**8imLesson 8.1: Two-Way Frequency Tables**

**Essential Question**: How can categorical data for two categories be summarized?

Data that can be expressed with numerical measurements are **quantitative data**. In this lesson, you will examine qualitative data, or **categorical data**, which cannot be expressed using numbers. Data describing animal type, model of car, or favorite song are examples of categorical data

If a data set has two categorical variables, you can list the frequencies of the paired values in a **two-way frequency table.**

**Example 1:**

A high school’s administration asked 100 randomly selected students in the 9th and 10th grades about what fruit they like best. Complete the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade**  | **Apple** | **Orange** | **Banana** | **Total** |
| **9th**  | 19 | 12 | 23 |  |
| **10th**  | 22 | 9 | 15 |  |
| **Total** |  |  |  |  |

**Example 2:**

Jenna asked some randomly selected students whether they preferred dogs, cats, or other pets. She also recorded the gender of each student. The results are shown in the two-way frequency table below. Each entry is the frequency of students who prefer a certain pet and are a certain gender. For instance, 8 girls prefer dogs as pets. Complete the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender**  | **Dog**  | **Cat** | **Other** | **Total** |
| **Girl** | 8 | 7 | 1 |  |
| **Boy** | 10 | 5 | 9 |  |
| **Total** |  |  |  |  |

**Example 3:**

One hundred students were surveyed about which beverage they chose at lunch. Some of the results are shown in the two-way frequency table below. Complete the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Juice** | **Milk** | **Water** | **Total** |
| **Girl**  | 10 |  | 17 |  |
| **Boy** | 15 | 24 | 21 | 60 |
| **Total** |  |  |  |  |

**Homework:** Workbook Page 352 & Workbook Page 355.

**Lesson 8.2: Relative Frquency**

**Essential Question**: How can you recognize possible associations and trends between two categories of categorical data?

 To show what portion of a data set each category in a frequency table makes up, you can convert the data to relative frequencies. The **relative frequency** of a category is the frequency of the category divided by the total of all frequencies.

Two types of relative frequencies are found in a relative frequency table: 1. A **joint relative frequency** is found by dividing a frequency that is not in the Total row or the Total column by the grand total. It tells what portion of the total has both of the two specified characteristics. 2. A **marginal relative frequency** is found by dividing a row total or a column total by the grand total. It tells what portion of the total has a specified characteristic.

**Example 1:**

For her survey about sports preferences, Kenesha also recorded the gender of each student. The results are shown in the two-way frequency table for Kenesha’s data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Basketball** | **Football** | **Soccer** | **Total** |
| **Girls**  | 6 | 12 | 18 | 36 |
| **Boys** | 14 | 20 | 10 | 44 |
| **Total** | 20 | 32 | 28 | 80 |

**Step One: Take all your table values and create fractions using your totals. Find each categories percentage.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Basketball** | **Football** | **Soccer** | **Total** |
| **Girl** |  |  |  |  |
| **Boy** |  |  |  |  |
| **Total** |  |  |  |  |

**Example 2:**

Millie performed a survey of students in the lunch line and recorded which type of fruit each student selected along with the gender of each student. The two-variable frequency data she collected is shown in the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Apple** | **Bananas** | **Orange** | **Total** |
| **Girl** | 16 | 10 | 14 | 40 |
| **Boy** | 25 | 13 | 14 | 52 |
| **Total** | 41 | 23 | 28 | 92 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Apple** | **Bananas** | **Orange** | **Total** |
| **Girl** |  |  |  |  |
| **Boy** |  |  |  |  |
| **Total** |  |  |  |  |

A **conditional relative frequency** describes what portion of a group with a given characteristic also has another characteristic. A conditional relative frequency is found by dividing a frequency that is not in the Total row or the Total column by the total for that row or column.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gender** | **Basketball** | **Football** | **Soccer** | **Total** |
| **Girls**  | 6 | 12 | 18 | 36 |
| **Boys** | 14 | 20 | 10 | 44 |
| **Total** | 20 | 32 | 28 | 80 |

1). Find the conditional relative frequency that a a person in Kenesha's survey prefers soccer, given that the person is a girl.

2). Find the conditional relative frequency that a student in Kensha’s survey prefer basketball, given that the student is a girl.

3). Compare your results:

**Homework:** Workbook Pages 365-367 (1-19)

**Lesson 9.1: Measures of Center and Spread**

**Essential Question**: How can you describe and compare data sets?

Two commonly used measures of center for a set of numerical data are the mean and median. Measures of center represent a central or typical value of a data set. The **mean** is the sum of the values in the set divided by the number of values in the set. The **median** is the middle value in a set when the values are arranged in numerical order.

Measures of spread are used to describe the consistency of data values. They show the distance between data values and their distance from the center of the data. Two commonly used measures of spread for a set of numerical data are the range and interquartile range (IQR). The **range** is the difference between the greatest and the least data values. **Quartiles** are values that divide a data set into four equal parts. **The first quartile** (Q 1)is the median of the lower half of the set, the **second quartile** (Q 2) is the median of the whole set, and the **third quartile** (Q 3) is the median of the upper half of the set. The interquartile range (IQR) of a data set is the difference between the third and first quartiles. It represents the range of the middle half of the data.

**Example 1:**

The number of text messages that Isaac received each day for a week is shown. 47, 49, 54, 50, 48, 47, 55.

 **Mean:** Add up all the given numbers and divide by the total numbers in the sequence

 **Medium:** List the numbers in order from least to greatest.

 **Range & Quartiles:** Use your medium set-up to solve the Range and Quartiles.

**Example 2:**

The amount of money Elise earned in tips per day for 6 days is listed below. $75, $97, $360, $84, $119, $100.

**Mean:** Add up all the given numbers and divide by the total numbers in the sequence

 **Medium:** List the numbers in order from least to greatest.

 **Range & Quartiles:** Use your medium set-up to solve the Range and Quartiles.

Another measure of spread is the **standard deviation**, which represents the average of the distance between individual data values and the mean.

The formula for finding the standard deviation of the data set {x1, x2, x2, x2 ∙∙∙, xn}, with n elements and mean x, is shown below.

**Example 3:**

Find the standard deviation of 77, 86, 84, 93, 90.

**Step 1: Find the mean**

**Step 2:** Fill out the graph

|  |  |  |
| --- | --- | --- |
| **Data Value, x** | **Deviation from mean** | **Squared Deviation** |
| **77** |  |  |
| **84** |  |  |
| **86** |  |  |
| **90** |  |  |
| **93** |  |  |

**Step 3:**  Find the mean for the squared deviations.

**Step 4:**  take the square root of the standard deviation mean.

**Example 2:** Find the standard deviation of 4, 8, 12, 9, &, 15

**Step 1: Find the mean**

**Step 2:** Fill out the graph

|  |  |  |
| --- | --- | --- |
| **Data Value, x** | **Deviation from mean** | **Squared Deviation** |
| **4** |  |  |
| **8** |  |  |
| **9** |  |  |
| **12** |  |  |
| **15** |  |  |

**Step 3:**  Find the mean for the squared deviations.

**Step 4:**  take the square root of the standard deviation mean.

**Homework:**  Workbook Pages 384-385 (1-10)

**Homework:**  Workbook Page 386 (16-21)

**Lesson 9.2: Data Distributions and Outliers**

**Essential Question**: What statistics are most affected by outliers, and what shapes can data distributions have?

A **dot plot** is a data representation that uses a number line and Xs, dots, or other symbols to show frequency. Dot plots are sometimes called line plots.

An **outlier** is a value in a data set that is much greater or much less than most of the other values in the data set. Outliers are determined by using the first or third quartiles and the IQR.

**Example 1:**

Twelve employees at a small company make the following annual salaries (in thousands of dollars): 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 150.

**Step 1: Choose the number line with the most appropriate scale for this problem. Explain your reasoning.**

**Step 2: Create and label a dot plot of the data. Put an X above the number line for each time that value appears in the data set.**

**Step 3: Finding the Outlier- Find the quartiles and the IQR to determine whether 150 is an outlier.**

150 >Q3 + 1.5(IQR)

**Example 2:**

Suppose that the salaries from Part A were adjusted so that the owner’s salary is $65,000. Now the list of salaries is 25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, and 65.

**Step 1: Choose the number line with the most appropriate scale for this problem. Explain your reasoning.**

**Step 2: Create and label a dot plot of the data. Put an X above the number line for each time that value appears in the data set.**

**Step 3: Finding the Outlier- Find the quartiles and the IQR to determine whether 150 is an outlier.**

65 > Q3 + 1.5(IQR)

**Homework:** Workbook Page 396 (1-6)

**Homework:** Workbook Page 398-399 (13-15)

**Comparing Distribution Data**

A **data distribution** can be described as **symmetric, skewed to the left, or skewed to the right,** depending on the general shape of the distribution in a dot plot or other data display.

 **Skewed to the Left**  **Symmetric** **Skewed to the Right**

**Example 1:**

The data table shows the number of miles run by members of two track teams during one day.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Miles**  | **3** | **3.5** | **4** | **4.5** | **5** | **5.5** | **6** |
| **Team A** | 2 | 3 | 4 | 4 | 3 | 2 | 0 |
| **Team B** | 1 | 2 | 2 | 3 | 3 | 4 | 3 |

**Step 1: Create your distribution charts.**

 **Team A** **Team B**

**Example 2:**

The table shows the number of free throws attempted during a basketball game. Make a dot plot and determine the type of distribution. Then explain what the distribution means for the data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Shots**  | **0** | **2** | **4** | **6** | **8** |
| **Team A** | 2 | 2 | 4 | 2 | 2 |
| **Team B** | 3 | 4 | 2 | 2 | 1 |

**Step 1: Create your distribution charts.**

 **Team A** **Team B**

**Lesson 9.3 Histograms and Box Plots**

**Essential Question:** How can you interpret and compare data sets using data displays?

A **histogram** is a bar graph that is used to display the frequency of data divided into equal intervals. The bars must be of equal width and should touch but not overlap. The heights of the bars indicate the frequency of data values within each interval

When creating a histogram, make sure that the bars are of equal width and that they touch without overlapping. Create a frequency table to help organize the data before constructing the histogram. Consider the range of the data values when creating intervals.

**Example 1:**

Listed are the ages of the 100 U.S. senators at the start of the 112th Congress on January 3, 2011. 39, 39, 42, 44, 46, 47, 47, 47, 48, 49, 49, 49, 50, 50, 51, 51, 52, 52, 53, 53, 54, 54, 55, 55, 55, 55, 55, 55, 56, 56, 57, 57, 57, 58, 58, 58, 58, 58, 59, 59, 59, 59, 60, 60, 60, 60, 60, 60, 60, 61, 61, 62, 62, 62, 63, 63, 63, 63, 64, 64, 64, 64, 66, 66, 66, 67, 67, 67, 67, 67, 67, 67, 68, 68, 68, 68, 69, 69, 69, 70, 70, 70, 71, 71, 73, 73, 74, 74, 74, 75, 76, 76, 76, 76, 77, 77, 78, 86, 86, 86

**Step 1:** Create a frequency table. **Step 2**: Use the frequency table to The data values range from 39 to 86, create your histogram. so use an interval width of 10 and start the first interval at 30.

|  |  |
| --- | --- |
| **Age Interval**  | **Frequency** |
| **30-39** |  |
| **40-49** |  |
| **50-59** |  |
| **60-69** |  |
| **70-79** |  |
| **80-89** |  |

**Example 2:**

Listed are the scores from a golf tournament. 68, 78, 76, 71, 69, 73, 72, 74, 76, 70, 77, 74, 75, 76, 71, 74

|  |  |
| --- | --- |
| **Age Interval**  | **Frequency** |
| **30-39** |  |
| **40-49** |  |
| **50-59** |  |
| **60-69** |  |
| **70-79** |  |
| **80-89** |  |

 **Step 2:** Use the frequency table to create your histogram.

**Estimating Histograms**

**Example 1:** The histogram shows the ages of teachers in a high school.

**Step 1:** To estimate the mean, first find the midpoint of each interval, and multiply by the frequency.

**Step 2:** Find the mean of the frequencies

**Example:**

The histogram shows the 2012 Olympic results for women’s weightlifting

**Step 1:** To estimate the mean, first find the midpoint of each interval, and multiply by the frequency.

**Step 2:** Find the mean of the frequencies

**Homework:** Workbook Pages 410-412 (5-11)

**Box Plots**

A box plot can be used to show how the values in a data set are distributed. You need 5 values to make a box plot: the minimum (or least value), first quartile, median, third quartile, and maximum (or greatest value).

**Example 1:**

3, 4, 8, 12, 7, 5, 4, 12, 3, 9, 11, 4, 14, 8, 2, 10, 3, 10, 9, 7

**Step 1:** Order the data from least to greatest

**Step 2:** Identify the 5 needed values. Those values are the minimum, first quartile, median, third quartile, and maximum

**Step 3:** Draw a number line and plot a point above each of the 5 needed values. Draw a box whose ends go through the first and third quartiles, and draw a vertical line segment through the median. Draw horizontal line segments from the box to the minimum and maximum.

**Example 2:**

13, 14, 18, 13, 12, 17, 15, 12, 13, 19, 11, 14, 14, 18, 22, 23

**Step 1:** Order the data from least to greatest

**Step 2:** Identify the 5 needed values. Those values are the minimum, first quartile, median, third quartile, and maximum

**Step 3:** Draw a number line and plot a point above each of the 5 needed values. Draw a box whose ends go through the first and third quartiles, and draw a vertical line segment through the median. Draw horizontal line segments from the box to the minimum and maximum.

**Homework:** Workbook Pages 412-413 (13-18)

**Lesson 9.4***:* Normal Distribution

Essential Question: How can you use characteristics of a normal distribution to make estimates and probability predictions about the population that the data represents?

A bell-shaped, symmetric distribution with a tail on each end is called a **normal distribution**.

**Using Properties of Normal Distribution**

Use the top of workbook page 420.

**Example 1:**

The masses (in grams) of pennies minted in the United States after 1982 are normally distributed with a mean of 2.50 g and a standard deviation of 0.02 g.

**Example 1:** Find the percent of pennies that have a mass between 2.46 g and 2.54 g.

**Step 1:** Find the difference between the given amount and the mean.

**Step 2**: Divide the difference by the standard deviation.

**Step 3:** What does that mean?!?!

**Example 2:** Find the percent of pennies that have a mass between 2.48 g and 2.52 g

**Step 1:** Find the difference between the given amount and the mean.

**Step 2**: Divide the difference by the standard deviation.

**Step 3:** What does that mean?!?!

**Estimating Probabilities in Approximately Normal Distribution**

 **Example 1**: The masses (in grams) of pennies minted in the United States after 1982 are normally distributed with a mean of 2.50 g and a standard deviation of 0.02 g.

1). Estimate the probability that a randomly chosen penny has a mass greater than 2.52 g.

**Step 1:** Find the different between the mean and the given amount.

**Step 2:** Divide your answer by the standard deviation.

**Step 3**: Identify the percent greater than 2.52g. Recall normal distribution values.

2). Estimate the probability that a randomly chosen penny has a mass greater than 2.56 g.

**Step 1:** Find the different between the mean and the given amount.

**Step 2:** Divide your answer by the standard deviation.

**Step 3**: Identify the percent greater than 2.56g. Recall normal distribution values.