## Lesson 23: Using Sample Data to Compare the Means of Two or

## More Populations

## Classwork

In the previous lesson, you described how far apart the means of two data sets are in terms of the MAD (mean absolute deviation), a measure of variability. In this lesson, you will extend that idea to informally determine when two sample means computed from random samples are far enough apart from each other to imply that the population means also differ in a meaningful way. Recall that a meaningful difference between two means is a difference that is greater than would have been expected just due to sampling variability.

## Example 1: Texting

With texting becoming so popular, Linda wanted to determine if middle school students memorize real words more or less easily than fake words. For example, real words are food, car, study, swim; whereas fake words are stk, fonw, cqur, ttnsp. She randomly selected 28 students from all middle school students in her district and gave half of them a list of 20 real words and the other half a list of 20 fake words.
a. How do you think Linda might have randomly selected 28 students from all middle school students in her district?
b. Why do you think Linda selected the students for her study randomly? Explain.
c. She gave the selected students one minute to memorize their lists after which they were to turn the lists over and, after two minutes, write down all the words that they could remember. Afterward, they calculated the number of correct words that they were able to write down. Do you think a penalty should be given for an incorrect word written down? Explain your reasoning.

## Exercises 1-4

Suppose the data (number of correct words recalled) she collected were as follows
For students given the real words list: $8,11,12,8,4,7,9,12,12,9,14,11,5,10$.
For students given the fake words list: $3,5,4,4,4,7,11,9,7,7,1,3,3,7$.

1. On the same scale, draw dot plots for the two data sets.
2. From looking at the dot plots, write a few sentences comparing the distribution of the number of correctly recalled real words with the distribution of number of correctly recalled fake words. In particular, comment on which type of word, if either, that students recall better. Explain.
3. Linda made the following calculations for the two data sets:

|  | Mean | MAD |
| :--- | :---: | :---: |
| Real words recalled | 9.43 | 2.29 |
| Fake words recalled | 5.36 | 2.27 |

In the previous lesson, you calculated the number of MADs that separated two sample means. You used the larger MAD to make this calculation if the two MADs were not the same. How many MADs separate the mean number of real words recalled and the mean number of fake words recalled for the students in the study?
4. In the last lesson, our work suggested that if the number of MADs that separate the two sample means is 2 or more, then it is reasonable to conclude that not only do the means differ in the samples, but that the means differ in the populations as well. If the number of MADs is less than 2 , then you can conclude that the difference in the sample means might just be sampling variability and that there may not be a meaningful difference in the population means. Using these criteria, what can Linda conclude about the difference in population means based on the sample data that she collected? Be sure to express your conclusion in the context of this problem.

## Example 2

Ken, an eighth-grade student, was interested in doing a statistics study involving sixth-grade and eleventh-grade students in his school district. He conducted a survey on four numerical variables and two categorical variables (grade level and gender). His Excel population database for the 265 sixth graders and 175 eleventh graders in his district has the following description:

| Column | Name | Description |
| :---: | :--- | :--- |
| 1 | ID | ID numbers are from 1 through 440 |
|  |  | $1-128$ Sixth-grade females |
|  |  | $129-265$ Sixth-grade males |
|  |  | 266 -363 Eleventh-grade females |
|  |  | $364-440$ Eleventh-grade males |
| 2 | Texting | Number of minutes per day text (whole number) |
| 3 | ReacTime | Time in seconds to respond to a computer screen stimulus (two decimal places) |
| 4 | Homework | Total number of hours per week spent on doing homework (one decimal place) |
| 5 | Sleep | Number of hours per night sleep (one decimal place) |

a. Ken decides to base his study on a random sample of 20 sixth graders and a random sample of 20 eleventh graders. The sixth graders have IDs 1-265, and the eleventh graders are numbered 266-440. Advise him on how to randomly sample 20 sixth graders and 20 eleventh graders from his data file.

Suppose that from a random number generator:
The random ID numbers for his 20 sixth graders:
23115192068618323325314236195139752105640661141279
The random ID numbers for his 20 eleventh graders:
391319343426307360289328390350279283302287269332414267428
b. For each set, find the homework hours data from the population database that corresponds to these randomly selected ID numbers.
c. On the same scale, draw dot plots for the two sample data sets.
d. From looking at the dot plots, list some observations comparing the number of hours per week that sixth graders spend on doing homework and the number of hours per week that eleventh graders spend on doing homework.
e. Calculate the mean and MAD for each of the data sets. How many MADs separate the two sample means? (Use the larger MAD to make this calculation if the sample MADs are not the same.)

|  | Mean (hr.) | MAD (hr.) |
| :---: | :--- | :--- |
| Sixth Grade |  |  |
| Eleventh Grade |  |  |

f. Ken recalled Linda suggesting that if the number of MADs is greater than or equal to 2 , then it would be reasonable to think that the population of all sixth-grade students in his district and the population of all eleventh-grade students in his district have different means. What should Ken conclude based on his homework study?

## Lesson Summary

To determine if the mean value of some numerical variable differs for two populations, take random samples from each population. It is very important that the samples be random samples. If the number of MADs that separate the two sample means is 2 or more, then it is reasonable to think that the populations have different means. Otherwise, the population means are considered to be the same.

## Problem Set

1. Based on Ken's population database, compare the amount of sleep that sixth-grade females get on average to the amount of sleep that eleventh-grade females get on average.

Find the data for 15 sixth-grade females based on the following random ID numbers:
651671011068785951204647410231128
Find the data for 15 eleventh-grade females based on the following random ID numbers:
348313297351294343275354311328274305288267301
2. On the same scale, draw dot plots for the two sample data sets.
3. Looking at the dot plots, list some observations comparing the number of hours per week that sixth graders spend on doing homework and the number of hours per week that eleventh graders spend on doing homework.
4. Calculate the mean and MAD for each of the data sets. How many MADs separate the two sample means? (Use the larger MAD to make this calculation if the sample MADs are not the same.)

|  | Mean (hr.) | MAD (hr.) |
| :--- | :--- | :--- |
| Sixth-grade females |  |  |
| Eleventh-grade females |  |  |

5. Recall that if the number of MADs in the difference of two sample means is greater than or equal to 2 , then it would be reasonable to think that the population means are different. Using this guideline, what can you say about the average number of hours of sleep per night for all sixth-grade females in the population compared to all eleventhgrade females in the population?
