

Chapter 5

Descriptive Statistics (Part 1)

By the end of this chapter you should be able to:

- Explain what descriptive statistics are and what they are designed to do
- Define the 3 classes, or types, of descriptive statistics
- Define the term “Frequency” as it is applied in statistics
- Construct and interpret a “Frequency Table”
- Compare and contrast “Bar Charts”, “Pie Charts”, “Histograms” and “Scatterplots” and explain the circumstances under which each one should be used.
- Define the term “Normal Distribution” and identify the characteristics by which it may be recognized
- Define the term “Skewness” as it relates to a distribution of scores and explain what causes it.
- Determine whether a distribution is “Negatively Skewed” or “Positively Skewed”
- Define the term “Kurtosis” and describe the three “levels” of kurtosis.
- Define the term “Central Tendency” as it relates to a distribution of scores
- Define the term “Variability” as it relates to a distribution of scores

Rocket Boy

When I was in the eighth grade, I went through a “Rocket Phase”. I wanted to build a rocket in the worst way. Unfortunately, my parents never felt inclined to help me in such endeavors. While many of my friends had parents who would help them build and fly model rockets, I was left to improvise as best I could.

I used to explore all around the house looking for potential supplies that could be employed to build a rocket that would actually fly. I could describe hundreds of attempts that, let me just say, were less than successful.

Perhaps my most memorable attempt at rocket engineering began one day when I was exploring the bathroom cabinets for something that was roughly in the shape of a rocket. Have you ever noticed that a cardboard tampon applicator actually looks a lot like a two stage rocket? If you haven’t, take a look! I bet you will agree. Anyway, my eighth grade mind exclaimed “Eureka!” and I was sure that I was on my way into the history books!

“Borrowing” a few tampons, I headed off to my room which doubled as my laboratory. My first step was to glue a few fins onto the base of the tampon to give it aerodynamic stability. I also glued a cone to the top.

Next, I needed to decide on an appropriate fuel. I felt that regular gunpowder wouldn’t work because it burned too slowly. I needed something that burned fast!

Having a small supply of chemicals that I purchased on my meager allowance, I whipped up a batch of black powder. This is the fast burning—and unpredictable—stuff that people used in muskets before modern bullets were invented.

I packed my once tampon now rocket with black powder and felt like some kind of mad genius. Like Mr. Scott, the Chief Engineer on the old Star Trek television series, it seemed to me like I had worked a technological miracle! I could see the news headlines flash before my mind's eye. “Wonder boy converts household products into rockets. Launches first tampon into deep space! NASA offers big royalties to use the new technology!” Looking back I wonder, “What was I thinking?”

What really happened was quite different. I lit my homemade fuse and then moved safely away (lucky for me). When the fuse set off the black powder, I was witness to what was possibly the world's first exploding tampon. It never even left the launch pad. Instead, it simply vanished into a giant ball of flames.

The moral of this story is that even with the best intentions my efforts were doomed to failure. Why? Because I did not have the correct tools for the job and I did not have an adequate understanding of the basics of rocketry.

Many people (I tend to believe that a greater proportion of them, like myself, are male) like to skip the basics and go straight to the good stuff. Whether one is talking about relationships, business skills, hobbies, or assembling things, we don't like to bother with developing a solid foundation on which we can build. Instead, we just jump to the main event and hope for the best. This is a bad idea!

The same is true with statistics. You need to learn to walk before you can run. Did you notice that here we are in Chapter 5 of a statistics book and we have not even done any math yet? This is because I wanted to try and give you an understanding of the basic concepts that underlie science in general and the field of statistics in particular.

In this chapter, you will begin to develop a basic understanding of some of the most basic statistical tools used in applied statistics. These statistical tools are called “Descriptive Statistics”. There is very little math. It is mostly about using pictures to make the interpretation of data easier. In the next chapter, we will begin delving deeper into statistics. At each step I will try to give you the information you need to take the next step and so on until, before you know it, you will be doing some pretty complex stuff. For now, just try to master these basics. If you do, it will make the rest of the book much easier!

I need to warn you that the next five chapters are going to be a little tedious and perhaps a little boring. They are, however, very important to your understanding of the rest of the book. So, if you find yourself wanting to skim or skip, “Remember the exploding tampon” and don't make the same mistake in your study of statistics.

What Are Descriptive Statistics?

While they are not glamorous and don't allow you to do some of the things that the more exotic *inferential statistics* allow you to do, chances are that you will use *descriptive statistics* far more often than any other type of statistical analysis.

What are descriptive statistics? The term “descriptive statistics” refers to a set of tools that you can use to organize and summarize large (and sometimes small) amounts of data into clear and easily understandable bite-sized pieces.

Descriptive statistics are especially powerful in the business and consumer environments because they help people who either do not have the technical skill or else do not have the time to analyze the data themselves to get the big picture really fast. I still remember a manager I worked for at the California Department of Mental Health who wanted a little card that would fit in his wallet that contained all of the relevant information about California's public mental health system. This little card, he said, should have information such as:

- A count of how many mental health patients are being served
- A breakdown of those patients by gender and ethnicity
- A count of how many people spent time in a mental hospital
- How many days, on average, patients stay when they are admitted to a mental hospital
- How much money the State of California allocates to the mental health system and
- How much money the Federal government allocates to California's mental health program.

Each of the above examples illustrates the use of descriptive statistics. In addition to communicating summary information to managers and consumers, descriptive statistics are also used by researchers. They help researchers and other consumers of research information to better understand the general population as well as the specific group of people who participated in the research study. I will discuss this later in the book and your instructor will undoubtedly address it in detail in class.

There are basically three kinds of descriptive statistics that you need to be aware of:

1. Tabular Statistical Tools
 - a. Frequency Tables
2. Graphical Tools
 - a. Pie Charts
 - b. Bar Charts
 - c. Histograms
 - d. Scatterplots
3. Summary Statistics (these will be discussed in detail in later chapters)
 - a. Measures of Central Tendency (See chapter 6)
 - b. Measures of Variation (See chapter 7)

Tabular Statistical Tools

Tabular statistics are essentially numbers that have been rearranged and placed into a table in a manner that makes them much easier to understand. Let me give you a real world example that someone in business might use.

Suppose that you own a small clothing business in the mall. In order to help make sure that you are targeting your advertising to the right audience, you decide that you need to find out how old your typical customers are. After all, if your customers are mostly under 21 years of age, you don't want to advertise in *Sunset Magazine* (a magazine that your grandparents are probably very familiar with). On the other hand, if most of your customers are over 30, you probably don't want to advertise in *Skateboard Death Magazine*.

So, how are you going to find out how old your typical customers are? One way would be to simply ask every customer who buys something what their age is and write their age down on a sheet of paper.

Why don't you want to ask *everyone* who comes into your store their age? Remember, you are interested in your actual customers, not just people who come in and browse around. After all, you are a small business and you don't have a lot of money to waste on advertising. You want to know about the people who actually buy stuff.

Anyway, suppose that you asked 100 customers their ages and you got the information below (See Table 5-1).

TABLE 5-1
Ages of 100 Customers Who Purchased Something in Your Store

18	27	22	17	19	22	17	16	22	35
35	20	15	31	17	18	18	22	28	22
21	48	34	22	17	38	44	15	17	41
31	41	37	27	21	22	38	35	65	21
45	22	22	18	26	22	18	42	22	33
12	15	20	37	30	39	24	17	13	26
16	18	19	15	25	19	22	39	20	19
28	21	31	22	20	24	31	44	22	25
13	35	24	19	16	29	17	22	22	48
65	31	22	24	20	22	23	19	41	22

Just by looking at this table of numbers, what could you conclude? If you are like me, you would probably say, “Not a lot”. It looks like your customers come from a pretty wide variety of ages.

These kinds of data are called “raw numbers” or “raw data”. Why are they “raw”? It is because you have not mathematically manipulated them in any way. You have not transformed them by multiplying, dividing, subtracting, adding them or performing any other mathematical manipulations on them. Anytime you get away from the actual numbers and summarize them in some way, you are no longer dealing with “raw data” but you have transformed or summarized these scores in some manner.

Raw Score

An individual’s score on a variable that has not been changed in any way. It has not been mathematically altered by adding, subtracting, multiplying or dividing it.

As is probably clear to you, the main problem with the numbers that we collected in Table 5-1 is that there are so many of them you can’t really see any trends or get any information to help you decide how you should advertise your store.

What should you do? You have several options. The first and easiest thing you could do is simply arrange these data into some kind of order, say from the youngest person to the oldest person. Look at Table 5-2, below, where I do just that.

TABLE 5-2
Customer Ages Ordered From Youngest to Oldest

12	17	18	20	22	22	24	29	35	41
13	17	18	20	22	22	24	30	35	41
13	17	18	20	22	22	25	31	35	42
15	17	19	20	22	22	25	31	37	44
15	17	19	21	22	22	26	31	37	44
15	17	19	21	22	22	26	31	38	45
15	17	19	21	22	22	27	31	38	48
16	18	19	21	22	23	27	33	39	48
16	18	19	22	22	24	28	34	39	65
16	18	20	22	22	24	28	35	41	65

Table 5-2 should be much clearer than what you saw in Table 5-1. Do you agree? Just look at it! Right away you can see that the vast majority of your customers are less than 40 years of age. As a matter of fact, only 11 of your customers were over 40. You

even had a couple of 65 years-olds who may have been buying something for their kids or grandkids.

You could potentially stop right here because it seems pretty obvious that your target customer is a young person. You could use this information to place advertisements in magazines or on radio stations that cater to people who are similar in age to your typical customer.

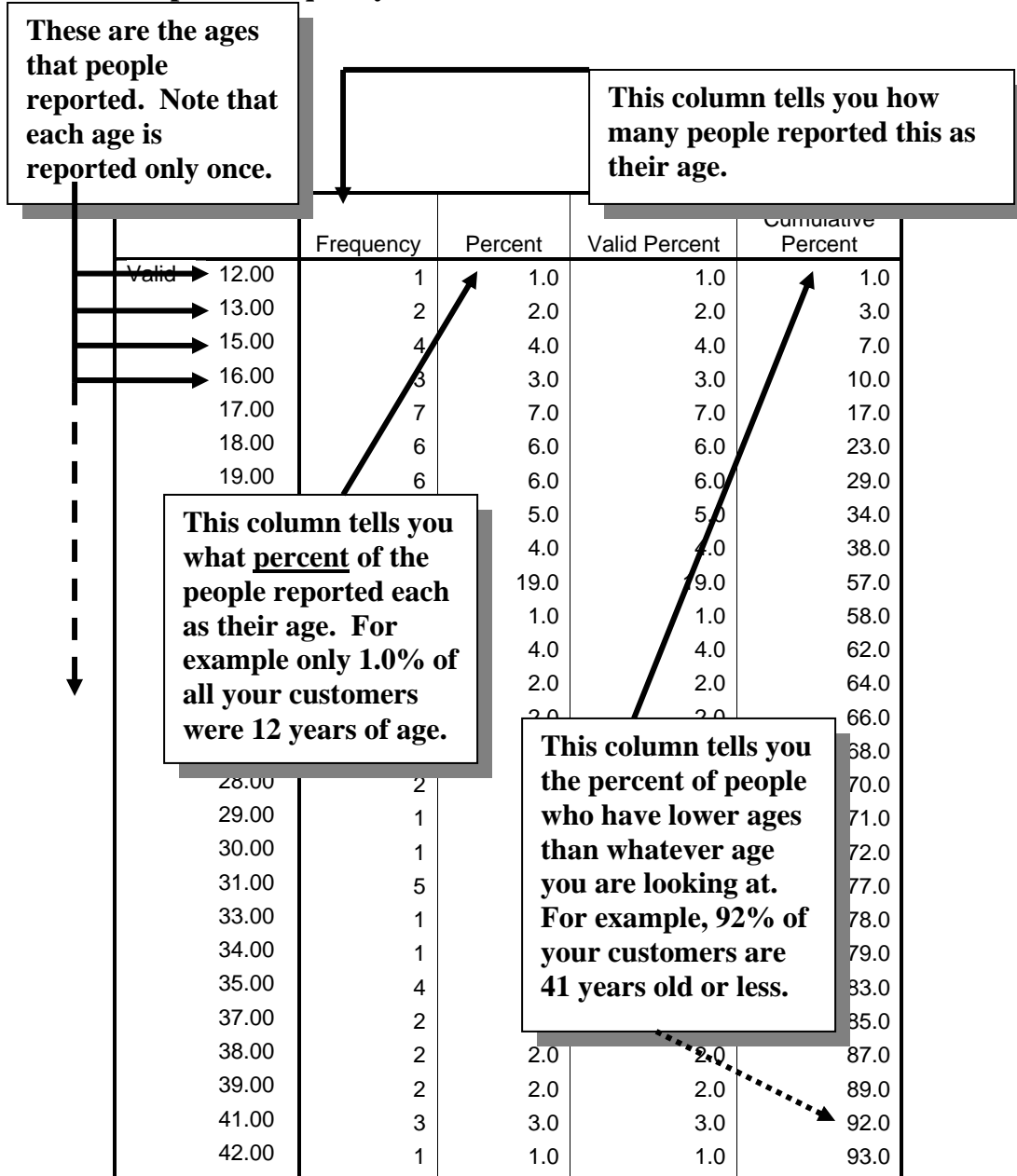
I however would take things one step farther. There is a special kind of table called a “Frequency Table” that summarizes the data in a way that makes it even easier to understand. Look at Table 5-3, below. It presents the same data as seen in Tables 5-1 and 5-2 but this time the information is presented in a Frequency Table.

TABLE 5-3
Frequency Table of Ages For Our 100 Customers

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	12.00	1	1.0	1.0	1.0
	13.00	2	2.0	2.0	3.0
	15.00	4	4.0	4.0	7.0
	16.00	3	3.0	3.0	10.0
	17.00	7	7.0	7.0	17.0
	18.00	6	6.0	6.0	23.0
	19.00	6	6.0	6.0	29.0
	20.00	5	5.0	5.0	34.0
	21.00	4	4.0	4.0	38.0
	22.00	19	19.0	19.0	57.0
	23.00	1	1.0	1.0	58.0
	24.00	4	4.0	4.0	62.0
	25.00	2	2.0	2.0	64.0
	26.00	2	2.0	2.0	66.0
	27.00	2	2.0	2.0	68.0
	28.00	2	2.0	2.0	70.0
	29.00	1	1.0	1.0	71.0
	30.00	1	1.0	1.0	72.0
	31.00	5	5.0	5.0	77.0
	33.00	1	1.0	1.0	78.0
	34.00	1	1.0	1.0	79.0
	35.00	4	4.0	4.0	83.0
	37.00	2	2.0	2.0	85.0
	38.00	2	2.0	2.0	87.0
	39.00	2	2.0	2.0	89.0
	41.00	3	3.0	3.0	92.0
	42.00	1	1.0	1.0	93.0
	44.00	2	2.0	2.0	95.0
	45.00	1	1.0	1.0	96.0
	48.00	2	2.0	2.0	98.0
	Total	100	100.0	100.0	

At first glance, table 5-3 may seem harder to understand than Table 5-2. It takes a little getting used to, but believe it or not, this thing can be really helpful. Look at Figure 5-1, below where I break the table down and guide you through how to interpret it.

FIGURE 5-1
How to Interpret a Frequency Table



Okay, let's talk about this a little to make sure you get it. Look at back at Table 5-3. Specifically, look at people who are 22 years of age. I am going to pull that row out of the table so we can take a closer look at it. See Table 5-4 below.

TABLE 5-4
Focusing on Customers who are 22 Years of Age (Extracted From Table 5-3)

Valid	Frequency	Percent	Valid Percent	Cumulative Percent
22	19	19	19	57

Table 5-4 shows the information from the frequency table for people who are 22 years of age. What does this tell you?

- First, by looking in the “Frequency” column” you can see that 19 people gave 22 as their age.
- Second, by looking under the “Percent” column, you can tell that 19% of all the people gave 22 as their age.
- Finally, by looking under the “Cumulative Percent” column you can tell that 57% of your customers are 22 years of age or younger.

That’s all there is too it. A frequency table is useful because it takes up less room, it gives you a count of how many people were represented by each age and it even gives you the percent of people for each age.

Frequency Table

A table into which data has been entered in such a manner that the data are summarized to show the frequency (a count of how many times each number appears) as well as the percentage of time each number appears.

Graphical Statistical Tools

It has been said that a picture is worth a thousand words. Why is this true? Because a picture can take a lot of complex information and present it in a way that makes sense to people—even “technically challenged” people.

This chapter will present you with a quick overview of some of the most common graphical tools for summarizing data as well as what kinds of data you should use with each one.

Specifically, you will learn about:

- Pie Charts
- Bar Charts
- Histograms
- Scatter Plots

Pie Charts

Most people are acquainted with pie charts. Maybe it has something to do with the fact that it sounds like they are related to food.

Remember those holidays when there was a yummy pie for desert? Perhaps you were a child and as you sat there looking at it you knew that someone was going to cut the pie into slices. Speaking for myself, I always hoped that somehow I would get a big piece. Perhaps it is true that “bigger is better!”

A pie chart works in the same way. For example, sticking with our customer example that we have been working with, suppose that in addition to asking their age, we also asked each customer their ethnicity.

A pie chart allows you to take a variable, like Ethnicity, and cut it up into slices with the size of each slice representing how many people reported belonging to each ethnic group.

For example, suppose that for our 100 customers, we learned that:

- 25 reported being “White”
- 15 reported being “Hispanic”
- 20 reported being “African American”
- 12 reported being “Chinese”
- 10 reported being “Native American”
- 13 reported being “Filipino”
- 5 reported their ethnicity as “Other”

Now, it would be perfectly acceptable to use a table to present these data (just like in Table 5-3). You could however, use a pie chart such as is seen in Figure 5-2, below. Take a close look at it. Notice how much information there is in this one simple picture.

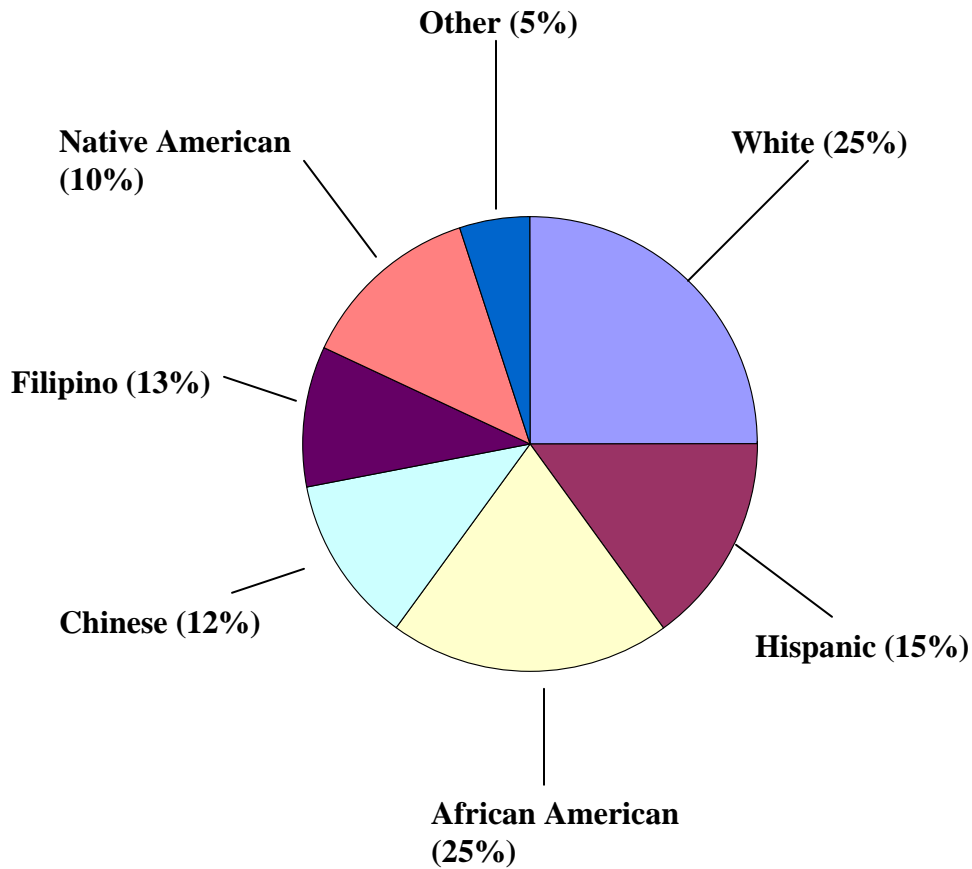
Notice that the pie chart in Figure 5-2 helps you to easily see who makes up the largest proportion of your customers. At a glance you can easily see that Whites and African Americans make up the greatest proportion of your customers (a total of 45%) followed by Hispanics who account for 15% of our customers.

Some Tips for Using Pie Charts

1. Pie Charts work best with *Nominal* and *Ordinal* Level Variables (e.g., gender, ethnicity, highest degree earned in school, eye color, etc.)
2. Pie charts work best when your variable has less than 10 to 12 categories. (If you have too many categories, the pie chart becomes so cluttered that it becomes hard to understand).
3. The goal is to make data understandable. Try a pie chart but remember, if it doesn't make interpretation of your data easier, try something else.

FIGURE 5-2

Pie Chart of Ethnicity of Our Customers

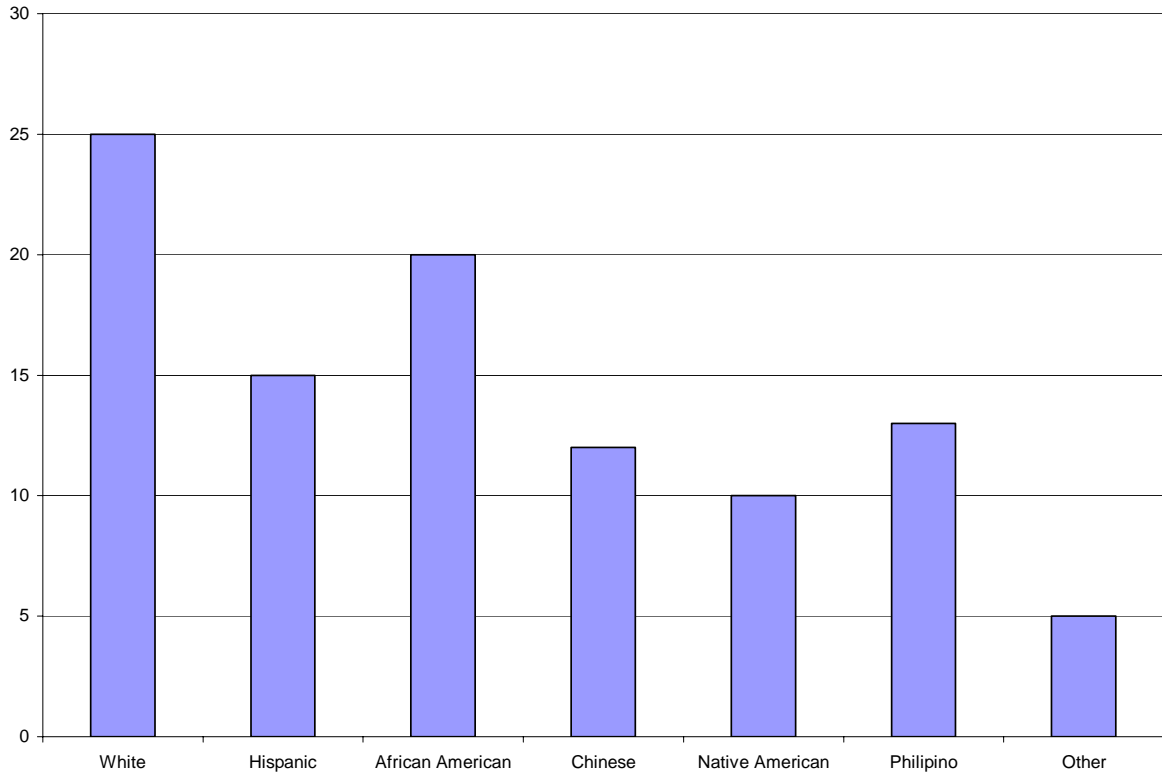


Bar Charts

Bar charts are another kind of useful graph. Take a look at Figure 5-3 for an example of a bar chart that is based on the same data as we used in the pie chart in Figure 5-2.

The bar chart is so easy to understand that really nothing much needs to be said about it. Notice that each bar represents one of our ethnic categories. If you look along the left side of the bar chart you can see the numbers that represent the percent that each bar represents. For example, if you look at the bar for “White” you can see that it accounts for 25% of the customers.

FIGURE 5-3
Bar Chart of Ethnicity of Our Customers



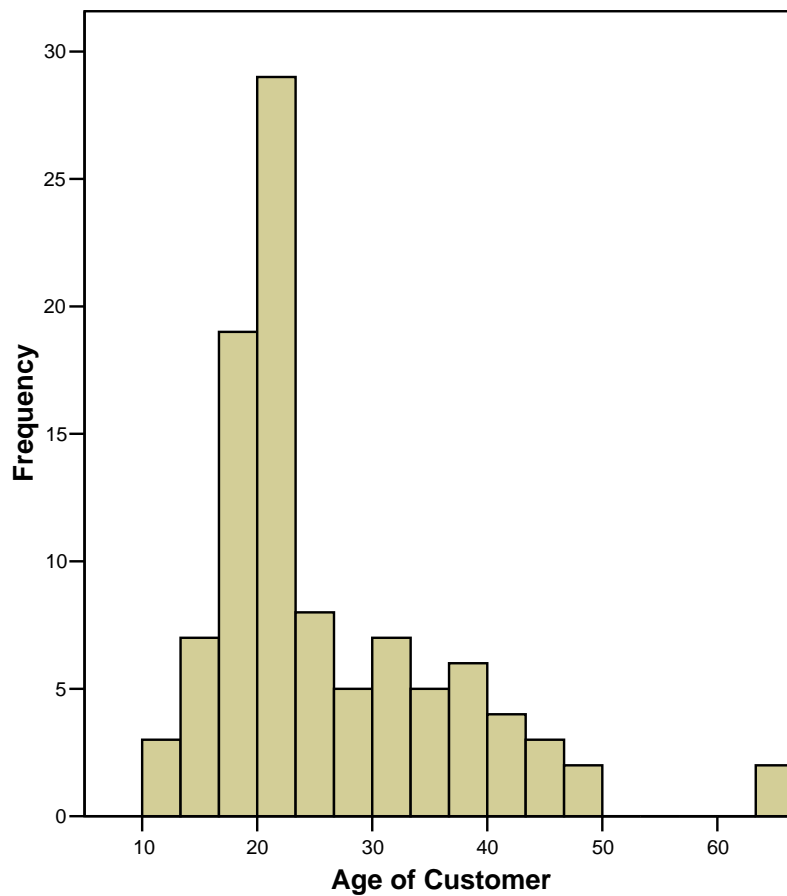
Some Tips for Using Bar Charts

1. Bar Charts work best with *Nominal* and *Ordinal* Level Variables (e.g., gender, ethnicity, highest degree earned in school, eye color, etc.)
2. Bar charts work best when your variable has less than 10 to 12 categories. (If you have too many categories, the bar chart becomes so cluttered that it becomes hard to understand).
3. The goal is to make data understandable. Try a bar chart but remember, if it doesn't make interpretation of your data easier, try something else.

Histograms

Histograms are quite a bit different from the pie charts and bar charts we just covered. For one thing, where pie and bar charts are for use with *Nominal* and *Ordinal* data, **Histograms are for use with interval or ratio scales** of measurement. Remember, Age is measured on a Ratio scale. See Figure 5-4 for an example of a histogram. This time, I am going to use the age data we used way back in the beginning of this chapter in Table 5-1. Take a look at Figure 5-4, below.

FIGURE 5-4
Histogram of Ages For Our Customers



Notice that a histogram looks a lot like a bar chart but with quite a few more bars. Also notice that the bottom of the chart (the X axis) is a Ratio scale rather than a Nominal or Ordinal scale. The scale along the Y axis is a count of how many people were of each particular age. Look at the bar that is around 22.

This histogram tells you that approximately 28 people said that they were 22 years of age. By looking at the chart you can quickly get an idea about how many of

your customers are less than, say, 30. If you wanted to take the time, you could even add the total number represented by each bar and get the actual total number of people who were less than 30 years old.

IMPORTANT POINT!

Look at the histogram as a whole. Notice that the scores are spread out or distributed from lowest to highest. When scores are spread out like this, it is called a “Distribution”. In other words, the histogram in Figure 5-4 shows the distribution of ages for our customers.

This concept of a “distribution of scores” is extremely important. Any time you arrange scores from lowest to highest you are showing the distribution of those scores. A histogram is one way of presenting this information in a graph format so that you can easily see what shape the distribution is taking.

Please make sure that you understand this concept. If you do not, please talk to your instructor to clear up your confusion.

Some Tips for Histograms

1. Histograms are not for use with Nominal or Ordinal variables.
2. Histograms are designed to be used with a single interval or ratio-level variable.
3. As with all graphical tools for use with data, the goal is to make data understandable. Try a histogram, if it doesn't make interpretation of your data easier, try something else. Perhaps simply a frequency table is what you should use.

Before we move on from the histogram, I want to digress for just a bit into another topic. After this, I will talk about Scatterplots.

The Shape of Things to Come

Every distribution has three major characteristics that you must learn to recognize and understand. These characteristics are:

- Shape
- Central Tendency
- Variability

In order to fully understand a distribution of scores, you need to take a close look at each of these characteristics. This is very important because if a distribution of scores does not have a certain shape there are limits on which statistical tests you can use on your data. Each of these characteristics of distributions is discussed below.

Characteristic #1: The Shape of a Distribution

The first characteristic that you need to look at is the shape of a distribution. The best way to get an idea about how a distribution is shaped is to use a Histogram. Just by looking at the Histogram you should be able to determine how the distribution is shaped.

There are two ways that a distribution's shape can differ. One way is called "Skew" and the other way is called "Kurtosis". It will be much easier for you if I just show you what Skew and Kurtosis are rather than just trying to explain it. Let's look at Skew first.

What is a Skewed Distribution?

A skewed distribution is what you get when one of the "tails" or ends of the distribution is pulled farther to the left or right than the other tail is. Look at Figure 5-5, below.

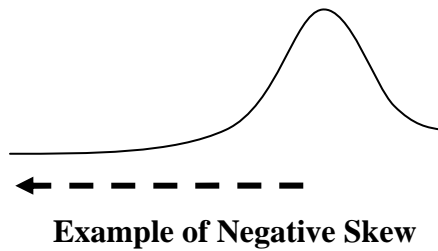
If you look back at Figure 5-4 where I show you a histogram of customer ages, can you see that the distribution is skewed? Which tail is being pulled off to the side? What type of skew is it showing?

In Figure 5-4 the positive tail is being pulled farther to the right because of some scores that are extreme compared to the rest of the sample. As a result, the distribution is "positively skewed" because it is pulled in the positive direction.

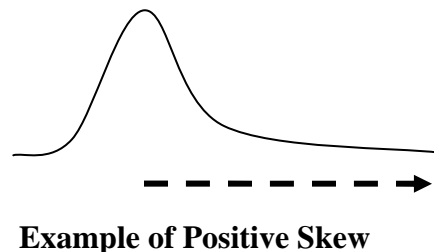
It is extremely important to find out if your distribution is skewed before you try to do any statistical analysis with your data. This is because if a distribution is significantly skewed to the left (negative) or to the right (positive) there are certain statistical tests that you will learn that should not be used with the data because you will get inaccurate results.

If, on the other hand, your distribution of scores is not skewed and is a normal distribution, then a whole world of statistical tests is available to help you answer a wide variety of research questions. You will learn much more about these tests and what they can do for you later. For now, it is critical that you understand what a skewed distribution is and what it means (for example, a distribution is skewed when one of the tails is pulled way off to the side by some extreme scores). Also you need to understand what a normal distribution is and why it is important. I will explain more about the normal distribution in Chapters 5, 6 and 7.

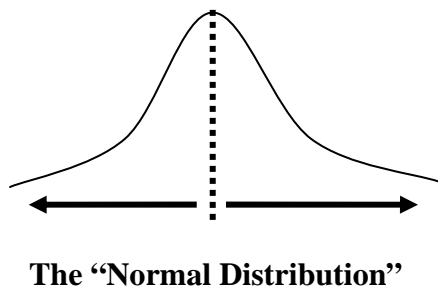
FIGURE 5-5
Examples to Illustrate How Distributions Can be Skewed



Notice that a distribution is **negatively skewed** when the left-hand **tail is pulled to the left**. This means that you have a few abnormally low scores in your distribution.



A distribution is **positively skewed** when the right-hand **tail is pulled to the right**. This means that you have a few abnormally high scores in your distribution.



This is a good thing!
 When you have a distribution that has a single peak with half the scores above it and half below it, your distribution is not skewed. This kind of distribution has a special name—**The Normal Distribution**. (or Bell Curve)

The Normal Distribution

The most powerful statistical tests, such as those designed to find out if two or more variables are related (for example, smoking and lung cancer) or if there are differences between variables (for example, did people who took Prozac experience less depression than those who took a placebo) require that the data be normally distributed. In other words, each variable must have a normal distribution.

What is Kurtosis?

Remember that there are two ways that a distribution’s shape can differ from other distributions. One of the ways distributions can differ is related to whether or not

they are skewed. By now, you should have a good basic understanding of what is meant by a skewed distribution.

The second way a distribution's shape can change is related to what is called "Kurtosis". The term "Kurtosis" refers to how a distribution's height compares to its width. Let me show you a picture that will help you understand the concept of Kurtosis. See Figure 5-6, below.

Looking at Figure 5-6 you can see that there are three basic types of Kurtosis. These are Leptokurtic, Messokurtic, and Platykurtic. As you will learn in the chapter 6, kurtosis is important to understanding how much variability there is in your data.

At this point, you are probably saying, "Yeah, so what!?" Let me assure you that this is important. Once we get past chapter 6, you will begin to see how all of this ties together and how it really does create the foundation you need to understand the rest of statistics.

Let me tell you how I remember the names of the three types of kurtosis and what they mean:

Leptokurtic: When I think of the name "Leptokurtic", I think of "Leaping" which is another word for jumping. In a Leptokurtic Distribution, the center is leaping high up into the air and as it stretches up, it gets really narrow.

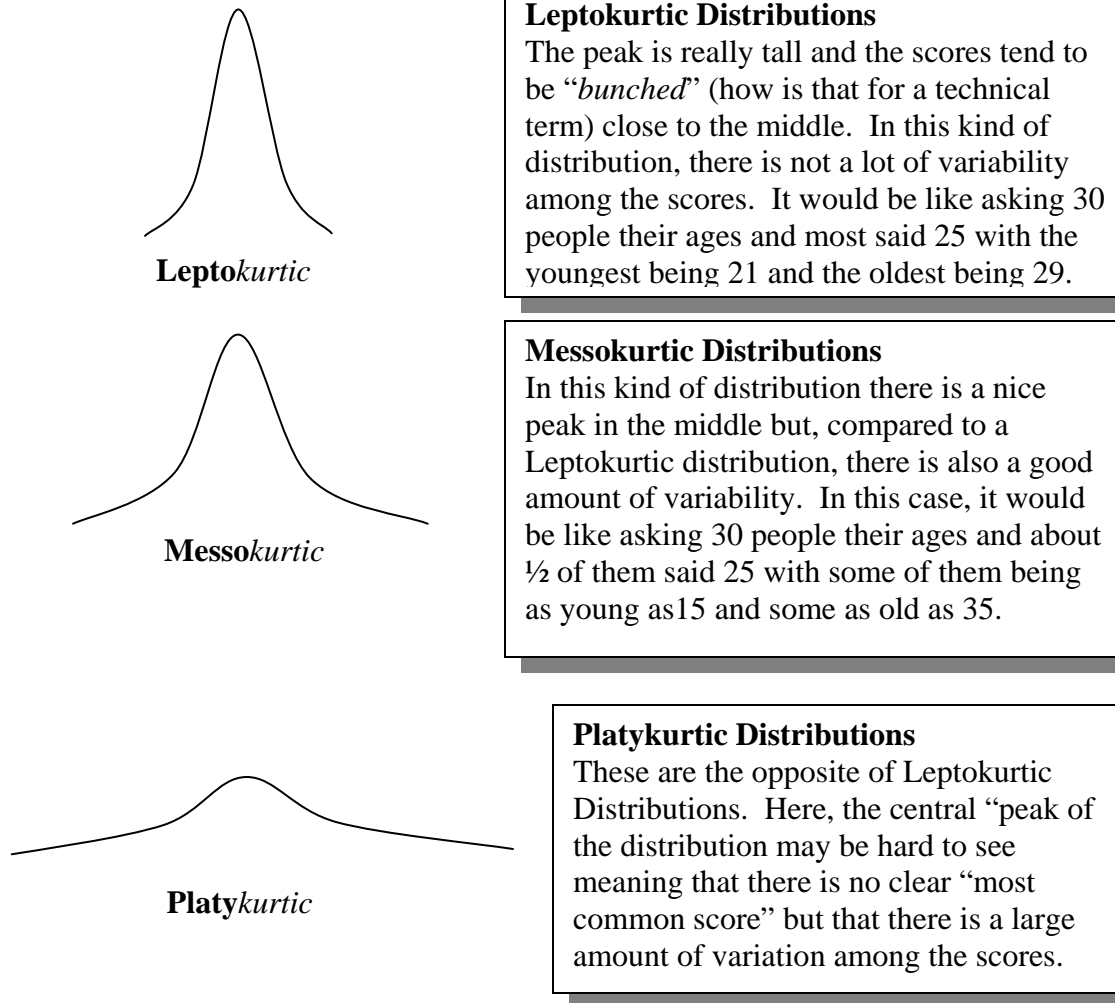
Platykurtic: I remember what a Platykurtic Distribution is by thinking about a "Plate" (You know, "Platy" sounds kind of similar to "Plate"). A plate is flat and wide (a good thing, too, so you can get lots of pie onto it!). So, a Platykurtic distribution is kind of flat and spread out without a clearly distinct "peak" in the center.

Messokurtic: Okay, I have not been able to think of an easy memory device to help with Messokurtic so I try to remember that, since there are only three kinds, once I have got Platykurtic and Leptokurtic down, Messokurtic must be what is left over! One of my former students suggested that "Messo" kind of sounds like "Middle" and so he used this as a hint to remind himself that Messokurtic Distributions are somewhere between being flat and wide on the one hand and being narrow and tall on the other hand.

Kurtosis

Kurtosis is a measure of how "peaked" or "spread out" a distribution is. It ranges from Platykurtic, or very flat and spread out with lots of variation to Leptokurtic, or very narrow but tall with little variation.

FIGURE 5-6
Examples Illustrating the Meaning of Kurtosis as it Relates to Distributions of Scores



Back to the Three Characteristics of Distributions

Several pages ago I told you that all distributions have three basic characteristics. These were:

- **Shape** – Which is composed of:
 - *Skew* (negative, positive and not skewed (normal))
 - *Kurtosis* (Leptokurtic, Messokurtic, Platykurtic)
- **Central Tendency**
- **Variability**

We have not yet talked at length about Central Tendency or Variability. I will write only briefly about these two characteristics because we will cover each one in detail in Chapter 5 (Central Tendency) and Chapter 6 (Variability).

You will need to look back at Figure 5-4 for just a minute. I recommend that you put your finger or a marker in the book so you can quickly flip back to it as you read the next two paragraphs.

- **Central Tendency:** Looking at Figure 5-4, notice that there is a “peak” in the distribution somewhere around 22 years of age. In other words, more people are somewhere around 22 years of age than any other age. The fact that this peak exists means that there is a “tendency” for most of our customers to be somewhere around 22 years old. This “tendency” is called “Central Tendency”.
- **Variability:** The amount that scores spread out around the “central tendency” is called “variability”. The more spread out scores are, the more variability there is. If scores are not spread out much, then there is not much variability. Notice that in Figure 5-4 you can see that some of our customers are as young as 10 while a few were over 60 years of age. So, while most of our customers are around 22, there is a certain amount of variability in our customer ages with some being much younger and some much older.

Just like with the term “distribution”, you are going to read about “central tendency” and “variability” almost constantly in this book. If you are taking a statistics class, you will hear about them almost daily.

You must understand the concepts of central tendency and variability. If these concepts are unclear to you, please speak to your instructor. If you are not comfortable asking questions in class, see him or her during an office hour. If you fail to understand each of these concepts as you work with them, you will be lost when you get to the more advanced concepts.

Meanwhile...back at graphical statistical tools!

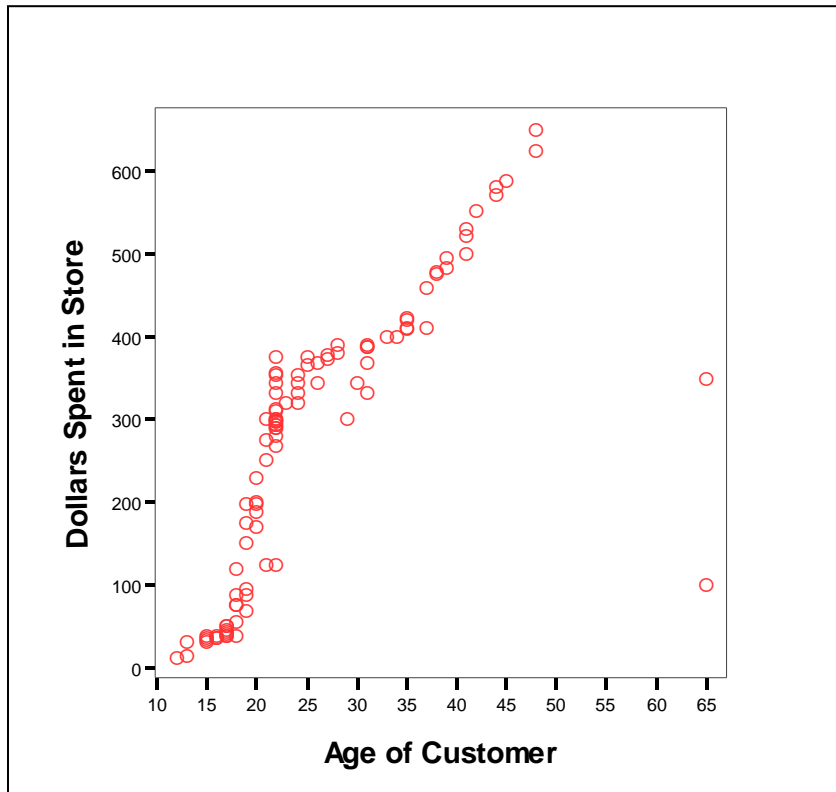
Scatter Plots

The Scatterplot is another graphical tool for presenting data. It is the most different of the graphical tools that I am going to show you because it actually allows you to evaluate **two variables at once**. A scatterplot allows you to see how two variables are related to each other.

Suppose, for a moment, that we not only want to know how old our typical customers are, but we are interested in how much money they spend in our store. Why would we want to do this? Well, suppose that while most of our customers are young, they only spend a small amount in our store while older customers, even though there are fewer of them, tend to spend tons of cash. If this is the case we want more older customers, right?

So, look at Figure 5-7, below, for an example of a Scatterplot that compares the ages of our customers by how much they spend.

FIGURE 5-7
Scatterplot of How Much Each Customer Spends According to Their Age



Think about what you see in Figure 5-7. What does it seem to tell you?

Well, for one thing it looks like as a person's age goes up, the amount of money they tend to spend in your store goes up also. Whereas most of your customers are young (around 22 years old) it looks like they only spend around \$50 to \$100.

On the other hand, while people over 40 years of age make up only a small portion of your customers, it looks like they spend, on average, over \$400 when they come in. So, you tell me, how you would spend your advertising dollars.

As long as all we were looking at is "who is our most common customer?" it looked like we needed to ignore older people because they represented such a small percentage of our total customers. Now, however, when we take into account the additional information of how much customers spend, it looks like we can't afford to simply ignore older customers.

What this chart is telling you is that **"there is a relationship** between customer age and the amount of money a customer spends. Specifically, the amount spent goes up as the age of the customer goes up." **Scatter Plots tell you whether relationships exist between two variables.**

I am not going to try and explain this now (that is what a lot of this class will teach you to do). Rather, I want you to realize that when you get more data—and use it wisely—it helps you see that things are often more complicated than you thought. Sometimes you will find that your preconceptions were completely wrong!

Additionally, when you take this additional information into account, you are much more likely to make correct—and profitable—decisions.

Some Tips for Scatter Plots

1. Scatter Plots are **not** for use with *Nominal* or *Ordinal* variables.
2. Scatter Plots require you to use two (2) *interval* or *ratio-level* variables.
3. As with all graphical tools for use with data, the goal is to make data understandable. Try a scatterplot, if it doesn't make interpretation of your data easier, try something else.

Future chapters will focus on the other two kinds of descriptive statistics that I opened this chapter with: Measures of Central Tendency and Measures of Variability. These two types of statistics are so important that you we will deal with them in-dept in class.

Terms to Learn

You should be able to define the following terms based on what you have learned in this chapter.

Bar Chart

Central Tendency

Distribution

Frequency

Histogram

Kurtosis

Leptokurtic Distribution

Messokurtic Distribution

Negative Skew

Normal Distribution

Pie Chart

Platykurtic Distribution

Positive Skew

Scatterplot

Skew (Skewness)

Variability

HOMEWORK

Instructions: Answer the questions below on a separate sheet of paper. Don't forget to write your name on your paper as well as the hour of the class that you attend. Also, don't forget to clearly label your answers.

1. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

10, 10, 11, 11, 11, 12, 13, 13, 14, 14, 14, 15, 15, 16

2. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

2, 3, 3, 3, 5, 5, 6, 7, 9, 9, 9, 9, 15, 15

3. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

122, 135, 136, 136, 136, 140, 140, 147, 149, 149, 149, 151, 155, 161

4. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

19, 29, 29, 37, 37, 37, 37, 39, 41, 45, 45

5. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

100, 157, 165, 165, 192, 192, 231, 231, 235, 300, 300, 300

6. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

6.1, 6.1, 6.1, 6.1, 6.9, 6.9, 6.9, 7.5, 7.5, 8.3, 8.3, 9.0

7. Create a frequency table that shows frequency, percent, and cumulative percent for the following data set.

.375, .380, .380, .451, .555, .551, .551, .551, .619, .619

What **graph** (I am not talking about frequency tables) might be appropriate for each of the following variables?

8. Age _____
9. Gender _____
10. Ethnicity _____
11. Highest Year of School Completed? _____
12. College Major _____
13. Annual Salary _____
14. Grade Point Average _____
15. Marital Status _____
16. Population of California's 10 largest cities _____
17. Favorite Color _____
18. Shoe Size _____
19. Hours worked per week _____
20. Favorite kind of pet _____

Draw a picture to illustrate each of the following:

21. A slightly negatively skewed distribution:
22. A very positively skewed distribution:
23. A Leptokurtic distribution:
24. A Platykurtic distribution:
25. A "normal" distribution:
26. A Messokurtic Distribution:
27. A very negatively skewed distribution:
28. A slightly positively skewed distribution:
29. Using the Internet, conduct a search to find out what a "Box and Whiskers Plot" is and how it is used. Draw an example of a Box and Whiskers Plot and explain what each part of the graph tells you.