DATA SCIENCE
TRUMAN STATE UNIVERSITY

6TH GRADE MATH MODULE: OUTLINE

Rationale

The purpose of this module is to introduce students to the basic concepts within data science while also providing an introductory activity for the instruction of related topics contained in the Missouri Learning Standards. This module will allow students to investigate the distribution of data, including measures of center and variation. Students will display and interpret data using dot plots, histograms, boxplots and circle graphs.

This module serves as a starting point for instruction related to the following Missouri Learning Standards:

Math:

- 6.DSP.A.1 - Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.
- 6.DSP.A.2 - Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread and overall shape.
- 6.DSP.A.3 - Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary from a single number.
- 6.DSP.B.4 - Display and interpret data.
  a. Use dot plots, histograms and box plots to display and interpret numerical data.
  b. Create and interpret circle graphs.
- 6.DSP.B.5 - Summarize numerical data sets in relation to the context.
  a. Report the number of observations.
  b. Describe the nature of the attribute under investigation, including how it was measured and its units of measurement.
  c. Give quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context of the data.
  d. Analyze the choice of measures of center and variability based on the shape of the data distribution and/or the context of the data.

Science

- 6-8.ESS3.B.1 - Analyze and interpret data on natural hazards to forecast future catastrophic events and inform the development of technologies to mitigate their effects.

MoExcel Data Science Standards

- MoExc1: **Identify** issues, problems, questions, or claims that can be addressed using large datasets.
  *The expectation is that students be able to **identify** statements, claims, or questions that can be refined into testable hypotheses.*
- MoExc2: **State** data-driven investigative questions.
  *The expectation is that students be able to **state** investigative questions based on quantitative data.*
- MoExc3: **Construct** visual representations of real-life data from publicly available datasets and **describe** patterns observed.
  *The expectation is that students are familiar with large datasets of publicly available data that allow users simple but rich manipulation of bivariate data and **describe** patterns that result from purposeful manipulation of the information.*
- MoExc4: **Suggest** and **discuss** the possible interactions among data.
  The expectation is that students can provide and consider alternative explanations to the relationships (or lack thereof) among data.

TRUMAN STATE UNIVERSITY

- MoExc5: **Identify** and **discuss** potential factors that can influence relationships between the independent and dependent variables.

  *The expectation is that students reflect on the complexity of real-life problems and consider it when attempting analyses or problem-solving. This includes identifying and accounting for different forms of control variables (intervening, confounding, or antecedent). Discussion of the differences among control variables is **not** expected.*

- MoExc6: **Interpret** real-life data by using patterns and relationships among data.

  The expectation is that students are able to construct stories that provide plausible explanations for relationships that have been identified among data.

Standards for Mathematical Practice

Standard#:	Standard:
MP1	Making sense of problems and persevere in solving them.
MP2	Reason abstractly and quantitatively.
MP3	Construct viable arguments and critique the reasoning of others.
MP4	Model with mathematics.
MP5	Use appropriate tools strategically.
MP6	Attend to precision.
MP7	Look for and make use of structure.
MP8	Look for and express regularity in repeated reasoning.

Prior Knowledge & Possible Misconceptions:

Prior Knowledge:

This module assumes that previous instruction has covered the 4th and 5th grade data standards, including:

4.DS.A.3 Analyze the data in a frequency table, line plot, bar graph or picture graph.

5.DS.A.1 Create a line graph to represent a data set, and analyze the data to answer and solve problems.

5.DS.A.2 Create a line plot to represent a given or generated data set, and answer questions and solve problems, recognizing the outliers and generating the median.

Possible Misconceptions:

1. Students may not recognize the difference between a bar graph and a histogram. Be sure to emphasize that histograms display quantitative or numerical data while bar graphs are used for categorical data.

6th Grade Data Science Math Module

Example: Tornadoes in Missouri

Question: What do we know about tornadoes in Missouri?

Data: Tornado Tracks Tool - MRCC using data from NOAA

Goal: The goal of this lesson is to get students thinking about data and statistical questions. It will help them see the need for ways to summarize information from a dataset.

Materials & Tech Requirements: Students will need a device (computer, iPad, etc.) to complete the tornado map exploration. One device per student is preferable. If necessary, students may share a device in groups no larger than three.

Discussion:

Tell students that they are going to start their statistics and data unit by looking at some information about tornadoes in Missouri. Split students into groups of 2-3 and have them work together to complete the first two columns of the tornadoes KWL chart.

Ask students to share some of the information they listed in the first two columns of their chart, and make a summary of some of their information on the board.

Now, show the National Geographic Tornadoes 101 Video so that students can become familiar with facts about tornadoes. Ask them to look out for information that may answer their "Want to Know" questions.

Give students some time in their small groups to complete the "Learned" column of their KWL chart, and then discuss it together as a class. Be sure to discuss the Fujita scale, as they will need to be familiar with the fact that it measures the strength of a tornado.

Next, ask students: "What questions could you ask about tornadoes in Missouri?" Give them a few examples ("Does Missouri have more or less tornadoes than Kansas?" "Has Missouri ever had a tornado in December?") to get started. Give groups a stack of index cards, and have them take turns writing questions. One group member should write a question on an index card, and put it in the middle, then the next person writes a new question on a new index card and adds it to the middle. Their goal should be to brainstorm at least 10 questions. Have groups share a couple of questions with the class and see if there were any common questions. Jot them on the board, then collect and keep the question cards, as you will return to them later in the unit.

Possible questions students may have:

- Should we be worried about tornadoes in Missouri?
- How many tornadoes have happened this year?
- How many tornadoes happen each year in Missouri?
- Is the amount of tornadoes increasing?
- How much damage do tornadoes cause in Missouri?
- How many people does a tornado usually injure in Missouri?
- How many people does a tornado usually kill in Missouri?
- What month usually has the most tornadoes in Missouri?
- What time of day do tornadoes happen?
- Has the yearly amount of tornadoes been changing over time?
- How strong are tornadoes in Missouri?
- What was the strongest tornado ever to happen in Missouri?
- How long do tornadoes last?
- Has a tornado ever come close to my house?
- Where do the most tornadoes happen?

Tell students that they are going to work to answer some of these questions during the unit. Right now, they are going to explore a map with tornado data and see if they can start answering some of these questions. Give students time to complete the **exploration handout**. It would be best if students complete the exploration individually, but if necessary, they could complete it in their small group that they have been working with. If students are in groups make sure to ask them to switch roles frequently so that each student gets to engage with and explore the map data.

After students have completed the handout, discuss responses together. Students should find that it was difficult to answer the last three questions using the map, and their estimates may be very different. Tell them that they are going to learn about some measures that data scientists and mathematicians use to answer those questions (mean, median, mode) and some different ways data can be summarized besides on a map.

Suggestions for Unit Integration

Throughout the unit, refer back to the Missouri tornadoes dataset to reinforce and relate to the concepts that are being taught. Resources and ideas are given below.

Statistical Questions

Statistical Questions

Look through the questions that students brainstormed during the introductory activity. Discard any that are unclear or don't make sense. After you have discussed the difference between statistical and non-statistical questions with students, select a few question cards and discuss together whether or not each question is a statistical question. For example, "How many tornadoes happened in Missouri last year?" is not statistical because there is only one answer, but "How many tornadoes happen each year in Missouri?" would be statistical because the response will vary by year. Then take the remaining cards and pass out two question cards to each student. Split the class into random groups of four and have students sort their question cards into statistical questions and non-statistical questions.

Note: Depending on the questions your students wrote, you may want to add a few of your own questions to ensure that students have a good mix of statistical and non-statistical questions to sort.

A few suggestions for statistical questions:

- What time of day do tornadoes happen in Missouri?
- How strong are tornadoes in Missouri?
- How much property damage do tornadoes cause in Missouri?

Suggestions for non-statistical questions:

- What scale do we use to measure tornadoes?
- When was the strongest tornado to ever happen in Missouri?
- Has there ever been a tornado in _____ County? (give a specific county name)

Displays of Data

Creating Displays

Ask students to use the **tornado data** to create various displays, including dot plots, histograms, box plots, and circle graphs. Directions for creating each type of display are **here**. *Note*: The data file contains many variables. Looking at multiple variables and relationships between them is a key part of data science, so it is recommended that you discuss the full dataset with students initially. However, if you then want students to focus on a specific variable or to decrease the amount of information that is visible, you can hide some of the columns in the data set.

Ideas:

- **Histograms** showing total number of tornadoes broken down by year, start time, total property damage over time, width of tornadoes, or length of tornadoes
- For **dot plots**, consider plotting some variable for a smaller time frame, like one year or a single county so that the number of data values to be plotted is reasonable. This would be a good opportunity to discuss sample size and how many observations of tornadoes are given in the dataset.
- **Box plots** showing number of tornadoes per year, number of injuries per tornado, length, width or magnitude
- **Circle graphs** showing breakdown of tornadoes by month or magnitude

Analyzing Distribution

Once students have created dot plots and histograms for various quantities, discuss how the shape of the distribution varies for different variables. For example, there are many more low magnitude tornadoes than there are high magnitude tornadoes. If plotted by "begin time" it becomes clear that tornadoes are increasingly likely to start in the afternoon, peak around 6 pm and then decrease in likelihood.

Statistical Measures

Measures of Center

Use Google sheets and the **tornado data file** to have students make and compare values for measures of center.

To calculate the mean, type =AVERAGE(J:J) where J is the column of data you wish to use. You can click this column instead of typing it.

To calculate the median, type =MEDIAN(J:J) where J is the column of data you wish to use. You can click this column instead of typing it.

To calculate the mode, type =MODE(J:J) where J is the column of data you wish to use. You can click this column instead of typing it.

There are many different quantities that could be used, including:

- Tornado Width or Length (shown below)
- Number of Injuries
- Number of Deaths
- Cost of Property Damage

Example: Tornado Length (mi)

Mean Length	3.875047136
Median Length	1.5
Mode of Length	0.2

Discuss what would cause these measures of center to vary, and which might be most representative for the data set. Example: In the tornado length measures given above, the mode shows us that the most frequently recorded tornado length is very short – only 0.2 miles. The only thing the mode tells us is which length has been recorded the most often, without taking into account any of the other data values. The median, or middle length of the tornado data is 1.5 miles. This means half of tornado lengths were less than 1.5 miles long and half were greater than 1.5 miles. Looking at the original data values, there was one very long tornado (160 miles) which had a much greater length than the rest. Since every value is used in calculating the mean, this large value causes the mean to be greater than the median, despite the fact that shorter tornadoes occur more frequently. Due to this, the median is probably the most representative measure of tornado length. It is less skewed by this one large data value.

Measures of Variation

The students can use the same quantities listed above and calculate the interquartile range and mean absolute deviation.

To calculate the interquartile range, use the formulas below (where B is replaced with your data column) to first find Q1 and Q3. Then subtract them.

	J	K	L
2	=QUARTILE(B:B,1)	=QUARTILE(B:B,3)	=K2-J2

To calculate the mean absolute deviation, type =AVEDEV(J:J) where J is the column of data you wish to use. You can click this column instead of typing it.

Extension/Summary Activity

Have students select one statistical question about tornadoes to answer. For example: "What time of day do tornadoes start in Missouri?" You could also assign them a question instead of having them choose it. Ask them to create a box plot, as well as one other display of data (circle graph, dot plot, histogram) for the data associated with their question. In addition, have them calculate the measures of central tendency and variation for the data. They should summarize their results and answer their question in a Google doc or Google slides presentation. Presenting data, especially in a compelling way, is a key part of data science. Being able to summarize and tell the story that the data shows is an important skill.

Links and Sources

National Geographic. (2015, December 23). *Tornadoes 101*. YouTube. Retrieved August 16, 2022, from https://www.youtube.com/watch?v=iMWoYPaQx0Q

Purdue University. (n.d.). *Tornado Tracks Tool*. Midwestern Regional Climate Center. Retrieved August 16, 2022, from https://mrcc.purdue.edu/gismaps/cntytorn.htm

Storm events database. National Centers for Environmental Information. (n.d.). Retrieved August 16, 2022, from https://www.ncdc.noaa.gov/stormevents/choosedates.jsp?statefips=29%2CMISSOURI