Overview for Families

Mathematics in Context unit: ***Insights into 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

Insights into Data

Section A Patterns in Data

Section Focus

Students set up a one-week experiment to collect plant growth data. They plant bean sprouts in different solutions; this is an experiment that simulates the effects of different chemicals on crop growth. Later in the unit (Section D), these data are analyzed and presented, and students compare the variability in the lengths of the sprouts over seven days.

Students investigate statements about data represented in a graph. They explore a scatter plot with census data that compare states by per capita income and the percent of the population that live in cities. Students analyze underlying patterns in the data. They find that it may be necessary to collect additional data in order for them to make appropriate inferences.

Learning Lines

In this section, students set up an experiment to collect their own data on plant growth. An experiment is one way to collect data. Sampling, conducting a survey, or doing a simulation are other ways of collecting data. Students explore and compare different data-collection methods in this unit. In the context of data gathered by the U.S. Census Bureau, students think about reasons for collecting certain data. Data collected and studied in this section are data on the length of bean sprouts and data on per capita income and percent of population living in cities for all U.S. states. The terms *population* and *sample*, introduced in the unit *Dealing with Data*, are revisited here.

Patterns in Scatter Plots

Students analyze a scatter plot that shows the results of research conducted by the U.S. Census Bureau about per capita income in each state and the percentage of people who live in urban areas in that state. They look for patterns in the data, try to explain these patterns, and draw conclusions.

Scatter plots were introduced and used in the unit *Dealing with Data.* A scatter plot can be used to investigate the relationship between two quantitative variables. The variable you begin with (sometimes called the *independent variable*) is generally associated with the horizontal axis. The second variable (sometimes called the *dependent variable*) is generally placed on the vertical axis. There are often underlying patterns in scatter plots of paired data. Students look for clusters of points, outliers, and trends in the scatter plot.

Identifying the data points as belonging to different categories can also help reveal patterns. For example, in the scatter plot on page 4, each data point can be circled in a color that indicates which region of the country that state is in. In this way, regional trends can be discovered. Critical review of representations of data may uncover more complex relationships than are initially noticed.

Statistics

The meaning of *the mean* is addressed in this section connecting the scatter plot with the census data. The per capita income per state is a "mean," and the nationwide per capita income is a "mean" as well. Students reflect on the meaning of this "mean" nationwide per capita income in relation to the per capita income per state.

Insights into Data

Section A Patterns in Data

Learning Outcomes

Students will have conducted an experiment and collected and recorded data from it. They will have analyzed representations of data—in particular, data represented in a scatter plot—and should be more aware of questions that should be asked when analyzing data sets and representations of data. Students will also be able to describe the relationship between two variables represented in a scatter plot; in their description, they can use terms like *clusters, patterns*, and *outliers*.

Insights into Data

Section B Selecting Samples

Section Focus

In this section, students investigate the concepts of sample and simulation. They discuss the process of sampling and how to collect "fair" data. They learn how to select a random sample of the whole population. They use a simulation and graph the results to investigate how likely certain outcomes in a random sample are. In doing so, students investigate the variation in samples. Students investigate possible bias in survey results.

Learning Lines

The focus of this section is on how to get good samples from a population. The terms *population* and *random* are defined in this section. (Students have been informally introduced to these terms in the unit *Dealing with Data*.) This section introduces students to the use of simulations to gather data to investigate how likely it is that certain outcomes occur in a random sample or in other words: how representative of a population random samples can be.

Students carry out a simulation by actually drawing random samples from a population of "notes" in a box. The notes in the box represent the distribution of the population of Delaware: 80% living in cities (8 notes in the box) and 20% living in rural areas (2 notes in the box). Later students repeat this simulation using random numbers.

Variation in Samples

Students take a number of random samples and record the class results both in a table and a histogram. By taking a number of samples, students create a sampling distribution. The table and histogram show the variation in samples. Students learn that the more samples they add to the study, the more they begin to approach the expected 80% figure for urban dwellers. As a mathematical rule, the results of the simulation will vary, but by adding additional trials, students find a cluster of samples that center around 16 out of 20 students in the sample (80%) coming from cities. The distribution in the collection of samples, if larger samples are taken, will approach the normal distribution or bell curve.

Conclusions from Samples and Bias

Inferences about the population are based on the analysis of the data from the sample. In a random sample, each person has the same chance of being selected for the survey. It is important for students to recognize that any outcome is possible but that some outcomes are more likely than others. There is a range, or cluster, of likely values. In the example in this section, the cluster of likely outcomes centers around 16 students from cities, which is 80% of the sample. Bias can result if the selection procedure for determining the sample was not objective and impartial. When the selection procedure is biased, taking a larger sample does not help. Many sources of bias are investigated in this section.

Insights into Data

Section B Selecting Samples

Learning Outcomes

Students recognize possible causes for bias in sample surveys. They understand how to set up and carry out a simulation to collect data. They will be able to use random samples in gathering data and can represent collected data in tables and graphs. Students will be able to recognize that in random samples variation in outcomes is normal. They understand that some outcomes are more likely then others.

Insights into Data

Section C Interpreting Graphs

Section Focus

This section introduces the importance of developing a critical approach to analyzing graphical representations. Students investigate several graphs to determine whether they represent data clearly and accurately. They investigate the influence of factors such as the scaling of the axes, the location of the origin, and the dimensions of three-dimensional representations. At the end of this section, students design and conduct their own statistical survey.

Learning Lines

Conclusions from Graphs

A statistical investigation can be split in three parts: collecting data, representing data, and drawing conclusions. This section introduces misleading graphical representations of data. In the unit *Dealing with Data* and in the previous sections, students have seen and used different types of graphs to represent data. Now the focus is on the (mis)representation of data in such graphs and the effect this has on the drawing of conclusions. Some common methods for misrepresentation are using inconsistent scaling of the axes, not beginning each axis at the origin, and using dimensions for pictographs or bar charts that do not match the data. Each of the methods can make a change in data values seem dramatic or minimal, depending on the desired effect.

First, students encounter a graph with no quantitative information in it and, consequently, no scaling of the axes. Then, students look at a pair of graphs that present the same data, but whose *y*-axes are scaled differently and begin with different numbers. Histograms and pictographs are then explored for misleading representations of data. And, finally, students design their own statistical surveys and represent the data graphically.

Learning Outcomes

Students will

- be able to represent data graphically in a correct way;
- analyze representations of data;
- determine whether representations of data are appropriate and accurate;
- know and understand methods used for misrepresentation of data in different types of graphs;
- describe the effect misrepresentations of data can have; and
- be able to design and conduct a statistical survey, represent the data graphically, and draw conclusions.

Insights into Data

Section D Using Data

Section Focus

This section emphasizes the use of graphs and numbers to display the results of the bean sprouts experiment that was conducted in Section A. Students analyze their data from the experiment. They select appropriate plots to summarize the data and use histograms and box plots to compare information. They also study and analyze other plant growth data. They review the use of mean, median, and mode.

Learning Lines

Representation of Data: Graphs and Statistics

In this section, several plots and graphs are used to represent data about the growth rate of bean sprouts in various solutions. Box plots and histograms were introduced in the unit *Dealing with Data* and used in Section B. Most of the concepts from statistics (measures and graphs) that appear in this section were introduced in the unit *Dealing with Data.*

Students make histograms and box plots that show the lengths of sprouts grown in various solutions. A histogram best presents clusters and gaps in the data. A box plot provides a summary of the five major data points: the minimum, the first quarter, the median, the third quarter, and the maximum.

Students discuss using a plot over time (usually a line graph) to show the data points as they change from day to day. There are many ways to make a plot over time: plot the growth of a representative member of the population; plot the range in growth per day; plot the mean or median length per day; or plot the growth of each individual in the sample. A plot over time allows you to look for trends in the data.

Drawing Conclusions

Students study and interpret different representations of the same data. They do so in the context of plant growth, connecting different types of graphs to different conclusions and statements about the plant data. In doing so, they also revisit the use of one-number summaries: mean, mode, and median. For the plant growth data collected from the bean growth experiment, students summarize the results and draw conclusions.

Learning Outcomes

Students will

• be able to represent data graphically in different ways, depending on the type of data and the statements they want to make;

• be able to describe data numerically using one-number summaries such as mean, mode, and median in combination with summaries, such as range, that tell about the spread of the data;

• know how to analyze representations of data;

• draw conclusions based on given data sets and representations of data; and

• be able to connect different representations of data to statements about these data.

Insights into Data

Section E Correlating Data

Section Focus

In this section, students investigate the relationship between two variables in scatter plots, tables, and three-dimensional plots. They determine if there is a strong, weak, positive, negative, linear, non-linear, or no apparent correlation between the two variables. They also look for linear patterns in scatter plots. Students draw conclusions based on scatter plots and correlations and learn that a relationship between two variables is not necessarily a cause-effect relationship.

Learning Lines

Correlation in Scatter Plots

In Section A, students were introduced to studying patterns in data represented in scatter plots. A scatter plot can be used to describe the relationship between two quantitative variables. In this section, students continue studying patterns in two variable data in different representations. In doing so, students are informally introduced to the idea of linear correlation. Correlation is a measure of the linear association between the two variables.

Students are not expected to learn how to compute the strength of the correlation (called *the correlation coefficient*) in this unit. An informal, intuitive understanding of the strength and type of the correlation is sufficient.

If the cloud of points forms a tight pattern, the correlation is referred to as strong. If the points form a loose pattern, the correlation is referred to as weak. When the data in the scatter plot form a "cloud" of points that are evenly spread across an area, there is no correlation between the data being graphed. If the cloud of points in a scatter plot slopes upward and to the right, there is a positive correlation between the variables. If the cloud of points slopes downward and to the right, there is a negative correlation between the variables.

In Section F, students' intuitive understanding is extended by including straight lines to describe a linear relationship.

Correlation and Cause-Effect

Students learn that the existence of a correlation between two variables, however, does not mean that a cause-effect relationship exists. Sometimes a third variable can be found that explains the relationship. Reasoning alone is sometimes enough to recognize a relationship and tell whether there is a strong correlation and/or a cause-effect relationship. At other times, more data is needed. Students should be able to contribute examples of strong cause-effect relationships.

Insights into Data

Section E Correlating Data

Learning Outcomes

Students will

- be able to analyze two variable data in tables, graphs, and three-dimensional plots;
- be able to describe the relationship between two variables;
- be able to identify the degree of correlation between two variables; and
- know that a strong correlation between two variables does not imply a cause-effect relationship.

Insights into Data

Section F Lines That Summarize Data

Section Focus

Students study the relationship between the lengths and widths of bird eggs by making scatter plots for various families of birds. They draw lines of best fit on the scatter plots in order to predict widths for given lengths or vice versa. This leads to an exploration of the slope of the line of best fit. They also compare data sets using the best-fit lines.

Learning Lines

Finding Lines of Best Fit in Scatter Plots

In Sections A and E, students have explored patterns in two variable data represented in scatter plots, and they have described the type and strength of the linear correlation between the two variables. In this section, students use best-fit lines to investigate the correlation between the two variables further.

A linear pattern in a scatter plot can be summarized by a line through the data. Different lines may be used, including an "eyeball" line drawn freehand to capture the trend, a least squares linear regression line based on the mean of the data, and a median fit line based on finding medians as representative points. In this section, students only use the "eyeball" line.

Students discuss criteria to decide if a line of best fit they drew is a "good" line. These criteria can include things like the following: a line going through as many points as possible or a line that has as many points above as below it, and so on.

Students are expected to be able to draw a line by eye and to find the equation of this line, as they learned in *Graphing Equations*, by finding the slope and the *y*-intercept.

Using Lines of Best Fit

Students use the lines of best fit to summarize the pattern, to check existing values, to predict unknown values, and to compare data sets. The accuracy of the predictions made using a best fit line depends on how the data are scattered around the line. When the data are very close to the line, you are more confident about the predictions based on that line.

Other Models of Fit for Scatter Plots

Some relationships between variables cannot be summarized with straight lines. The pattern in the scatter plot of the data in that case is clearly not a linear one, but a pattern seems to exist anyway. A curved model may then be a good one to use. Growth, for example, is often associated with curves. Note that statistical software and graphing calculators have the capability of fitting a lot of different regression models on scatter plots. Every model will always generate a graph on the scatter plot even if no relationship between the variables exists. Often such a model does not really fit the data or is meaningless in the context. Therefore, it is important to have students first decide whether a relationship will exist and what kind of relationship it will be.

Insights into Data

Section F Lines That Summarize Data

Learning Outcomes

Students will

• be able to describe the relationship between two variables;

• be able to identify the degree of correlation between variables;

• be able to draw a line of best fit in a scatter plot and be able to write the equation of this line;

• know how to use a line of best fit to check existing values, to predict unknown values, and to compare different values;

• be able to interpret the slope of a line of best fit in terms of the context;

• know informal criteria for drawing a good line of best fit; and

• know that sometimes a curved model can better be used to summarize a pattern in data than a straight line.