

STATISTICS IS THE GRAMMAR OF SCIENCE

PROBABILITY AND STATISTICS

LECTURE – 2

**PRESENTATION OF
QUALITATIVE DATA**

PREPARED BY
HAZBER SAMSON
FAST NUCES ISLAMABAD

REPRESENTATION OF DATA

After collection of sample data, we must “get acquainted” with them. The best way to become acquainted with the data is to use an initial exploratory data-analysis. For this we need to organize the data first and then visualize and analyze it.

There are two major components of data Representation

- Data Organization
- Data Visualization

DATA ORGANIZATION

In order to visualize and analyze the data first step is to organize the data, we can organize data in two ways.

- Organizing Qualitative Data
- Organizing Quantitative Data

DATA VISUALIZATION

When we organize our data, we sometimes begin to discover patterns or relationships in our data. To better explore and discover patterns and relationships, we can visualize your data by creating various charts and special “displays.” As is the case when organizing data, the techniques we use to visualize our data depend on the type of variable (categorical or numerical) of our data.

We can visualize data in two ways

- Qualitative Data Visualization
- Quantitative Data Visualization

Before going into details of data organization let’s study the idea of frequency distribution

THE FREQUENCY DISTRIBUTION

FREQUENCY The number of times a particular distinct value occurs is called its frequency (or count).

DISTRIBUTION The pattern of variability displayed by the data of a variable. The distribution displays the frequency of each value of the variable.

FREQUENCY DISTRIBUTION A frequency distribution provides a table of the values of the observations and how often they occur.

A frequency distribution is the organization of raw data in table form, using classes and frequencies.

TYPES OF FREQUENCY DISTRIBUTION There are two types of frequency distributions that are mostly used

- Categorical frequency distribution
- Grouped frequency distribution.

ORGANIZING QUALITATIVE DATA

We organize categorical data by tallying responses by categories and placing the results in tables. Typically, you construct a summary table to organize the data for a single categorical variable and you construct a contingency table to organize the data from two or more categorical variables.

There are two simple ways of organizing categorical data

- The Summary Table
- The Contingency Table

THE SUMMARY TABLE

A summary table presents tallied responses as frequencies or percentages for each category. A summary table helps you see the differences among the categories by displaying the frequency, amount, or percentage of items in a set of categories in a separate column. Table 2.1 shows a summary table that tallies the responses to a recent survey that asked adults how they pay their monthly bills.

TABLE 2.1 Types of Bill Payment

Form of Payment	Percentage (%)
Cash	15
Check	54
Electronic/online	28
Other/don't know	3

From Table 2.1, you can conclude that more than half the people pay by check and 82% pay by either check or by electronic/online forms of payment.

THE CONTINGENCY TABLE

A contingency table allows you to study patterns that may exist between the responses of two or more categorical variables. This type of table cross-tabulates, the responses of the categorical variables.

TABLE 2.2 Contingency Table Displaying Type of Fund and Whether a Fee Is Charged

TYPE	FEE		Total
	Yes	No	
Intermediate government	34	53	87
Short-term corporate	20	77	97
Total	54	130	184

CATEGORICAL FREQUENCY DISTRIBUTIONS

The categorical frequency distribution is used for data that can be placed in specific categories, such as nominal- or ordinal-level data. For example, data such as political affiliation, religious affiliation, or major field of study would use categorical frequency distributions.

Frequency Distribution of Qualitative Data

A **frequency distribution** of qualitative data is a listing of the distinct values and their frequencies.

PROCEDURE OF CONSTRUCTING FREQUENCY DISTRIBUTION OF QUALITATIVE DATA

To Construct a Frequency Distribution of Qualitative Data

Step 1 List the distinct values of the observations in the data set in the first column of a table.

Step 2 For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.

Step 3 Count the tallies for each distinct value and record the totals in the third column of the table.

Relative-Frequency Distribution of Qualitative Data

A **relative-frequency distribution** of qualitative data is a listing of the distinct values and their relative frequencies.

PROCEDURE OF CONSTRUCTING RELATIVE FREQUENCY DISTRIBUTION OF QUALITATIVE DATA

To Construct a Relative-Frequency Distribution of Qualitative Data

Step 1 Obtain a frequency distribution of the data.

Step 2 Divide each frequency by the total number of observations.

RELATIVE FREQUENCY DISTRIBUTION Relative frequency is the ratio of the frequency to the total number of observations

$$\text{Relative Frequency} = \frac{\text{Frequency}}{\text{Number of Observations}}$$

So a relative-frequency distribution provides a table of the values of the observations and (relatively) how often they occur.

EXAMPLE-1 FREQUENCY DISTRIBUTION OF QUALITATIVE DATA

Professor Weiss asked his introductory statistics students to state their political party affiliations as Democratic (D), Republican (R), or Other (O). The responses of the 40 students in the class are given in Table below.

(a) Determine a frequency distribution of these data.

(b) Determine a relative frequency distribution of these data.

TABLE Political party affiliations of the students in introductory statistics

D	R	O	R	R	R	R	R
D	O	R	D	O	O	R	D
D	R	O	D	R	R	O	R
D	O	D	D	D	R	O	D
O	R	D	R	R	R	R	D

SOLUTION We apply Procedure discussed above

STEP-1 *List the distinct values of the observations in the data set in the first column of a table.* The distinct values of the observations are Democratic, Republican, and Other, which we list in the first column of Table 2.7.

STEP-2 *For each observation, place a tally mark in the second column of the table in the row of the appropriate distinct value.* The first affiliation listed in Table 2.1 is Democratic, calling for a tally mark in the Democratic row of Table 2.7. The complete results of the tallying procedure are shown in the second column of Table 2.7.

STEP-3 *Count the tallies for each distinct value and record the totals in the third column of the table.* Counting the tallies in the second column of Table 2.2 gives the frequencies in the third column of Table 2.7. The first and third columns of Table 2.7 provide a frequency distribution for the data in Table 2.6.

TABLE 2.7 Table for constructing a frequency distribution for the political party affiliation data in Table 2.6

Party	Tally	Frequency
Democratic		13
Republican		18
Other		9
		40

Interpretation From Table 2.7, we see that, of the 40 students in the class, 13 are Democrats, 18 are Republicans, and 9 are Other.

By simply glancing at Table 2.7, we can easily obtain various pieces of useful information. For instance, we see that more students in the class are Republicans than any other political party affiliation.

(B) RELATIVE FREQUENCY DISTRIBUTION OF QUALITATIVE DATA**STEP-1** Obtain a frequency distribution of the data.

We obtained a frequency distribution of the data in part (a)

Party	Tally	Frequency
Democratic		13
Republican		18
Other		9
		40

STEP-2 Divide each frequency by the total number of observations.

Dividing each entry in the third column of the above Table by the total number of observations, 40, we obtain the relative frequencies displayed in the second column of Table below. The two columns of Table below provide a relative-frequency distribution for the data in Table 2.6.

Relative-frequency distribution for the political party affiliation data

Party	Relative frequency	
Democratic	0.325	← 13/40
Republican	0.450	← 18/40
Other	0.225	← 9/40
	1.000	

Interpretation From the above Table, we see that 32.5% of the students in Professor Weiss's introductory statistics class are Democrats, 45.0% are Republicans, and 22.5% are other.

NOTE: Relative-frequency distributions are better than frequency distributions for comparing two data sets. Because relative frequencies always fall between 0 and 1, they provide a standard for comparison.

VISUALIZING QUALITATIVE DATA

In this section we shall discuss graphs that are used to summarize qualitative, or attribute, or categorical data. For a single categorical variable we have three types of graphs

- Bar Chart
- Pie Chart

BAR CHART

The most common graphic form to present a qualitative variable is a bar chart. A bar chart compares different categories by using individual bars to represent the tallies for each category. Actually Bar graphs show the amount of data that belong to each category as a proportionally sized rectangular area. The length of a bar represents the amount, frequency, or percentage of values falling into a category.

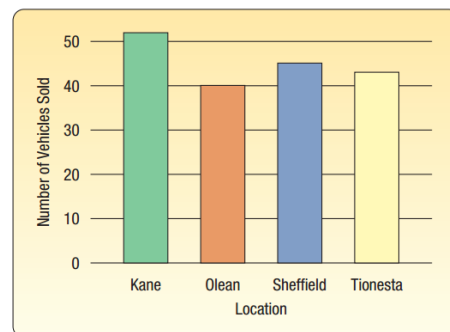
In most cases, the horizontal axis shows the variable of interest. The vertical axis shows the frequency or fraction of each of the possible outcomes. A distinguishing feature of a bar chart is there is distance or a gap between the bars. That is, because the variable of interest is qualitative, the bars are not adjacent to each other. Thus, a bar chart graphically describes a frequency table using a series of uniformly wide rectangles, where the height of each rectangle is the class frequency.

BAR CHART A graph that shows qualitative classes on the horizontal axis and the class frequencies on the vertical axis. The class frequencies are proportional to the heights of the bars.

EXAMPLE-2 Frequency Table for Vehicles Sold Last Month at Apple wood Auto Group by Location

Bar Chart Number of Vehicles Sold by Location

Location	Number of Cars
Kane	52
Olean	40
Sheffield	45
Tionesta	43
Total	180



The variables of interest are the location where the vehicle was sold and the number of vehicles sold at each location. We label the horizontal axis with the four locations and scale the vertical axis with the number sold. In Figure below, the locations are listed alphabetically. The locations could also be in order of decreasing or increasing frequencies.

EXAMPLE-3 Table below shows a summary table that tallies the responses to a recent survey that asked adults how they pay their monthly bills.

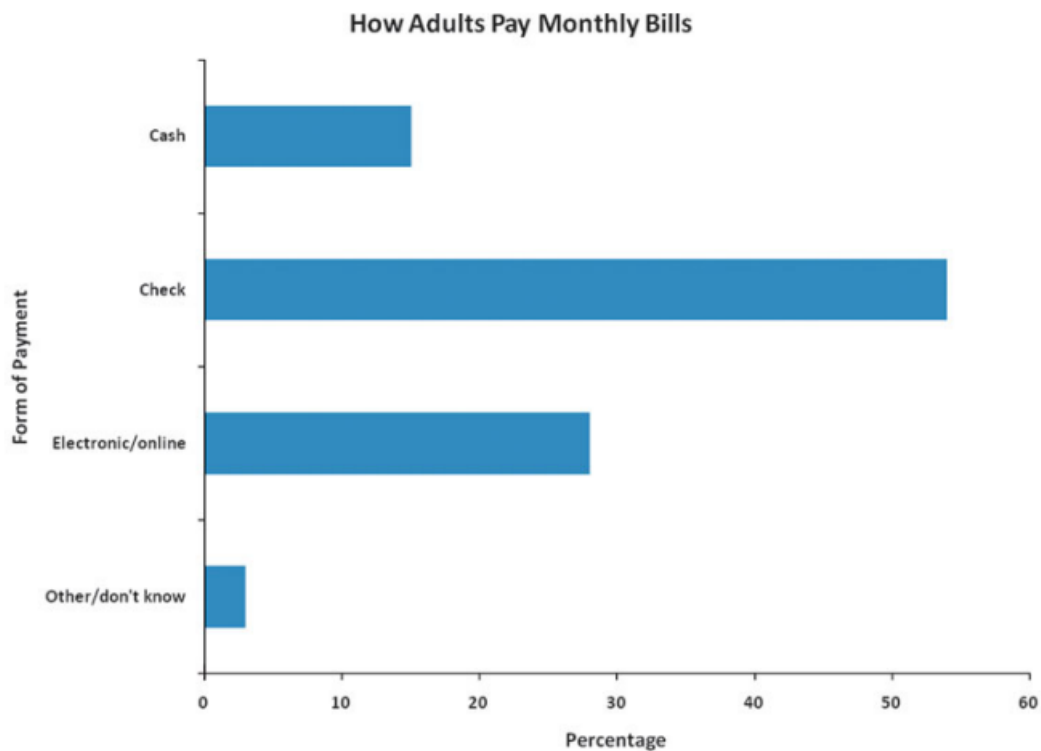
Types of Bill Payment

Form of Payment	Percentage (%)
Cash	15
Check	54
Electronic/online	28
Other/don't know	3

Source: Data extracted from "How Adults Pay Monthly Bills," USA Today, October 4, 2007, p. 1.

Figure 2.2 displays the bar chart for the data of Table 2.1 on page 30, which is based on a recent survey that asked adults how they pay their monthly bills

Bar Chart for how people pay their bills



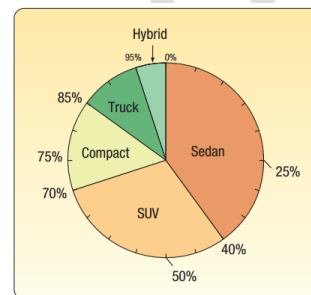
PIE CHART

Pie chart another useful type of chart for depicting qualitative information is a. A pie chart uses parts of a circle to represent the tallies of each category. The size of each part, or pie slice, varies according to the percentage in each category.

PIE CHART A chart that shows the proportion or percentage that each class represents of the total number of frequencies.

EXAMPLE-4 Construct a pie chart using the information in Table below, which shows the frequency and percent of cars sold by the Applewood Auto Group for each vehicle type.

Vehicle Type	Number Sold	Percent Sold
Sedan	72	40
SUV	54	30
Compact	27	15
Truck	18	10
Hybrid	9	5
Total	180	100



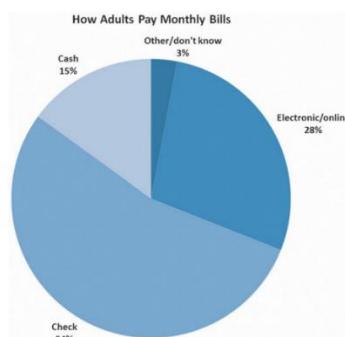
EXAMPLE-5 Table below shows a summary table that tallies the responses to a recent survey that asked adults how they pay their monthly bills.

Types of Bill Payment

Form of Payment	Percentage (%)
Cash	15
Check	54
Electronic/online	28
Other/don't know	3

Source: Data extracted from "How Adults Pay Monthly Bills," USA Today, October 4, 2007, p. 1.

Pie Chart for how people pay bills



EXAMPLE-6

SkiLodges.com is test marketing its new website and is interested in how easy its website design is to navigate. It randomly selected 200 regular Internet users and asked them to perform a search task on the website. Each person was asked to rate the relative ease of navigation as poor, good, excellent, or awesome. The results are shown in the following table:

Awesome	102
Excellent	58
Good	30
Poor	10

1. What type of measurement scale is used for ease of navigation?
2. Draw a bar chart for the survey results.
3. Draw a pie chart for the survey results.

SOLUTION

The data are measured on an ordinal scale. That is, the scale is ranked in relative ease of navigation when moving from “awesome” to “poor.” The interval between each rating is unknown so it is impossible, for example, to conclude that a rating of good is twice the value of a poor rating.

We can use a bar chart to graph the data. The vertical scale shows the relative frequency and the horizontal scale shows the values of the ease-of-navigation variable.

A pie chart can also be used to graph these data. The pie chart emphasizes that more than half of the respondents rate the relative ease of using the website awesome.

FIGURE 2.5 Bar Chart and Pie Chart