Overview for Families

Mathematics in Context unit: ***Dealing with Data***

Mathematical strand: **Data Analysis and Probability**

The following pages will help you to understand the mathematics that your child is currently studying as well as the type of problems (s)he will solve in this unit.

Each page is divided into three parts:

- *Section Focus*
  Identifies the mathematical content of each section.

- *Learning Lines*
  Describes the mathematical flow of each section.

- *Learning Outcomes*
  Outlines what students should know and be able to do at the end of each section.

"From the very beginning of his education, the child should experience the joy of discovery."

Alfred North Whitehead

Dealing with Data

Section A Are People Getting Taller?

Section Focus

Students explore a large data set that compares the heights of fathers to the heights of their sons. This is a historical data set from around 1900, of Pearson and Lee, two scientists. This data set will be further explored and analyzed throughout the rest of the unit. Students understand the need to organize and summarize a large data set; they evaluate statements about the data; and they are introduced to the concept of sampling.

Learning Lines

Using Data

In this section, students begin to understand how data are used to answer questions. Students begin to think about what data is, where it comes from, and how it is used. Students evaluate various arguments based on the presented data about the heights of fathers and their sons to see whether sons are taller than their fathers. The data set explored in this section is a paired data set: Each value in one data set (a father's height) is related to one value in the other data set (a son's height). This data set reappears throughout the unit.

Sampling

Collecting data is the beginning of the process that renders statistics. In some cases, it is important to ensure that the data collected are representative of the population of interest. If the whole population cannot be studied, a sample can be taken. In this section, students are introduced to the idea of a representative sample. This is done in an informal way. The term *sample* is used, and a definition is given. The term *representative* is not used. Students reflect on the way a sample is chosen.

Sampling

Students learn to understand that sampling should be done in a "proper way" in order to make valid conclusions possible. Students develop a critical attitude towards surveys and how data are collected, including the process of sampling. Sampling is revisited and formalized in the units *Insights into Data* and *Great Predictions*.

Organizing and Representing Large Data Sets

If a data set is very large, organizing techniques must be used to make sense of the data in some way. The data set students study in Appendix A has too many entries for students to get an idea of trends or patterns in the data by just looking at the numbers. Students think about ways to organize these data. In this section, students describe data in words (a qualitative description) or they use numbers.

Numbers that summarize sets of data are called *statistics*; these are used to make sense of data. Graphs can also be used to make sense of data by giving a global visual picture of the data set. In this section, no graphs are used yet. In the rest of this unit, students will learn to use a variety of graphs and statistics to make sense of data.

Dealing with Data

Section A Are People Getting Taller?

Learning Outcomes

Students will

• understand the importance of a sample being chosen in a proper way;

• understand the need to organize and summarize a large set of data in order for it to be useful; and

• develop a critical attitude toward surveys and sampling.

Dealing with Data

Section B Scatter Plots

Section Focus

Students learn to make a scatter plot by plotting matched pairs on a coordinate system. They then examine and evaluate the complete scatter plot of the Pearson and Lee data on the heights of fathers and their sons.

Learning Lines

Organizing and Representing Large Data Sets

This section introduces students to the scatter plot. Each point on a scatter plot represents one data pair. Students learn how to make a scatter plot on a coordinate grid. They have made graphs on a coordinate system in the unit *Expressions and Formulas*; the coordinate system will be formally introduced in the unit *Operations*.

Students learn to interpret a scatter plot and draw conclusions from it about the data. Scatter plots are useful in seeing patterns in the data and finding clusters or identifying trends or outliers (points whose values or coordinates are considerably different from those of all the other points). Students learn to think carefully about the plots and about what a given point represents.

The $y = x$ line, through all points whose coordinates are equal, can help students make inferences about the data. This line is defined in terms of the situation: In this case, it is the line where each son's height equals his father's height. This line can help to see which group was taller. Any data points representing pairs in which sons were taller than their fathers will be on one side of the line. Data points for pairs in which fathers were taller will be on the other (opposite) side.

Scatter plots are further studied in the unit *Insights into Data*.

Learning Outcomes

Students will

- be able to plot points to create a scatter plot from paired data;
- understand and interpret scatter plots; and
- understand the meaning of the line $y = x$ in a scatter plot if this line is defined in terms of the situation.

Dealing with Data

Section C Stem-and-Leaf Plots and Histograms

Section Focus

Students represent data using dot plots, stem-and-leaf plots, and histograms, and they use these plots to answer questions about the data. The mode is formally introduced in this section. Students use a data set of the ages of U.S. presidents at inauguration, as well as the data set of Pearson and Lee about the heights of fathers and sons. Students also collect and explore their own data set on estimates of the size of the teacher's head.

Learning Lines

Organizing Large Data Sets
Students study the data set of the ages of U.S. presidents at inauguration. This data set is presented as a large table. Finding specific information from the data set is easier if the information is ordered in some way. Students have seen this in Section A for the data set of Pearson and Lee. Students organize the president data in a new list or a diagram that makes it easier to see the distribution and to answer questions about the data.

Representing Data in Graphs
This section introduces dot plots, frequency tables, stem-and-leaf plots, and histograms, offering examples using the presidents' ages at inauguration. Students have seen some of these types of graphs in the unit *Picturing Numbers*.

Using Statistics to Represent Data
Data sets can be represented using numbers that summarize them—that is, statistics. Students have been informally introduced to the measures of central tendency in the unit *Picturing Numbers*. In this section, students think about what typical number may represent the age of a president at inauguration. The mode (one of the common one-number summaries) is formally introduced in this section, as "the value that occurs most often." The other statistics commonly used are the mean (see Section D) and the median (see Section E).

Learning Outcomes

Students will

- understand the need to organize and summarize a large set of data in order for it to be useful;
- be able to create and interpret a stem-and-leaf plot and a histogram;
- be able to find and interpret the mode of a data set;
- compare different representations and understand their differences and similarities; and
- be able to build an argument based on statistical measures and graphs, specifically: They can use the mode and dot plots, stem-and-leaf-plots, and histograms.

Dealing with Data

Section D Histograms and the Mean

Section Focus

Students learn to use the mean and the range to describe data sets. They explore different data sets. The mean is formally introduced in this section. Students learn how to find and calculate the mean and the range. They connect the mean to different types of graphs especially to histograms. Students also investigate situations in which the mean is not a meaningful statistic.

Learning Lines

Using Statistics to Represent Data

In the unit *Picturing Numbers*, students were introduced to a compensation strategy for finding the mean. In this section, they should recognize that while compensation, or taking away and adding to values to make them equal, is adequate in finding the mean for a small data set, it is difficult to use that strategy to find the mean for 1,000 values. By the end of this section, students should be able to find the mean using the calculation rule or algorithm (adding the data values and dividing by the number of data points). The mean is not always a data point itself.

In addition, students learn to see that the mean by itself is not a sufficient way of comparing two groups because it says nothing about the spread of the data, and it can be distorted by extreme values, or outliers.

In the context of mean yearly temperatures, students discover that the mean in itself does not tell much because the variation of the data is not reflected in the mean. In general, any single measure of central tendency will provide a weak description of a data set. A single number in combination with other numbers (e.g., range, minimum, or maximum) and/or graphs gives a better picture.

Representing Data in Graphs

In this section, histograms that were introduced in Section C are explored further. Students try to estimate the mean using a histogram of the data set.

Learning Outcomes

Students will

- be able to find, calculate, and interpret the mean and the range of a data set;
- be able to estimate the mean from a histogram;
- develop a critical attitude toward using statistical methods to solve problems and make a decision; specifically they learn that the mean is not always a meaningful statistic;
- build an argument based on statistical measures, specifically the mean and the range; and
- solve problems by choosing appropriate statistical measures and graphs.

Dealing with Data

Section E Box Plots and the Median

Section Focus

Students use a number line dot plot to explore the growth of the United States. They investigate the median and quartiles and use these to create a box plot out of the number line dot plot. These plots show how data are clustered or spread out. Students learn how to make and interpret box plots and how to find and use the median. Students also use box plots to compare data sets; for example, they compare the heights of fathers to the heights of sons.

Learning Lines

Using Statistics to Represent Data

In this section, the median is formally introduced as the middle value of a data set. The median is a common statistic used to describe a data set using one number. The median in itself will not be enough to describe a data set in a meaningful way—information about the spread of the data is needed as well. In a data set, 50% of the values are above the median and 50% are below the median. The medians of each half of a data set are called quartiles. Students do not need to know this word, although they must be able to determine the middle values of the upper and lower half of a data set in order to be able to create a box plot.

Representing Data in Graphs

The section begins with a number line plot of the dates of admission to the Union for all 50 states. The middle value (median) and middle value of each half (quartiles) can be found by dividing the data on the number line into four groups of equal size. Then a box plot can be made using the median, the quartiles, and the greatest and least values.

A box plot is a useful way to summarize a data set with only five numbers. The box (reaching from the 25th percentile to the 75th percentile) shows the distribution of the center (the middle 50%) of the data, and the whiskers show the spread of the first and last 25% of the data.

A box plot shows how data are clustered or spread out, but it does not show individual data points. Box plots are especially useful to compare the distribution of two or more large data sets.

Learning Outcomes

Students will

- be able to create and interpret box plots;
- find and interpret the median and quartiles of a data set;
- compare two data sets by using box plots;
- build an argument using the median, the spread, and box plots;
- develop a critical attitude toward using statistical methods to solve problems and make a decision; and
- solve problems by choosing appropriate statistical measures and graphs.