Name: _____   Date: _____ |

Prediction Games

Instructions:
For each of the games below, calculate the statistics and determine which one works best.
- 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.
- 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.

Game 1

Given the following statistics from randomly selected height training data, determine the best predictor for the following testing data.

Here are the heights from the testing data – 66, 67, 73, 68, 68, 73, 69, 64, 66, 67

	Training Data Summary (Heights in Inches)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	64.20	66.40	67.76	68.22	69.13	73.15
MSE						
MAE						

Which statistic did best with MSE?

Which statistic did best with MAE?

Game 2

Given the following statistics from a FitBit's daily steps, predict the number of steps, as counted by a FitBit, this person will take in the future. Choose your prediction from these values:

	Summary (daily steps)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	0	0	4370	7708	13220	27900
MSE						
MAE						

Outcomes: here are the actual daily steps that this person took – 0, 27903, 6044, 0, 0, 17436, 2697, 14944, 8060, 0.

Which of these numbers did best? Compare your score using the mean squared errors.

Prediction Games

Game 3

Predict the number of minutes it took 10 randomly selected teenagers to run the Cherry Blossom 10 Mile Race in Washington, DC. Choose your prediction from these values:

	Summary (race in minutes)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	70.52	73.95	85.28	90.87	102.10	123.30
MSE						
MAE						

Outcomes: here are the actual race times of the teenagers – 74, 123, 121, 103, 75, 72, 85, 71, 86, 101.

Which of these numbers did best? Compare your score using the mean squared errors.

Using the mean squared errors, which statistic/ team is the winner that made the best predictions in all three games?

Prediction Games
ANSWER KEY

Game 1

Given the following statistics from randomly selected height training data, determine the best predictor for the following testing data.

Here are the heights from the testing data – 66, 67, 73, 68, 68, 73, 69, 64, 66, 67

	Training Data Summary (Heights in Inches)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	64.20	66.40	67.76	68.22	69.13	73.15
MSE	22.9	10.58	7.8056	7.7044	8.7509	33.1925
MAE	3.94	2.34	2.1	2.188	2.578	5.05

Which statistic did best with MSE?
Mean

Which statistic did best with MAE?
Median

Game 2

Given the following statistics from a FitBit's daily steps, predict the number of steps, as counted by a FitBit, this person will take in the future. Choose your prediction from these values:

	Summary (daily steps)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	0	0	4370	7708	13220	27900
MSE	141468199	141468199	931933683	82048768	112426503	489749479
MAE	7708.4	7708.4	7169	7501.8	9636.2	20192.2

Outcomes: here are the actual daily steps that this person took – 0, 27903, 6044, 0, 0, 17436, 2697, 14944, 8060, 0.

Which of these numbers did best? Compare your score using the mean squared errors.
Student comparisons will vary depending on which value they chose as their prediction. Overall, mean is the best predictor when using MSE and median is the best predictor when using MAE.

Game 3

Predict the number of minutes it took 10 randomly selected teenagers to run the Cherry Blossom 10 Mile Race in Washington, DC. Choose your prediction from these values:

	Summary (race in minutes)					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
	70.52	73.95	85.28	90.87	102.10	123.30
MSE	777.0264	647.6125	387.3624	353.5429	474.49	1390.33
MAE	20.58	18.13	15.7	16.674	19.14	32.2

Outcomes: here are the actual race times of the teenagers – 74, 123, 121, 103, 75, 72, 85, 71, 86, 101.

Which of these numbers did best? Compare your score using the mean squared errors.
Student comparisons will vary depending on which value they chose as their prediction. Overall, mean is the best predictor when using MSE and median is the best predictor when using MAE.

Using the mean squared errors, which statistic/ team is the winner that made the best predictions in all three games?
The MEAN is the statistic/ team that was the best predictor when using MSE in all three games.