Introduction to Statistics

Chapter 2: Displaying and Describing Categorical Data

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Objectives:

## Students will be able to analyze categorical variables.

## Students will be able to see how two categorical variables are related by putting counts and percentages into a contingency table.

|  |  |
| --- | --- |
| **Main Idea** | **Notes** |
| **Three Rules to Data Analysis:** | There are three rules to data analysis:   1. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ - things may be revealed that are not obvious in the raw data - things to *think* about. 2. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ - important features and patterns in the data will *show* up. You may also see things that you did not expect. 3. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_- the best way to *tell* others about your data is with a well-chosen picture. |
| **Frequency Tables:**  **Example: Drawing a Relative Frequency Table** | For categorical variables, we often compile data by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  the number of values in each \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and display  these counts in a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Example:    A relative frequency table is similar, but shows \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  instead of \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Now make a relative frequency table of the Titanic passengers below: |
| **The Area Principle** | You might think that a good way to show the *Titanic* data is with this display:  03_02  Write what you think is wrong with this display:  This shape display violates the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. |
| **The Area Principle (Cont.)** | The ship display makes it look like most of the people on the *Titanic* were  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, with a few passengers along for the ride.  When we look at each ship, we see the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ taken up by  the ship, instead of the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the ship.  Question: How do you think we can fix this graph? |
| **Bar Charts** | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ display the distribution of a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable, showing the counts side-by-side.  Draw a bar chart of the Titanic data below:  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ display  the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of counts.  Draw a relative frequency bar chart of the Titanic Data below: |
| **Pie Charts** | When you want to display parts of a whole, you can use a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Pie charts display the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ *or* \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Draw a pie chart of the Titanic data below: |
| **Classwork: Violation of the Area Principle** | 1. Find a graph on the internet that is an example of a violation of the area principle. 2. Explain how the graph is misleading and what should be changed to improve it. 3. Create a new graphical display of the data that does not violate the area principle. For example, you can create a well-drawn bar graph or pie chart. 4. We will share these with the class. |
| **Contingency Tables**  **Contingency Tables (Cont.)** | Contingency tables allow us to look at \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ categorical variables at the same time.  They show how individuals are distributed across each \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_,  contingent on the second \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  ta03-04a  What two *variables* are we looking at here??  The \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (right and bottom) of a contingency table should have row and column totals (ADD THEM).  We use these totals to calculate \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  What is the marginal distribution of alive?  What is the marginal distribution of second class?  What is the marginal distribution of dead?  What is the marginal distribution of third class?  What is the marginal distribution of crew?  Each cell of the table gives a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_for a combination of variables.  ta03-04a  What does the second cell in the crew column tells us?  What does the first cell in the second-class column tells us?  What does the second cell in the third-class column tells us? |
| **Conditional Distribution** | A \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ shows the distribution of one variable for just the individuals who satisfy some condition on another variable.  The following is the conditional distribution of ticket *Class*, conditional on being A*live*:  ta03-07a  Now, create the conditional distribution of ticket *Class*, conditional on being *Dead:* |
| **Conditional Distribution (Cont.)** | Is there a difference in class for those who survived and those who perished?  Now you can make two pie charts for the two distributions.  03-06a  Is the distribution of *Class* for the survivors different from that of the non-survivors? Explain.  Do you think that Class and Survival are associated? Explain.  So, is Class and Survival independent of each other? Why or why not?  **Independence**: The variables are considered independent when the distribution of one variable is the same for all categories of the other variable. |
| **Segmented Bar Charts** | A \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ displays the same  information as a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, but in the form of bars instead of circles.  Each bar is treated as the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and is divided into segments  corresponding to the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in each group.  Draw the segmented bar chart for ticket *Class* by *Survival* status:  Is *Class* and *Survival* independent of each other? Why or why not? Use the segmented bar graph to explain your answer. |
| **Classwork/**  **Homework:** | Chapter 2 Example Problems  Chapter 2 Review for Test |