Lesson Plan: Basic Descriptive Stats for 5th grade

The following is a lesson on basic descriptive statistics for Juniata’s MSP Grant, aimed at 5th grade Elementary School teachers. Some residual benefits include exposure to the StatKey web-site.

Lock5 homepage:  http://www.lock5stat.com/
StatKey linked off Lock5 homepage:  http://www.lock5stat.com/statkey/index.html
Descriptive Stats for one quantitative variable:
http://www.lock5stat.com/statkey/descriptive_1_quant/descriptive_1_quant.html
Descriptive Stats for one quantitative and one categorical variable:
http://www.lock5stat.com/statkey/descriptive_1_quant_1_cat/descriptive_1_quant_1_cat.html

Begin the lesson by asking everyone in class to take their temperature and write it down on a notecard. Write a number line on the board, and then ask the class to place their cards, in position, on the chalk tray, and place a dot on the number line as well.

96.0  96.25  96.50  96.75  97.0  97.25  97.50  97.75  98.0  98.25  98.50  98.75  99.0

Define the median, and then with the data on the board, locate the median temperature value. Spend some time discussing calculating the median for data with even N or odd N.

Next, write an obvious outlier, 95.0 or 104.0 for example, on a card, and place in the proper place. How does the outlier affect the median? Remove the outlier, but save the card for later.

Try to estimate the median of the values to the left (smaller) of the overall median. Calculate it with and without the “middle” value. Discuss how Q1 is the median of the left half of the data, but note that there isn’t a consensus whether the middle value is included. Then calculate the median of the larger half of the data and compare that to Q3. Have a discussion on why Q1 and Q3 are called that (note that sometimes they are referred to as the upper and lower quartiles). Ask the students what another name for Q2 is.

Draw a box plot on the board, using the median, Q1 and Q3 calculated above. Write it high enough above the number line to leave space for a histogram.

The mean is another way to calculate average. Define it (the sum of the data values, divided by the number of data values). Without calculating the mean, estimate it. It’s sometimes referred to as the “balance-beam.” How does the mean change if the outlier is added to the board? Compare the change in the mean to the change in the median when the outlier is included.
Next, demonstrate how to enter the temperatures into the “1 Quantitative Variable” choice in StatKey:
http://www.lock5stat.com/statkey/descriptive_1_quant/descriptive_1_quant.html

Ask how the dotplot compares to the picture on the board. Next, click on the box plot and compare it to the picture on the board. How do the Q1, median, and Q3 calculated in StatKey compare to the values on the board.

Finally, choose the histogram options in StatKey. Spend a bit of time showing how the histogram is constructed, using the equal intervals and counting the dots in them, and show this on the board. StatKey will use the min and max as the endpoints for the first and last intervals (StatKey calls them “buckets”), and the Set Limits button allows the user more control over the intervals. Remind the students that histograms are calculated on one quantitative variable, and that they have equally spaced intervals, so the height of the bars can be compared.

Ask the students to enter the data in StatKey. Let the students find the web-page (Google “Lock 5”) and enter the data. Have the class edit the data and add the outlier temperature. Note the changes in the mean and median calculated by StatKey.

For more practice in StatKey, consider the temperature data included in StatKey. How does it compare to the class’ data? The data can be found in the section “Confidence Interval for a Mean, Median, Std. Dev,” at the URL: http://www.lock5stat.com/statkey/bootstrap_1_quant/bootstrap_1_quant.html. To find it, click on the upper-left button for the data, and select Temperature. Then to capture it, click on “Edit Data” and copy it and paste it into the data for descriptive stats.

Next, we’d like to explore two different datasets, and compare their box plots.

One convenient set of data to use is scores from the National Football League (NFL). These are available on the www.nfl.com web-page, and it’s nice to compare two teams. Ask the class to choose a team and enter their point totals from each of their sixteen regular season games. For example, the number of points the Chicago Bears scored in each of their sixteen 2013 regular season games are listed below. To find these, click on a team’s icon at the top of www.nfl.com, find their schedule, and be sure to choose the 2013 season. Enter each score into StatKey, and calculate the descriptive statistics for their point totals, as done above for the temperature data.

Next, we would like to compare the box plots of the scores of two different of the teams, and maybe make a prediction on how those teams might have fared vs. each other. To do this, we’ll need to use StatKey’s:
Descriptive Stats for one quantitative and one categorical variable
http://www.lock5stat.com/statkey/descriptive_1_quant_1_cat/descriptive_1_quant_1_cat.html

This requires two columns of data. The first is categorical, which in this case is team name, and the second column will be the score. StatKey wants these delimited somehow, usually by spaces, commas,
or tabs. Note that if the data is originally entered in with spaces or tabs as delimiters, StatKey will add in commas.

So if we wanted to compare the Chicago Bears and the Minnesota Vikings, the data in StatKey would look like:

Bears, 11
Bears, 18
Bears, 19
Bears, 20
Bears, 21
Bears, 23
Bears, 24
Bears, 27
Bears, 27
Bears, 28
Bears, 31
Bears, 32
Bears, 38
Bears, 40
Bears, 41
Bears, 45
Vikings, 7
Vikings, 10
Vikings, 14
Vikings, 14
Vikings, 20
Vikings, 23
Vikings, 23
Vikings, 24
Vikings, 26
Vikings, 26
Vikings, 27
Vikings, 30
Vikings, 31
Vikings, 34
Vikings, 34
Vikings, 48
Here’s the comparison dotplot produced by StatKey for these two teams:

Help the students analyze the two box plots. Ask leading questions like:
Which of the two teams do you think had a better season?
If these two teams played, who do you expect to win?
What on the box plots did you use to help predict this?
Did your two teams play? If so, what was the result?
Does this agree with your prediction? Why or why not?

If there is still time, discuss more descriptive stats for data spread, like range and IQR, and compare these.
### Appendix: Points Scored by 2013 NFL Teams

<table>
<thead>
<tr>
<th>Team</th>
<th>Points Scored</th>
<th>Team</th>
<th>Points Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bears</td>
<td>24</td>
<td>Vikings</td>
<td>24</td>
</tr>
<tr>
<td>Bears</td>
<td>31</td>
<td>Vikings</td>
<td>30</td>
</tr>
<tr>
<td>Bears</td>
<td>40</td>
<td>Vikings</td>
<td>27</td>
</tr>
<tr>
<td>Bears</td>
<td>32</td>
<td>Vikings</td>
<td>34</td>
</tr>
<tr>
<td>Bears</td>
<td>18</td>
<td>Vikings</td>
<td>10</td>
</tr>
<tr>
<td>Bears</td>
<td>27</td>
<td>Vikings</td>
<td>7</td>
</tr>
<tr>
<td>Bears</td>
<td>41</td>
<td>Vikings</td>
<td>31</td>
</tr>
<tr>
<td>Bears</td>
<td>27</td>
<td>Vikings</td>
<td>23</td>
</tr>
<tr>
<td>Bears</td>
<td>19</td>
<td>Vikings</td>
<td>34</td>
</tr>
<tr>
<td>Bears</td>
<td>23</td>
<td>Vikings</td>
<td>20</td>
</tr>
<tr>
<td>Bears</td>
<td>21</td>
<td>Vikings</td>
<td>26</td>
</tr>
<tr>
<td>Bears</td>
<td>20</td>
<td>Vikings</td>
<td>23</td>
</tr>
<tr>
<td>Bears</td>
<td>45</td>
<td>Vikings</td>
<td>26</td>
</tr>
<tr>
<td>Bears</td>
<td>38</td>
<td>Vikings</td>
<td>48</td>
</tr>
<tr>
<td>Bears</td>
<td>11</td>
<td>Vikings</td>
<td>14</td>
</tr>
<tr>
<td>Bears</td>
<td>28</td>
<td>Vikings</td>
<td>14</td>
</tr>
</tbody>
</table>