Name:_____ Date:_____

A Tale of Two Rules

Background:

How consistent were our rules for the contest in determining a winner?

Statisticians and Data Scientists use rules, like mean squared error, for determining the accuracy of their predictions.

- The **mean squared error** rule says: the score is determined by finding the average of the squared differences between the prediction and the actual values.

Instructions:

For each of the test datasets below, calculate the mean squared error and determine which prediction is best.

Dataset A

Here are the heights from Dataset A – 70.1, 61, 70.1, 68.1, 63, 66.1, 61, 70.1, 72.8, 70.9

	Summary of 40 Students (heights in inches)					
			Team A	Team B	Team C	
	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum
	59.1	65	67.9	68.1	70.9	76
MSE		22.05				

Which team/statistic made the best prediction using MSE?

Dataset B

Here are the heights from Dataset B – 70.1, 72, 68.9, 61.8, 70.9, 59.8, 72, 65, 66.1, 68.9

	Summary of 40 Students (heights in inches)					
			Team A	Team B	Team C	
	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum
	59.1	65	67.9	68.1	70.9	76
MSE						

Which team/statistic made the best prediction using MSE?

Using the mean squared errors, which team/statistic made the best prediction for both test datasets?

Name:_____ Date:_____

A Tale of Two Rules

Background:

Statisticians and Data Scientists use another rule called the mean absolute error for determining the accuracy of their predictions.

- The **mean absolute error** rule says: the score is determined by finding the average of the absolute value of the differences between the prediction and the actual values.

Rules:

Now you will calculate the mean absolute error and determine which prediction is best. Use what you learned in class to write the lines of code to calculate the MAE then run them in RStudio.

1. _____
2. _____
3. _____
4. _____
5. _____

Dataset A

Here are the heights from Dataset A – 70.1, 61, 70.1, 68.1, 63, 66.1, 61, 70.1, 72.8, 70.9

	Summary of 40 Students (heights in inches)					
			Team A	Team B	Team C	
	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum
	59.1	65	67.9	68.1	70.9	76
MAE						

Which team/statistic made the best prediction using MAE?

Dataset B

Here are the heights from Dataset B – 70.1, 72, 68.9, 61.8, 70.9, 59.8, 72, 65, 66.1, 68.9

	Summary of 40 Students (heights in inches)					
			Team A	Team B	Team C	
	Minimum	1st Quartile	Mean	Median	3rd Quartile	Maximum
	59.1	65	67.9	68.1	70.9	76
MAE						

Which team/statistic made the best prediction using MAE?

Using the mean absolute errors, which team/statistic made the best prediction for both test datasets?