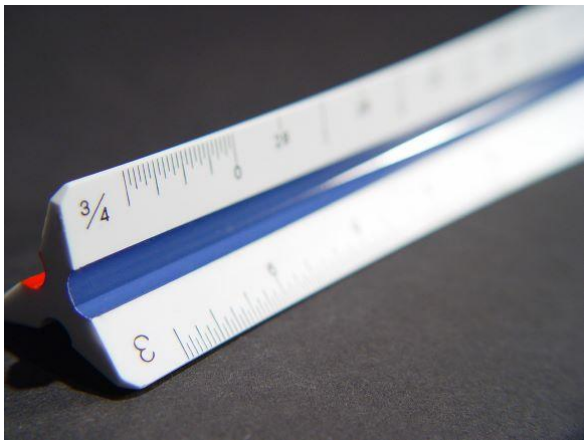
Data Analysis: Introduction to Measurement Error and Outliers

Setting the Stage

Many times students are tasked with collecting data for an investigation, and use their data to make sense of some kind of phenomena. In some instances, data is compared across the class; however, the challenges of analyzing class data is not always addressed. For example, students may not be provided the opportunity to discuss experimental error. This lesson may be used to support learners who are new to data collection and analysis. They measure and compare the rate of a falling object (a penny) at different heights and while doing so, learn about statistical error.



Scale Ruler Photo credit:filmingilman

Activity Overview

- *Part 1 – Engagement & Exploration (15 minutes) Errors in Data Collection*
  Students collect and share data to explore the quality of their data and learn about error in data collection.
- *Part 2 – Explanation (15 minutes) Variation in Data Collection*
  Students continue the exploration of the concept of variation in data.
- *Part 3 – Elaboration & Evaluation (15 minutes) Penny Drop Data Analysis*
  Students decide on a method to analyze class data and explain why. They discuss the different types of error in data collection.

Instructional Overview	
Grade Level	Middle School
Instructional Time	45 minutes
NGSS Standards	Science and Engineering Practice: MS--Analyzing and Interpreting Data
Driving Question	How do we account for error in data collection? Can data with statistical errors be used to support science?
Learning Goals	Students will be able to define and apply the statistical concepts of measurement, error, and outliers.
Materials	For each team of students: ❏ Penny ❏ Stopwatch ❏ Meter stick ❏ Notebook, graph paper, or spreadsheet program
Vocabulary	Measurement error is the difference between a measured value and the true, theoretical value. Sampling error is error that arises from only measuring a portion of a population, rather than the whole population. Outlier is a measurement that's far away from other measurements, an extreme measurement.
Instructional Strategies	The *think-pair-share* strategy provides time for students to think about a question by themselves for a few minutes (they may write their thoughts in a notebook), and then they discuss their ideas with another student for a few minutes. In the final step, the pair of students share their ideas with the class.

	Web Links for Lesson Resources Note: all resources are downloaded as pdfs in the *Activity Resources Folder*
Part 2	Video: *From Dog walking to Weather and Climate* http://spark.ucar.edu/dog-walking-weather-and-climate

Part 1 (Engage & Explore)
Errors in Data Collection (15 minutes)

Ask students if there is a relationship between the height at which a penny is dropped and the amount of time it takes for the penny to hit the ground. Listen for their ideas and challenge them to collect the data to demonstrate their thinking.

Pair students and provide them with a penny, a stopwatch, and a meter stick. Ask students to use a stopwatch to measure how long it takes for a penny to fall from a set distance. While working in pairs, one student drops the penny and holds the meterstick, while the other uses the stopwatch to record how long the penny takes to hit the ground when dropped from 25 cm, 50 cm, 75 cm, and 100 cm. Their data should be organized in their notebooks or a spreadsheet so that it can be easily analyzed later in this activity.

Create a class data table that looks something like this, but includes columns for data from each team

	Time (seconds)				
Drop Height (cm)	**Team 1**	**Team 2**	**Team 3**	**Team 4**	**Team 5**
25					
50					
75					
100					

Ask students to analyze the data they see in the table. They will likely see that there is a lot of variation in the measured times.

Think-pair-share: *Why didn't they all get the same number?*

Hold a class discussion that explores the possible reasons for the variations in their data. There are many sources of error in this experiment, such as:

- Pennies may not all have been dropped from exactly the same height or may not have been dropped in the same way (e.g., some horizontally, some vertically).
- The reaction times for starting and stopping the stopwatch for the various teams may be quite different.
- The pennies themselves might be slightly different shapes or weights.
- The stopwatches are not perfect! They are only so accurate and have limitations.

Compile their ideas, and ask students if they see categories for their ideas. Try to guide the discussion so three important categories of error are highlighted:

- Human error (e.g., method of dropping, drop height, reaction times)
- Instrument error (e.g., the stopwatches are only so accurate)
- Sampling error (e.g., the pennies are not identical)

Part 2 (Explain)
Variation in Data Collection (15 minutes)

Lead the discussion on sources of variation/error by highlighting these points.

- Science is a search for the answers to questions or reality of a situation (e.g., the actual time it takes for the penny to hit the ground).

- Because of measurement error, the theoretical answer to questions can never be exactly known.
    - Measurement error is the difference between a measurement and the actual value.
    - From the earlier discussion, both human error and instrument error are part of measurement error.
    - Even with the best instruments in the world, there will still be some measurement error, although it may be very small. All instruments have limitations.

- Sampling error is another reason the answers can't be exactly known with 100% certainty.
    - All pennies are very similar, but they're not exactly the same. We can't measure every single penny!

It's important to discuss that although all data have error, that doesn't mean that scientists are unable to draw strong conclusions from data. Stress that the more data collected, the more we understand and the closer we come to an accepted theoretical understanding.

> Consider showing UCAR's short video (1:04) *From Dog Walking to Weather and Climate*
>
> From the website: "Do you know the difference between weather and climate? And do you know how weather and climate are like walking a dog? In this animated short, the relationship between trend and variation are explained with an excellent analogy to a man walking his dog. There is much more variation in the path that the dog takes as compared with the man, but they are both headed the same way. Similarly, weather can be highly variable and climate means long term trends."

Debrief the video by asking students to think about how the video modeled helps them to understand the variation in their penny drop data. Menton that they will explore the data again in the next part of this activity.

Part 3 (Elaborate & Evaluate)
Penny Drop Data Analysis (15 minutes)

Come back to the measurements of how long it takes for a penny to fall. Ask the students what they would do with those measurements, if someone asked them for just one number representing the fall-time of a penny. Most likely, someone will suggest taking the average.

> Think-pair-share: *Why is the average a better representation of the measurements than simply choosing one of them?*

Assist students as they determine the average for each row of data. Ask them if there were any data points in each row that may have had a greater influence on the average for the row. Mention that those data points may be considered outliers and that an outlier is a measurement that is far away from other measurements or an extreme measurement. Ask students:

> *What may have caused outliers in their data collection and therefore data points?*
>
> *What effect do outliers have on conclusions?*

Ask the students to imagine that one of the penny measurements was an outlier. Go ahead and replace one of the measurements with something way off (if the average was about 0.5, replace one of the measurements with something like 3).

Recalculate the average. It will be very different!

> Outliers have an extra-large effect on our conclusions. This is why we should be suspicious of outliers – IF we have reason to think the outlier data point is wrong (say, we remember being distracted when taking the measurement). In this case, we should exclude that data point from our analysis.

Wrap this activity up by asking students to analyze their penny drop data AND the class penny drop data. In their conclusion, they are to address error in their data collection and error in the class data using some of the terminology they learned in this activity.