Lesson 22: Using Sample Data to Compare the Means of Two or More Populations

Classwork

In previous lessons, you worked with one population. Many statistical questions involve comparing two populations. For example:

- On average, do boys and girls differ on quantitative reasoning?
- Do students learn basic arithmetic skills better with or without calculators?
- Which of two medications is more effective in treating migraine headaches?
- Does one type of car get better mileage per gallon of gasoline than another type?
- Does one type of fabric decay in landfills faster than another type?
- Do people with diabetes heal more slowly than people who do not have diabetes?

In this lesson, you will begin to explore how big of a difference there needs to be in sample means in order for the difference to be considered meaningful. The next lesson will extend that understanding to making informal inferences about population differences.

Examples 1–3

Tamika's mathematics project is to see whether boys or girls are faster in solving a KenKen-type puzzle. She creates a puzzle and records the following times that it took to solve the puzzle (in seconds) for a random sample of 10 boys from her school and a random sample of 11 girls from her school:

													Mean	MAD
Boys	39	38	27	36	40	27	43	36	34	33			35.3	4.04
Girls	41	41	33	42	47	38	41	36	36	32	46		39.4	3.96

1. On the same scale, draw dot plots for the boys' data and for the girls' data. Comment on the amount of overlap between the two dot plots. How are the dot plots the same, and how are they different?

2. Compare the variability in the two data sets using the MAD (mean absolute deviation). Is the variability in each sample about the same? Interpret the MAD in the context of the problem.

3. In the previous lesson, you learned that a difference between two sample means is considered to be meaningful if the difference is more than what you would expect to see just based on sampling variability. The difference in the sample means of the boys' times and the girls' times is 4.1 seconds (39.4 seconds $-$ 35.3 seconds). This difference is approximately 1 MAD.

   a. If 4 sec. is used to approximate the values of 1 MAD for both boys and for girls, what is the interval of times that are within 1 MAD of the sample mean for boys?

   b. Of the 10 sample means for boys, how many of them are within that interval?

   c. Of the 11 sample means for girls, how many of them are within the interval you calculated in part (a)?

   d. Based on the dot plots, do you think that the difference between the two sample means is a meaningful difference? That is, are you convinced that the mean time for all girls at the school (not just this sample of girls) is different from the mean time for all boys at the school? Explain your choice based on the dot plots.

Examples 4–7

How good are you at estimating a minute? Work in pairs. Flip a coin to determine which person in the pair will go first. One of you puts your head down and raises your hand. When your partner says "start," keep your head down and your hand raised. When you think a minute is up, put your hand down. Your partner will record how much time has passed. Note that the room needs to be quiet. Switch roles except this time you talk with your partner during the period when the person with his head down is indicating when he thinks a minute is up. Note that the room will not be quiet.

Group	Estimate for a minute													
Quiet														
Talking														

Use your class data to complete the following.

4. Calculate the mean minute time for each group. Then, find the difference between the "quiet" mean and the "talking" mean.

5. On the same scale, draw dot plots of the two data distributions, and discuss the similarities and differences in the two distributions.

6. Calculate the mean absolute deviation (MAD) for each data set. Based on the MADs, compare the variability in each sample. Is the variability about the same? Interpret the MADs in the context of the problem.

7. Based on your calculations, is the difference in mean time estimates meaningful? Part of your reasoning should involve the number of MADs that separate the two sample means. Note that if the MADs differ, use the larger one in determining how many MADs separate the two means.

Lesson Summary

Variability is a natural occurrence in data distributions. Two data distributions can be compared by describing how far apart their sample means are. The amount of separation can be measured in terms of how many MADs separate the means. (Note that if the two sample MADs differ, the larger of the two is used to make this calculation.)

Problem Set

1. A school is trying to decide which reading program to purchase.
   a. How many MADs separate the mean reading comprehension score for a standard program ($\text{mean} = 67.8$, $\text{MAD} = 4.6$, $n = 24$) and an activity-based program ($\text{mean} = 70.3$, $\text{MAD} = 4.5$, $n = 27$)?
   b. What recommendation would you make based on this result?

2. Does a football filled with helium go farther than one filled with air? Two identical footballs were used: one filled with helium and one filled with air to the same pressure. Matt was chosen from your team to do the kicking. You did not tell Matt which ball he was kicking. The data (yards) follow.

Air	25	23	28	29	27	32	24	26	22	27	31	24	33	26	24	28	30
Helium	24	19	25	25	22	24	28	31	22	26	24	23	22	21	21	23	25

	Mean	MAD
Air		
Helium		

   a. Calculate the difference between the sample mean distance for the football filled with air and for the one filled with helium.
   b. On the same scale, draw dot plots of the two distributions, and discuss the variability in each distribution.
   c. Calculate the MAD for each distribution. Based on the MADs, compare the variability in each distribution. Is the variability about the same? Interpret the MADs in the context of the problem.
   d. Based on your calculations, is the difference in mean distance meaningful? Part of your reasoning should involve the number of MADs that separate the sample means. Note that if the MADs differ, use the larger one in determining how many MADs separate the two means.

3. Suppose that your classmates were debating about whether going to college is really worth it. Based on the following data of annual salaries (rounded to the nearest thousands of dollars) for college graduates and high school graduates with no college experience, does it appear that going to college is indeed worth the effort? The data are from people in their second year of employment.

College Grad	41	67	53	48	45	60	59	55	52	52	50	59	44	49	52
High School Grad	23	33	36	29	25	43	42	38	27	25	33	41	29	33	35

a. Calculate the difference between the sample mean salary for college graduates and for high school graduates.

b. On the same scale, draw dot plots of the two distributions, and discuss the variability in each distribution.

c. Calculate the MAD for each distribution. Based on the MADs, compare the variability in each distribution. Is the variability about the same? Interpret the MADs in the context of the problem.

d. Based on your calculations, is going to college worth the effort? Part of your reasoning should involve the number of MADs that separate the sample means.