(http://csmatters.org)   3 - 8

0b11 - 0b1000

*Optional*
# Basic Statistics with Excel

**Unit 3. Information and the Internet**

**Revision Date:** Jul 22, 2019
**Duration:** 1 50-minute session

## Lesson Summary

**Summary:** This lesson is designed for students to review basic statistics, including calculations of the mean, median, mode, and standard deviation. It will also give the students some experience using spreadsheet software to calculate the statistics and to create histograms.  *Note: This lesson is intended primarily as a review and a reminder of material that should already be familiar to the students. If your students have little familiarity or experience with using Excel to compute statistics or generate plots, you may wish to extend this lesson to two sessions, and provide more scaffolding and instruction on the basic mechanisms.*

**Outcomes:**

- Students will review the basic statistical concepts of mean, median, mode, and standard deviation.
- Students will use spreadsheet software to calculate the statistics and to create histograms.

**Overview:**

1. Getting Started (5 min)
2. Introduction of Content (10 min) - Statistics Introduction and Review
3. Guided Activity (30 min) - Students Create Plots and Calculate Candy Statistics
4. Wrap Up (5 min) - Journal

**Source:** This lesson was adapted from Unit 2: The Engineering Design Process, Lesson 2: Collecting and Processing Information ©2013 International Technology and Engineering Educators Association Foundations of Technology, Third Edition/ Technology, Engineering, and Design

## Learning Objectives

# CSP Objectives

- *EU DAT-2 - Programs can be used to process data, which allows users to discover information and create new knowledge.*
  - LO DAT-2.D - Extract information from data using a program.
  - LO DAT-2.E - Explain how programs can be used to gain insight and knowledge from data.

# Math Common Core Practice:

- MP5: Use appropriate tools strategically.

# Common Core Math:

- S-ID.1-4: Summarize, represent, and interpret data on a single count or measurement variable

# Common Core ELA:

- RST 12.10 - Read and comprehend science/technical texts

# NGSS Practices:

- 5. Using mathematics and computational thinking

# Key Concepts

The students must understand the basic statistical concepts of mean, median, mode, and standard deviation. They must also be able to use spreadsheet software to calculate the statistics and to create histograms.

Students often have some initial difficulty learning how to use formulas in the spreadsheet software to do the calculations.

---

# Essential Questions

- How can computation be employed to help people process data and information to gain insight and knowledge?
- How can computation be employed to facilitate exploration and discovery when working with data?

# Teacher Resources

Student computer usage for this lesson is: **required**

**For Each Student:**

- Package of colored candy. Alternatively, you can ask students at the end of the previous lesson to collect some distributional data to use for the exercise (ideas: colors of cars in the school parking lot, colors of shirts worn by students in the room, favorite sports teams or

bands of the students), or you can simply provide some data for the students to use. This could be either representing colors of candy (using the test data in the lesson if you would like), or similar distributional statistics.

- Excel software or other spreadsheet software
- Web resource with information about measures of central tendency: https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php (https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php)

## Lesson Plan

## Getting Started (5 min)

- Students should describe what they know about statistics in their journals.
- Have students share what they know about statistics and introduce the lesson.

## Introduction of Content (10 min)

**Review of Statistics:**

Present a review of basic statistics (min, max, mean, median, mode, and range), and use the following board exercise to have the class review their understanding of these basic concepts:

- Ask eight or so randomly selected students for their birth date (day of the month).
- Write these numbers on the board.
- On the side of the board, list the key terms "min," "max," "mean," "median," "mode," and "range."
- Ask the class as a group to compute each of these values:
    - **Min:** The smallest number (what if there is more than one? - no problem!)
    - **Max:** The largest number (what if there is more than one? - no problem!)
    - **Mean:** The average (sum of the numbers, divided by how many numbers there are)
    - **Median**: The center value in a sorted list of numbers.
        1. Have the students help you to rewrite the values from smallest to largest.
        2. Which is the middle number?
        3. Since there are 8 numbers, there is no middle number!
        4. In this case, the median is the mean (average) of the two center numbers = 4th number + 5th number / 2.
    - **Range:** The difference between the largest and smallest value (max - min).
    - **Mode:** The most frequently appearing value.  In such a small set, there is likely to not be a mode, unless two students happen to share the same birth date.  You might wish to poll the students for another number (e.g., the students' grade) that's likely to have more repeated values, and then compute the mode (and, optionally, the other statistics).

**Discussion:**

Ask the class to come up with situations where it might be most useful to compute the mean, median, or mode of a set of values. Encourage them to understand that each of these statistics can be useful in different situations, but may be misleading. Have them generate sets of data that would give "misleading values" for mean (if there is an "outlier value"), median (if the values have a longer "tail" on one side than the other), or mode (if there is a frequent value that happens to occur at one end or the other of a wider range).

- This web resource has a good discussion of different measures of central tendency and when they are appropriate: https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php (https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php)

## Guided Activity (30 min) - Candy Statistics

**Note:** The teacher may want to do this activity along with the students, displaying the spreadsheet on a screen so that the students may ask questions and see how to do the statistical calculations using the spreadsheet software. Students who do not have much experience with spreadsheets may need more scaffolding and instruction. (If you have many such students, you may wish to spread this lesson out over two class sessions.)

Students will use spreadsheet software, such as Excel, to calculate the average number and standard deviation of candy color in an individual-sized bag of M&Ms, Skittles, or other colored candy. Optionally, students may compare their results to other online published statistics for each candy.

1. Have the students predict how many individual candy pieces are in their bag of candy and write their predictions in their journals.
2. Have the students open their bag of candy and sort the candy into categories based on color.
3. Have the students note the difference in the total number of candies predicted versus the actual number that was in the packet. They should note the difference in their journals.
4. Open an Excel program and create a spreadsheet like the following. Each Trial Number in the example below corresponds to a student or group in the class. *(**Note:** If you do not have candy to do the counting exercise, you may simply give the sample spreadsheet below to the students.)*

Candy
Statistics

| Trial Number | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Yellow | 17 | 20 | 24 | 19 | 19 | 17 |
| Red | 21 | 13 | 19 | 21 | 15 | 18 |
| Blue | 10 | 18 | 16 | 18 | 21 | 20 |
| Brown | 7 | 12 | 5 | 12 | 12 | 14 |
| Green | 26 | 26 | 16 | 17 | 22 | 18 |
| Orange | 24 | 16 | 20 | 15 | 15 | 16 |
| Package Total | 105 | 105 | 100 | 102 | 104 | 103 |

The students will also need to create columns further to the right labeled Mean, Median, Mode, and Standard Deviation.

| Mean | Median | Mode | Standard Deviation |
|------|--------|------|--------------------|
| 19.375 | 19 | 19 | 2.199837656 |
| 18.125 | 19 | 19 | 2.799872446 |
| 17 | 17.5 | 18 | 3.338091842 |
| 9.5 | 10 | 12 | 3.380617019 |
| 21.125 | 22 | 22 | 3.833592124 |
| 18.25 | 17.5 | 16 | 3.284161124 |

5. Each student will enter their own data for each color and the data from another student or group into the table.
6. Using their data, students will:
   1. Calculate the mean value for each color category within the experiment. They should use the Average function to do the calculation.
   2. Calculate the median, mode, and standard deviation for all color categories. They should use the appropriate functions to do the calculations.
   3. Calculate the package total for each trial by using the `SUM` function.
   4. Create a ± 3ϭ histogram for each candy color.
   5. Create a frequency distribution table for each candy color, as illustrated below.
   6. Create a histogram for each candy color, using your bin and frequency data.

| Yellow Candy σ = | 2.199837656Get on |
|------------------|--------------------|
| 3Cs | 25.97451297 |
| 2S | 23.77467531 |
| 1s | 21.57483766 |
| Mean | 19.375 |
| -1s | 17.17516234 |
| -2s | 14.97532469 |
| -3s | 12.77548703 |

# Wrap Up (5 min)

Students will answer the following question in their journals:

- Why is it important to use statistics to understand large data sets? When are different measures of central tendency appropriate or inappropriate?

## Options for Differentiated Instruction

Learners may be paired to assist each other in the use of the spreadsheet software.

## Evidence of Learning

## Formative Assessment

The teacher should frequently check the students' work for accuracy as the lesson progresses so that misunderstandings may be quickly resolved.

## Summative Assessment

- Have the students calculate the mean, median, mode and standard deviation of a set of data.
- Have the students use a spreadsheet to do statistical calculations and create a histogram.

(http://www.umbc.edu/) (http://www.umd.edu/)

(http://www.nsf.gov/)

*Authored by:* CS Matters in Maryland
*Website:* csmatters.org (http://csmatters.org)
*Email:* csmattersinmaryland@gmail.com (mailto:csmattersinmaryland@gmail.com)