**College Success?**

### Alignments to Content Standards

### 7.SP.B. Draw informal comparative inferences about two populations.

### 7.SP.B.3. Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

### 7.SP.B.4. Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.

### Overview: In this task, students are able to conjecture about the differences in the two groups from a strictly visual perspective and then support their comparisons with appropriate measures of center and variability. This will reinforce that much can be gleaned simply from visual comparison of appropriate graphs, particularly those of similar scale. Students are also encouraged to consider how certain measurements and observation values from one group compare in the context of the other group.

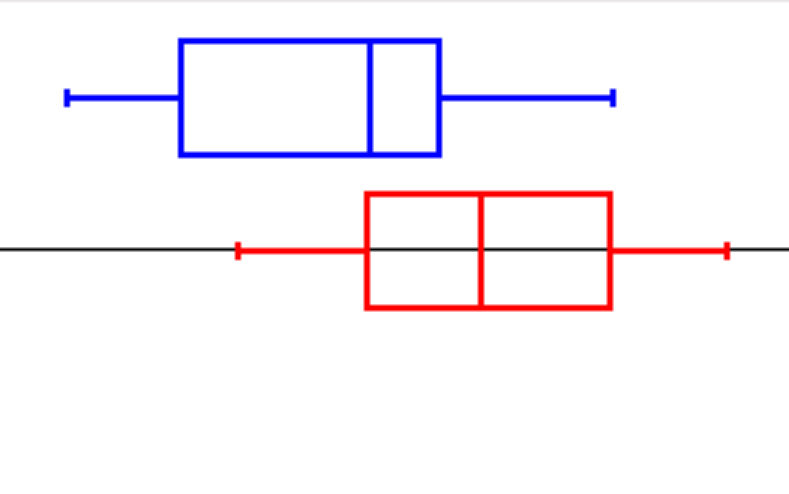
### 

**Student Activity**: Below are the GPAs of graduates from Valley of the Sun High School from the 2010-2011 academic year. The data has been separated into two groups. Group 1 is of students that have successfully graduated from a college (or university) and Group 2 is comprised of students that did not graduate from a college (or university). Note that the data has been sorted from lowest GPA to highest.

|  |  |
| --- | --- |
| **GPA of High School Students Who Graduated from College/University** | **GPA of High School Students Who Did Not Graduate from College/University** |
| 2.30 | 1.70 |
| 2.35 | 1.74 |
| 2.47 | 1.75 |
| 2.73 | 1.78 |
| 2.74 | 2.00 |
| 2.75 | 2.20 |
| 2.83 | 2.21 |
| 2.87 | 2.35 |
| 2.88 | 2.60 |
| 3.13 | 2.75 |
| 3.16 | 2.76 |
| 3.24 | 2.79 |
| 3.35 | 2.80 |
| 3.42 | 2.84 |
| 3.59 | 2.90 |
| 3.60 | 3.10 |
| 3.68 | 3.31 |
| 3.71 | 3.36 |
| 3.80 | 3.45 |
| 4.00 | 3.60 |

1. Compare and contrast the two distributions by their SOCS. Be sure to point out any similarities **and** differences.

*Teacher Note: The students will have to create two different histograms and two different box plots to answer question a. and the following questions. Be sure to prompt them to do so if they don’t do so on their own.*



**DNG**: **G:**

Min: 1.70 2.30

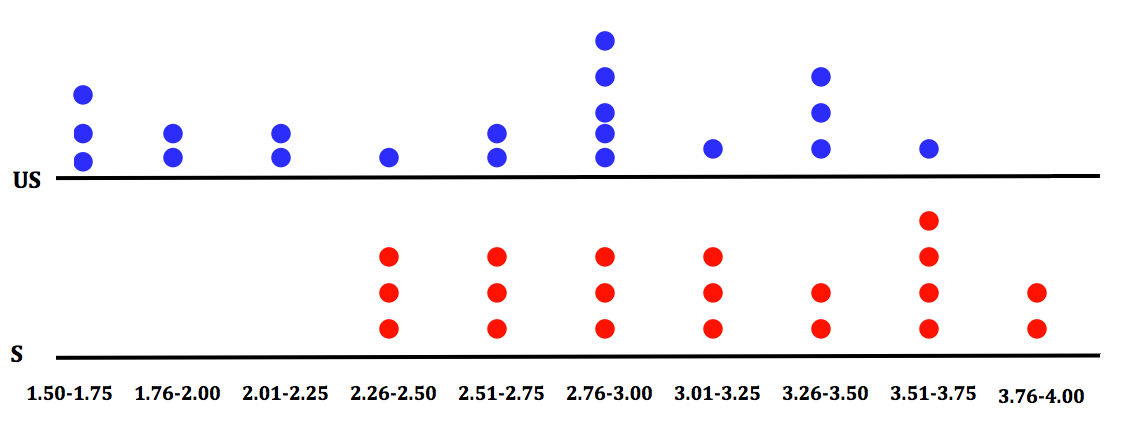
Q1: 2.10 2.75

Med: 2.76 3.15

Q3: 3.00 3.60

Max: 3.60 4.00

*Teacher Note: In order to answer the shape aspect of the distribution the students will need a dot plot or a histogram. We are not able to determine shape from a box plot. For example, the Graduate box plot could represent a distribution of data that could be bi-modal, tri-modal, or just one mode (among many other possibilities).*



G

DNG

***Did Not Graduate******Graduated***

***Shape****: Bi-modal? Normal? Uniform? Normal?*

***Outlier (+-1.5 \* IQR)****: No No*

***Center****: Median: 2.76 Mean: 2.60 Median: 3.15 Mean: 3.13*

***Spread****: Range: 1.9 MAD: 0.51 Range: 1.7 MAD: 0.43*

*The shape is very hard to determine with so few data points and them being spread pretty uniformly. The center of the graduated students is much higher on the number line than the center of those who did not graduate, so at first glance, it appears that the graduated group has a higher average. There are no outliers. As far as spread goes, they both appear to be about the same.*

1. What percent of the graduated group exceeds the median value of the group that did not graduate?

*Median of did not graduate is 2.76 so 14/ 20 graduated GPAs exceed this or 70%.*

1. The minimum GPA of the graduated group is approximately at the \_\_\_ percentile of the group who did not graduate.

*35th percentile*

1. The maximum GPA of the group who did not graduate is approximately at the \_\_\_ percentile of the graduated group.

*70th percentile*

1. Based on visual inspection of the dotplots, which group of students appears to have the larger average GPA? Which group of students appears to have the greater variability in the GPAs?

*Larger average: Graduated. Greater variability: Did not graduate.*

1. Compute the mean and mean absolute deviation (MAD) for each group. Do these values support your answers in part (e)?

\*\*\*Teacher Note: Give the students three data points and have them begin to calculate the MAD to help them recall the procedure but to save time…give them the MAD.

***Did Not Graduate******Graduate***

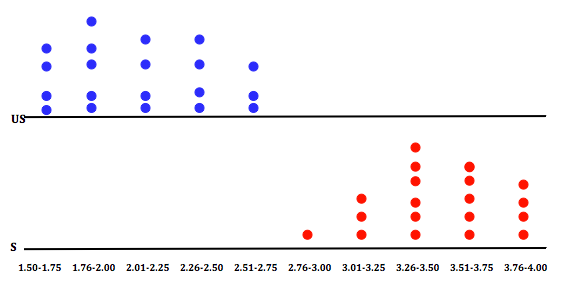
*Center: Mean: 2.60 Mean: 3.13*

*Spread: MAD: 0.51 MAD: 0.43*

*Yes & Yes*

1. Assume that you are counseling three high school students who are considering whether they should go to college, Albert has a GPA of 2.23, Delvin has a GPA 2.9, and Connie has a GPA of 3.6. Based on the **hypothetical** distribution of GPA data provided here, do you anticipate successful or unsuccessful completion of college for each student? Explain your reasoning.

\*Hypothetical GPAs for Valley of the Sun High School Graduates 2010-2011



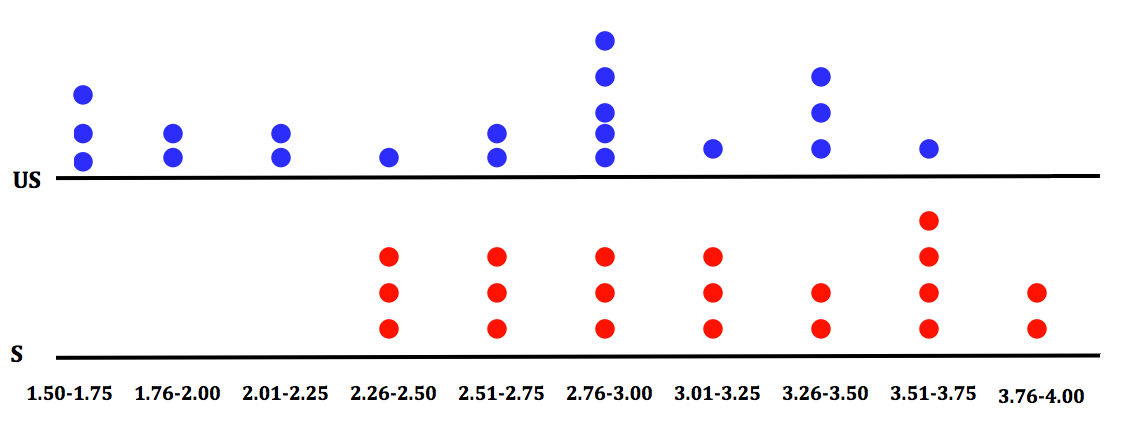
G

DNG

*We would anticipate that Albert with a GPA of 2.23 and Delvin with a GPA 2.9 would not be successful, but Connie with a GPA of 3.6 would be successful in graduating. The first two students are clearly in the data that comes from those who were unsuccessful at graduating college but the last student’s GPA is similar to those students who were successful in doing so.*

1. Consider the same three high school students (Albert has a GPA of 2.23, Delvin has a GPA 2.9, and Connie has a GPA of 3.6) as in the previous problem but use the actual distribution of GPA data (see below) to answer this question: Do you anticipate successful or unsuccessful completion of college for each student? Explain your reasoning and be sure to utilize the data analysis you have performed on the GPA data to support your conjecture.

**GPAs for Valley of the Sun High School Graduates 2010-2011**



DNG

G

*We would anticipate that Albert with a GPA of 2.23 would not be successful at graduating from college, anticipating what Delvin’s outcome will be with a GPA 2.9 is difficult, but Connie with a GPA of 3.6 appears that she would be successful in graduating. The first student is clearly in the data that comes from those who were unsuccessful and the last student’s GPA is similar to those students who were successful. However, since Delvin’s GPA is near the center of the unsuccessful group of students but is also fairly close to the center of the successful students it is very hard to make a confident hypothesis.*

1. Which data set from problem g. or h. was easier to make conjectures about the three students’ successful completion from college? Why?

*The second data set is much more difficult to make conjectures from due to the overlap of data where students with similar GPAs were both successful and unsuccessful in completing college.*

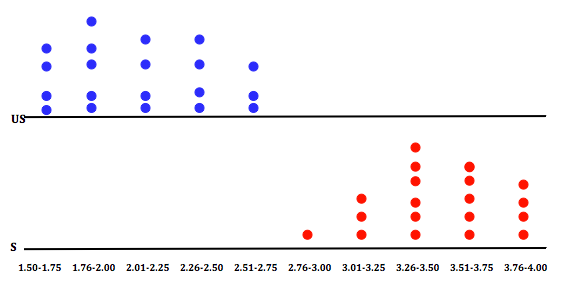
1. If the variability is relatively similar, thedifference between the centers of a distribution is one way to determine the amount of overlap. Calculate the difference between the centers of the student GPA data by expressing it as a multiple of the MAD. Use the diagram below in your answer. What does this numerical value tell us about using the GPA of Valley of the Sun High School students to anticipate their success of graduating from college?

*(3.15 – 2.76)/.50 = 0.78 MADs. This tells that the data overlaps pretty much since the centers are not even a whole MAD apart from each other.* When the distributions are far apart there would be “many” MADs between their centers but as the distance between the centers decreases the number of MADs will also decrease.



1. Take another look at the hypothetical GPA data (see below). If we would calculate the difference between the centers of this data by expressing it as a multiple of the MAD, would the numerical value be greater than or less than the answer we got in part j. Justify your answer.

\*Hypothetical GPAs for Valley of the Sun High School Graduates 2010-2011



DNG

G

*If we would calculate the distance between the centers in terms of the MAD, the numerical value would be greater. This is because the distance between the data is greater and the MAD is the same value. This tells us that the data does not overlap as much and if we have a specific data value we will have an easier time choosing which distribution of data it comes from.*

**A Look Ahead**: Informally assessing the degree of visual overlap between two distributions leads to a more formal statistical analysis through the two-sample t-test (when comparing two sample means) or the two sample z-test (when comparing two sample proportions). These statistical tools are used in investigating and formally comparing parameters between two populations or treatment groups. The primary goal of these two tests is to determine if there is a statistically significant difference between the two means or proportions rather than going by “visual overlap” as we have done in this activity.

For example, let’s assume that you want to determine if people tend to scoop out different amounts of ice cream for themselves if they are given different sized bowls. Assume two randomly assigned groups are formed and one group is given a 17-ounce bowl to use and another is given a 34-ounce bowl. Sample results for the actual volumes taken (in ounces) are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Sample Size | Sample Mean | Sample Standard Deviation |
| 17-ounce bowl | 20 | 4.38 | 2.05 |
| 34-ounce bowl | 17 | 5.81 | 2.26 |

A typical question that could be asked is: Is the difference in ice cream volumes served between the two bowl sizes statistically significant at the 0.05 level? How about at the 0.01 level?

How do think this problem connects to the concepts we discussed in this activity? What do you think the concept of being “statistically significant” has to do with the distance between the means of ice cream helpings?

**Homework - Offensive Linemen**: (from NCTM Illuminations)

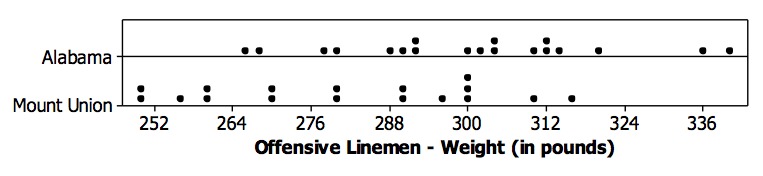
College football teams are grouped with similar teams into "divisions" (and in some cases, "subdivisions") based on many factors such as game attendance, level of competition, athletic department resources, and so on. Schools from the Football Bowl Subdivision (FBS, formerly known as Division 1-A) are typically much larger schools than schools of any other division in terms of enrollment and revenue. "Division III" is a division of schools with typically smaller enrollment and resources.

One particular position on a football team is called "offensive lineman," and it is generally believed that the offensive linemen of FBS schools are heavier on average than the offensive linemen of Division III schools.

For the 2012 season, the University of Mount Union Purple Raiders football team won the Division III National Football Championship while the University of Alabama Crimson Tide football team won the FBS National Championship. Below are the weights of the offensive linemen for both teams from that season.

(Accessed at <http://athletics.mountunion.edu/sports/fball/2012-13/roster>, <http://www.rolltide.com/sports/m-footbl/mtt/alab-m-footbl-mtt.html> on 1/14/13)

|  |  |
| --- | --- |
| **Alabama** | **Mount Union** |
| 277 | 250 |
| 265 | 250 |
| 292 | 290 |
| 303 | 260 |
| 303 | 270 |
| 320 | 270 |
| 300 | 310 |
| 313 | 290 |
| 267 | 280 |
| 288 | 315 |
| 311 | 280 |
| 280 | 295 |
| 302 | 300 |
| 335 | 300 |
| 310 | 260 |
| 290 | 255 |
| 312 | 300 |
| 340 |  |
| 292 |  |



1. Based on visual inspection of the dotplots, which group appears to have the larger average weight? Does one group seem to have greater variability in its weights than the other, or do the two groups look similar in that regard?
2. Compute the mean and mean absolute deviation (MAD) for each group. Do your measures support your answers in part (a)?
3. Choose from the following to fill in the blank: "The average Alabama offensive lineman's weight is about \_\_\_\_\_\_\_\_\_\_ than the average Mount Union offensive lineman's weight."
   1. 20 pounds lighter
   2. 15 pounds lighter
   3. 15 pounds heavier
   4. 20 pounds heavier

"This difference in average weights is approximately \_\_\_\_\_\_\_\_\_\_\_\_ of either team."

* 1. About half of the MAD
  2. Slightly more than 1 MAD
  3. Twice the MAD

1. The offensive linemen on the Alabama team are not a random sample from all FBS offensive linemen. Similarly, the offensive linemen on the Mount Union Team are not a random sample from all Division III offensive linemen. However, for purposes of this task, suppose that these two groups can be regarded as random samples of offensive linemen from their respective divisions/subdivisions. If these were random samples, would you think that offensive linemen from FBS schools are typically heavier than offensive linemen from Division III schools? Explain your decision using answers to the previous questions and/or additional analysis.