Dare to dive into

Susan Troutman
Rice University School Mathematics Project
Director of Secondary Programs
troutman@rice.edu

Carolyn White
Rice University School Mathematics Project
Director of Elementary Programs
clwhite@rice.edu
NCTM Standards – Data Analysis

• Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them

  Grades 6–8 Expectations for students:
  - formulate questions, design studies, and collect data about a characteristic shared by two populations or different characteristics within one population;
  - select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.

• Select and use appropriate statistical methods to analyze data

  Grades 6–8 Expectations for students:
  - find, use, and interpret measures of center and spread, including mean and interquartile range;
  - discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.
6.12 Measurement and data. The student applies mathematical process standards to use numerical or graphical representations to analyze problems. The student is expected to:

(A) represent numeric data graphically, including dot plots, stem-and-leaf plots, histograms, and box plots;
(B) use the graphical representation of numeric data to describe the center, spread, and shape of the data distribution;
(C) summarize numeric data with numerical summaries, including the mean and median (measures of center) and the range and interquartile range (IQR) (measures of spread), and use these summaries to describe the center, spread, and shape of the data distribution; and
(D) summarize categorical data with numerical and graphical summaries, including the mode, the percent of values in each category (relative frequency table), and the percent bar graph, and use these summaries to describe the data distribution.

6.13 Measurement and data. The student applies mathematical process standards to use numerical or graphical representations to solve problems. The student is expected to:

(A) interpret numeric data summarized in dot plots, stem-and-leaf plots, histograms, and box plots; and
(B) distinguish between situations that yield data with and without variability.
7.12 Measurement and data. The student applies mathematical process standards to use statistical representations to analyze data. The student is expected to:

(A) compare two groups of numeric data using comparative dot plots or box plots by comparing their shapes, centers, and spreads;
(B) use data from a random sample to make inferences about a population; and
(C) compare two populations based on data in random samples from these populations, including informal comparative inferences about differences between the two populations.
8.11 Measurement and data. The student applies mathematical process standards to use statistical procedures to describe data. The student is expected to:

(A) construct a scatterplot and describe the observed data to address questions of association such as linear, non-linear, and no association between bivariate data;

(B) determine the mean absolute deviation and use this quantity as a measure of the average distance data are from the mean using a data set of no more than 10 data points; and

(C) simulate generating random samples of the same size from a population with known characteristics to develop the notion of a random sample being representative of the population from which it was selected.
How many letters are in your first name?

- Write your name at the top of a post-it note.
- Write the number of letters in your first name.
- Draw a box for each letter of your first name.

Example: Susan

<table>
<thead>
<tr>
<th>Susan</th>
<th>5</th>
</tr>
</thead>
</table>
Measures of Center and Measures of Spread

**Range (R-e)**

**Median**

**Interquartile Range**

**Mode**

http://www.ncesd.org
Measures of Central Tendency

https://www.youtube.com/watch?v=oNdVynH6hcY
### Measures of Spread

<table>
<thead>
<tr>
<th>Range</th>
<th>Interquartile Range (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

### Measures of Center

<table>
<thead>
<tr>
<th>Mode</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2.3</td>
</tr>
</tbody>
</table>

A measure of how much a collection of data is spread out. Commonly used types include range and quartiles. (Also known as spread or dispersion.)

An average; a single value that is used to represent a collection of data. Three commonly used types of averages are mode, median, and mean. (Also called measures of central tendency or measures of average.)
Think: _______________
To find the **mean**:
1. ____________ all values.
2. ____________ by the number of values.

Think: _______________
To find the **median**:
1. Put the data set in ____________, from least to greatest.
2. Mark off highest and lowest value, starting from the edges until you reach the ____________
3. If there are two middle values, add them together and ____________ by two.

Think: _______________
To find the **mode**:
1. Put data set in ____________, from least to greatest.
2. Find the number that appears the ____________.
3. There may be ________ mode, ________ mode, ________ one mode, or there may be ________ mode.

Think: _______________
To find the **range**:
1. Put the data set in ____________, from least to greatest.
2. ____________ the lowest value from the highest value.

Think: _______________
To find the **Interquartile Range (IQR)**:
1. Put the data set in ____________, from least to greatest.
2. Find the medians of the ____________ half (Q1) and the ____________ half (Q3) of the data.
3. ____________ quartile one (Q1) from quartile three (Q3).

- Cut these apart.
- Glue next to the appropriate term.

<table>
<thead>
<tr>
<th>6, 8, 11, 13, 20, 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>23 – 6 = 17</td>
</tr>
<tr>
<td>Describes the middle value of a data set</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8, 11, 12, 18, 18, 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
</tr>
<tr>
<td>Describes the spread of the middle half of the data set</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>17, 20, 14, 18, 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>17+20+14+18+16 = 85</td>
</tr>
<tr>
<td>85 ÷ 5 = 17</td>
</tr>
<tr>
<td>Describes the general spread of the data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2.4, 5, 7, 8, 9, 10, 12, 15, 16, 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Quartile (Q1) = 5</td>
</tr>
<tr>
<td>Median = 9</td>
</tr>
<tr>
<td>Upper Quartile (Q3) = 15</td>
</tr>
<tr>
<td>15 – 5 = 10</td>
</tr>
<tr>
<td>3, 4, 8, 9</td>
</tr>
<tr>
<td>(4+8) ÷ 2 = 6</td>
</tr>
<tr>
<td>Describes the average value of a data set</td>
</tr>
</tbody>
</table>
Measures of Center

Think: **Average**
To find the **mean**:
1. **Add** all values.
2. **Divide** by the number of values.

Think: **Middle**
To find the **median**:
1. Put the data set in **order** from least to greatest.
2. Mark off the greatest and least values, starting from the edges until you reach the **middle**.
3. If there are two middle values, add them together and **divide** by two.

Think: **Most frequent**
To find the **mode**:
Put data set in **order** from least to greatest.
Find the number that appears the **most**.
There may be **one** mode, **more** than one mode, or there may be **no** mode.

### Examples

**Describes the average value of a data set**

17, 20, 14, 18, 16

\[
\frac{17+20+14+18+16}{5} = 85
\]

\[
85 \div 5 = 17
\]

**Describes the middle value of a data set**

7, 11, 13, 14, 16, 17, 50

\[
\frac{7, 11, 13, 14, 16, 17, 50}{7, 11, 13, \underline{14}, 16, 17, 50}
\]

or

3, 4, 8, 9

\[
\frac{3+4+8+9}{4} = 6
\]

**Describes the most frequent value of a data set**

8, 11, 12, 18, 18, 20

\[
\underline{18}
\]
Measures of Spread

Think: __Spread of data________
To find the **range**:
1. Put the data set in __order___ from least to greatest.
2. **Subtract** the least value from the greatest value.

Think: __Spread of the middle half________
To find the **Interquartile Range (IQR)**:
1. Put data set in ____order____ from least to greatest.
2. Find the medians of the __lower___ half (Q1) and the __upper___ half (Q3) of the data.
3. **Subtract** quartile one (Q1) from quartile three (Q3).

Describes the general spread of the data

6, 8, 11, 13, 20, 23
23 – 6 = 17

Describes the spread of the middle half of the data set

2, 4, 5, 7, 8, 9, 10, 12, 15, 16, 18
Lower Quartile (Q1) = 5
Median = 9
Upper Quartile (Q3) = 15
15 – 5 = 10
Q3 – Q1
What a Deal!

Value of cards:
A = 1    2 = 2    3 = 3    4 = 4    5 = 5    6 = 6    7 = 7    8 = 8    9 = 9    10 = 10    J = 11    Q = 12    K = 13

Example:

3    5    6    6    Q

Range = Q - 3
       12 - 3 = 9

Median = 6
         3    5    6    6    12

Mode = 6

Mean =
       3 + 5 + 6 + 6 + 12 = 32/5 = 6.4
# What a Deal!

Value of cards:
A = 1   2 = 2   3 = 3   4 = 4   5 = 5   6 = 6   7 = 7   8 = 8   9 = 9   10 = 10   J = 11   Q = 12   K = 13

## Score Sheet for What a Deal!

<table>
<thead>
<tr>
<th>Names of players</th>
<th>Player 1</th>
<th>Player 2</th>
<th>Player 3</th>
<th>Player 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haley</td>
<td></td>
<td>Susan</td>
<td>Juan</td>
<td>Linda</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round 1</th>
<th>Range</th>
<th>Mode</th>
<th>Median</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Famous Celebrity</td>
<td>Estimated Height in inches</td>
<td>Actual Height in inches</td>
<td>Write these numbers as an ordered pair</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>----------------------------</td>
<td>-------------------------</td>
<td>----------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Beyoncé Knowles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jay Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennifer Lopez</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. J. Watt</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lady Gaga</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michael Jordan</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oprah Winfrey</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selena Gomez</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shaquille O’Neal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor Swift</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Beyoncé Knowles

5 ft. 7 in. or 67 inches
Jay Z

6 ft. 2 in. or 74 inches
Jennifer Lopez

5 ft. 5 in. or 65 inches
J. J. Watt

6 ft. 5 in. or 77 inches
Lady Gaga

5 ft. 1 in. or 61 inches
Michael Jordan

6 ft. 6 in. or 78 inches
Oprah Winfrey

5 ft. 7 in. or 67 inches
Selena Gomez

5 ft. 5 in. or 65 inches
Shaquille O’Neal

7 ft. 1 in. or 85 inches
Taylor Swift

5 ft. 10 in. or 70 inches
Scatterplots

• Use the data from the completed table to create a scatter plot. Use the estimated height as the label for the horizontal axis and the actual height as the label for the vertical axis.

• For which celebrity was your estimate closest to the actual height?

• For which celebrity did you have the greatest difference between your estimate and the actual height?

• How well were you able to estimate the heights of the celebrities listed?

• Does there seem to be a relationship between your estimates of the celebrities’ heights and their actual heights?
Investigating Data of MLB Pitchers

Each table will be given a roster for the pitchers of a Major League Baseball team.

- What information is listed on the roster?
- What types of graphical representations can be created with the data?
Attributes of MLB Pitchers

At each table, have different groups select a different attribute to investigate.

• Age
• Weight
• Height
Graphical Representations

Each group will create a **dot plot** and a **box plot**, for the numeric data of their selected attribute:

- Age
- Weight
- Height (as fractions – 6’6” would be recorded as 6 ½ feet)
Word Problems

Create one- and two-step problems using data from the graphical representations.
Gallery Walk

• Each group will post their graphical representations and word problems.
• Groups will compare and contrast the different types of graphical representations for each of the different teams.
Handouts available electronically at https://rusmp.rice.edu/
\[(\text{THANK YOU})^n\]

\[n \in \mathbb{N}, n > 1\]

\[\text{YOU} \in \{\text{Awesome People Set}\}\]

Handouts available electronically at
https://rusmp.rice.edu/