Welcome MYP 9 Mathematics!

	Assignment Effort Grade (Circle One)	Comments (What was interesting or challenging?)
Monday Date: 4/16 Topic: NO SCHOOL	0 1 2	
Tuesday Date: 4/17 Topic: 10FG Five Number Summary & Box Plots	0 1 2	
Wednesday Date: _____ Topic: _____	0 1 2	
Thursday Date: _____ Topic: _____	0 1 2	
Friday Date: _____ Topic: _____	0 1 2	

Class Plan

1. Warm-up
2. Investigation

-How do we interpret the different measures of spread?

-How can we verify we have outliers?

F MEASURING THE SPREAD OF A DATA SET

3. Practice

C^2O Fair	8:05-9:30
Lunch A Schedule	
1	9:35-10:12
2	10:17-10:54
3	10:59-11:36
LUNCH	11:41-12:11
4	12:16-12:53
5	12:58-1:35
6	1:40-2:17
7	2:22-3:00

C^2O Fair	8:05-9:30
Lunch B Schedule	
1	9:35-10:12
2	10:17-10:54
3	10:59-11:36
4	11:41-12:18
LUNCH	12:23-12:53
5	12:58-1:35
6	1:40-2:17
7	2:22-3:00

C^2O Fair	8:05-9:30
Lunch C Schedule	
1	9:35-10:12
2	10:17-10:54
3	10:59-11:36
4	11:41-12:18
5	12:23-1:00
LUNCH	1:05-1:35
6	1:40-2:17
7	2:22-3:00

Warm-up:

Create a 5 # Summary.

(Use your graphing calculator)

Which players could be outliers?

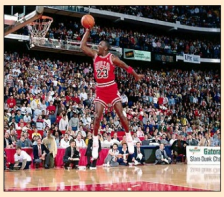
```
1-Var Stats
↑n=10
 minX=198
 Q1=375
 Med=524.5
 Q3=841
 maxX=2357
```

Points Scored by Chicago Bulls Players Who Played over 40 Games (1997-98 Season)

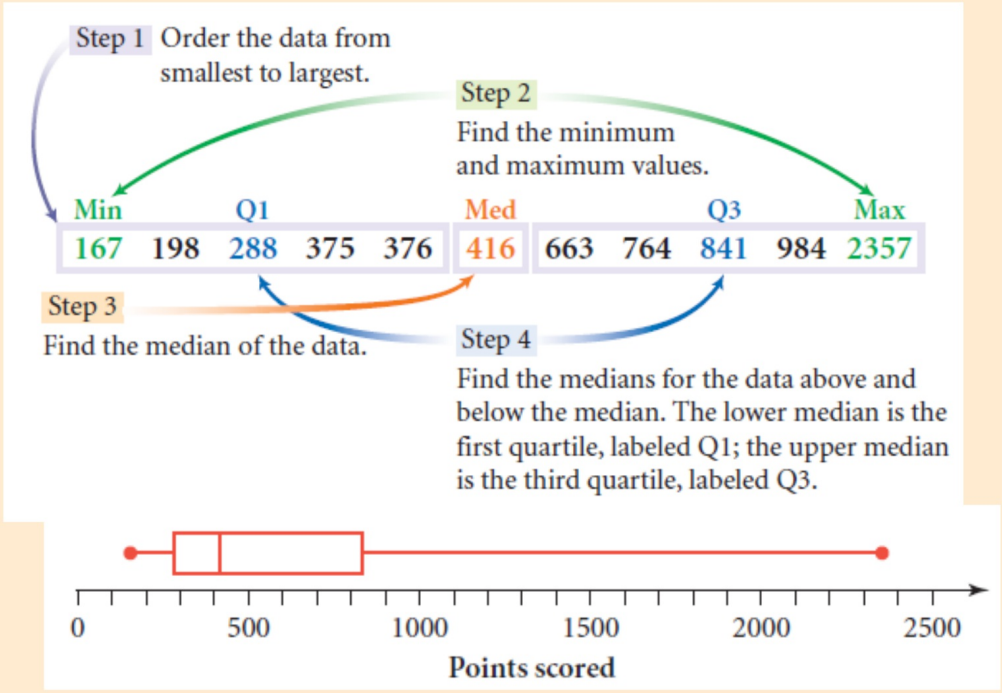
Chicago Bulls	Points
Michael Jordan	2357
Toni Kukoe	984
Scottie Pippen	841
Ron Harper	764
Lue Longley	633
Scott Burrell	416

Chicago Bulls	Points
Steve Kerr	376
Dennis Rodman	375
Randy Brown	288
Jud Bucchler	198

National Basketball Association (www.nba.com)



Chicago Bulls Data 5 # Summary and Box Plot



What is an outlier?

Is there a formula to determine if a data point is an outlier?

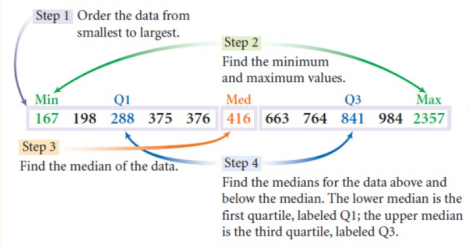
***A data point that is distinctly separate from the rest of the data.**

***One definition of outlier is any data point more than 1.5 interquartile ranges (IQRs) below the first quartile or above the third quartile.

$$\text{Outlier} < Q_1 - 1.5IQR$$

$$\text{Outlier} > Q_3 + 1.5IQR$$

Examining Outliers of Chicago Bulls



IQR = 841-288 = 553

1.5(553) = 829.5

Q1 - 1.5IQR = 288-829.5 = -541.5

Q3 + 1.5IQR = 841+829.5 = 1670.5

Did anyone score a number of points outside of -541.5 to 1670.5?

Michael Jordan	2357

Investigation: Minnesota Lynx Data

Minnesota Lynx Players and their minutes played (2016 season)

Maya Moore	263
Lindsay Whalen	213
Sylvia Fowles	251
Seimone Augustus	215
Rebekkah Brunson	234
Natasha Howard	104
Jia Perkins	141
Renee Montgomery	101
Keisha Hampton	8
Anna Cruz	51
Janel McCarville	19

1. Five Number Summary
2. Outliers
3. Standard Deviation (Measure of Spread)

Google

Interpret & Analyze Measures

***Come back together at end to go over solution & discuss standard deviation

Done: Help others!

Investigation: Minnesota Lynx Data

1) Calculate using the Lynx player's time.

Minimum 8 Maximum 263 Range 255

2) Analyze and interpret what the range says about the player's time on the basketball court.

Maya Moore	263
Lindsay Whalen	213
Sylvia Fowles	251
Seimone Augustus	215
Rebekkah Brunson	234
Natasha Howard	104
Jia Perkins	141
Renee Montgomery	101
Keisha Hampton	8
Anna Cruz	51
Janel McCarville	19

Investigation: Minnesota Lynx Data SOLUTION

Minnesota Lynx Players and their minutes played (2016 season)

Maya Moore	263
Lindsay Whalen	213
Sylvia Fowles	251
Seimone Augustus	215
Rebekkah Brunson	234
Natasha Howard	104
Jia Perkins	141
Renee Montgomery	101
Keisha Hampton	8
Anna Cruz	51
Janel McCarville	19

1) Calculate using the Lynx player's time.

Minimum 8 Maximum 263 Range 255

2) Analyze and interpret what the range says about the player's time on the basketball court.

There are 255 minutes between the players who have the most and least number of minutes played on the court.

Investigation: Minnesota Lynx Data

3) Identify any possible outliers and explain how they could affect the data.

Maya Moore	263
Lindsay Whalen	213
Sylvia Fowles	251
Seimone Augustus	215
Rebekkah Brunson	234
Natasha Howard	104
Jia Perkins	141
Renee Montgomery	101
Keisha Hampton	8
Anna Cruz	51
Janel McCarville	19

4) Calculate using the Lynx player's time.

Min_____ Q_1_____ Median_____ Q_3_____ Max_____ IQR = 183

5) Analyze and interpret what the **IQR** says about the player's time on the basketball court.

Investigation: Minnesota Lynx Data SOLUTION

3) Identify any possible outliers and explain how they could affect the data.

Maya Moore, Keisha Hampton, and Jane McCarville could affect the average time playing in the season

4) Calculate using the Lynx player's time.

$Q_3 - Q_1 = IQR$

Min 8 Q_1 51 Median 141 Q_3 234 Max 263 IQR = 234-51 = 183

5) Analyze and interpret what the **IQR** says about the player's time on the basketball court.

50% of the Lynx players have played between 51 minutes and 234 minutes in the season.

183 minutes of time is represented in this 50%.

^ Spread

Investigation: Minnesota Lynx Data

6) Verify that the possible outliers you identified in #3) ***are in fact*** outliers.

Outlier $< Q_1 - 1.5IQR$
Outlier $> Q_3 + 1.5IQR$

IQR = 183 min $Q_3 = 234$

$Q_1 = 51$

$1.5(183) = 274.5$

$51 - 274.5$

-223.5 minutes

Below

$234 + 274.5$

$= 508.5$ min.

anyone above?

No outliers

Calculating OUTLIERS

"1.5IQR above Q_3 or below Q_1"

IQR(Interquartile Range) = Q3 – Q1

Any point that falls outside the interval calculated by

Q1- 1.5(IQR) and Q3 + 1.5(IQR)

is considered an outlier.

Investigation: Minnesota Lynx Data SOLUTION

6) Verify that the possible outliers you identified in #3) ***are in fact*** outliers.

Outlier $< Q_1 - 1.5IQR$
Outlier $> Q_3 + 1.5IQR$

$1.5IQR = 1.5(183) = 274.5$

$Q_1 - 1.5IQR$
$51 - 274.5$
-223.5
No players below -223.5

$Q_3 + 1.5IQR$
$234 + 274.5$
508.5
No players above 508.5 minutes

Conclusion
No outliers of time played

Investigation: Minnesota Lynx Data

Finding the Standard Deviation

What is a standard deviation?

Deviation just means how far from the normal

Standard Deviation

The Standard Deviation is a measure of how spread out numbers are.

Its symbol is σ (the greek letter sigma)

Investigation: Minnesota Lynx Data SOLUTION

Finding the Standard Deviation

What is a standard deviation?

A measure of how far the data points are on average from the mean. σ: Sigma

Investigation: Minnesota Lynx Data

How can we find the average distance from the mean ("norm")?

Challenge Yourself.

Find the algorithm for standard deviation by hand.

(average distance from the mean...)

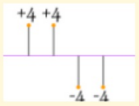
Challenge Yourself.

Algorithm for standard deviation.

(average distance from the mean...)

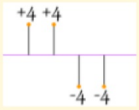
***Footnote: Why *square* the differences?**

If we just add up the differences from the mean ... the negatives cancel the positives:



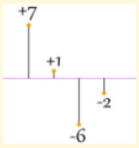
$$\frac{4 + 4 - 4 - 4}{4} = 0$$

So that won't work. How about we use absolute values?



$$\frac{|4| + |4| + |-4| + |-4|}{4} = \frac{4 + 4 + 4 + 4}{4} = 4$$

That looks good (and is the Mean Deviation), but what about this case:



$$\frac{|7| + |1| + |-6| + |-2|}{4} = \frac{7 + 1 + 6 + 2}{4} = 4$$

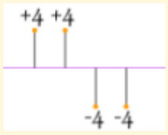
Oh No! It also gives a value of 4, Even though the differences are more spread out.

Challenge Yourself.

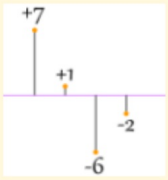
Algorithm for standard deviation.

(average distance from the mean...)

So let us try squaring each difference (and taking the square root at the end):



$$\sqrt{\left(\frac{4^2 + 4^2 + 4^2 + 4^2}{4}\right)} = \sqrt{\left(\frac{64}{4}\right)} = 4$$



$$\sqrt{\left(\frac{7^2 + 1^2 + 6^2 + 2^2}{4}\right)} = \sqrt{\left(\frac{90}{4}\right)} = 4.74...$$

That is nice! The Standard Deviation is bigger when the differences are more spread out ... just what we want.

In fact this method is a similar idea to distance between points, just applied in a different way.

And it is easier to use algebra on squares and square roots than absolute values, which makes the standard deviation easy to use in other areas of mathematics.

Standard deviation by hand.

(average distance from the mean...)

Test Scores	Mean $\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
72	69	3	9
55	69	−14	196
53		−16	256
51		−18	324
74		5	25
71		2	4
83		14	196
95		26	676
67		−2	4
		Sum	

Subtract

$n = 9$

$\bar{x} = 69$

$\frac{sum}{n} =$

$\sigma = \sqrt{\frac{sum}{n}} =$

Standard deviation =

Standard deviation by hand.

(average distance from the mean...)

Test Scores	Mean $\bar{x}$	$x_i - \bar{x}$ Sum=0	$(x_i - \bar{x})^2$
72	69	3	9
55	69	-14	196
53	69	-16	256
51	69	-18	324
74	69	5	25
71	69	2	4
83	69	14	196
95	69	26	676
67	69	-2	4
		Sum	1690

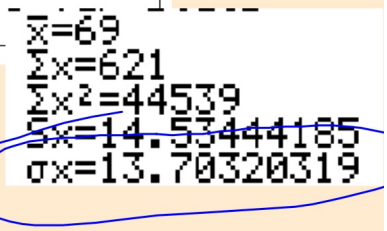
$n = 9$

$\bar{x} = 69$

$\frac{sum}{n} = \frac{1690}{9}$

$\sigma = \sqrt{\frac{sum}{n}} = \sqrt{187.8}$

Standard deviation = 13.7



Standard deviation by calculator:

Investigation: Minnesota Lynx Data

7) Calculate using the Lynx player's time. Mean = 145.5

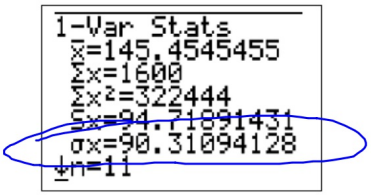
8) Use the steps below to find the standard deviation of the player's time using a calculator. (TI-83/84)

i. **STAT, 1:EDIT**, then enter the data into a list.

ii. **STAT, CALC, 1:1-var stats, ENTER**

You should see a picture like:

Find **σ** = 90.3 min



9) Interpret what this measure of spread means.

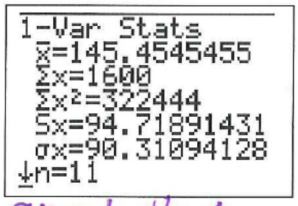
Investigation: Minnesota Lynx Data SOLUTION

7) Calculate using the Lynx player's time. Mean = 145.45 minutes

8) Use the steps below to find the standard deviation of the player's time using a calculator. (TI-83/84

i. **STAT, 1:EDIT,** then enter the data into a list.

ii. **STAT, CALC, 1:1-var stats, ENTER**



You should see a picture like:

Find σ = 90.3 minutes

9) Interpret what this measure of spread means.

On average, each Player is about 90 minutes from the average time of 145.45 minutes.

Investigation: Minnesota Lynx Data

10) Suppose an all-star team is compiled of players that are within 1 standard deviation ***above*** the mean. Which Lynx players would qualify for the all-star team?

$\bar{x} \approx 145.5 \text{ min}$ $\sigma \approx 90 \text{ min}$

$\bar{x} + 1\sigma = 145.5 + 90$

≈ 235.5

Did anyone play above??

Investigation: Minnesota Lynx Data SOLUTION

10) Suppose an all-star team is compiled of players that are within 1 standard deviation ***above*** the mean. Which Lynx players would qualify for the all-star team?

1 s.d = 90.3 mean + 1σ = 145.45 + 90.3 ≈ 235.75

Sylvia Fowles and Maya Moore have minutes over 235.75.

Exercises...

Analyzing Data Spread Practice Worksheet

Analyzing Data Spread Practice Name_____

1. The lengths in minutes of nine playlists are 45, 63, 74, 69, 72, 53, 72, 73, and 50. $\sigma \approx 10.6$ minutes

a) Find the 5 # Summary of the data and create a box plot on the line below to show the data spread.

Exercises...

b) Calculate and interpret the IQR. Use the IQR to verify that there are ***no outliers.***

$\text{Outlier} < Q_1 - 1.5IQR$

$\text{Outlier} > Q_3 + 1.5IQR$

c) Suppose playlists that are 1 standard deviation ***above*** the mean require too much memory on our phones. Which playlist(s) would be too ***big*** for our phone?

Exercises...

2. Apply the meaning of standard deviation.

APPLICATION The mean diameter of a Purdy Goode Compact Disc is 12.0 cm, with a standard deviation of 0.012 cm. No CDs can be shipped that are more than one standard deviation from the mean. How would the company's quality control engineer use those statistics?

Exercises...

3. The students in four classes recorded their resting pulse rates in beats per minutes. The class means and standard deviations are given in the table.

Resting Pulse Rates (beats/min)

Class	Mean	Standard deviation
First period	79.4	3.2
Third period	74.6	5.6
Fifth period	78.2	4.1
Sixth period	80.2	7.6

a) Which class has students with pulse rates most alike? How can you tell?

b) Which class has the students with the fastest pulse rates? How could you show this information?

c) Which class had the biggest range of pulse rates? How can you tell?

Exercises...

4. Thirty students were given an English vocabulary test out of 100 points. The number of questions the students ***got wrong*** are listed below.

{5, 2, 3, 45, 13, 12, 10, 4, 4, 5, 6, 7, 9, 3, 0, 2, 11, 23, 29, 11, 23, 22, 23, 24, 25, 6, 17, 18, 9, 2}

a) Find the 5 # Summary and IQR of the data.

b) Calculate and verify whether there are ***any outliers.***

$$\text{Outlier} < Q_1 - 1.5IQR$$
$$\text{Outlier} > Q_3 + 1.5IQR$$

c) Students who were **upper** outliers received an award. How many students won an award?

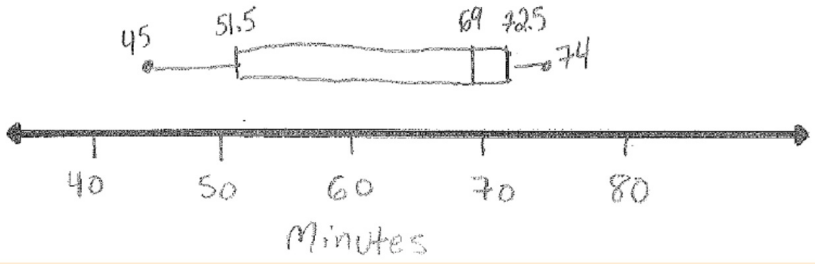
Exercise Solutions...

1. The lengths in minutes of nine playlists are 45, 63, 74, 69, 72, 53, 72, 73, and 50. $\sigma \approx 10.6$ minutes

a) Find the 5 # Summary of the data and create a box plot on the line below to show the data spread.

{45, 51.5, 69, 72.5, 74}

Playlist of Music



Exercise Solutions...

b) Calculate and interpret the IQR. Use the IQR to verify that there are ***no outliers.***

Outlier $< Q_1 - 1.5IQR$

Outlier $> Q_3 + 1.5IQR$

$1.5IQR = 1.5(21) = 31.5$

NO OUTLIERS!

$IQR = Q_3 - Q_1$

$IQR = 72 - 51.5$

IQR = 21 min

$Q_1 - 31.5$

$51.5 - 31.5 = 20$ min

No playlist under 20 min

$Q_3 + 31.5$

$72 + 31.5 = 103.5$ min

No playlist above 103.5 min

c) Suppose playlists that are 1 standard deviation ***above*** the mean require too much memory on our phones. Which playlist(s) would be too ***big*** for our phone?

$\sigma \approx 10.6$ m

$\bar{X} \approx 63$ min

$\sigma + \bar{X} \approx 10.6 + 63 \approx 73.6$ minutes

The 74 min playlist is slightly too big for the available memory.

2. Apply the meaning of standard deviation

Note: Answer could be different for **c,** depending how you rounded the mean.

Exercise Solutions...

2. Apply the meaning of standard deviation.

APPLICATION The mean diameter of a Purdy Goode Compact Disc is 12.0 cm, with a standard deviation of 0.012 cm. No CDs can be shipped that are more than one standard deviation from the mean. How would the company's quality control engineer use those statistics?

$1\sigma + \bar{X} \approx .012 + 12 \approx 12.012\,cm$

The control engineer should NOT ship CD's with diameters larger than 12.012 cm. The engineer should make sure all packaged CD's diameter's are measured.

Exercise Solutions...

3. The students in four classes recorded their resting pulse rates in beats per minutes. The class means and standard deviations are given in the table.

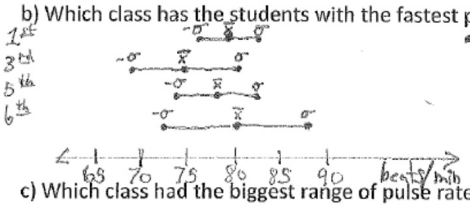
Resting Pulse Rates (beats/min)

Class	Mean	Standard deviation
First period	79.4	3.2
Third period	74.6	5.6
Fifth period	78.2	4.1
Sixth period	80.2	7.6

a) Which class has students with pulse rates most alike? How can you tell?

Firts period! Their SD is the smallest, so they have less deviation from the mean beats/min.

b) Which class has the students with the fastest pulse rates? How could you show this information?



- Periods 3 and 5 would NOT have the fastest pulse rates.
- Period 6 could have the fastest pulse rates because their mean is the highest and their positive deviations ↑,
- Period 1 could have the fastest rates, because a student chosen at random will consistently have a high rate, ↓ deviation

c) Which class had the biggest range of pulse rates? How can you tell?

6th period, On average, each student has the highest difference from their average pulserate, Each student is about 7.6 beats/min above or below 80 beats/min.

Exercise Solutions...

4. Thirty students were given an English vocabulary test out of 100 points. The number of questions the students ***got wrong*** are listed below. $\bar{x} = 12.433$ $\sigma = 10.26366364$

{5, 2, 3, 45, 13, 12, 10, 4, 4, 5, 6, 7, 9, 3, 0, 2, 11, 23, 29, 11, 23, 22, 23, 24, 25, 6, 17, 18, 9, 2}

a) Find the 5 # Summary and IQR of the data.

Min = 0
$Q_1 = 4$
Median = 9.5
$Q_3 = 22$
Max = 45

IQR = 18

$4 - 1.5(18) = 4 - 27 = -23$ Lower Outlier Boundary

b) Calculate and verify whether there are ***any outliers.***

Outlier $< Q_1 - 1.5IQR$
Outlier $> Q_3 + 1.5IQR$

$22 + 1.5(18) = 22 + 27 = 49$ Upper Outlier Boundary

Although 45 appears much higher than the rest of the data, it does NOT fall outside the upper outlier boundary (as defined $Q_3 + 1.5$ IQR).

There are no lower or upper outliers in the data set.

c) Students who were **upper** outliers received an award. How many students won an award?

None. No students recieved the award. Additionally, this doesn't make sense in the context of questions answered WRONG. An award should be given to students who got a lower number of questions wrong. I would suggest giving awards to students in the lowest quartile or the 25% of students getting the least wrong.