1

It Takes Two

Creating and Interpreting Frequency Distributions

Warm Up

1. Create a bar graph to display the data in the frequency table.

Favorite School Subjects in Mr. Luft's Class	
Subject	**Frequency**
Math	⊥⊥⊥ III
History	IIII
Science	⊥⊥⊥ I
Art	⊥⊥⊥ II

Learning Goals

- Construct and interpret frequency and marginal frequency distributions displayed in two-way tables for two-variable categorical data.
- Create and interpret graphs of frequency distributions displayed in two-way tables.

Key Terms

- categorical data
- two-way frequency table
- frequency distribution
- joint frequency
- marginal frequency distribution

You have explored the relationship between two variables of numerical data. How can you follow the statistical process to determine whether there are any associations between two variables of categorical data?

Survey Says . . .

Recall that the first step of the statistical process is to formulate a question. A statistical question anticipates an answer based on data that vary.

Cut out the survey questions located at the end of the lesson. Read each question and consider how you and other people might answer it. Sort the questions into groups based on the types of answers that could be given for each.

1. **Record your groups and the questions in each group.**

2. **What observations can you make about the types of data that can be collected by the questions?**

Two types of variable data that can be collected from a statistical question are numerical and *categorical data*. **Categorical data** are data that can be grouped into categories. Numerical data are data that can be placed on a numerical scale and compared.

3. **For the survey questions that have categorical answers, is there a way to group the data collected by the question into more than one category? Explain your reasoning.**

Categorical Data in Two Variables

Ms. Seymour is the school cafeteria supervisor at Williams High School. She has been asked to cut her food budget for the upcoming school year. One idea she has is to cut the number of meal choices during the week. Ms. Seymour decides to survey the students in Mr. Kolbe's gym class, which consists of 9th and 10th graders. She recorded the results of her survey in the table shown.

The first step of the statistical process is to formulate a statistical question. The second step is to collect the data.

1. Consider the table.

    a. What type of collection method did Ms. Seymour use?

    b. What question(s) did she ask the students?

    c. Describe the data she collected as either numerical or categorical. Explain your reasoning.

Grade	Favorite Meal
9	Salad bar
10	Burgers
10	Pizza
10	Chicken nuggets
10	Chicken nuggets
9	Burgers
10	Salad bar
9	Salad bar
10	Chicken nuggets
9	Burgers
10	Pizza
9	Salad bar
9	Burgers
10	Burgers
9	Chicken nuggets
9	Salad bar
10	Chicken nuggets
10	Chicken nuggets
10	Salad bar
10	Burgers
10	Salad bar
9	Burgers
9	Pizza
10	Chicken nuggets
10	Salad bar
9	Salad bar
10	Pizza
9	Pizza
10	Chicken nuggets
9	Pizza

The third step of the statistical process is to analyze the data.

2. Analyze Ms. Seymour's data table. Can you see any trends in the data just by looking at her data table? Explain why or why not.

Frequency Distribution Tables

Previously, you analyzed two-variable data sets that were quantitative, or numerical. You displayed those as a scatter plot and modeled them with a regression curve. In this topic you will explore the relationship between two-variable data sets that are qualitative, or categorical.

> Categorical data can also be called qualitative data.

One method of organizing categorical data is to use a *two-way frequency table*. A **two-way frequency table** displays categorical data by representing the number of occurrences that fall into each group for two variables. On the table, one variable is divided into rows and the other is divided into columns.

Consider the favorite meal data collected by Ms. Seymour in the previous activity.

1. **The first variable is the grade level. Identify the groups for this variable.**

> **Remember:**
>
> There is a difference between the variables in a data set and the groups in a data set.

2. **The second variable is the favorite meal. Identify the groups for this variable.**

Grade	Favorite Meal
9	Salad bar
10	Burgers
10	Pizza
10	Chicken nuggets
10	Chicken nuggets
9	Burgers
10	Salad bar
9	Salad bar
10	Chicken nuggets
9	Burgers
10	Pizza
9	Salad bar
9	Burgers
10	Burgers
9	Chicken nuggets
9	Salad bar
10	Chicken nuggets
10	Chicken nuggets
10	Salad bar
10	Burgers
10	Salad bar
9	Burgers
9	Pizza
10	Chicken nuggets
10	Salad bar
9	Salad bar
10	Pizza
9	Pizza
10	Chicken nuggets
9	Pizza

The third step of the statistical process is to analyze the data numerically and graphically. Creating a two-way frequency table helps to organize the data so that they can be analyzed numerically.

3. Create a two-way frequency table of the data.

   a. Enter the name of each group.

   b. Record the favorite meal for each student in the appropriate row using tally marks. Then, write the frequency of each meal for each grade level.

Favorite Meals

Grade Level				

4. Are there any associations between grade level and favorite meal? If so, explain what trends you think exist in these data.

The table you created is a *frequency distribution*. A **frequency distribution** displays the frequencies for categorical data in a two-way table. Each time you determined the frequency of one favorite meal of one of the grade levels, you recorded a *joint frequency*. Any frequency you record within the body of a two-way frequency table is known as a **joint frequency**.

A two-way frequency table is helpful in organizing each group's frequency in an efficient way. However, it is common to determine the total number of people surveyed just to ensure that a good survey was taken. Determining this total is also helpful to ensure that you recorded the data accurately within the table. For example, if you know 50 people took part in the survey, and the sum of the joint frequencies is 47, then you know that you are missing three data points from the data set.

© Carnegie Learning, Inc.

5. **Use the data from your frequency distribution to determine the total number of 9th graders and 10th graders, and to determine the total number of frequencies for each favorite meal category.**

Favorite Meals

	Burgers	Chicken Nuggets	Pizza	Salad Bar	Total
9th grade					
10th grade					
Total					

Grade Level

You just created a *marginal frequency distribution* of the data by determining the totals for each group.

> A **marginal frequency distribution** displays the total of the frequencies of the rows or columns of a frequency distribution.

6. **Analyze the marginal frequency distribution to answer each question.**

   a. **How many 9th graders participated in the survey?**

   b. **How many students prefer burgers?**

   c. **How many students prefer chicken nuggets?**

   d. **How many 10th graders participated in the survey?**

   e. **How many students prefer the salad bar?**

7. How can you use the totals to determine whether you correctly created the frequency distribution?

8. Use the marginal frequency distribution to answer each question.

   a. Which meal did the least number of students say was their favorite meal?

   b. Which meal did the least number of ninth grade students say was their favorite meal?

   c. Which meal is the most favorite of all students?

   d. Which meal is the most favorite of the 10th graders?

Representing Data

While a two-way table shows a numerical summary of the data, a graph can help relay information about a survey in a visual way. Remember, every graph tells a story.

Recall that Ms. Seymour is trying to determine ways to cut the cafeteria budget for the upcoming school year. She would like to use a graph to visually display the ideas she has for cutting the cafeteria budget.

Ms. Seymour has gathered her data and organized them in the frequency distribution table shown.

Favorite Meals

|		Burgers	Chicken Nuggets	Pizza	Salad Bar
Grade Level	9th grade	4	1	3	5
	10th grade	3	7	3	4

1. **Analyze the frequency distribution table.**

   a. **Determine which graphical display(s) would be appropriate to represent Ms. Seymour's data. Justify your response.**

   b. **Determine which graphical display(s) would not be appropriate to represent Ms. Seymour's data. Justify your response.**

You are still completing the third step of the statistical process, but now you are analyzing the data graphically by creating a display.

2. Construct two bar graphs of the frequencies. Be sure to include a key to identify what each bar represents.

Favorite Meal

Grade Level
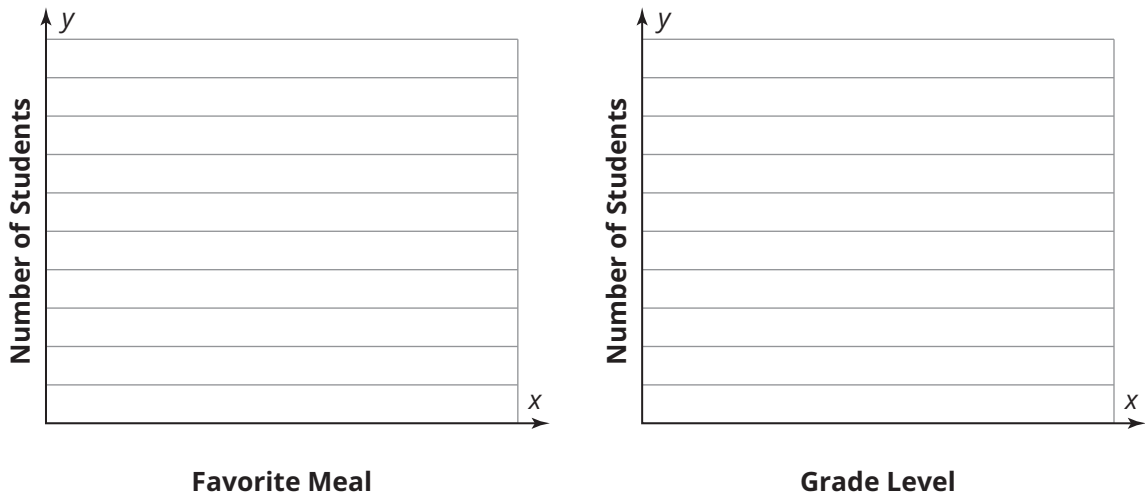
3. What conclusions can you draw by examining the graphs?

4. Use the graphs to determine whether you represented the data from the frequency distribution table accurately. Explain how you verified that the data in the graphs match the data in the frequency distribution table.

5. Does it matter which graph Ms. Seymour's uses to display her survey data? Explain your reasoning.

Ms. Seymour must decide on a plan for the upcoming school year. The principal of the school would like Ms. Seymour to present her data and a graph to justify her decision to cut costs.

6. Which meal choice would you cut according to the data? Explain why you would discontinue that meal choice. Then explain which graph you would recommend Ms. Seymour use when she presents her plan.

> The fourth step of the statistical process is to interpret the data.

Ms. Seymour just thought of an idea, and she thinks it will help cut the cafeteria costs. She is recommending that two lunch periods be created: one for the 9th graders and one for the 10th graders. She thinks that if two lunch periods exist, she can keep all four meal choices, but just cook a lesser amount of certain choices; thus cutting costs.

7. Do you think Ms. Seymour should present this idea to the principal? Use the data to justify your reasoning.

8. Which graph would you recommend Ms. Seymour use to justify her solution? Explain your reasoning.

TALK the TALK

Frequent Favorites

An art teacher asks his students what their favorite color is and records the results in the table shown.

Student	Favorite Color
Paul	red
Rachel	blue
Jennifer	blue
Sean	red
Adam	green
Vicki	blue
Spencer	red
Josh	red
Katie	green
Gina	green
Matt	red
James	blue
Corinne	red
Ben	red
Troy	blue
Zach	blue

1. **Explain how the art teacher can represent these data in a way that others could interpret it.**

2. **List some advantages and disadvantages of using a table, a marginal frequency distribution, and a graphical display to represent categorical data.**

   table:

   marginal frequency distribution:

   graphical display:

Survey Questions

A

How many pets do you own?

B

What is your favorite color?

C

What is your favorite sport?

D

What score did you earn on the last math test?

E

How many texts do you send per day?

F

What type of music do you prefer?

G

How many glasses of water do you drink each day?

H

How far do you live from school in miles?

I

What is your favorite snack?

J

How many hours do you watch television per day?

K

What mascot would you prefer for a sports team?

L

What grade are you in?

Write

Match each definition to its corresponding term.

1. displays the total of the frequencies of the rows or columns of a frequency distribution
2. displays the frequencies for categorical data in a two-way table
3. non-numerical data that can be grouped into categories
4. displays categorical data by representing the number of occurrences that fall into each group for two variables
5. any frequency you record within the body of a two-way frequency table

a. categorical data
b. two-way frequency table
c. frequency distribution
d. joint frequency
e. marginal frequency distribution

Remember

A frequency distribution table is helpful in organizing categorical data in two variables in order to see any associations and trends in the data.

You can use a double bar graph to visually represent categorical data in two variables.

Practice

1. Forty workers arriving at an office building in a city were asked how they got to work that day. They were also asked if they were less than 40 years old or older. The survey results are shown in the table.

   a. Identify the variables for this survey. Are the variables categorical or quantitative? Explain your reasoning.

   b. Construct and analyze a marginal frequency distribution for the survey data. What was the most commonly used transportation method for each age group? Explain how you determined your answer.

   c. Construct two bar graphs of the frequencies. In one, let the *x*-axis represent the transportation method, and in the other, let the *x*-axis represent the age levels. Let the *y*-axis represent the number of workers in both graphs. What conclusion(s) can you draw by examining each graph?

Age	Transportation Method	Age	Transportation Method
<40	Subway	<40	Bus
<40	Bus	<40	Bus
40+	Walk	<40	Subway
<40	Bus	40+	Car
<40	Subway	<40	Walk
40+	Car	40+	Taxi
40+	Car	40+	Walk
40+	Walk	<40	Subway
<40	Subway	40+	Car
40+	Taxi	<40	Taxi
<40	Walk	40+	Taxi
<40	Bus	<40	Bus
<40	Subway	<40	Bus
40+	Bus	<40	Subway
<40	Bus	40+	Walk
40+	Walk	40+	Car
40+	Taxi	40+	Subway
<40	Subway	40+	Bus
40+	Car	<40	Subway
<40	Car	40+	Taxi

d. The manager of a firm in the building where the survey was taken has noticed that a number of his employees have been coming in late. The late employees often say they are late because of subway problems, but he also notices it is mostly younger workers who are using this excuse. He thinks these employees may be irresponsible because most of his older employees are not coming in late. Which graph could be used to show the manager that his thinking may be wrong?

Stretch

1. Analyze the transportation survey data from the marginal frequency distribution in the Practice.

	Percent Distribution	
	<40	40+
Subway		
Bus		
Walk		
Car		
Taxi		
Total	100%	100%

a. Complete the table to show the percent of the total in each category that used the different forms of transportation.

b. Construct a bar graph of the percentages. Let the *x*-axis represent the transportation method, and let the *y*-axis represent the percent of workers. What conclusions can you make based on the graph?

Review

1. A group of 45 adults were asked how many times they dined out the previous week. Their responses are shown in the dot plot.

Number of Times Dining Out
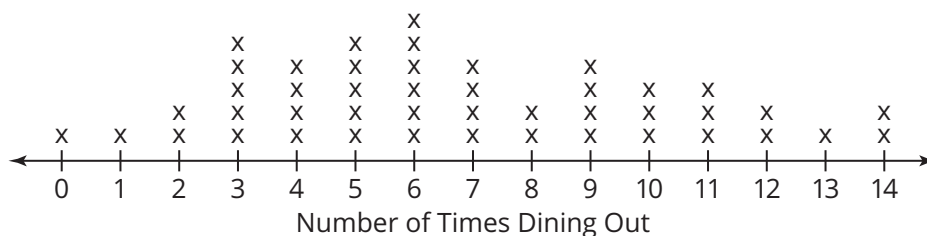
a. Describe the distribution of the dot plot.
b. How do you think the mean and median of the data set compare? Explain your reasoning.
c. Calculate the mean and median. Explain what they mean in terms of the problem situation.
d. Which measure of center do you think best represents these data? Explain your reasoning.

2. A clothing store has two checkout methods. In Method A, the customer chooses a line at any of the cashiers' stations. In Method B, the customers wait in one line and then get called to the next available cashier. Data were collected for customers using both methods. The dot plots show the average wait times in minutes for 15 customers for each method of checkout.

   a. Predict whether Method A or Method B has the greater standard deviation in wait times. Explain your reasoning.

   b. Determine the standard deviation for Method A and Method B. Round your answers to the nearest tenth. Explain what the standard deviations mean in terms of the problem situation.

   c. Which waiting line method would you prefer if you were in a big hurry to checkout? Explain your reasoning.

Method A

Wait Time (minutes)

Method B

Wait Time (minutes)
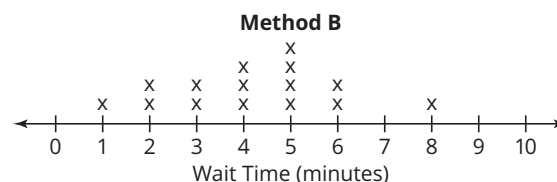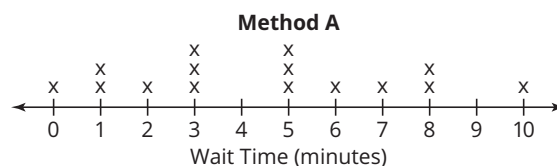
3. Use the graph of the exponential function to determine the domain, whether it is increasing or decreasing, the y- and x-intercept, and the horizontal asymptote.
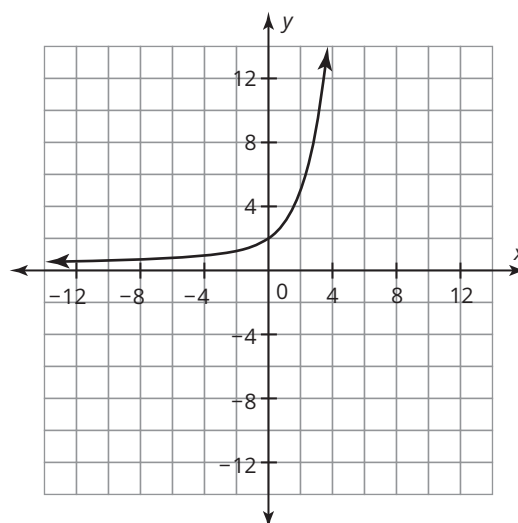
4. Is the graph of the function $f(x) = \left(\frac{1}{2}\right)^x$ increasing or decreasing? Explain your reasoning.