Mean Absolute Deviation Guided Notes

Background

Merriam Webster defines **data as facts about something that can be used for calculating, reasoning, or planning.** Scientists use several tools to analyze data for patterns and features to determine meaning. Engineers use data analysis to establish which solution best meets the criteria and keeps within the constraints. Tools to examine data include creating tables and graphs; finding the mean, median, and mode; plus looking at the range. A tool to analyze data more fully is computing the mean absolute deviation.  The mean absolute deviation surveys data for its variability or how consistent were the measurements.

**Mean absolute deviation or MAD of a set of data is the average distance between a data point and the mean.** To compute the MAD, students complete three steps. First, the average of the distances each data point is from mean is computed. Second, find the difference between each data value and the average. It is important to take the absolute value of that difference. Third, is to take the mean of each of the differences.

**Example:**

A pair of students investigated the following question: Does mass affect the time an object takes to fall? They used two identical quart size plastic storage bags. In one bag, they measured and added four scoops of plastic beads. In the other bag, they measured and added eight scoops of plastic beads. The bags were held 1.5 meters above the ground and dropped one at a time. To insure consistency, one student was in charge of all the drops. A second student from the group timed the drops and recorded the data. The pair decided to complete 5 trials.

**Data**

Four scoops (seconds): 0.95, 0.82, 0.85, 0.59, and 0.68

Eight scoops (seconds): 0.92, 0.76, 0.74, 0.88, and 0.94

One of the first tools to data analysis is to organize data into a table format.  Organize the data above into a table. The proper format for a data table is as follows:

**Title (states the purpose of the experiment)**

|  |  |  |
| --- | --- | --- |
| **Independent Variable (unit)** | **Dependent Variable (unit)** | **Derived Quantity (unit)** |
| **Trial 1** | **Trial 2** | **Trial 3** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Notice the data requires 5 trials and only two levels of an independent variable, so an adjustment to the data table structure will need to be made to account for this difference. **What do you think the derived quantity should be?** Average of the times

**Table:**

|  |  |  |
| --- | --- | --- |
| Independent Variable (scoops) | Dependent Variable (seconds) | Derived quantity (seconds) |
| Trial 1 | Trial 2  | Trial 3 | Trial 4 | Trial 5 |
| 4 | 0.95 | 0.82 | 0.85 | 0.59 | 0.68 | 0.78 |
| 8 | 0.92 | 0.76 | 0.74 | 0.88 | 0.94 | 0.85 |

Next tool for data analysis is constructing a graph. **For this data set, which type of graph? Bar or Line** Remember, graphs require a title, both axis labeled (with units), and a scale that presents the data clearly. Graphs allow for a visual comparison of data to discern patterns or major features.

**Graph**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Study the data table and graph. **Using the information provided, is there a difference between the mean drop time for 4 scoops and 8 scoops? Provide evidence supporting your answer.** **What confidence do you have in your data?** Yes, there is a difference between the mean drop time for 4 scoops and 8 scoops. The mean drop time for 4 scoops is 0.78 seconds and the mean drop time for 8 scoops is 0.85 seconds. I do not have confidence that the difference is significant because there is a great deal of overlap between the data sets.

**Mean Absolute Deviation**

**Step 1:** Determine the mean for each set of data (already completed from above)

|  |  |  |
| --- | --- | --- |
| Plastic bag containing beads (scoops) | Time (seconds) | Average Time (seconds) |
| Trial 1 | Trial 2 | Trial 3 | Trial 4 | Trial 5 |
| 4 | .95 | .82 | .85 | .59 | .68 | .78 |
| 8 | .92 | .76 | .74 | .88 | .94 | .85 |

**Step 2:** Find the difference each data point is from the mean. Be sure to take the absolute value of the difference.

**Step 3:** For each set of differences, find the average difference. This the mean absolute deviation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Trial  | Plastic Bag containing 4 scoops (seconds) | Subtract the average for 4 scoops | Absolute value of the difference | Plastic Bag containing 8 scoops (seconds) | Subtract the average for 8 scoops | Absolute value of the difference |
| 1 | 0.95 | 0.78 | 0.17 | 0.92 | 0.85 | 0.07 |
| 2 | 0.82 | 0.78 | 0.04 | 0.76 | 0.85 | 0.09 |
| 3 | 0.85 | 0.78 | 0.07 | 0.74 | 0.85 | 0.11 |
| 4 | 0.59 | 0.78 | 0.19 | 0.88 | 0.85 | 0.03 |
| 5 | 0.68 | 0.78 | 0.10 | 0.94 | 0.85 | 0.09 |
|  |  | MAD | 0.11 |  | MAD | 0.08 |

**Revisit the Graph** - Go back to your bar graph and add the MAD as bars to show variation.

**Questions:**

1. **What is the average time for four scoops? What was the MAD?**

0.78 seconds, 0.11 seconds

1. **Does the MAD show the data as tightly grouped (less than 10% variation) or spread (more than 10% variation) around the mean?**

The data is spread more than 10% variation around the mean.

1. **What is the average time for eight scoops? What was the MAD?**

0.85 seconds, 0.08 seconds

1. **Does the MAD show the data as tightly grouped (less than 10% variation) or spread (more than 10% variation) around the mean?**

The data is within 10% variation around the mean.

1. **Based on the variability in each data set, what confidence do you have in the accuracy of your mean?**

For the mean of 8 scoops, I have confidence in the number. For the mean of 4 scoops, there might have been error in measurement.

1. **Compare the two means, with the new information regarding the variability of the data would you claim the two means are different or similar?**

I doubt the means are significantly different. The data has overlap between the datasets.

**Comparison of the Means**

To more accurately compare two means, it is important to determine the distance between the means which explains why the first step is subtraction. The question becomes, at that point, is the difference significant?  To answer that question, we must compare the variability of the data to the difference. The answer is a ratio. The ratio, when greater than 1, usually indicates the means of each data set are different. When the ratio is less than 1, the means of each data set are similar. This sounds confusing, but there is a chart that will help us.

**A two-step process:**

**Step one** - find the difference between the means.

Difference between the means = mean of 8 scoops - mean of 4 scoops

0.85 seconds – 0.78 seconds = 0.07 seconds

**Step two** - Divide the difference of the means by the MAD.

0.07 seconds/0.11 seconds = 0.63 0.07 seconds/0.08 seconds = 0.875

Use the chart to determine if the difference of the means is significant. To use the chart, find the number of trials along the left side. For our example, we will use 5 trials. Then, move right along the 5 trials line on the chart until the ratio of the difference between the means divided by the MAD that was computed. The color of the square of the intersection indicates if the difference is significant or not. Notice for 5 trials, the ratio needs to be greater than 2.93 in order for the difference to be significant.



**More questions:**

1. **According to the ratio of the difference of the means, are the averages for the two data sets significantly different?  Why or why not?**

The averages of the two data sets are not significantly different because when we compare the difference of the means with the MAD, the overlap in datasets is too great.

1. **Looking at the chart, describe the effect of increasing the number of trials has on the ratio of the difference of means divided by MAD.**

Increasing the number of trials makes a lower ratio between averages possibly significant.

1. **Taking your answer from question 2, how many trials would you recommend running to determine if the averages of the data sets are significantly different?**

In this case, increasing the number of trials to 20 and possibly having the ratio move closer to 1 would make the difference significant. If the ratio does not increase (which should not happen considering that mass does not affect fall time), then the number of trials will not change the significance.

1. **After calculating, why is it important to take accurate data measurements? How does the range of your data affect your MAD?**

Smaller ranges (no more than 10 % variability in the data) indicate more confidence in the average. Small ranges also creates a smaller MAD, a smaller MAD means the divisor is small which increases the ratio. A larger ratio means more confidence in the difference of means.

1. **So, taking this back to our original question, does the mass of the object affect the time of fall?  Support your answer with evidence.**

The mass of the object does not affect time of fall. Even though there is a difference in the two means, the difference is not significant which aligns with accepted science explanations.

Problem:

A student wanted to determine if more light passed through salt water or fresh water. The student took 5 samples of each type of water. The results are in the data table below:

|  |  |
| --- | --- |
|  | Type of water versus light transmittance -Transmittance (%T) |
| Type of water | Trial 1 | Trial 2 | Trial 3 | Trial 4  | Trial 5 |
| Fresh Water | 77.23 | 74.50 | 64.88 | 75.27 | 54.66 |
| Ocean Water | 85.23 | 92.82 | 78.94 | 60.71 | 57.96 |

What claim can the student make and support with evidence? This will require you to calculate the MAD and complete a comparison of the means.