



Lesson 3: Teams & Players by the Numbers

Objective: Students will be able to:

- Review how to find the mean, median and mode of a data set.
- Calculate the standard deviation of a data set.
- Evaluate data using a stemplot.

Time Required: 1-2 class periods

Materials Needed:

- Copies of the “Batting Stats by Team for the 2011 Season” table (included) – 1 for each student
- Copies of the “MLB Scouting Reports” packet (included) – 1 for each student or pair of students
- Calculators
- Scrap Paper
- Pencils
- Internet access for student research

Vocabulary:

Mean – The average of a set of data

Median - The middle number in a set of data when the data is arranged in numerical order

Mode - The number that appears the most in a data set

Range - The difference between the largest and smallest number in a data set

Standard Deviation - A number that measures how far away each number in a data set is from the mean

Statistics - A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of numerical data

Stemplot - A device for presenting quantitative data in a graphical format so that it is easy to see the shape of a distribution

Applicable Common Core State Standards:

CCSS.Math.Content.HSN-Q.A.1 Use units as a way to understand problems and to guide the solution of multi-step problems; choose and interpret units consistently in formulas; choose and interpret the scale and the origin in graphs and data displays.

CCSS.Math.Content.HSN-Q.A.2 Define appropriate quantities for the purpose of descriptive modeling.

CCSS.Math.Content.HSA-REI.A.1 Explain each step in solving a simple equation as following from the equality of numbers asserted at the previous step, starting from the assumption that the original equation has a solution. Construct a viable argument to justify a solution method.

CCSS.Math.Content.HSA-REI.B.3 Solve linear equations and inequalities in one variable, including equations with coefficients represented by letters.

CCSS.Math.Content.HSS-ID.A.1 Represent data with plots on the real number line (dot plots, histograms, and box plots).

CCSS.Math.Content.HSS-ID.A.2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.

CCSS.Math.Content.HSS-ID.A.3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).



Lesson

1. To begin this lesson, provide students with the “Batting Stats by Team for the 2011 Season” table (included).
2. Have students spend a minute or two examining the information available on the table. Ask students, what type of data is available on this table? Is there anything that stands out to you? Is it easy or difficult to try to compare teams just by looking at the numbers?
3. Explain that teams could be compared using every column of data on this table, but since that would take a very long time, start by comparing teams using a single data set: number of hits (H). Explain that it is often helpful to start evaluating data by creating a graph. For this type of data, a **stemplot** will allow for easy comparison.
4. Demonstrate how to create a stemplot. A basic stemplot contains two columns separated by a vertical line. The left column contains the “stems” and the right column contains the “leaves.” For hits, it will be easiest to create a stem using the ones and tens places, and a leaf of the hundreds and thousands places. For example, Boston had 1600 hits. The stem and leaf would look like this:

16|00

5. Next write down all of the possible stems (12-16), and record each hit total by writing down the leaf on the line corresponding to the stem.
6. Work with students to create a stemplot for all 30 teams. Have students copy this example into their notebooks.

12|8463
13|8457248094873095455819572527
14|523829227709342342
15|99401360
16|00

7. Once the stemplot is finished, encourage students to draw conclusions from the shape of the data on the plot. For example, although they haven't calculated it yet, students should be able to determine that the mean and median will fall in the 1300-1400 range. Students should also be able to identify the spread of data as well as any outliers.
8. Now, have students go over the stemplot again, this time replacing the leaf with an abbreviation for the team's name (using different colored pens or pencils for each league helps).

```

12|SD SEA
13|TOR AZ TB CLE LAA CHW OAK LAD ATL FL WA MIN PIT SF
14|NYY CIN COL MIL NYM PHI BAL CHI HOU
15|TEX DET STL KC
16|BOS

```

9. Have students draw further conclusions from this stemplot. With the team names in place of numbers, it's easier to see that there was a greater spread between hits for AL teams than hits for NL teams. Other observations might include:
 - Since the "Batting Stats" table is ranked by Runs, we can determine that Toronto was able to achieve the 6th most overall runs with a lower number of hits than other top run-scoring teams. It could be that Toronto was better able to get players into scoring position. Also, a quick examination of Toronto's HR data will show that the team had the 5th highest HR total, and a .413 team SLG. Although they had fewer hits than many teams, it appears that they had more powerful ones.
10. Review the definitions for **mean**, **median**, and **mode**.
 - **Mean** is the average of a set of data. To calculate the mean, find the sum of the data and then divide by the number of data.
 - **Median** is the middle number in a set of data when the data is arranged in numerical order. First arrange the data in numerical order. Then find the number in the middle or the average of the two numbers in the middle.
 - **Mode** is the number that appears the most. Sometimes a data set will have more than one mode (bimodal), and sometimes there is no mode in a data set (when all numbers occur only once).
11. Have students figure out the mean, median, and mode for the "hits" column on the 2011 Batting Stats table. *Answers:* Mean = 1408.9, Median = 1394.5, Mode = 1357

12. While it is useful to know that the average number of hits by all teams in Major League Baseball, the mean does not reveal anything about the distribution or variation of the data set. So we need to come up with some way of measuring not just the average, but also the spread of the distribution of data.
13. Explain that the **standard deviation** is a number that measures how far away each number in a data set is from the mean. If the standard deviation is large, it means the numbers are spread out from their mean. If the standard deviation is small, it means the numbers are close to their mean.
14. Walk students through the process of determining the standard deviation for hits in the American League in 2011. Have students copy down this example into their notebooks for future reference.
15. Listed below is the number of hits for all American League teams.
Step 1 - Determine the mean. *Answer:* $(20004/14)$ Mean = 1428.9

H
1600
1452
1599
1540
1384
1560
1434
1324
1380
1394
1387
1330
1357
1263

16. Step 2 – For each team's hit total, determine the distance from the mean.

H	Distance from Mean
1600	171.1
1452	23.1
1599	170.1
1540	111.1
1384	-44.9
1560	131.1
1434	5.1
1324	-104.9
1380	-48.9
1394	-34.9
1387	-41.9
1330	-98.9
1357	-71.9
1263	-165.9

17. Step 3 - Square each distance.

H	Distance from Mean	Distance Squared
1600	171.1	29275.21
1452	23.1	533.61
1599	170.1	28934.01
1540	111.1	12343.21
1384	-44.9	2016.01
1560	131.1	17187.21
1434	5.1	26.01
1324	-104.9	11004.01
1380	-48.9	2391.21
1394	-34.9	1218.01
1387	-41.9	1755.61
1330	-98.9	9781.21
1357	-71.9	5169.61
1263	-165.9	27522.81

18. Step 4 - Add up the sum of the distances squared. Answer = 149157.7

19. Step 5 - Divide by (n-1) where n represents the amount of numbers you have. (In this case 14)

$$\frac{149157.7}{(14-1)} = 11473.7$$

20. Step 6 - Take the square root of the average distance.

$$\sqrt{11473.7} = 107.1$$

21. The standard deviation for this data set is 107.1.

22. Have students work on their own to determine the standard deviation for hits in the National League.

Teacher's Reference for NL:

H	Distance from Mean	Distance Squared
1513	121.6	14786.56
1438	46.6	2171.56
1429	37.6	1413.76
1357	-34.4	1183.36
1422	30.6	936.36
1477	85.6	7327.36
1409	17.6	309.76
1423	31.6	998.56
1395	3.6	12.96
1345	-46.4	2152.96
1358	-33.4	1115.56
1319	-72.4	5241.76
1442	50.6	2560.36
1325	-66.4	4408.96
1284	-107.4	11534.76
1327	-64.4	4147.36

Sum of Distances Squared = 60301.96

$$60301.96/(16-1) = 4020.1$$

$$\sqrt{4020.1} = 63.4 \quad \text{Standard Deviation} = 63.4$$

23. Have students compare the standard deviations of the two data sets.
AL = 107.1, NL = 63.4.
24. Encourage students to draw conclusions from the standard deviation of hits.
Basically, the standard deviation of this data set tells us that National League teams are more equally balanced in terms of hits than teams in the American League.
25. Introduce the activity.



Activity

1. For this activity, students may work in pairs or individually.
2. Explain the instructions for this activity. Each student (or pair of students) has been hired as a scout for a Major League Baseball team. The team owner is in the market for a new first baseman. As a scout, you are responsible for evaluating two candidates for the position of first baseman. You are to create a scouting report on each candidate and make a recommendation for which of the two candidates the team owner should hire.
3. Explain that for each player, a scouting report must include the following:
 - A stemplot showing the player's batting average over the course of his career
 - The mean, median, and mode of his average
 - The standard deviation of his average
4. Player statistics can be easily found at <http://espn.go.com/mlb/> or at <http://www.baseball-reference.com/>. You may choose to assign students two current first basemen to evaluate or have students choose for themselves.
5. Finally, students must choose which of the two players they would like to hire for their teams and explain their reasoning using the player's statistics as evidence.
6. Provide students with the "MLB Scouting Reports" packet (included). Have students complete this project in class or as a homework assignment.

Conclusion:

To conclude this lesson and check for understanding, review the results of each student's (or pair of students') scouting reports. Students should share their statistical findings as well as their recommendation for which player to hire. After all students have had an opportunity to share, discuss the results with the class. Which players were more likely to be recommended? Players with consistently good averages over their careers, or players who show variance in their careers, but appear to be improving? Why?

For homework, have students write journal entries in which they address the importance of statistics. Do statistics tell a manager or an owner everything he or she needs to know about a player? What are some skills that *can't* be revealed through statistics?