV](image)

a) Identify the overall shape of each distribution.

b) How would you characterize the center(s) of each of these distributions?

c) Name at least two statistics that would most likely be the same for all four of these distributions.

d) Which of these distributions has the smallest standard deviation? Which of these distributions has the largest standard deviation? Explain.

e) For which of these distributions would it be appropriate to use the mean and standard deviation as numerical summaries? For which would the five-number summary be more appropriate?
5.7 Chapter 5 Review

Chapter 5 Summary

In this chapter, we have learned that when working with a set of data it is important to choose an appropriate type of graphical display so that we can see what the data looks like. Bar graphs and pie charts are useful ways to display categorical data. Time plots are line graphs that help us to see how a given variable has changed over a specified period of time. When working with numerical data we have learned how to make dot plots, stem plots, histograms, and box plots. It is also possible to make graphs so that comparisons can be made between more than one data set. Back-to-back stem plots and parallel box plots are two such types of graphs.

Our next step was to analyze the data sets by calculating numerical statistics. We began by looking at the mean, median, and mode. These statistics are measures of central tendency and give us an idea of where the center of the data set is located. The range, interquartile range (IQR), and the standard deviation are all measures of the spread or variability of a data set. We have calculated the five-number summary, which divides a set of data into quarters and allows us to construct a box plot.

Once the graphs were constructed and the statistics were calculated, we learned to describe what the data showed. When describing a numerical set of data, in addition to explaining where the center is located and what the spread is, we also describe the shape of the distribution and whether or not any outliers are present in the data. The shapes that we focused on are symmetrical distributions and skewed distributions, remembering that the direction of the skew is toward the tail or outliers. We learned to make appropriate conclusions and comparisons that are based on the data, graphs, and statistics. Statisticians should avoid opinions and judgment statements as much as possible.

We learned that the $1.5 \times (IQR)$ Criterion can be used to determine whether or not any data values are outliers. In addition, we found that the mean and standard deviation are easily influenced by unusual values or skewed data sets. We avoid using these statistics when working with data that contains outliers or is skewed.

Review Exercises

1) Multiple-Choice: Which of the following can be inferred from this histogram?

   a) The mode is 12.5
   b) mean < median
   c) median < mean
   d) The distribution is skewed left.
   e) None of the above can be inferred from this histogram.
2) The owner of a small company is trying to determine whether he should go with a different company for his shipping needs. He needs to analyze the weights of the packages that his company ships out. This graph shows the distribution of the weights of packages that were shipped during the last month.

   a) Calculate the mean, standard deviation, mode, and range for this data. Use a calculator.

   b) Determine the five number summary for this data. Construct a box plot for this data.

   c) Which of these two graphs – the dot plot or the box plot – is more informative? Explain.

   d) The owner figures that he will save money with the new company on any package that weighs less than 16.75 ounces. What percent of packages will he lose money on (those weighing more than 16.75 ounces)?

3) After some bullying issues were brought to light in a big high school, a committee was formed to study the issue. A questionnaire was designed that contained several questions related to bullying and safety. A stratified random sample was selected that included students from all four grade levels. The table that follows shows the responses to one of the questions on the questionnaire.

<table>
<thead>
<tr>
<th>Student Responses to the Question:</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I Feel Safe at School”</td>
</tr>
<tr>
<td>n = sample size</td>
</tr>
<tr>
<td>1003</td>
</tr>
</tbody>
</table>

   a) Create a pie chart that shows the results of this survey question. Be sure to include labels, percentages, a title, and a key if needed. You may do this by hand or with technology.

   b) Describe what the graph shows in context. Be sure to include percentages to support your observations.

   c) Comment on whether or not the committee should be concerned. Explain.
In questions 4-7, match the distribution with the choice of the correct real-world situation that best fits the graph.

4) [Graph]

5) [Graph]

6) [Graph]

7) [Graph]

a) Andy collected and graphed the heights of all the 12th grade students in his high school.

b) Brittany asked each of the students in her statistics class to bring in 20 pennies selected at random from their pocket or piggy bank. She created a plot of the dates of the pennies.

c) Maya asked her friends what their favorite movie was this year and graphed the results.

d) Jeno bought a large box of doughnut holes at the local pastry shop, weighed each of them, and then plotted their weights to the nearest tenth of a gram.
Questions 8 - 17 are multiple-choice questions. Select the best answer from the choices given.

8) Which of the box plots on the right matches the histogram on the left?

9) Identify the 5 number summary for this set of numbers:

<table>
<thead>
<tr>
<th>12,356</th>
<th>16,564</th>
<th>15,684</th>
<th>12,358</th>
<th>15,987</th>
<th>13,556</th>
<th>18,564</th>
<th>18,965</th>
<th>19,683</th>
<th>18,432</th>
<th>18,563</th>
<th>19,352</th>
</tr>
</thead>
</table>

a) \{12,356; 14,600; 17,498; 18,000; 19,683\}
b) \{12,356; 14,620; 17,498; 18,764.5; 19683\}
c) \{12,356; 14,650.5; 17,498; 18,700.5; 19683\}
d) \{12,356; 14,683; 17,500; 18,800; 19683\}
e) \{12,356; 14,695.5; 17,900; 18,888; 19683\}

10) Thirty students took a statistics examination having a maximum of 50 points. The grade distribution is given in the stem-and-leaf plot:

The median grade is equal to:

a) 30.5
b) 30.0
c) 25.0
d) 28.5
e) 44.0
11) Ms. Davis conducted a survey of the 44 students in her stats classes and asked how tall each student was in inches. The five-number summary of the data is \{57, 64, 67, 69, 79\}.

Approximately how many people are shorter than 64 inches tall?

a) 8
b) 21
c) 22
d) 11
e) 18

12) In which scenario(s) would it be better to use the 5-number summary versus the mean and standard deviation?

a) a graph that is skewed
b) a graph that is fairly symmetric
c) a graph that is symmetric but has several high outliers
d) Both choice (b) and (c)
e) Both choice (a) and (c)
f) All of (a), (b) and (c)

13) Suppose the lowest score on an English exam was 35% and the highest score was 90%. If the teacher of the class was to examine her students’ test scores, which shape of the distribution would she prefer to see?

a) skewed to the right
b) skewed to the left
c) fairly symmetric
d) none of the above

14) Several people were surveyed as they were leaving a movie theatre. Among other things, they were asked how much money they had spent. Their answers were: $14, $17.50, $16, $16, $19.25, $12.75, $16, $37.75, $13.50 and $17. It was later discovered that the person who answered “$37.75” actually spent $17.75. Which of the following would not change as a result?

a) the box plot
b) the mean & the mode
c) the median & the mode
d) the standard deviation
e) they all change
15) What does the following five-number summary below tell you about the shape of the distribution?
{15, 16.7, 18.7, 22.3, 32}

   a) It is skewed to the right.
   b) It is skewed to the left.
   c) It is symmetric.
   d) It is uniform.
   e) We cannot determine anything about the shape of the distribution.

16) According to the 1.5*(IQR) Criterion, what are the two cut-off values for determining whether the data set in question #15 contains any outliers?

   a) 7.5 & 24.6
   b) 7.7 & 27.9
   c) 11.3 & 23.9
   d) 4.5 & 24.1
   e) 8.3 & 30.7

17) A class survey was conducted to determine students’ preferences. One question asked about each student’s favorite sport to watch on TV. Of those surveyed, 9 said football, 12 said hockey, 5 said basketball, 6 said baseball, and 3 said some other sport. What would the central angle be for “hockey” in a pie chart of this data?

   a) 65°
   b) 123°
   c) 90°
   d) 34°
   e) 111°

18) The AHS Tornadoes and the BHS Bengals are big rivals! Every year students try to prove that their school is better at sports than the other school. The table to the right shows the number of points scored by each school’s basketball team during the last 15 games played against other teams this year.

   a) Construct a back-to-back stem plot for the data.
   b) Calculate the five number summary, mean and standard deviation for both teams.
   c) Construct parallel box plots for the data.
   d) Compare the two distributions. This should be done in context and include at least three distinct comparisons.
   e) What other information would you like to know when comparing these two basketball teams? Explain.

<table>
<thead>
<tr>
<th>Tornadoes</th>
<th>Bengals</th>
</tr>
</thead>
<tbody>
<tr>
<td>58</td>
<td>74</td>
</tr>
<tr>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>71</td>
<td>73</td>
</tr>
<tr>
<td>64</td>
<td>63</td>
</tr>
<tr>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>63</td>
<td>84</td>
</tr>
<tr>
<td>60</td>
<td>92</td>
</tr>
<tr>
<td>72</td>
<td>38</td>
</tr>
<tr>
<td>48</td>
<td>77</td>
</tr>
<tr>
<td>59</td>
<td>84</td>
</tr>
<tr>
<td>72</td>
<td>95</td>
</tr>
<tr>
<td>62</td>
<td>66</td>
</tr>
<tr>
<td>57</td>
<td>70</td>
</tr>
<tr>
<td>64</td>
<td>68</td>
</tr>
<tr>
<td>49</td>
<td>72</td>
</tr>
</tbody>
</table>
19) The following two graphs are based from the US Census Bureau in 2008. Note that ‘per capita’ means per person. The dots represent actual data values, and the red curves represent models that can be used to predict future trends. Study the two graphs and answer the questions that follow.

a) What type of graphs are these?

b) Describe the trend that each graph shows separately. This should be in context.

c) Notice that the horizontal scales are the same. Compare and contrast the trends that are shown in the two different graphs in context.

d) Approximately how many cell phones were there per person in 1997? In 2005? How many will there be, if the trend continues as the model indicates, in 2018?

e) Approximately how many landlines were there per person at the peak? What year did this occur? How many landlines did the model predict per person in 2015?

20) The table that follows shows the percent of people, 25 years and older, who are high school graduates for several states in the central United States according to the 2010 U. S. Census website.

a) Construct a histogram (use Xmin = 79%, and bin width = 1%).

b) Calculate the five number summary.

c) Identify any outliers. Use the outlier test.

d) Accurately sketch a box plot.

e) Calculate the range, IQR, and mode.

f) Calculate the mean and standard deviation.

g) Compare the mean and the median. (i.e. Which is larger? How different are they?)

h) In this case would the 5-number summary or the mean & standard deviation be more appropriate? Why?

i) Describe the distribution. Be thorough! Don’t forget your S.O.C.C.S! (Shape, Outliers, Center, Context, & Spread)

j) Among this Census data, where does Minnesota fall?
21) An employer in Minneapolis was interested in determining how much money his employees were spending on parking each week. An SRS of 50 employees was selected to complete a questionnaire about parking. Several questions were asked including where they park, how much they spend per week, how often they have difficulty finding spots, and if they pay daily, weekly, or monthly. The following table is the average weekly expenditure for parking for this sample of 50 employees.

<table>
<thead>
<tr>
<th>Weekly Expenditure ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 40 22 21 20 10 20 20</td>
</tr>
<tr>
<td>20 13 18 50 20 18 15 8</td>
</tr>
<tr>
<td>22 10 20 22 21 15 23 30</td>
</tr>
<tr>
<td>9 20 40 22 22 19 15 20 20</td>
</tr>
<tr>
<td>20 15 19 21 14 22 21 35</td>
</tr>
<tr>
<td>20</td>
</tr>
</tbody>
</table>

a) Construct a split-stem plot.

b) Calculate the five number summary.

c) Identify any outliers. Use the outlier test

d) Accurately sketch a box plot. Be sure to scale and label the graph.

e) What is the range? The IQR? The mode?

f) Calculate the mean and standard deviation.

g) Compare the mean and the median.

h) In this case would the 5-number summary or the mean & standard deviation be more appropriate? Why?

i) Describe the distribution. Be thorough! Remember your S.O.C.C.S!

Image References:

- Cars. http://www.icoachmath.com
- School Lunch pictogram. http://alex.state.al.us
- Stem plot example. http://www.basic-mathematics.com/
- Weight of Jessica graph. http://stat.psu.edu/
- Crowded stem plot. http://illuminations.nctm.org/
- Three histogram example. http://classes.cec.wustl.edu
- Package weight graph. http://flylib.com
Appendices

Appendix A – Tables

Appendix A, Part 1 - Random Digit Table

<table>
<thead>
<tr>
<th>Line 101</th>
<th>19223</th>
<th>95034</th>
<th>05756</th>
<th>28713</th>
<th>96409</th>
<th>12531</th>
<th>42544</th>
<th>82853</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line 102</td>
<td>73676</td>
<td>47150</td>
<td>99400</td>
<td>01927</td>
<td>27754</td>
<td>42648</td>
<td>82425</td>
<td>36290</td>
</tr>
<tr>
<td>Line 103</td>
<td>45467</td>
<td>71709</td>
<td>77558</td>
<td>00095</td>
<td>32863</td>
<td>29485</td>
<td>82226</td>
<td>90056</td>
</tr>
<tr>
<td>Line 104</td>
<td>52711</td>
<td>38889</td>
<td>93074</td>
<td>60227</td>
<td>40011</td>
<td>85848</td>
<td>48767</td>
<td>52573</td>
</tr>
<tr>
<td>Line 105</td>
<td>95592</td>
<td>94007</td>
<td>69971</td>
<td>91481</td>
<td>60779</td>
<td>53791</td>
<td>17297</td>
<td>59335</td>
</tr>
<tr>
<td>Line 106</td>
<td>68417</td>
<td>35013</td>
<td>15529</td>
<td>72765</td>
<td>85089</td>
<td>57067</td>
<td>50211</td>
<td>47487</td>
</tr>
<tr>
<td>Line 107</td>
<td>82739</td>
<td>57890</td>
<td>20807</td>
<td>47511</td>
<td>81676</td>
<td>55300</td>
<td>94383</td>
<td>14893</td>
</tr>
<tr>
<td>Line 108</td>
<td>60940</td>
<td>72024</td>
<td>17868</td>
<td>24943</td>
<td>61790</td>
<td>90656</td>
<td>87964</td>
<td>18883</td>
</tr>
<tr>
<td>Line 109</td>
<td>36009</td>
<td>19365</td>
<td>15412</td>
<td>39638</td>
<td>85453</td>
<td>46816</td>
<td>83485</td>
<td>41979</td>
</tr>
<tr>
<td>Line 110</td>
<td>38448</td>
<td>48789</td>
<td>18338</td>
<td>24697</td>
<td>39364</td>
<td>42006</td>
<td>76688</td>
<td>08708</td>
</tr>
<tr>
<td>Line 111</td>
<td>81486</td>
<td>69487</td>
<td>60513</td>
<td>09297</td>
<td>00412</td>
<td>71238</td>
<td>27649</td>
<td>39950</td>
</tr>
<tr>
<td>Line 112</td>
<td>59636</td>
<td>88804</td>
<td>04634</td>
<td>71197</td>
<td>19352</td>
<td>73089</td>
<td>84898</td>
<td>45785</td>
</tr>
<tr>
<td>Line 113</td>
<td>62568</td>
<td>70206</td>
<td>40325</td>
<td>03699</td>
<td>71080</td>
<td>22553</td>
<td>11486</td>
<td>11776</td>
</tr>
<tr>
<td>Line 114</td>
<td>45149</td>
<td>32992</td>
<td>75730</td>
<td>66280</td>
<td>03819</td>
<td>56202</td>
<td>02938</td>
<td>70915</td>
</tr>
<tr>
<td>Line 115</td>
<td>61041</td>
<td>77684</td>
<td>94322</td>
<td>24709</td>
<td>73698</td>
<td>14526</td>
<td>31893</td>
<td>32592</td>
</tr>
<tr>
<td>Line 116</td>
<td>14459</td>
<td>26056</td>
<td>31424</td>
<td>80371</td>
<td>65103</td>
<td>62253</td>
<td>50490</td>
<td>61181</td>
</tr>
<tr>
<td>Line 117</td>
<td>38167</td>
<td>98532</td>
<td>62183</td>
<td>70632</td>
<td>23417</td>
<td>26185</td>
<td>41448</td>
<td>75532</td>
</tr>
<tr>
<td>Line 118</td>
<td>73190</td>
<td>32533</td>
<td>04470</td>
<td>29669</td>
<td>84407</td>
<td>90785</td>
<td>65956</td>
<td>86382</td>
</tr>
<tr>
<td>Line 119</td>
<td>95857</td>
<td>07118</td>
<td>87664</td>
<td>92099</td>
<td>58806</td>
<td>66979</td>
<td>98624</td>
<td>84826</td>
</tr>
<tr>
<td>Line 120</td>
<td>35476</td>
<td>55972</td>
<td>39421</td>
<td>65850</td>
<td>04266</td>
<td>35435</td>
<td>43742</td>
<td>11937</td>
</tr>
<tr>
<td>Line 121</td>
<td>71487</td>
<td>09984</td>
<td>29077</td>
<td>14863</td>
<td>61683</td>
<td>47052</td>
<td>62224</td>
<td>51025</td>
</tr>
<tr>
<td>Line 122</td>
<td>13873</td>
<td>81598</td>
<td>95052</td>
<td>90908</td>
<td>73592</td>
<td>75186</td>
<td>87136</td>
<td>95761</td>
</tr>
<tr>
<td>Line 123</td>
<td>54580</td>
<td>81507</td>
<td>27102</td>
<td>56027</td>
<td>55892</td>
<td>33063</td>
<td>41842</td>
<td>81868</td>
</tr>
<tr>
<td>Line 124</td>
<td>71035</td>
<td>09001</td>
<td>43367</td>
<td>49497</td>
<td>72719</td>
<td>96758</td>
<td>27611</td>
<td>91596</td>
</tr>
<tr>
<td>Line 125</td>
<td>96746</td>
<td>12149</td>
<td>37823</td>
<td>71868</td>
<td>18442</td>
<td>35119</td>
<td>62103</td>
<td>39244</td>
</tr>
<tr>
<td>Line 126</td>
<td>96927</td>
<td>19931</td>
<td>36089</td>
<td>74192</td>
<td>77567</td>
<td>88741</td>
<td>48409</td>
<td>41903</td>
</tr>
<tr>
<td>Line 127</td>
<td>43909</td>
<td>99477</td>
<td>25330</td>
<td>64359</td>
<td>40085</td>
<td>16925</td>
<td>85117</td>
<td>36071</td>
</tr>
<tr>
<td>Line 128</td>
<td>15689</td>
<td>14227</td>
<td>06565</td>
<td>14374</td>
<td>13352</td>
<td>49367</td>
<td>81982</td>
<td>87209</td>
</tr>
<tr>
<td>Line 129</td>
<td>36759</td>
<td>58984</td>
<td>68288</td>
<td>22913</td>
<td>18638</td>
<td>54303</td>
<td>00795</td>
<td>08727</td>
</tr>
<tr>
<td>Line 130</td>
<td>69051</td>
<td>64817</td>
<td>87174</td>
<td>09517</td>
<td>84534</td>
<td>06489</td>
<td>87201</td>
<td>97245</td>
</tr>
<tr>
<td>Line 131</td>
<td>05007</td>
<td>16632</td>
<td>81194</td>
<td>14873</td>
<td>04197</td>
<td>85576</td>
<td>45195</td>
<td>96565</td>
</tr>
<tr>
<td>Line 132</td>
<td>68732</td>
<td>55259</td>
<td>84292</td>
<td>08796</td>
<td>43165</td>
<td>93739</td>
<td>31685</td>
<td>97150</td>
</tr>
<tr>
<td>Line 133</td>
<td>45740</td>
<td>41807</td>
<td>65561</td>
<td>33302</td>
<td>07051</td>
<td>93623</td>
<td>18132</td>
<td>09547</td>
</tr>
<tr>
<td>Line 134</td>
<td>27816</td>
<td>78416</td>
<td>18329</td>
<td>21337</td>
<td>35213</td>
<td>37741</td>
<td>04312</td>
<td>68508</td>
</tr>
<tr>
<td>Line 135</td>
<td>66925</td>
<td>55658</td>
<td>39100</td>
<td>78458</td>
<td>11206</td>
<td>19876</td>
<td>87151</td>
<td>31260</td>
</tr>
<tr>
<td>Line 136</td>
<td>08421</td>
<td>44753</td>
<td>77377</td>
<td>28744</td>
<td>75592</td>
<td>08563</td>
<td>79140</td>
<td>92454</td>
</tr>
<tr>
<td>Line 137</td>
<td>53645</td>
<td>66812</td>
<td>61421</td>
<td>47836</td>
<td>12609</td>
<td>15373</td>
<td>98481</td>
<td>14592</td>
</tr>
<tr>
<td>Line 138</td>
<td>66831</td>
<td>68908</td>
<td>40772</td>
<td>21558</td>
<td>47781</td>
<td>33586</td>
<td>79177</td>
<td>06928</td>
</tr>
<tr>
<td>Line 139</td>
<td>55588</td>
<td>99404</td>
<td>70708</td>
<td>41098</td>
<td>43563</td>
<td>56934</td>
<td>48394</td>
<td>51719</td>
</tr>
<tr>
<td>Line 140</td>
<td>12975</td>
<td>13258</td>
<td>13048</td>
<td>45144</td>
<td>72321</td>
<td>81940</td>
<td>00360</td>
<td>02428</td>
</tr>
<tr>
<td>Line 141</td>
<td>96767</td>
<td>35964</td>
<td>23822</td>
<td>96012</td>
<td>94591</td>
<td>65194</td>
<td>50842</td>
<td>53372</td>
</tr>
<tr>
<td>Line 142</td>
<td>72829</td>
<td>50232</td>
<td>97892</td>
<td>63408</td>
<td>77919</td>
<td>44575</td>
<td>24870</td>
<td>04178</td>
</tr>
<tr>
<td>Line 143</td>
<td>88565</td>
<td>42628</td>
<td>17797</td>
<td>49376</td>
<td>61762</td>
<td>16953</td>
<td>88604</td>
<td>12724</td>
</tr>
<tr>
<td>Line 144</td>
<td>62964</td>
<td>88145</td>
<td>83083</td>
<td>69453</td>
<td>46109</td>
<td>59505</td>
<td>69680</td>
<td>00900</td>
</tr>
<tr>
<td>Line 145</td>
<td>19687</td>
<td>12633</td>
<td>57857</td>
<td>95806</td>
<td>09931</td>
<td>02150</td>
<td>43163</td>
<td>58636</td>
</tr>
<tr>
<td>Line 146</td>
<td>37609</td>
<td>59057</td>
<td>66967</td>
<td>83401</td>
<td>60705</td>
<td>02384</td>
<td>90597</td>
<td>93600</td>
</tr>
<tr>
<td>Line 147</td>
<td>54973</td>
<td>86278</td>
<td>88737</td>
<td>74351</td>
<td>47500</td>
<td>84552</td>
<td>19909</td>
<td>67181</td>
</tr>
<tr>
<td>Line 148</td>
<td>00694</td>
<td>05977</td>
<td>19664</td>
<td>65441</td>
<td>20903</td>
<td>62371</td>
<td>22725</td>
<td>53340</td>
</tr>
<tr>
<td>Line 149</td>
<td>71546</td>
<td>05233</td>
<td>53946</td>
<td>68743</td>
<td>72460</td>
<td>27601</td>
<td>45403</td>
<td>88692</td>
</tr>
<tr>
<td>Line 150</td>
<td>07511</td>
<td>88915</td>
<td>41267</td>
<td>16853</td>
<td>84569</td>
<td>79367</td>
<td>32337</td>
<td>03316</td>
</tr>
</tbody>
</table>
Appendix A, Part 2 – The Normal Distribution Table

For z-scores with z less than or equal to zero

<table>
<thead>
<tr>
<th>z</th>
<th>0.00</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.0</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0011</td>
<td>0.0010</td>
<td>0.0010</td>
</tr>
<tr>
<td>-2.9</td>
<td>0.0019</td>
<td>0.0018</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0014</td>
</tr>
<tr>
<td>-2.8</td>
<td>0.0026</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0022</td>
<td>0.0021</td>
<td>0.0021</td>
<td>0.0020</td>
<td>0.0019</td>
</tr>
<tr>
<td>-2.7</td>
<td>0.0035</td>
<td>0.0034</td>
<td>0.0033</td>
<td>0.0032</td>
<td>0.0031</td>
<td>0.0030</td>
<td>0.0029</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0026</td>
</tr>
<tr>
<td>-2.6</td>
<td>0.0047</td>
<td>0.0045</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0041</td>
<td>0.0040</td>
<td>0.0039</td>
<td>0.0038</td>
<td>0.0037</td>
<td>0.0036</td>
</tr>
<tr>
<td>-2.5</td>
<td>0.0062</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0057</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.0054</td>
<td>0.0052</td>
<td>0.0051</td>
<td>0.0049</td>
</tr>
<tr>
<td>-2.4</td>
<td>0.0082</td>
<td>0.0080</td>
<td>0.0078</td>
<td>0.0075</td>
<td>0.0073</td>
<td>0.0071</td>
<td>0.0069</td>
<td>0.0068</td>
<td>0.0066</td>
<td>0.0064</td>
</tr>
<tr>
<td>-2.3</td>
<td>0.0107</td>
<td>0.0104</td>
<td>0.0102</td>
<td>0.0099</td>
<td>0.0096</td>
<td>0.0094</td>
<td>0.0091</td>
<td>0.0089</td>
<td>0.0087</td>
<td>0.0084</td>
</tr>
<tr>
<td>-2.2</td>
<td>0.0139</td>
<td>0.0136</td>
<td>0.0132</td>
<td>0.0129</td>
<td>0.0125</td>
<td>0.0122</td>
<td>0.0119</td>
<td>0.0116</td>
<td>0.0113</td>
<td>0.0110</td>
</tr>
<tr>
<td>-2.1</td>
<td>0.0179</td>
<td>0.0174</td>
<td>0.0170</td>
<td>0.0166</td>
<td>0.0162</td>
<td>0.0158</td>
<td>0.0154</td>
<td>0.0150</td>
<td>0.0146</td>
<td>0.0143</td>
</tr>
<tr>
<td>-2.0</td>
<td>0.0228</td>
<td>0.0222</td>
<td>0.0217</td>
<td>0.0212</td>
<td>0.0207</td>
<td>0.0202</td>
<td>0.0197</td>
<td>0.0192</td>
<td>0.0188</td>
<td>0.0183</td>
</tr>
<tr>
<td>-1.9</td>
<td>0.0287</td>
<td>0.0281</td>
<td>0.0274</td>
<td>0.0268</td>
<td>0.0262</td>
<td>0.0256</td>
<td>0.0250</td>
<td>0.0244</td>
<td>0.0239</td>
<td>0.0233</td>
</tr>
<tr>
<td>-1.8</td>
<td>0.0359</td>
<td>0.0351</td>
<td>0.0344</td>
<td>0.0336</td>
<td>0.0329</td>
<td>0.0322</td>
<td>0.0314</td>
<td>0.0307</td>
<td>0.0301</td>
<td>0.0294</td>
</tr>
<tr>
<td>-1.7</td>
<td>0.0446</td>
<td>0.0436</td>
<td>0.0427</td>
<td>0.0418</td>
<td>0.0409</td>
<td>0.0401</td>
<td>0.0392</td>
<td>0.0384</td>
<td>0.0375</td>
<td>0.0367</td>
</tr>
<tr>
<td>-1.6</td>
<td>0.0548</td>
<td>0.0537</td>
<td>0.0526</td>
<td>0.0516</td>
<td>0.0505</td>
<td>0.0495</td>
<td>0.0485</td>
<td>0.0475</td>
<td>0.0465</td>
<td>0.0455</td>
</tr>
<tr>
<td>-1.5</td>
<td>0.0668</td>
<td>0.0655</td>
<td>0.0643</td>
<td>0.0630</td>
<td>0.0618</td>
<td>0.0606</td>
<td>0.0594</td>
<td>0.0582</td>
<td>0.0571</td>
<td>0.0559</td>
</tr>
<tr>
<td>-1.4</td>
<td>0.0808</td>
<td>0.0793</td>
<td>0.0778</td>
<td>0.0764</td>
<td>0.0749</td>
<td>0.0735</td>
<td>0.0721</td>
<td>0.0708</td>
<td>0.0694</td>
<td>0.0681</td>
</tr>
<tr>
<td>-1.3</td>
<td>0.0968</td>
<td>0.0951</td>
<td>0.0934</td>
<td>0.0918</td>
<td>0.0901</td>
<td>0.0885</td>
<td>0.0869</td>
<td>0.0853</td>
<td>0.0838</td>
<td>0.0823</td>
</tr>
<tr>
<td>-1.2</td>
<td>0.1151</td>
<td>0.1131</td>
<td>0.1112</td>
<td>0.1093</td>
<td>0.1075</td>
<td>0.1056</td>
<td>0.1038</td>
<td>0.1020</td>
<td>0.1003</td>
<td>0.0985</td>
</tr>
<tr>
<td>-1.1</td>
<td>0.1357</td>
<td>0.1335</td>
<td>0.1314</td>
<td>0.1292</td>
<td>0.1271</td>
<td>0.1251</td>
<td>0.1230</td>
<td>0.1210</td>
<td>0.1190</td>
<td>0.1170</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.1587</td>
<td>0.1562</td>
<td>0.1539</td>
<td>0.1515</td>
<td>0.1492</td>
<td>0.1469</td>
<td>0.1446</td>
<td>0.1423</td>
<td>0.1401</td>
<td>0.1379</td>
</tr>
<tr>
<td>-0.9</td>
<td>0.1841</td>
<td>0.1814</td>
<td>0.1788</td>
<td>0.1762</td>
<td>0.1736</td>
<td>0.1711</td>
<td>0.1685</td>
<td>0.1660</td>
<td>0.1635</td>
<td>0.1611</td>
</tr>
<tr>
<td>-0.8</td>
<td>0.2119</td>
<td>0.2090</td>
<td>0.2061</td>
<td>0.2033</td>
<td>0.2005</td>
<td>0.1977</td>
<td>0.1949</td>
<td>0.1922</td>
<td>0.1894</td>
<td>0.1867</td>
</tr>
<tr>
<td>-0.7</td>
<td>0.2420</td>
<td>0.2389</td>
<td>0.2358</td>
<td>0.2327</td>
<td>0.2296</td>
<td>0.2266</td>
<td>0.2236</td>
<td>0.2206</td>
<td>0.2177</td>
<td>0.2148</td>
</tr>
<tr>
<td>-0.6</td>
<td>0.2743</td>
<td>0.2709</td>
<td>0.2676</td>
<td>0.2643</td>
<td>0.2611</td>
<td>0.2578</td>
<td>0.2546</td>
<td>0.2514</td>
<td>0.2483</td>
<td>0.2451</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.3085</td>
<td>0.3050</td>
<td>0.3015</td>
<td>0.2981</td>
<td>0.2946</td>
<td>0.2912</td>
<td>0.2877</td>
<td>0.2843</td>
<td>0.2810</td>
<td>0.2776</td>
</tr>
<tr>
<td>-0.4</td>
<td>0.3446</td>
<td>0.3409</td>
<td>0.3372</td>
<td>0.3336</td>
<td>0.3300</td>
<td>0.3264</td>
<td>0.3228</td>
<td>0.3192</td>
<td>0.3156</td>
<td>0.3121</td>
</tr>
<tr>
<td>-0.3</td>
<td>0.3821</td>
<td>0.3783</td>
<td>0.3745</td>
<td>0.3707</td>
<td>0.3669</td>
<td>0.3632</td>
<td>0.3594</td>
<td>0.3557</td>
<td>0.3520</td>
<td>0.3483</td>
</tr>
<tr>
<td>-0.2</td>
<td>0.4207</td>
<td>0.4168</td>
<td>0.4129</td>
<td>0.4090</td>
<td>0.4052</td>
<td>0.4013</td>
<td>0.3974</td>
<td>0.3936</td>
<td>0.3897</td>
<td>0.3829</td>
</tr>
<tr>
<td>-0.1</td>
<td>0.4602</td>
<td>0.4562</td>
<td>0.4522</td>
<td>0.4483</td>
<td>0.4443</td>
<td>0.4404</td>
<td>0.4364</td>
<td>0.4325</td>
<td>0.4286</td>
<td>0.4247</td>
</tr>
<tr>
<td>-0.0</td>
<td>0.5000</td>
<td>0.4960</td>
<td>0.4920</td>
<td>0.4880</td>
<td>0.4840</td>
<td>0.4801</td>
<td>0.4761</td>
<td>0.4721</td>
<td>0.4681</td>
<td>0.4641</td>
</tr>
</tbody>
</table>
For z-scores with $z$ greater than or equal to 0

Table 8.2

<table>
<thead>
<tr>
<th>$z$</th>
<th>0.0</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
<th>0.04</th>
<th>0.05</th>
<th>0.06</th>
<th>0.07</th>
<th>0.08</th>
<th>0.09</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.5000</td>
<td>0.5040</td>
<td>0.5080</td>
<td>0.5120</td>
<td>0.5160</td>
<td>0.5199</td>
<td>0.5239</td>
<td>0.5279</td>
<td>0.5319</td>
<td>0.5359</td>
</tr>
<tr>
<td>0.1</td>
<td>0.5398</td>
<td>0.5438</td>
<td>0.5478</td>
<td>0.5517</td>
<td>0.5557</td>
<td>0.5596</td>
<td>0.5636</td>
<td>0.5675</td>
<td>0.5714</td>
<td>0.5753</td>
</tr>
<tr>
<td>0.2</td>
<td>0.5793</td>
<td>0.5832</td>
<td>0.5871</td>
<td>0.5910</td>
<td>0.5948</td>
<td>0.5987</td>
<td>0.6026</td>
<td>0.6064</td>
<td>0.6103</td>
<td>0.6141</td>
</tr>
<tr>
<td>0.3</td>
<td>0.6179</td>
<td>0.6217</td>
<td>0.6255</td>
<td>0.6293</td>
<td>0.6331</td>
<td>0.6368</td>
<td>0.6406</td>
<td>0.6443</td>
<td>0.6480</td>
<td>0.6517</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6554</td>
<td>0.6591</td>
<td>0.6628</td>
<td>0.6664</td>
<td>0.6700</td>
<td>0.6736</td>
<td>0.6772</td>
<td>0.6808</td>
<td>0.6844</td>
<td>0.6879</td>
</tr>
<tr>
<td>0.5</td>
<td>0.6915</td>
<td>0.6950</td>
<td>0.6985</td>
<td>0.7019</td>
<td>0.7054</td>
<td>0.7088</td>
<td>0.7123</td>
<td>0.7157</td>
<td>0.7190</td>
<td>0.7224</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7257</td>
<td>0.7291</td>
<td>0.7324</td>
<td>0.7357</td>
<td>0.7389</td>
<td>0.7422</td>
<td>0.7454</td>
<td>0.7486</td>
<td>0.7517</td>
<td>0.7549</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7580</td>
<td>0.7611</td>
<td>0.7642</td>
<td>0.7673</td>
<td>0.7704</td>
<td>0.7734</td>
<td>0.7764</td>
<td>0.7794</td>
<td>0.7823</td>
<td>0.7852</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7881</td>
<td>0.7910</td>
<td>0.7939</td>
<td>0.7967</td>
<td>0.7995</td>
<td>0.8023</td>
<td>0.8051</td>
<td>0.8078</td>
<td>0.8106</td>
<td>0.8133</td>
</tr>
<tr>
<td>0.9</td>
<td>0.8159</td>
<td>0.8186</td>
<td>0.8212</td>
<td>0.8238</td>
<td>0.8264</td>
<td>0.8289</td>
<td>0.8315</td>
<td>0.8340</td>
<td>0.8365</td>
<td>0.8389</td>
</tr>
<tr>
<td>1.0</td>
<td>0.8413</td>
<td>0.8438</td>
<td>0.8461</td>
<td>0.8485</td>
<td>0.8508</td>
<td>0.8531</td>
<td>0.8554</td>
<td>0.8577</td>
<td>0.8599</td>
<td>0.8621</td>
</tr>
<tr>
<td>1.1</td>
<td>0.8643</td>
<td>0.8665</td>
<td>0.8686</td>
<td>0.8708</td>
<td>0.8729</td>
<td>0.8749</td>
<td>0.8770</td>
<td>0.8790</td>
<td>0.8810</td>
<td>0.8830</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8849</td>
<td>0.8869</td>
<td>0.8888</td>
<td>0.8907</td>
<td>0.8925</td>
<td>0.8944</td>
<td>0.8962</td>
<td>0.8980</td>
<td>0.8997</td>
<td>0.9015</td>
</tr>
<tr>
<td>1.3</td>
<td>0.9032</td>
<td>0.9049</td>
<td>0.9066</td>
<td>0.9082</td>
<td>0.9099</td>
<td>0.9115</td>
<td>0.9131</td>
<td>0.9147</td>
<td>0.9162</td>
<td>0.9177</td>
</tr>
<tr>
<td>1.4</td>
<td>0.9192</td>
<td>0.9207</td>
<td>0.9222</td>
<td>0.9236</td>
<td>0.9251</td>
<td>0.9265</td>
<td>0.9279</td>
<td>0.9292</td>
<td>0.9306</td>
<td>0.9319</td>
</tr>
<tr>
<td>1.5</td>
<td>0.9332</td>
<td>0.9345</td>
<td>0.9357</td>
<td>0.9370</td>
<td>0.9382</td>
<td>0.9394</td>
<td>0.9406</td>
<td>0.9418</td>
<td>0.9429</td>
<td>0.9441</td>
</tr>
<tr>
<td>1.6</td>
<td>0.9452</td>
<td>0.9463</td>
<td>0.9474</td>
<td>0.9484</td>
<td>0.9495</td>
<td>0.9505</td>
<td>0.9515</td>
<td>0.9525</td>
<td>0.9535</td>
<td>0.9545</td>
</tr>
<tr>
<td>1.7</td>
<td>0.9554</td>
<td>0.9564</td>
<td>0.9573</td>
<td>0.9582</td>
<td>0.9591</td>
<td>0.9599</td>
<td>0.9608</td>
<td>0.9616</td>
<td>0.9625</td>
<td>0.9633</td>
</tr>
<tr>
<td>1.8</td>
<td>0.9641</td>
<td>0.9649</td>
<td>0.9656</td>
<td>0.9664</td>
<td>0.9671</td>
<td>0.9678</td>
<td>0.9686</td>
<td>0.9693</td>
<td>0.9699</td>
<td>0.9706</td>
</tr>
<tr>
<td>1.9</td>
<td>0.9713</td>
<td>0.9719</td>
<td>0.9726</td>
<td>0.9732</td>
<td>0.9738</td>
<td>0.9744</td>
<td>0.9750</td>
<td>0.9756</td>
<td>0.9761</td>
<td>0.9767</td>
</tr>
<tr>
<td>2.0</td>
<td>0.9772</td>
<td>0.9778</td>
<td>0.9783</td>
<td>0.9788</td>
<td>0.9793</td>
<td>0.9798</td>
<td>0.9803</td>
<td>0.9808</td>
<td>0.9812</td>
<td>0.9817</td>
</tr>
<tr>
<td>2.1</td>
<td>0.9821</td>
<td>0.9826</td>
<td>0.9830</td>
<td>0.9834</td>
<td>0.9838</td>
<td>0.9842</td>
<td>0.9846</td>
<td>0.9850</td>
<td>0.9854</td>
<td>0.9857</td>
</tr>
<tr>
<td>2.2</td>
<td>0.9861</td>
<td>0.9864</td>
<td>0.9868</td>
<td>0.9871</td>
<td>0.9875</td>
<td>0.9878</td>
<td>0.9881</td>
<td>0.9884</td>
<td>0.9887</td>
<td>0.9890</td>
</tr>
<tr>
<td>2.3</td>
<td>0.9893</td>
<td>0.9896</td>
<td>0.9898</td>
<td>0.9901</td>
<td>0.9904</td>
<td>0.9906</td>
<td>0.9909</td>
<td>0.9911</td>
<td>0.9913</td>
<td>0.9916</td>
</tr>
<tr>
<td>2.4</td>
<td>0.9918</td>
<td>0.9920</td>
<td>0.9922</td>
<td>0.9925</td>
<td>0.9927</td>
<td>0.9929</td>
<td>0.9931</td>
<td>0.9932</td>
<td>0.9934</td>
<td>0.9936</td>
</tr>
<tr>
<td>2.5</td>
<td>0.9938</td>
<td>0.9940</td>
<td>0.9941</td>
<td>0.9943</td>
<td>0.9945</td>
<td>0.9946</td>
<td>0.9948</td>
<td>0.9949</td>
<td>0.9951</td>
<td>0.9952</td>
</tr>
<tr>
<td>2.6</td>
<td>0.9953</td>
<td>0.9955</td>
<td>0.9956</td>
<td>0.9957</td>
<td>0.9959</td>
<td>0.9960</td>
<td>0.9961</td>
<td>0.9962</td>
<td>0.9963</td>
<td>0.9964</td>
</tr>
<tr>
<td>2.7</td>
<td>0.9965</td>
<td>0.9966</td>
<td>0.9967</td>
<td>0.9968</td>
<td>0.9969</td>
<td>0.9970</td>
<td>0.9971</td>
<td>0.9972</td>
<td>0.9973</td>
<td>0.9974</td>
</tr>
<tr>
<td>2.8</td>
<td>0.9974</td>
<td>0.9975</td>
<td>0.9976</td>
<td>0.9977</td>
<td>0.9977</td>
<td>0.9978</td>
<td>0.9979</td>
<td>0.9979</td>
<td>0.9980</td>
<td>0.9981</td>
</tr>
<tr>
<td>2.9</td>
<td>0.9981</td>
<td>0.9982</td>
<td>0.9982</td>
<td>0.9983</td>
<td>0.9984</td>
<td>0.9984</td>
<td>0.9985</td>
<td>0.9985</td>
<td>0.9986</td>
<td>0.9986</td>
</tr>
<tr>
<td>3.0</td>
<td>0.9987</td>
<td>0.9987</td>
<td>0.9987</td>
<td>0.9988</td>
<td>0.9988</td>
<td>0.9989</td>
<td>0.9989</td>
<td>0.9989</td>
<td>0.9990</td>
<td>0.9990</td>
</tr>
</tbody>
</table>
Appendix A, Part 3 – A standard deck of 52 cards

<table>
<thead>
<tr>
<th>Black cards</th>
<th>Red cards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clubs</td>
<td>Spades</td>
</tr>
<tr>
<td>A ♠</td>
<td>A ♠</td>
</tr>
<tr>
<td>2 ♠</td>
<td>2 ♠</td>
</tr>
<tr>
<td>3 ♠</td>
<td>3 ♠</td>
</tr>
<tr>
<td>4 ♠</td>
<td>4 ♠</td>
</tr>
<tr>
<td>5 ♠</td>
<td>5 ♠</td>
</tr>
<tr>
<td>6 ♠</td>
<td>6 ♠</td>
</tr>
<tr>
<td>7 ♠</td>
<td>7 ♠</td>
</tr>
<tr>
<td>8 ♠</td>
<td>8 ♠</td>
</tr>
<tr>
<td>9 ♠</td>
<td>9 ♠</td>
</tr>
<tr>
<td>10 ♠</td>
<td>10 ♠</td>
</tr>
<tr>
<td>Jack ♠</td>
<td>Jack ♠</td>
</tr>
<tr>
<td>Queen ♠</td>
<td>Queen ♠</td>
</tr>
<tr>
<td>King ♠</td>
<td>King ♠</td>
</tr>
</tbody>
</table>

Appendix A, Part 4 - Results for the total of two 6-sided dice

<table>
<thead>
<tr>
<th>+</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(1,1)</td>
<td>(1,2)</td>
<td>(1,3)</td>
<td>(1,4)</td>
<td>(1,5)</td>
<td>(1,6)</td>
</tr>
<tr>
<td>2</td>
<td>(2,1)</td>
<td>(2,2)</td>
<td>(2,3)</td>
<td>(2,4)</td>
<td>(2,5)</td>
<td>(2,6)</td>
</tr>
<tr>
<td>3</td>
<td>(3,1)</td>
<td>(3,2)</td>
<td>(3,3)</td>
<td>(3,4)</td>
<td>(3,5)</td>
<td>(3,6)</td>
</tr>
<tr>
<td>4</td>
<td>(4,1)</td>
<td>(4,2)</td>
<td>(4,3)</td>
<td>(4,4)</td>
<td>(4,5)</td>
<td>(4,6)</td>
</tr>
<tr>
<td>5</td>
<td>(5,1)</td>
<td>(5,2)</td>
<td>(5,3)</td>
<td>(5,4)</td>
<td>(5,5)</td>
<td>(5,6)</td>
</tr>
<tr>
<td>6</td>
<td>(6,1)</td>
<td>(6,2)</td>
<td>(6,3)</td>
<td>(6,4)</td>
<td>(6,5)</td>
<td>(6,6)</td>
</tr>
</tbody>
</table>
Appendix B – Glossary and Index

95% confidence statement – Page 120, Section 4.4
A confidence statement is a summary statement of the findings of a study. All confidence statements have the form ‘We are 95% confident that the true proportion of (parameter of interest) will be between (low value of confidence interval) and (high value of confidence interval).’

Back to Back Stem Plots – Page 184, Section 5.6
A stem plot in which two sets of numerical data share the stems in the middle, with one set having its leaves going to the right and the other set having its leaves going to the left.

Bar Graph – Page 137, Section 5.1
A graph in which each bar shows how frequently a given category occurs. The bars can go either horizontally or vertically. Bars should be of consistent width and need to be equally spaced apart. The categories may be placed in any order along the axis.

Bias - Page 95, 103, Section 4.1, 4.2
Bias occurs when a measurement repeatedly reports values that are either too high or too low.

Bin Width
See Class Size

Bivariate Data - Page 200, Section 6.1
Numerical data that measures two variables.

Blind Study - Page 126, Section 4.5
A study in which the subject does not know exactly what treatment they are getting.

Block Design - Page 128, Section 4.5
A study in which subjects are divided into distinct categories with certain characteristics (for example, males and females) before being randomly assigned treatments in an experiment.
**Box Plot (Box and Whisker Plot)** - Page 171, Section 5.5
A display in which a numerical data set is divided into quarters. The ‘box’ marks the middle 50% of the data and the ‘whiskers’ mark the upper 25% and lower 25% of the data.

**Categorical Variable** - Page 93, 136, Section 4.1, 5.1
Variables that can be put into categories, like favorite color, type of car you own, your sports jersey number, etc...

**Census** - Page 97, 101, Section 4.1, 4.2
A special type of study in which data is gathered from every single member of the population.

**Center** - Page 147, 156, Section 5.2, 5.3
Typically, it is the mean, median, or the mode of a data set. In a normal distribution curve the mean, median, and mode all mark the center. If a data set is skewed or has outliers, it is standard practice to use the median as the center.

**Chance Behavior** - Page 26, Section 2.1
Events whose outcomes are not predictable in the short term, but have long term predictability.

**Class Size (Bin Width)** - Page 164, Section 5.4
A consistent width that all bars on a histogram have. A quick estimation of a reasonable class size is to roughly divide the range by a value from about 7 to 10.

**Coincidence** - Page 215, Section 6.2
A relationship between two variables that simply occurs by chance.

**Combination** - Page 15, Section 1.4
An arrangement of a set of objects in which the order does not matter.
\[
\binom{n}{r} = \frac{n!}{r!(n-r)!}
\]

**Common Response** - Page 214, Section 6.2
A situation in which two variables have a strong correlation but are actually responding to an additional lurking variable.

**Complement of an Event** - Page 26, Section 2.1
The probability of an event, ‘A’, NOT occurring. It can be thought of the opposite of an event and can be notated as Ac or A'. P(A') = 1 – P(A)
Compound Event - Page 33, Section 2.2
An event with two or more steps such as drawing a card and then rolling a die.

Conditional Probability - Page 54, Section 2.5
The probability of a particular outcome happening assuming a certain prerequisite condition has already been met. A clue that a conditional probability is being considered is the word ‘given’ or the vertical bar symbol, |.

Confidence Interval - Page 119, Section 4.4
The range of answers included within the margin of error. Typically, we use a 95% confidence interval meaning it is very likely (95% chance) that the parameter lies within this range.

Confounding - Page 215, Section 6.2
Occurs when two variables are related, but it is not a clear cause/effect relationship because there may be other variables that are influencing the observed effect.

Context - Page 156, 204 Section 5.3
The specific realities of the situation we are considering. We often consider the labels and units when defining the context.

Contingency Table
See Two-Way Table

Control - Page 125, Section 4.5, 6.1
A researcher in an experiment establishes control when one of the treatment groups receives either a placebo or the currently accepted treatment.

Control Group - Page 125, Section 4.5
A group in an experiment that does not receive the actual treatment, but rather receives a placebo or a known treatment.

Convenience Sample - Page 106, Section 4.2
A biased sampling method in which data is only gathered from those individuals who are easy to access or are conveniently located.

Correlation (r) - Page 210-213, Section 6.2
A statistic that is used to measure the strength and direction of a linear correlation whose values range from -1 to 1. The sign of the correlation (+/-) matches the sign of the slope of the regression equation. A correlation value of 0 indicates no linear relationship whatsoever.
Data - Page 93, Section 4.1
A collection of facts, measurements, or observations about a set of individuals.

Density Curve - Page 236, Section 7.1
A curve that gives a rough description of a distribution. The curve is smooth and always has an area equal to 1 or 100%.

Direct Cause and Effect - Page 214, Section 6.2
A situation in which one variable causes a specific effect to occur with no lurking variables.

Direction - Page 210, Section 6.2
One of three general results reported for a linear regression. It will be reported as either being positive, negative, or 0.

Disjoint
See Mutually Exclusive Events

Dot Plot - Page 154, Section 5.3
A simple display that places a dot above each marked value on the x-axis. There is a dot for each result, so results that occur more than once will be shown by stacked dots.

Double Blind - Page 126, Section 4.5
A study in which neither the person administering the treatments nor the subject knows which treatment is being given.

Empirical Rule (68-95-99.7 Rule) - Page 238, Section 7.1
A rule stating that in a normal distribution, 68% of the data is located within one standard deviation of the mean, 95% of the data is located within two standard deviations of the mean, and 99.7% of the data is located within three standard deviations of the mean.

Event - Page 1, Section 1.1
Any action from which a result will be recorded or measured.

Expected Value - Page 67, Section 3.1
The average result over the long run for an event if repeated a large number of times.

Experiment - Page 97, 124, Section 4.1, 4.5
A study in which the researchers impose a treatment on the subjects.
Explanatory Variable - Page 125, 200, Section 4.5, 6.1
The x-axis variable. It can often be viewed as the 'cause' variable or the independent variable.

Factorial - Page 7, Section 1.2
A number followed by an exclamation point indicated repeated multiplication down to 1. For example, 4! = 4 × 3 × 2 × 1.

Fair Game - Page 76, Section 3.2
A game in which neither the player nor the house has an advantage. An average player over the long run will neither gain nor lose money. In other words, the expected value of the game is the same as the cost to play the game.

Five-Number Summary - Page 171, Section 5.5
A description of data that includes the minimum, first quartile, median, third quartile, and maximum numbers which can be used to create a box plot.

Form - Page 204, Section 6.1
A general description of the pattern in a scatterplot. Typical descriptions include linear, curved, or random (no specific form).

Frequency Table - Page 137, Section 5.1
A table that shows the number of occurrences in each category.

Fundamental Counting Principle - Page 4, Section 1.2
A rule that states that in order to find the number of outcomes for a multi-step event, simply multiply the number of possibilities from each step of the event.

Histogram - Page 164, Section 5.4
A special bar graph for a numerical data set. In a histogram, each bar has the same bin width and there is no space between consecutive bars. Each bar tracks the number or frequency of results in its given range.

Independent Events - Page 33, Section 2.2
Two events are independent if the outcome of one event does not change the probability for the outcome for the other event.

Individual - Page 93, Section 4.1
This is the person, animal, or object being studied.

Interquartile Range (IQR) - Page 174, Section 5.5
The distance between the lower and upper quartiles. IQR = Q₃ - Q₁
**Instrument of Measurement** - Page 94, Section 4.1
This is the tool used to make measurements. Some examples of instruments include rulers, scales, thermometers, or speedometers.

**Intersection of Events** - Page 42, Section 2.3
In a Venn Diagram, it includes the results that are members of more than one group simultaneously. We use the symbol, ∩, to indicate the intersection and think of the intersection as those parts of the diagram that include both A and B.

**Law of Large Numbers** - Page 26, 84, Section 2.1, 3.3
A rule that states that we will eventually get closer to the theoretical probability as we greatly increase the number of times an event is repeated.

**Line Graph**
See Time Plot

**Lurking Variable** - Page 124, 214, Section 4.5, 6.2
An additional variable that was not taken into account in a particular situation.

**Margin of Error** - Page 119, Section 4.4
It is the distance we move above and below the mean to help establish a 95% confidence interval in which we believe the true parameter is located. An approximation for the margin of error for a 95% confidence interval is M.O.E = ± 1/√n where n represents the sample size.

**Mean (Average)** - Page 147, 237, Section 5.2, 7.1
The sum of all the numbers divided by the number of values in a data set. It is also located at the center of a normal distribution and is a good measure of center for symmetric data sets.

**Median** - Page 147, Section 5.2
The data result in the middle of a data list that has been organized from smallest to largest. If there are two middle data values, then the median is located halfway between those two values. Visually, it marks the spot where half of the area of a graph is below the median and half of the area is above the median. It is common to use the median as your measure of center for skewed data sets or data sets that contain outliers.
Mode - Page 147, Section 5.2
The result that appears most frequently in a data set. It also occurs at the highest point of a density curve.

Multistage Random Sample - Page 104, Section 4.2
A sampling technique that uses randomly selected sub-groups of a population before random selection of individuals occurs.

Mutually Exclusive Events (Disjoint) - Page 41, Section 2.3
Outcomes that cannot occur at the same time. For example, if a single card is drawn from a standard deck, the outcomes of a diamond and a black card are mutually exclusive.

Negative Linear Association - Page 205, Section 6.1
A situation such that as one numerical variable increases, another numerical variable decreases.

Non-Response - Page 108, Section 4.2
A non-sampling error in which individuals selected for a study do not participate or do not answer questions in a survey.

Normal Distribution Curve - Page 237, Section 7.1
A bell-shaped curve that describes a symmetrical data set such that the most frequent results occur near the mean and results become less frequent as you move further from the mean.

Numerical Variable - Page 93, Section 4.1
A variable that can be assigned a numerical value, such as height, distance, or temperature.

Observational Study - Page 97, 124, Section 4.1, 4.5
A study in which researchers do not impose a treatment on the individuals being studied. Data is collected by observing the individuals, surveying the individuals, or collecting data from the individuals from information that is already available. (Observe but do not disturb)

Outcome - Page 1, Section 1.1
A possible result of an event.

Outlier - Page 155, 178, 204, Section 5.3, 5.5, 6.1
A value that is unusual when compared to the rest of a data set. High outliers will be greater than \( Q_3 + 1.5 \times IQR \). Low outliers will be below \( Q_1 - 1.5 \times IQR \).
**Parallel Box Plots** - Page 183, Section 5.6

Multiple box plots graphed on the same axes to compare multiple data sets.

**Parameter** - Page 111, Section 4.2

A value that describes the truth about a population. The value is frequently unknown so a parameter is often given as a description of truth.

**Permutation** - Page 10, Section 1.3

A specific order or arrangement of a set of objects or items. In a permutation, the order in which the items are selected matters.

**Pictograph** - Page 141, Section 5.1

A bar graph that uses pictures instead of bars. These graphs can be misleading because pictures measure height and width, where bar graphs measure only height. To be effective, all the pictures used must be the same size.

**Pie chart** - Page 139, Section 5.1

A graph which shows each category as a part of the whole in a circle graph. Pie charts can be used if exactly 100% of the results from a particular situation are known.

**Placebo** - Page 126, Section 4.5

A fake treatment that is similar in appearance to the real treatment.

**Placebo Effect** - Page 126, Section 4.5

The placebo effect occurs when a subject starts to experience changes simply because they believe they are receiving a treatment.

**Population** - Page 101, Section 4.2

The entire group of individuals we are interested in. A population is often described using the word ‘all’.

**Positive Linear Association** - Page 205, Section 6.1

A situation in which as one numerical variable increases, the other numerical variable also increases.

**Prime Number** - Page 42, Section 2.3

A number that has exactly 2 factors. Remember, 1 is not a prime number!

**Probability** - Page 26, Section 2.1

The likelihood of a particular outcome occurring.
**Probability Model** - Page 49, Section 2.4
A table that lists all the values for the outcomes of an event and their respective probabilities. The sum of all the probabilities in a probability model must equal 1.

**Processing Errors** - Page 109, Section 4.2
An error commonly made due to issues like poor calculations or inaccurate recording of results.

**Prospective Studies** - Page 124, Section 4.5
A study which follows up with study subjects in the future in an effort to see if there were any long-term effects.

**Quartile 1** - Page 172, Section 5.5
The median of all the values to the left of the median. Do not include the median itself in this calculation if the median is one of the data points.

**Quartile 3** - Page 172, Section 5.5
The median of all the values to the right of the median. Do not include the median itself in this calculation if the median is one of the data points.

**Random Digit Table** - Pages 82, 114, Section 3.3, 4.3, Appendix A
A long list of randomly chosen digits from 0 to 9, usually generated by computer software or calculators. A table of random digits can be found in Appendix A, Part 1.

**Random Event** - Page 26, Section 2.1
An event is random if it does not have short-term predictability but it has long-term predictability. For example, a coin flip is a random event because we do not know what will happen on the next flip, but we can be reasonably sure that about 50% of a long series of flips will land on heads.

**Random Sampling Error** - Page 107, Section 4.2
Even though a sample is randomly selected, it is entirely possible that a particular result within the population will be over-represented causing us to be significantly different from the parameter. Larger sample sizes reduce random sampling error. The margin of error is stated with most studies to account for random sampling error.

**Range** - Page 148, 174, Section 5.2, 5.5
A basic description of how spread out a data set is. It is calculated by subtracting the smallest number from the largest number in a data set.
Reliability - Page 95, Section 4.1
How consistently a particular measurement technique gives the same, or nearly the same measurement.

Response Bias - Page 109, Section 4.2
Occurs when an individual responds to a survey with an incorrect or untruthful answer. This type of bias can frequently happen when questions are potentially sensitive or embarrassing.

Response Variable - Page 125, 200, Section 4.5, 6.1
This is the y-axis variable. It can often be thought of as the ‘effect’ variable or dependent variable.

Retrospective Study - Page 124, Section 4.5
A study in which information about a subject’s past is used in the study.

Sample - Page 102, Section 4.2
A representative subset of a population.

Sample Space - Page 1, Section 1.1
A list of all the possible outcomes that may occur.

Sample Survey - Page 97, Section 4.1
A survey that uses a subset of the population in order to try to make predictions about the entire population.

Sampling Frame - Page 103, Section 4.2
A list of all members of a population.

Scatterplot - Page 200, Section 6.1
Graphs that represent a relationship between two numerical variables where each data point is shown as a coordinate point on a scaled grid.

SCOFD - Page 203-206, Section 6.1
This is an acronym used for the description of a scatterplot and stands for Strength, Context, Outliers, Form, and Direction.

Simple Random Sample (SRS) - Page 103, Section 4.2
A sample where all possible groups of a particular size are equally possible. It can be thought of as putting names of all members of a population in a hat and randomly drawing until the desired sample size is reached.
**Simulation** - Page 82, Section 3.3
A model of a real situation that can be used to make predictions about what might really happen. Often, tables of random digits are used to carry out simulations.

**Skewed Distribution** - Page 155, 236, Section 5.3, 7.1
A distribution in which the majority of the data is concentrated on one end of the distribution. Visually, there is a ‘tail’ on the side with less data and this is the direction of the skew.

**SOCCS** - Page 154-156, Section 5.3
An acronym used to remember the key information to discuss for a distribution: Shape, Outliers, Center, Context, and Spread.

**Spread** - Page 156, Section 5.3
A way to measure variability of a data set. Common measures of spread are the range, standard deviation, and IQR.

**Standard Deviation** - Page 174, 237, Section 5.5, 7.1
A measure of spread relative to the mean of a data set. Use this measurement for any data set which is approximately normally distributed.

**Statistic** - Page 111, Section 4.2
A number that describes results from sample. This number is often a percentage and is used to make an approximation of the parameter.

**Stem Plot** - Page 157, Section 5.3
A method of organizing data that sorts the data in a visual fashion. The stem is made up of all the leading digits of a piece of data and the leaf is the final digit. No commas or decimal points should be used in a stem plot.

**Stratified Random Sample** - Page 104, Section 4.2
A sample in which the population is divided into distinct groups called strata before a random sample is chosen from each strata.

**Strength** - Page 203, 210, Section 6.1, 6.2
One of three measurements reported for a best-fit line that describes how close the data is to being perfectly linear.

**Subjects** - Page 125, Section 4.5
The individuals that are being studied in an experiment.
**Symmetrical Distribution** - Page 155, Section 5.3
A distribution in which the left side of the distribution looks like a mirror image of the right side of the distribution.

**Systematic Random Sample** - Page 104, Section 4.2
A sampling method in which the first selection is made randomly and then a 'system' is used to make the remaining selections. For example, randomly select one person from a list and then select every 14th person after that.

**Theoretical Model** - Page 26, 82 Section 2.1, 3.3
A model that gives a picture of exactly the frequencies of what should happen in a situation involving probability.

**Theoretical Probability** - Page 26, Section 2.1
A mathematical calculation of the likelihood that a given outcome will occur.

**Time Plot (Line Graph)** - Page 145, Section 5.2
A graph that shows how a numerical variable changes over time.

**Tree Diagram** - Page 2, 4, 48 Section 1.1, 1.2, 2.4
A visual representation of a multi-step event where each successive step branches off from the previous step.

**Two-Way Table (Contingency Table)** - Page 55, Section 2.5
A table which tracks two characteristics from a set of individuals. For example, we might track gender and grade of all the students in your high school.

**Undercoverage** - Page 107, Section 4.2
A sampling error in which an entire group or groups of subjects are left out or underrepresented in a study.

**Union of Events** - Page 41, Section 2.3
A union includes all results that are in either one category, another category, or both categories in a Venn diagram. We use the symbol $\cup$ and can think of a union as anything belonging to either A, B, or both A and B.

**Validity** - Page 95, Section 4.1
A measurement technique is valid if it is a reasonable way to collect data.

**Variables** - Page 93, Section 4.1
Characteristics about the individuals in a study in which researchers might have interest.
Venn Diagrams - Page 29, 42, Section 2.1, 2.3
Diagrams that represent outcomes or categories using intersecting circles.

Voluntary Response Survey - Page 105, Section 4.2
A biased sampling method in which participants get to choose whether or not to participate in the survey. The bias occurs because those who are most passionate about an issue will be more likely to respond.

Wording of a Question - Page 108, Section 4.2
The wording of a question can be used to manipulate individuals in a survey such that they are more likely to respond a certain way in the survey which causes bias.

Z-Score - Page 245, Section 7.2
A measure of the number of standard deviations a particular data point is away from the mean in a normal distribution. If a z-score is positive, the value is larger than the mean and if it is negative, it is less than the mean.
Appendix C – Calculator Help

This appendix is not meant to be a full guide for calculators common to students who take this course. Rather, it is intended to highlight some of the locations to access a variety of commands commonly used on a TI-30XS Multiview Scientific Calculator and a TI-84 Plus Graphing Calculator. One online source that can be helpful for those of you with graphing calculator issues can be found on the Prentice Hall website at http://www.prenhall.com/divisions/esm/app/calc_v2/.

Topic 1 - Combinations, Permutations, and Factorials

TI-30 XS Multiview

Access located in the prb menu. Enter the value for n, select nCr or nPr, and then enter the value for r.

TI-84 Plus

Access located in the Math, PRB menu. Enter the value for n, select nCr or nPr, and then enter the value for r.

Topic 2 – Random Number Generators

TI-30 XS Multiview

Access located in the prb menu. Select rand, enter lowest value, enter highest value.

TI-84 Plus

Access located in the Math, PRB menu. Select RandInt, enter lowest value, enter highest value, enter number of random values desired.
Topic 3 – Means and Standard Deviations

TI-30 XS Multiview

Enter data into L in the data menu. Press 2nd data (stat) and select 1-Var Stats. Arrow down to find the mean, \( \bar{x} \), and the standard deviation, \( s_x \).

TI-84 Plus

Enter data in L1 by selecting STAT and EDIT. Press STAT and CALC and then select 1-Var Stats. Arrow down to find the mean, \( \bar{x} \), and the standard deviation, \( s_x \).

Topic 4 – Correlations, Slopes, and Y-Intercepts

TI-30 XS Multiview

Enter data into L1 and L2 in the data menu. Press 2nd data (stat) and select 2-Var Stats for L1 and L2. Arrow down to find the slope (a), the y-intercept (b) and the correlation coefficient (r).

TI-84 Plus

Enter data in L1 and L2 by selecting STAT and EDIT. Press STAT and CALC and then select LinReg(ax+b). Be sure the Xlist and Ylist are L1 and L2. If you wish to store you equation into the Y= menu, press VARS, Y-VARS, Function, and Y1. If the correlation (r) does not show up, go to 2nd CATALOG and select DiagnosticOn.

Topic 5 – Normal Distributions

TI-30 XS Multiview

This calculator cannot perform normal distribution calculations.

TI-84 Plus

To find the percent of area in a normal curve, select 2nd DISTR and select normalcdf(. Enter the lower bound, upper bound, mean, and standard deviation. To find a value from a percentile in a normal distribution, select 2nd DISTR and select invNorm(. Enter the %tile, mean, and standard deviation.

Image References

Random Digit Table   http://uwsp.edu/math
NormalDistributionTable http://www.regentsprep.org
TI-30XS Multiview Calculator   http://education.ti.com
TI-84 Plus Graphing Calculator   http://education.ti.com
Appendix D – Selected Answers

Problem Set 5.1
1a) Bar Graph
1b) Possibly
1c) 12.3%, 44.3°; 14.5%, 27.0°; 12.0%, 43.3°; 6.0%, 21.6°; 10.1%, 36.4°
1d) Pie Chart
2) Answers will vary
3a) – c) Answers will vary
4a) – c) Answers will vary
5a) Answers will vary
5b) Bar Graph
6a) – c) Answers will vary
7a) Players
7b) # = Number, POS = Position, GP = Games Played, G = Goals, A = Assists, P = Points, +/- = Plus/Minus Rating, PIM = Penalty Minutes, PP = Power Play Goals, SH = Shorthanded Goals, GW = Game Winning Goals, S = Shots, S% = Shooting Percentage
7c) Cat. = POS, # = Neither, Num. = All Other Variables
8) 32,768
9) 0.0000305

Problem Set 5.2
1a) \( \mu = 34.33, \text{ Med} = 29, \text{ No Mode, Range} = 50 \)
1b) \( \bar{x} = 44.57, \text{ Med} = 48, \text{ Mode} = 22, \text{ Range} = 54 \)
1c) \( \bar{x} = 62.80, \text{ Med} = 62, \text{ No Mode, Range} = 100 \)
2) 171.6 lbs.
3a) 61.2 feet
3b) 38.4 feet
3c) 4.98 feet
4) Median = 32 or \( \bar{x} = 31 \)
5a) \( \bar{x} = 63.4, \text{ Med} = 70.5, \text{ No Mode, Range} = 72, D \)
5b) \( \bar{x} = 70.9, \text{ Med} = 70.5, \text{ Mode} = 70, \text{ Range} = 24, \text{ Mean Changed, C} \)
5c) \( \bar{x} = 70.3 \)
5d) 74.2
6a) $5.750; $5.150; $2.760; $3.100; $2.500; $5.450; $1.280; $3.130; $2.350; $3.675
6b) 30.7%; 24.2%; 13.9%; 14.0%; 17.6%; 24.6%; 9.8%; 12.3%; 8.6%; 12.8%
6c) \( \bar{x} = 16.9\%, \text{ Med} = 14.0\%, \text{ No Mode, Range} = 22.1\% \)
6d) Answers will vary
7) Answers will vary
8a) Time Plot
8b) Answers will vary
9) 30,240
10) 210
11) 220
12) 168
13) 990

Problem Set 5.3
1a) \( \bar{x} = 65, \text{ Med} = 70, \text{ Mode} = 70, \text{ Range} = 64 \)
1b) Answers will vary
2a) Stem Plot
2b) Answers will vary
3a) Split-Stem Plot
3b) Med = 72.5%
3c) Lower
3d) Answers will vary
4a) Split-Stem Plot
4b) Answers will vary
4c) Answers will vary
5a) \( \bar{x} = 75.62, \text{ Med} = 77, \text{ Mode} = 92 \)
5b) Median
6a) i) Skewed Left
   ii) Roughly Symmetrical
   iii) Roughly Symmetrical
   iv) Skewed Right
6b) i) Median ii) Median
   iii) Similar iv) Mean
6c) Answers will vary
7a) Stem Plot
7b) \( \bar{x} = 63.4, \text{ Med} = 71, \text{ Mode} = 82, \text{ Range} = 66 \)
7c) Answers will vary
8a) Plus/Minus
8b) Dot Plot
8c) Answers will vary
9a) All Springfield Res.
9b) True % who like Simp.
9c) 1,245 Springfield Res.

Problem Set 5.4
1a) 450
1b) $11,000
1c) Answers will vary
2a) Answers will vary
2b) 2 or 137 to 139 lbs.
2c) 76.7%
3a) Over 75 bin wrong
3b) Answers will vary
4a) Histogram
4b) Answers will vary
5a) – e) Sketches, Answers will vary
6a) Histogram
6b) Median
6c) \( \bar{x} = 71.52, \text{ Med} = 74, \text{ Mode} = 82, \text{ Range} = 35 \)
6d) Median
7a) Probability Model
7b) -.6
7c) No
8a) 0.191
8b) 0.375
8c) 0.078
9a) 0.083
9b) 0.525
9c) 0.55

Problem Set 5.5
1a) Med = 210, IQR = 130
1b) Med = 85, IQR = 55
2a) \{240, 340, 440, 600, 750\} Box Plot
2b) No Outliers
2c) Answers will vary
3a) Box Plot
3b) Med = 1030, IQR = 500
3c) \( \bar{x} = 992, SD = 291.43 \)
3d) Yes
4a) \{2, 11, 31.5, 56, 206\} Box Plot
4b) One High Outlier = 206
4c) \( \bar{x} = 48.8, SD = 54.7, \text{ Mean is larger than Median} \)

Appendix D – Selected Answers | 287
Appendices

4d) \( \{2, 11, 20, 54.5, 106\} \bar{x} = 36.7, SD = 31.9, \text{Median, Mean, Q3, Max, and SD all changed} \)
5a) Box Plot
5b) Answers will vary
5c) Answers will vary
6a) \( \{1142, 1215.5, 1371.5, 1601, 2717\} \) Box Plot
6b) Outlier is Burj Khalifa
6c) Skewed Right
6d) 1215.5 and 1601 feet
6e) Range = 109, IQR = 27
6f) Outlier is Burj Khalifa
7a) \( \{48, 66.5, 77, 86, 97\} \) Box Plot
7b) Answers will vary
7c) Answers will vary
7d) Answers will vary
8a) Bias
8b) Validity
8c) Reliability
9) Tree Diagram
10a) 0.125
10b) 0.375
10c) 0.875

Problem Set 5.6
1a) \% Fat = \{12, 22, 36, 45, 66\}, \% Sat. Fat = \{15, 29, 49, 63, 96\}
1b) Box Plots
1c) Answers will vary
2a) Boys: SD = 3.8, Range = 12, IQR = 7; Girls: SD = 3.2, Range = 11, IQR = 5
2b) Answers will vary
2c) Boys: \( \bar{x} = 69.1, \) Med = 69, Mode = 67 & 73; Girls: \( \bar{x} = 63.2, \) Med = 63.5, Mode = 61, 64, & 66
2d) Answers will vary
2e) Answers will vary
3a) Back to Back Stem Plot
3b) Class 3 = \{14, 24, 35, 38.5, 51\}; Class 4 = \{20, 27, 31, 38, 46\}
3c) Class 3: \( \bar{x} = 32.0, SD = 9.8, \) Mode = 35 & 37, Range = 37, IQR = 14.5; Class 4: \( \bar{x} = 32.9, SD = 6.5, \) Many Modes, Range = 26, IQR = 11
3d) Answers will vary
4a) BR: \( \bar{x} = 43.9, SD = 11.3, \) IQR = 19, \( \{22, 35, 46, 54, 60\} \); MM: \( \bar{x} = 38.7, SD = 18.1, \) IQR = 23, \( \{9, 29, 39, 52, 70\} \); BB: \( \bar{x} = 36.1, SD = 13.5, \) IQR = 19, \( \{16, 25, 34, 44, 73\} \); RM: \( \bar{x} = 26.1, SD = 15.6, \) IQR = 19, \( \{8, 14, 24.5, 33, 61\} \)
4b) Box Plots
4c) One outlier only for BB at 73 home runs
4d) Answers will vary
4e) Answers will vary
5a) Answers will vary
5b) Answers will vary
5c) Answers will vary
5d) Answers will vary
5e) Answers will vary
6a) Graph I: Sym. & Bell; Graph II: Sym. & Bell; Graph III: Sym. & Bimodall Graph IV: Sym. & Uniform
6b) Center = 52 for all
6c) Mean, Median, Range
6d) Graph I = Smallest SD
6e) Graph III = Largest SD
6f) Mean & SD on I and II, 5-# Summary on III and IV

Chapter 5 Review
1) C
2a) \( \bar{x} = 16.592, SD = 0.165, \) Mode = 16.6, Range = 0.8
2b) \( \{16.1, 16.5, 16.6, 16.7, 16.9\} \) Box Plot
2c) Dot Plot
2d) 11.5%
3a) Pie Chart
3b) Answers will vary
3c) Answers will vary
4) B
5) D
6) A
7) C
8) A
9) B
10) A
11) D
12) E
13) B
14) C
15) A
16) E
17) B
18a) Stem Plot
18b) Tornadoes: \( \{48, 58, 62, 71, 90\}, \bar{x} = 63.133, SD = 10.315; \) Bengals: \( \{38, 66, 73, 84, 95\}, \bar{x} = 73, SD = 14.147 \)
18c) Parallel Box Plots
18d) Answers will vary
18e) Answers will vary
19a) Time Plots
19b) Answers will vary
19c) Answers will vary
19d) 1997 = 0.2; 2005 = 0.75; 2018 = 1.15
19e) 2000 = 0.7; 2015 = 0.25
20a) Histogram
20b) \( \{79.6, 81.9, 87.7, 89.4, 91.3\} \)
20c) No Outliers
20d) Box Plot
20e) Range = 11.7, IQR = 7.5, Mode = 73 & 85
20f) \( \bar{x} = 86.27, SD = 4.23 \)
20g) Med is 1.4 larger than \( \bar{x} \)
20h) Answers will vary
20i) Answers will vary
20j) 1st
21a) Split Stem Plot
21b) \( \{8, 18, 20, 22, 50\} \)
21c) 8, 9, 10, 10, 29, 30, 35, 40, 40, 50
21d) Box Plot
21e) Range = 42, IQR = 4, Mode = 20
21f) \( \bar{x} = 20.90, SD = 7.65 \)
21g) \( \bar{x} \) is 0.9 larger than Med
21h) 5-number summary
21i) Answers will vary

Appendices