Algebra 1

Sept/Oct

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Mon | | Tue | | Wed | Thu | Fri |
|  |  |  | |  | |  |  |
|  |  | |  | |  |  |  |
|  |  | |  | |  |  |  |
|  |  | |  | |  |  |  |

|  |  |
| --- | --- |
| **Unit 4: Statistics** | |
| **ENDURING UNDERSTANDINGS** | |
| * Visual displays of qualitative and quantitative data can be understood in terms of their general shape, center, and spread. * The general shape of data can help us select appropriate statistics. * Sets of data can be compared with respect to shape, center, and spread to help us make decisions. | |
| ESSENTIAL QUESTIONS | KEY CONCEPTS |
| * Given a set of data, what is the best way to summarize and represent the data? * How can the shape and spread of two or more data sets be compared? * How can the shape of the data help determine the appropriate measure of center and spread? * How can statistics and data representations be used to support or refute a claim? * How can categorical data be summarized? * How can relative frequencies be interpreted in the context of the data? | 1. Data can be represented with the use of dot plots, box plots, and histograms. 2. If data is skewed or when meaningful outliers are present, the median and 5-Number summaries should be used to describe the data. 3. If the data is unimodal or symmetric, the mean and standard deviation are appropriate to describe the data. 4. The standard deviation tells us how much an individual measurement should be expected to deviate from the mean; a smaller standard deviation demonstrates consistency in the data. 5. Relative frequencies can reveal possible associations and trends in data. |
| STUDENT FRIENDLY OBJECTIVES | ACADEMIC VOCABULARY |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| * TSW create and interpret dot plot, box plots, and histograms. * TSW compare the shape, center, and spread of two or more data sets in various representations. * TSW examine the shape of data to determine the appropriate measure of center and spread. * TSW calculate statistics, including standard deviation, using technology. * TSW create and interpret two-way frequency tables. | | |  |  |  | | --- | --- | --- | | shape | outlier | unimodal distribution | | center | dot plot | bimodal distribution | | lower quartile | histogram | quantitative data | | upper quartile | skew | variation | | interquartile range | skewed lift | spread | | 5 number summary | left-tailed | two-way frequency | | mean | skewed right | table | | median | right-tailed | joint frequencies | | standard deviation | symmetric distribution | marginal frequencies | | statistic | range | conditional relative | | minimum | maximum | frequencies | |
| STANDARDS COVERED | | |
| Statistics & Probability | **S-ID.1.** Represent data with plots on the real number line (dot plots, histograms, and box plots). | |
| **S-ID.2.** Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. | |
| **S-ID.3.** Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers). | |
| **S-ID.5.** Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data. | |

**Exit Outcomes Skills:**

* I keep track of my assignments and important dates on my calendar or on my planner.
* I take notes (teacher prepared & impromptu), complete assignments & vocabulary on time.
* I set goals in my planner.
* I manage my time using my planner.
* I keep an organized notebook.
* I self reflect on my performance in my class using my notes, check for understanding, reflections, and activities.
* I ask for help when I need it (teacher, learning team, parent/guardian, HW Lab).
* I spent extra time using the additional resources provided.
* I am an active participate and an asset within my learning team.

**Goal Tracking Sheet:** \*This should be filled in each time feedback is given.  \* Use the exit outcome skills as a guideline to reflect on your results

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ #:\_\_\_\_\_\_\_\_

**PLAN**

**Student Goal** My goal is to achieve a\_\_\_\_\_\_\_\_\_\_ score on the “Chapter 4: Statistics” unit test.

**Action Steps** List at least 3 things you plan on doing to be successful in reaching your goal this unit.

(i.e. use flashcards to study and understand vocabulary, reflect on CFU problems that I miss, seek extra help when I do not understand a concept, ask question in class when I can’t find my mistake, try/preserver through all my CFU problems, use additional resources provided to make sure I understand topic, review pretest, etc.)

**STUDY**

**Act**

**Do**

|  |
| --- |
| **Suggestions for improvement:** |
| **Student (What will I do)** |

|  |  |  |
| --- | --- | --- |
| **Activities/**  **Assignments** | **Unit Concepts** | [https://encrypted-tbn0.google.com/images?q=tbn:ANd9GcQdpVrlS9P3GVQ2sGKC13_xyr8FydBnepuuiVQQvN_dUIAmc2ei7A](http://www.google.com/imgres?um=1&hl=en&biw=1024&bih=587&tbm=isch&tbnid=2nocVW5ycs7bkM:&imgrefurl=http://en.wikipedia.org/wiki/File:OCR-A_char_Plus_Sign.svg&docid=JD5k8ALWiXzKZM&imgurl=http://upload.wikimedia.org/wikipedia/commons/3/30/OCR-A_char_Plus_Sign.svg&w=744&h=1052&ei=YdbYT4v_MYHg2QXL6o2kDw&zoom=1&iact=hc&vpx=219&vpy=2&dur=906&hovh=267&hovw=189&tx=99&ty=181&sig=117890709682966998579&page=1&tbnh=109&tbnw=74&start=0&ndsp=21&ved=1t:429,r:1,s:0,i:155)**Was it helpful**  **/** |
| Reflection 4.1 | #2 |  |
| Reflection 4.2 | #2 |  |
| Reflection 4.4 | #1, #2, #5 |  |
| Reflection 4.5 | #1 – 4 |  |
| Reflection 4.6 | #1 - 4 |  |
| LT Task #1 |  |  |
| LT Task #2 |  |  |

**STUDY**

Was I on track to meet my goal (use the tracking paper to support your response)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Did I meet my goal? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What concepts #’s were my best scores? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ What concepts #’s were my lowest scores? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What action steps caused there to be a difference? Why? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What should I try next unit to be more successful? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Give feedback for the teacher. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |
| --- | --- |
| **4.1 Measures of Central Tendencies** | |
| **Notes** | DVUSD Algebra Flexbook: pg 41 – 48.  Watch the video for an explanation of mean, median and mode.  <http://www.ck12.org/flx/render/embeddedobject/53382>  Vocabulary  Mode: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Mean \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Median \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Outlier\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Resistant\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Example A  Find the mode for the number of children per house in the data set at the beginning of the Concept.    More Than One Mode: If there is more than one mode in a data set it is called \_\_\_\_\_\_\_\_\_\_\_  When a data set is described as being \_\_\_\_\_\_\_\_\_\_\_\_, it is clustered about \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  If there were more than two, they would \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  However, the more of them there are, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. If there is an equal number of eqch data value, the mode is \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_ in helping us understand the data, and thus we say \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Example B  Find the mean for the number of children per house.  Statisticians use the symbol \_\_\_\_\_\_\_\_ to represent the mean when \_\_\_\_\_\_\_\_ is the symbol for a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Read \_\_\_\_\_\_\_\_\_\_ as \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  The formula for the sample is: =  \_\_\_ is a statistic, since it is a \_\_\_\_\_\_\_\_\_\_\_, and \_\_ \_\_\_is a parameter, since it is a \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Example C  Consider the following quiz scores: 91, 83, 97, 89  Place them \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  To find the median, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The median will be the data point that is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ In this case we take the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the two middle numbers.  **Mean vs. Median**  Both the mean and the median are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_measures of center. Suppose you got and 85 and a 93 on your first two statistics quizzes, but then you had a really bad day and got a 14 on your next quiz!  The mean of your three grades would be \_\_\_\_\_\_\_\_\_\_\_. Which is a better measure of your performance? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The middle does not change if the lowest grade is an 84, or if the lowest grade is a 14. However, when you add the three numbers to find the \_\_\_\_\_\_\_\_\_ the sum will be much smaller if the lowest grade is a 14.  **Outliers and Resistance**  The mean and the median are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in this example because there is on grade that is \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ from the rest of the data. In statistics we call such extreme values \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. A statistic that is not affected by outliers is called \_\_\_\_\_\_\_\_\_\_\_\_\_. We say the median is a resistant measure of center, and the mean is not resistant. In a sense, the median is able to \_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_ of a far away value, but the mean is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_. It cannot resist the influence of outlier values. As a result, when we have a data set that \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_, it is often better to use the \_\_\_\_ \_\_\_\_\_\_\_ to describe the center, rather than the \_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **Check for Understanding** | 1. In Lois’ 2nd grade class, all of the students are between 45 and 52 inches tall, except one boy, Lucas, who is 62 inches tall. Which of the following statements is true about the heights of all the students? 2. The mean height and the median height are about the same. 3. The mean height is greater than the median height. 4. The mean height is less than the median height. 5. More information is needed to answer this question. 6. None of the above is true. 7. Enrique has a 91, 87, and 95 for his statistics grades for the first three quarters. His mean grade for the year must be a 93 in order for him to be exempt from taking the final exam. Assuming grades are rounded following valid mathematical procedures, what is the lowest whole number grade he can get for the 4th quarter and still be exempt from taking the exam? 8. The chart below shows data from the Galapos tortoise preservation program with just the number of individual tortoises that were bred in captivity and reintroduces into their native habitat.  |  |  | | --- | --- | | Island or Volcano | # of Ind. Repatriated | | Wolf | 40 | | Darwin | 0 | | Alcedo | 0 | | Sierra Negra | 286 | | Cerro Azul | 357 | | Santa Cruz | 210 | | Espanola | 1293 | | San Cristobal | 55 | | Santiago | 498 | | Pinzon | 552 | | Pinta | 0 |   Find the:   1. Mode b. Median c. Mean 2. In the previous question, why is the mean significantly higher than the median? 3. The mean of 10 scores is 12.6. What is the sum of the scores? 4. While on vacation John drove an average of 262 miles per day for a period of 12 days. How far did John drive in total while he was on vacation? 5. Find x if 5, 9, 11, 12, 13, 14, 15, and x have a mean of 13. 6. Find *a* given that 3, 0, *a, a,* 4, *a,* 6, *a,* and 3 have a mean of 4. 7. The table below shows the results when 3 coins were tossed simultaneously 30 times. The number of tails appearing was recorded. Calculate the mode, median, and mean.  |  |  | | --- | --- | | # of Tails | # of times occurred | | 3 | 4 | | 2 | 12 | | 1 | 11 | | 0 | 3 | | Total | 30 |  1. Compute the mean, median and mode for each of the following sets of numbers:  |  |  |  |  | | --- | --- | --- | --- | | **Data Set** | **Mean** | **Median** | **Mode** | | a. 3, 16, 3, 9, 5, 7, 11 |  |  |  | | b. 5, 3, 3, 7, 5, 5, 16, 9, 3, 18, 11, 5, 3, 7 |  |  |  | | c. 7, -4, 0, 12, 8, 121, -3 |  |  |  | | d. 65, 69, 73, 77, 81, 87 |  |  |  | | e. 31, 11, 41, 31 |  |  |  |  1. How do you determine which measure of center best describes a particular data set? |
| **Additional Resources** | * <http://www.virtualnerd.com/tutorials/?id=Alg1_14_01_0009> * <http://www.mathsisfun.com/mean.html> |
| **4.2 Measures of Dispersion** | |
| **Notes** | DVUSD Algebra Flexbook: pg 49-56  **Example A**  Given data set: 75, 80, 90, 94, 96  **Range** = 96 - 75 = 21  **Interquartile Range (IQR):** the difference between the\_\_\_\_ and \_\_\_    **Q1 =**  **Q3 =**  **IQR =**  **Example B**  A recent study proclaimed Mobile, Alabama the wettest city in America. The following table lists measurements of the approximate annual rainfall in Mobile over a 10 year period. Find the range and IQR for this data.  Rainfall (inches)  **Range** =  **Q1 =**  **Q3 =**  **IQR =**  1998 - 90  1999 - 56  2000 - 60  2001 - 59  2002 - 74  2003 - 76  2004 - 81  2005 - 91  2006 - 47  2007 – 59  **Outliers:** an \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ value.  Identified mathematically using the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  To determine if your data set has outliers: take your \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Using the original rainfall data above, the IQR is 22. Any observation that is below Q1 - 1.5 x IQR or above Q3 + 1.5 x IQR is considered an outlier. Any rainfall amount below 59 - 1.5(22) = 26 inches would be an outlier. Any rainfall amount above 81 + 1.5(22) = 114 inches would be an outlier.  After examining the data, you see that we do not have any outliers in this data set.  **Standard Deviation:**  an extremely important measure of spread that is based on the mean.  Recall that the mean is the numerical balancing point of the data.  One way to measure how the data are spread is to look at how far away each of the values is from the mean.  The difference between a data value and the mean is called the deviation.  **Technology Notes:**  ***Calculating Standard Deviation on the TI-83/84 Graphing Calculator***  Enter the data 9.5, 11.5, 12 in list L1 (see first screen below).  Then choose ’1-Var Stats’ from the CALC submenu of the STAT menu (second screen).  Enter L1 (third screen) and press [ENTER] to see the fourth screen.  In the fourth screen, the symbol sx is the sample standard deviation. |
| **Check for Understanding** | 1. Following are bowling scores for two people: Luna 1 1 2 10 12 1 9 6 7 8 Chris 4 3 4 1 4 1 6 7 11 5  a. Show that Chris and Luna have the same mean and range.  b. Whose performance is more variable? Explain.  2. Use the rainfall data to answer this question.   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Year** | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | | **Rainfall (in)** | 90 | 56 | 60 | 59 | 74 | 76 | 81 | 91 | 47 | 59 |   a. Calculate and record the sample mean:  b. Calculate the standard deviation using your calculator.  c. Are there any outliers in the data set? Explain  3. Calculate the range and the IQR for this data.    a. Calculate the sample standard deviation for this data using the calculator.  b. Are there any outliers in the data set above? Explain.  4. How do you determine which measure of spread best describes a particular data set?  5. What information does the standard deviation tell us about the specific, real data being observed?  6. What are the effects of outliers on the various measures of spread?   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Mean** | **Median** | **Q1** | **Q3** | **IQR** | |  |  |  |  |  |   7. How does altering the spread of a data set affect its visual representation(s)? |
| **Additional Resources** | * <http://www.youtube.com/watch?v=zATQYfMQ8dA> |
| **4.4 Histograms** | |
| **Notes** | DVUSD Algebra Flexbook: pg 63-74  **Vocabulary**  A **frequency table** is useful to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **Relative frequency** shows \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  A **histogram** is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **Technology Notes: Histograms on the TI-83/84 Graphing Calculator**  **Enter the data in a list**   * In the home screen, press [2ND][{], * Then enter the data separated by commas (see right) * When all the data have been entered, press [2ND][}][STO * Then press [2ND][L1][ENTER].   Plot the histogram.   * Press [2ND][STAT PLOT] to enter the STAT-PLOTS menu. {You can plot up to three statistical plots at one time}. * Choose Plot1. Turn the plot on, change the type of plot to a histogram (see sample screen below), and choose L1. * Enter ’1’ for the Freq by pressing [2ND][A-LOCK] to turn off alpha lock, which is normally on in this menu, because most of the time you would want to enter a variable here. * An alternative would be to enter the values of the variables in L1 and the frequencies in L2 as we did in Chapter 1.     **Set a window**   * **Press [WINDOW] and enter an appropriate window to display the plot.** * **In this case, ’XSCL’ is what determines the bin width.** * **Also notice that the maximum x value needs to go up to 9 to show the last bin, even though the data values stop at 8. Enter all of the values shown below.**   **Graph**   * **Press [GRAPH] to display the histogram.** * **If you press [TRACE] and then use the left or right arrows to trace along the graph, notice how the calculator uses the notation to properly represent the values in each bin.**   o  **Example A**  Take an informal poll of your class. Ask each member of the class, on average, how many beveragebottles they use in a week. Once you collect this data, the first step is to organize it so it is easier to understand. A frequency table is a common starting point. Frequency tables simply display each value of the variable, and the number of occurrences (the frequency) of each of those values. In this example, the variable is the number of plastic beverage bottles of water consumed each week.  Consider the following raw data:  6, 4, 7, 7, 8, 5, 3, 6, 8, 6, 5, 7, 7, 5, 2, 6, 1, 3, 5, 4, 7, 4, 6, 7, 6, 6, 7, 5, 4, 6, 5, 3  **When creating a frequency table, it is often helpful to use** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**as a running total to avoid** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_    **The following data set shows the countries in the world that consume the most bottled water per person per year.**    **These data values have been measured at the ratio level. The values range from** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. **By examining the data, it seems appropriate to create the frequency table in** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **A bracket,** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**, indicates that** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**in the class.**  **A parenthesis,** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**indicates that** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_    **Histograms**  Once you can create a frequency table, you are ready to create our first graphical representation, called a histogram. Let’s revisit our data about student bottled beverage habits.  **Here is the same data in a histogram:**    **In this case, the horizontal axis represents the** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, **and the vertical axis represents the** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**. Each vertical bar represents** \_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **On the Web:** [**http://illuminations.nctm.org/ActivityDetail.aspx?ID=78**](http://illuminations.nctm.org/ActivityDetail.aspx?ID=78)  **Here you can change the bin width and explore how it effects the shape of the histogram.**  **Guided Practice**  There is some question as to whether caloric content listed on food products is under-reported. Look at the following table of kinds of food products (Food) and the percentage difference between measured calories and labeled calories per item (Per Item).   |  |  | | --- | --- | | **Caloric Data on food items.** | | | Food Per item  noodles and alfredo sauce 2  cheese curls -28  green beans -6  mixed fruits 8  cereal 6  fig bars -1  oatmeal raisin cookie 10  crumb cake 13  crackers 15  blue cheese dressing -4  imperial chicken -4  vegetable soup -18  cheese 10  chocolate pudding 5  sausage biscuit 3  lasagna -7  spread cheese 3  lentil soup -0.5  pasta w/shrimp & tomato sauce -10  chocolate mousse 6 | meatless sandwich 41  oatmeal cookie 46  lemon pound cake 2  banana cake 25  brownie 39  butterscotch bar 16.5  blondie 17  oat bran snack bar 28  granola bar -3  apricot bar 14  chocolate chip cookie 34  carrot muffin 45  Chinese chicken 15  gyoza 60  jelly diet candy-reds flavor 250  jelly diet candy-fruit flavor 145  Florentine manicotti 6  egg foo young 80  hummus with salad 95  baby ghanoush with salad 3 |   **Draw a histogram of the percentage difference between observed and reported calories per item.**  **Solution:**  First, break up the percentage difference in calories per item, into intervals. By glancing at the data, it looks like using an interval length of 30 will work well, with the first interval being from -30 **to zero.** Count how many food items fall into each intervals, and then graph this frequency as the height on the vertical axis. For example, there are 10 food items that have a percentage difference of calories between -30 and 0, so we draw a bar with a height of 10 for that interval. |
| **Check for Understanding** | 1. The following frequency table contains exactly one data value that is a positive multiple of ten.  **What must that value be?**  a. 10  b. 20  c. 30  d. 40  e. There is not enough information to determine the answer.  Class Frequency  [0 - 5) 4  [5 - 10) 0  [10 - 15) 2  [15 - 20) 1  [20 - 25) 0  [25 - 30) 3  [30 - 35) 0  [35 - 40) 1  2. The following table includes the data from the same group of countries from the earlier bottled water consumption example, but is for the year 1999, instead.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Country** | **Liters of Bottled Water Consumed per Person per Year** |  | **Country** | **Liters of Bottled Water Consumed per Person per Year** | | Italy 154.8  Mexico 117.0  United Arab Emirates 109.8  Belgium and Luxembourg 121.9  France 117.3  Spain 101.8  Germany 100.7  Lebanon 67.8 | |  | Czech Republic 62.1  Austria 74.6  Portugal 70.4  Switzerland 90.1  Cyprus 67.4  United States 63.6  Saudi Arabia 75.3 | |   **a. Create a frequency table for this data set. b. Create the histogram for this data set.**   |  |  | | --- | --- | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  |   c. How would you describe the shape of this data set? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  d. Are there any outliers? If so, what are they?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **3. The figure below is a histogram of the salaries of CEOs.**    a. Are there any outliers? For any outlier, give a value for the salary and explain why you think it is an  outlier.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  b. What is the salary that occurs most often? Roughly, how many CEO’s report having this salary?  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  c. Roughly, how many CEOs report having $500,000? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **Additional Resources** | * <http://www.mathsisfun.com/data/histograms.html> * <http://www.shodor.org/interactivate/activities/Histogram/> * <http://quarknet.fnal.gov/toolkits/ati/histograms.html> |
| **4.5 Box & Whisker Plots** | |
| **Notes** | DVUSD Algebra Flexbook: pg. 75-82  **Box-and-Whisker Plot** : a graphically display a data set  What if your teacher recorded each of her student’s scores on the last math test? How could she display that data in such a way that it was broken up into four distinct segments? After completing this Concept, you’ll be able to make and interpret box-and-whisker plots for data such as this.  **Vocabulary**  • We call the division between the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The division between the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (the second quartile is, of course, the \_\_\_\_\_\_\_\_\_\_\_\_\_).  • A box-and-whisker plot is formed by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (These five numbers are often referred to as\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_) A box is drawn between the position of the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **Guidance**  Consider the following list of numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.  The median is the \_\_\_\_\_\_\_\_\_\_\_ th value. There are 10 values, so the median lies halfway between the 5th and the 6th value. The median is therefore\_\_\_\_\_\_\_\_\_\_\_\_. This splits the list cleanly into two halves. The **lower list** is: 1, 2, 3, 4, 5 And the **upper list** is: 6, 7, 8, 9, 10  The median of the lower half is \_\_\_\_\_\_. The median of the upper half is\_\_\_\_\_\_\_. These numbers, together with the median, cut the list into four quarters. We call the division between the lower two quarters the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The division between the upper two quarters is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_  (the second quartile is, of course, the median).  A box-and-whisker plot is formed by placing vertical lines at five positions, corresponding to the smallest value, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. (These five numbers are often referred to as the five number summary.) A box is drawn between the position of the first and third quartiles, and horizontal line segments (the whiskers) connect the box with the two extreme values.  The box-and-whisker plot for the integers 1 through 10 is shown.  **Example A**  Forty students took a college algebra entrance test and the results are summarized in the box-and-whisker plot below.  How many students would be allowed to enroll in the class if the pass mark was set at  a) 65%  b) 60%  **Solution**  From the plot, we can see the following information:  Lowest score =  First quartile =  Median score =  Third quartile =  Highest score =  Since the pass marks given in the question correspond with the median and the first quartile, the question is really asking how many students there are in: a) the upper half and b) the upper 3 quartiles.  a) Since there are 40 students, there are 20 in the upper half; that is, 20 students scored above 65%.  b) Similarly, there are 30 students in the upper 3 quartiles, so 30 students scored above 60%.  **Example B**  Harika is rolling 3 dice and adding the numbers together. She records the total score for each of 50 rolls, and the scores she gets are shown below. Display the data in a box-and-whisker plot, and find both the range and the inter-quartile range.  9, 10, 12, 13, 10, 14, 8, 10, 12, 6, 8, 11, 12, 12, 9, 11, 10, 15, 10, 8, 8, 12, 10, 14, 10, 9, 7, 5, 11, 15, 8, 9, 17, 12, 12, 13, 7, 14, 6, 17, 11, 15, 10, 13, 9, 7, 12, 13, 10, 12  **Solution**  First we’ll put the list in order.  Since there are 50 data points, find the median and it will be the mean of the 25th and 26th values and it will \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  The median will split the data into two lists of 25 values; we can write them as two distinct lists.  5,6,6,7,7,7,8,8,8,8,8,9,9,9,9,9,10,10,10,10,10,10,10,10,10,11,11  11,11,12,12,12,12,12,12,12,12,12,13,13,13,13,14,14,14,15,15,15,17,17  Since each sub-list has 25 values, the first and third quartiles of the entire data set can be found from the median of each smaller list. So the first Quartile is \_\_\_\_\_\_ and the third quartile is \_\_\_\_\_  From the ordered list we can see the five number summary:  • The lowest value is 5  • The first quartile is 9  • The median is 10.5  • The third quartile is 13  • The highest value is 17.  The box-and-whisker plot therefore looks like this:  The range is given by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  The inter-quartile range is given by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  **Representing Outliers in a Box-and-Whisker Plot**  Box-and-whisker plots can be misleading if we don’t take \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. An outlier is \_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ For box-and-whisker plots, we can define which points are outliers by how far they are from the box part of the diagram. Our basic measure of distance will be the inter-quartile range (IQR).  • A \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is a point that falls more than \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_the IQR outside of the box.  • An \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_is a point that falls more than \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_the IQR outside of the box.  When we draw a box-and-whisker plot, we don’t include the outliers in the “whisker” part of the plot; instead, we draw them as separate points.  **Example C**  Draw a box-and-whisker plot for the following ordered list of data:  1, 2, 5, 9, 10, 10, 11, 12, 13, 13, 14, 19, 25, 30  **Solution**  From the ordered list we see:  • The lowest value:  • The first quartile (Q1):  • The median (Q2):  • The third quartile (Q3):  • The highest value:  IQR: IQR = Q3-Q1  Looking back at the data we see:  • The value of 1 is less than 1.5, so it is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  • The value 2 is the lowest value that falls within the included\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  • The value 30 is greater than 21.5. In fact, it’s not just more than 7.5 units outside the box, it’s more than twice that far outside the box. Since it falls more than 3 times the IQR above the third quartile, it’s an \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  • The value 25 is also greater than 21.5, so it is a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  • The value 19 is the highest value that falls within the included \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  So when we draw our box-and-whisker plot, the whiskers will only go out as far as 2 and 19 respectively. The points outside of that range are all outliers. Here is the plot:    **Example D**  The ages of all the passengers traveling in a train carriage are shown below.  35, 42, 38, 57, 2, 24, 27, 36, 45, 60, 38, 40, 40, 44, 1, 44, 48, 84, 38, 20, 4, 2, 48, 58, 3, 20, 6, 40, 22, 26, 17, 18, 40, 51, 62, 31, 27, 48, 35, 27, 37, 58, 21  Use a graphing calculator to:  a) obtain the 5 number summary for the data.  b) create a box-and-whisker plot.  c) determine if any of the points are outliers.  **Solution**  **Enter the data in your calculator:**   * Press [START] then choose [EDIT]. * Enter all 43 data points in list L1.   **Find** **the 5 number summary:**   * Press [START] again. Use the right arrow to choose [CALU]. * Highlight the **1-Var Stats** option. Press [EDIT]. * The single variable statistics summary appears. * Note the mean ( ) is the first item given. * Use the down arrow to bring up the data for the five *number summary. n* is the number of data points, and the final fie numbers in the screen are the numbers we require.     **TABLE 2.21:**  Symbol Value  Lowest Value minX 1  First Quartile Q2 21  Median Med 37  Third Quartile Q3 45  Highest Value maxX 84    **Display the box-and-whisker plot:**   * Bring up the [STARTPLOT] option by pressing [2nd]. [Y=]. * Highlight 1:Plot1 and press [ENTER]. * There are two types of box-and-whisker plots available. The first automatically identifies outliers. Highlight it and press [ENTER]. * Press [WINDOW] and ensure that **Xmin** and **Xmax** allow for all data points to be shown. In this example, Xmin=0 and Xmax = 100. * Press [GRAPH] and the box-and-whisker plot should appear. * The calculator will automatically identify outliers and plot them as such. You can use the [TRACE] function along with the arrows to identify outlier values. In this case there is one outlier: 84. |
| **Check for Understanding** | 1. Draw a box-and-whisker plot for the following unordered data:  49, 57, 53, 54, 57, 49, 67, 51, 57, 56, 59, 57, 50, 49, 52, 53, 50, 58  2. The box-and-whisker plots below represent the percentage of people living below the poverty line by county in both Texas and California. Determine the 5-number summary for each state,      **California Texas\_\_\_\_\_\_\_\_\_\_**   1. Smallest value: Smallest value: 2. First quartile: First quartile: 3. Median: Median: 4. Third quartile: Third quartile: 5. Greatest value: Greatest value:   f) Explain the spread of each distribution.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  3. The 5-number summary for the average daily temperature in Atlantic City, NJ1 (given in degree F):  31, 39, 52, 68, 76.  Draw the box-and-whisker plot for this data and use it to determine which of the following, if any,  would be considered outliers if they were included in the data:  a. January’s record high temperature of 78  b. January’s record low temperature of -8  c. April’s record high temperature of 94  d. The all time record high of 106 |
| **Additional Resources** | * <http://www.purplemath.com/modules/boxwhisk.htm> * <http://www.regentsprep.org/regents/math/algebra/ad3/boxwhisk.htm> * <http://www.youtube.com/watch?v=GMb6HaLXmjY> |
| **4.6 Double Box & Whisker Plots** | |
| **Notes** | DVUSD Algebra Flexbook: pg 87 – 92  Watch multimedia link  Double box-and-whisker plots give you a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_of 2 sets of data.   * Means * Medians * maximums (upper range) * minimums (lower range)   Example A  Emma and Daniel are surveying the times it takes students to arrive at school from home. There are 2 main groups of commuters who were in the survey. There were those who drove their own cars to school, and there were those who took the school bus. Emma and Daniel collected the following data:  Bus times (min) 14 18 16 22 25 12 32 16 15 18  Car times (min) 12 10 13 14 9 17 11 10 8 11  Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the times it takes for students to arrive at school either by car or by bus.  When plotted, the box-and-whisker plots look like the following:    Using the medians, 50% of the cars arrive at school in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_or less, whereas 50% of the students arrive by bus in \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_or less. The range for the car times is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. For the bus times, the range is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Since the range for the driving times is smaller, it means the times to arrive by car are less spread out. This would, therefore, mean that the times are more predictable and reliable.  Example B  A new drug study was conducted by a drug company in Medical Town. In the study, 15 people were chosen at random to take Vitamin X for 2 months and then have their cholesterol levels checked. In addition, 15 different people were randomly chosen to take Vitamin Y for 2 months and then have their cholesterol levels checked. All 30 people had cholesterol levels between 8 and 10 before taking one of the vitamins. The drug company wanted to see which of the 2 vitamins had the greatest impact on lowering people’s cholesterol. The following data was collected:  Vitamin X 7:2 7:5 5:2 6:5 7:7 10 6:4 7:6 7:7 7:8 8:1 8:3 7:2 7:1 6:5  Vitamin Y 4:8 4:4 4:5 5:1 6:5 8 3:1 4:6 5:2 6:1 5:5 4:2 4:5 5:9 5:2  Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the 2 vitamins and provide a conclusion for the drug company. When plotted, the box-and-whisker plots look like the following:    Using the medians, 50% of the people in the study had cholesterol levels of \_\_\_\_\_ or lower after being on Vitamin X for 2 months. Also 50% of the people in the study had cholesterol levels of \_\_\_\_\_\_ or lower after being on Vitamin Y for 2 months. Knowing that the participants of the survey had cholesterol levels between 8 and 10 before beginning the study, it appears that Vitamin Y had a bigger impact on lowering the cholesterol levels. The range for the cholesterol levels for people taking Vitamin X was \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, while the range for the cholesterol levels for people taking Vitamin Y was \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. Therefore, the range is not useful in making any conclusions. |
| **Check for Understanding** | 1. International Baccalaureate has 2 levels of courses, which are standard level (SL) and higher level (HL). Students say that study times are the same for both the standard level exams and the higher level exams. The following data represents the results of a survey conducted to determine how many hours a random sample of students studied for their final exams at each level:  HL Exams 15 16 16 17 19 10 5 6 5 5 8 10 8 12 17  SL Exams 10 6 6 7 9 12 2 6 2 5 7 20 18 8 18   1. Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data.  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | **Smallest** | **Q1** | **Median** | **Q3** | **Largest** | | **HL Exams** |  |  |  |  |  | | **SL Exams** |  |  |  |  |  |  1. Compare the times students prepare for each level of exam.   2. Students in the AP math class at BCU High School took their SATs for university entrance. The following scores were obtained for the math and verbal sections:  Math 529 533 544 562 513 519 560 575 568 537 561 522 563 571  Verbal 499 509 524 530 550 499 545 560 579 524 478 487 482 570   1. Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data.  |  |  |  | | --- | --- | --- | |  | **Math** | **Verbal** | | **Smallest** |  |  | | **Q1** |  |  | | **Median** |  |  | | **Q3** |  |  | | **Largest** |  |  |  1. Compare the data for the 2 sections of the SAT using the five-number summary data.   3. The following box-and-whisker plots were drawn to analyze the data collected in a survey of scores for the doubles performances in the figure skating competitions at 2 Winter Olympic games. The box-and-whisker plot on the top represents the scores obtained at the 2010 winter games in Whistler, BC. The box-and-whisker plot on the bottom represents the scores obtained at the 2006 winter games in Torino, Italy.     |  |  |  | | --- | --- | --- | |  | **2010 Games** | **2006 Games** | | **Smallest** |  |  | | **Q1** |  |  | | **Median** |  |  | | **Q3** |  |  | | **Largest** |  |  |  1. What is the five-number summary for the games? 2. 50% of the judge’s scores were at least what for the \_\_\_\_\_\_ games?   2006: 2010:   1. What was the range for the scores of the \_\_\_\_\_\_ games?   2006: 2010:   1. In which of the winter games would you say that the data is more equally spread out? Explain. 2. In which of the winter games would you say that the scoring was more predictable and reliable? Explain. |
| **Additional Resources** | * <http://math.andyou.com/tools/boxandwhisker.html> * <http://www.ck12.org/statistics/Double-Box-and-Whisker-Plots/> |