

A1.1 One Variable Statistics

- Create dot plots, histograms, and box plots (S-ID.A.1).
- Use available classroom technology to create histograms and box plots and calculate measures of center and spread (S-ID.A.1).
- Use terms such as “flat,” “skewed,” “bell-shaped,” and “symmetric” to describe data distributions (S-ID.A.2).
- Analyze and compare data sets (S-ID.A.3).
- Understand relationships between mean and median for symmetrical and skewed data distributions (S-ID.A.2).
- Recognize outliers when they exist, and know to investigate their source (S-ID.A.3).
- Know that outliers affect the mean, but not the median of a data set (S-ID.A.3).
- Describe variability by calculating deviations from the mean (S-ID.A.2).
- Compare two data sets with the same means but different variabilities, and contrast them by calculating the deviation of each data point from the mean (S-ID.A.2).
- Understand that IQR is a description of variability better suited to a skewed distribution (S-ID.A.3).
- Work with two-way tables (S-ID.B.5).

The story before this unit:

In grades 6 to 8, students were introduced to data sets and different ways to represent data (histograms, dot plots, box plots). Statistics is introduced as a tool to answer questions about a population that have variability in the answer. Students learn about measures of center (median, mean) and measures of variability (interquartile range, mean absolute deviation), using them to draw informal comparative inferences about two populations.

The part of the story happening in this unit:

Students build on and expand their understandings of statistics in this unit. The key characteristics (measures of shape, center, and spread) are again seen and in addition, students may further describe the shape of a data distribution (symmetric, skewed, flat

or bell shaped) and summarize by a statistic measuring center and a statistic measuring spread. Instead of creating representations of data, the emphasis in high school is on interpreting representations and judiciously interpreting measures of center and spread.

Students develop more precise understanding of measures of center. They learn that mean and median are equal for symmetrical distributions, explain why mean and median are not equal in examples of skewed distributions, select median as the better measure of center for skewed distributions, and make generalizations about what kinds of distributions have means larger than medians and which have medians larger than means.

Students learn that standard deviation is a measure of spread, that a larger standard deviation means the data are more variable or spread out, and the meaning of standard deviation as “typical distance from the mean” for a symmetrical distribution. To aid in developing their understanding, students will calculate a standard deviation by hand for a small data set at least once. Given different visual representations of data (box plots, histograms, dot plots) students draw and justify significant and meaningful conclusions about the given situation. (All of these representations are frequency graphs, however, the Standards do not require students to know or use the term “frequency graph,” although they use the term “frequency tables.”) Students are introduced to two-way frequency tables and understand how to interpret relative frequencies in the context of the data represented in the tables. In unit S2 (which could take place either before or after this unit), students also build their statistics foundation by learning ways to determine whether two sets of data are correlated, and how strongly. Students identify linear association and interpret slope and intercept in the context of the data.

Given different visual representations of data (box plots, histograms, dot plots) students draw and justify significant and meaningful conclusions about the given situation. (All of these representations are frequency graphs, however, the Standards do not require students to know or use the term “frequency graph,” although they use the term “frequency tables.”)

Students are introduced to two-way frequency tables and understand how to interpret relative frequencies in the context of the data represented in the tables.

The story after this unit:

In unit S2 (which could take place either before or after this unit), students also build their statistics foundation by learning ways to determine whether two sets of data are

correlated, and how strongly. Students identify linear association and interpret slope and intercept in the context of the data. Given different visual representations of data (linear models) students draw and justify significant and meaningful conclusions about the given situation. Students begin to use technology as a means to plot data and generate correlation coefficients.

In S3, students revisit two-way frequency tables from a probability standpoint, and use them as a tool for conceptualizing and finding conditional probabilities.

In S4, students combine the ideas of distributions and probability. They learn about normal distributions and use them to solve problems, and use the distributions of probability models to find the likelihood of a particular outcome. In doing so, students build on their experience with standard deviations from S1, calculating standard deviations using technology, and interpreting the results.

Every high school statistics and probability standard is a modeling standard, hence modeling pervades the four units.

A1.1.0 Pre-unit diagnostic assessment

Diagnose students' recall of middle grades statistics, specifically their ability to:

- recognize a statistical question (6.SP.A.1);
- describe the distribution of data collected to answer a statistical question by its center, spread, and overall shape (6.SP.A.2);
- interpret statistical plots (6.SP.B.4);
- summarize numerical data sets in relation to their context (6.SP.B.5);
- informally assess the degree of visual overlap of two numerical data distributions with similar variabilities (7.SP.B.3).

A1.1.1 How can data be represented and summarized meaningfully?

- Revisit various ways to plot data: dot plots, histograms, and box plots (S-ID.A.1).
- Interpret plots of data within the context of the data (S-ID.A.3).
- Use the terms “symmetric” and “skewed” as descriptors of distributions (S-ID.A.2).

Students revisit the methods for representing and summarizing data that they learned in grades 6 to 8. They interpret dot plots, histograms, and box plots, attending to the referents of symbols used. Not just what are the quartile values, but what do they indicate about the data? Not just whether the graph is symmetric or skewed, but what does that tell us about the data?

A1.1.2 Analyze data distributions

- **Create dot plots, histograms and box plots (S-ID.A.1).**
- **Use available classroom technology to create histograms and box plots and calculate measures of center and spread (S-ID.A.1).**
- **Use terms such as “flat,” “skewed,” “bell-shaped,” and “symmetric” to describe data distributions (S-ID.A.2).**
- **Analyze and compare data sets (S-ID.A.3).**

Much of this section reviews ideas developed in grades 6 and 7 (but not 8), in order to allow the teacher to address any gaps in understanding revealed in the pre-unit assessment. If students are not already familiar with terms used to describe distributions (e.g., flat, skewed, bell-shaped, symmetric), then these terms should be introduced. With any task used, students should always be asked to interpret the statistical terms and measures in the context of the data set being described.

The teacher may also opt to use the tasks in this section to demonstrate how to use available technology to create histograms and box plots, and to calculate measures of center and spread, helping students to choose appropriate tools strategically when they analyze data in the future.

Tasks

[S-ID.1,2,3 Speed Trap](#)

A1.1.3 Measures of center

- **Recall how to calculate mean and median.**
- **Understand mean and median as a “typical value” that can answer a statistical question.**
- **Know that mean and median are equal for a symmetrical data distribution (S-ID.A.2).**
- **Explain why mean and median are unequal for a skewed data distribution**

(S-ID.A.2).

- **Select mean as the better measure for symmetrical distributions, and median as the better measure for skewed distributions (S-ID.A.2).**
- **Make generalization what kinds of distributions have means larger than medians, and what kinds have medians larger than means (S-ID.A.2).**
- **Recognize outliers when they exist, and know to investigate their source—that data point is way out there, why is that? Is there something weird about it that means we should disregard it (S-ID.A.3)?**
- **Know that outliers affect the mean, but not the median of a data set (S-ID.A.3).**

Students deepen their understanding of mean and median as measures of center, gaining a better understanding which to use in summarizing a given data distribution. They work with data sets where mean and median are equal and where they are different, and explain, using the context of the data, why this occurs. Teachers may continue to use the class’s available technology to calculate summary statistics and make plots.

Tasks

[S-ID Haircut Costs](#)

[Identifying Outliers](#)

[S-ID.3 Describing Data Sets with Outliers](#)

A1.1.4 Mid-unit assessment

Assess students’ ability to

- **describe a set of data given a graph or table (S-ID.A.2);**
- **identify and calculate spread, center, shape, outliers, quartiles, mean, median, mode (S-ID.A.2);**
- **construct and interpret a box plot (S-ID.A.1);**
- **compare, contrast, and draw conclusions when given two data sets (S-ID.A.3).**

A1.1.5 Standard deviation

- **Describe variability by calculating deviations from the mean (S-ID.A.2).**
- **Compare two data sets with the same means but different variabilities, and contrast them by calculating the deviation of each data point from the mean (S-ID.A.2).**

- **Interpret sets with greater deviations as having greater variability (S-ID.A.2).**
- **Calculate a standard deviation by hand for a small data set, and understand standard deviation as an indicator of a typical deviation from the mean of an element of the data set (S-ID.A.2).**

In the previous section, students interpreted the meaning of the various measures of center. Measures of center are important because they are single numbers that show what value is typical for a data set. However, data involves variation. How much data varies is an important question. In this section, students examine variability, the other major feature of measurements taken to answer a statistical question.

In grades 6 to 8, students learned that interquartile range (IQR) and mean absolute deviation (MAD) are ways to describe spread. In this section, they learn to calculate MAD's more sophisticated cousin, standard deviation. They start by looking at how much each data point deviates from the mean, and use these calculations to describe different data sets as more or less variable. Then, they go through the procedure for calculating standard deviation. (Although the Standards do not insist that students do such calculations or learn this procedure, doing the calculation a few times can help to illustrate what the standard deviation measures.) They come to understand standard deviation as "typical distance from the mean," and that higher values for standard deviation imply that a distribution is more spread out, whereas lower values imply that data are more closely clustered about the mean.

Tasks

[S-ID Understanding the Standard Deviation](#)

[S-ID Measuring Variability in a Data Set](#)

A1.1.6 Bringing it all together

- **Represent a data set in different ways and decide which way is most appropriate (S-ID.A.1).**
- **Select measures of center and spread appropriate to the shape of the distribution (S-ID.A.3).**
- **Compare and contrast two or more distributions by using appropriate measures to describe center, variability, and shape (S-ID.A.2).**

In this section, students are presented with summary statistics (mean, median,

standard deviation, Q1, Q3, and minimum and maximum values) for five sets of data. Students must decide how best to represent the data, interpret and understand differences in center and spread, and determine if there are any outliers in order to accurately compare the given data sets.

A1.1.7 Two-way frequency tables

- **Interpret a two-way table (S-ID.B.5).**
- **Understand that the choices made when organizing data can lead to different conclusions (S-ID.B.5).**

Sometimes, it is illuminating to categorize univariate data, especially when the categories might influence the values of the variable. For example, in health studies, participants are often categorized as smokers and non-smokers, or men and women.

Tasks

S-IC, S-ID Musical Preferences

A1.1.8 Summative Assessment

Assess students' ability to

- **calculate mean, median, and mode (S-ID.A.2);**
- **create box plots given data (S-ID.A.1);**
- **compare and contrast two frequency distributions (S-ID.A.3);**
- **articulate reasons to choose mean or use median as a measure of center (S-ID.A.2);**
- **read and interpret relative frequencies (S-ID.B.5).**



Unit Blueprint: One Variable Statistics

Typeset May 4, 2016 at 23:51:24. Licensed by Illustrative Mathematics under a Creative Commons Attribution 4.0 International License .