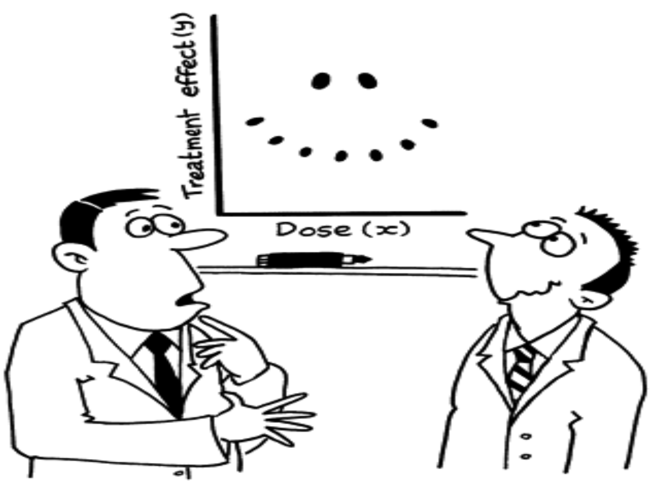
Analyzing Quantitative Data



"Oh my! The quantitative data doesn't support our hypothesis. But for some reason I'm very happy with the results."

CHAPTER

18

CHAPTER OUTLINE

After quantitative data are collected as illustrated in the previous chapter, they need to be analyzed—the purpose of this chapter. To be honest, a thorough understanding of quantitative statistical methods is far beyond the scope of this book and would necessitate more in-depth study, through taking one or more statistics courses. Instead, we briefly describe a select group of basic statistical analytical methods that are used frequently in many quantitative and qualitative social work research studies. Our emphasis is not on providing and calculating formulas, but rather on helping the reader to understand the underlying rationale for their use.

We present two basic groups of statistical procedures. The first group is called *descriptive statistics*; these procedures simply describe and summarize one or more variables for a sample or population. They provide information about only the group included in the study.

The second group is called *inferential statistics*; these procedures determine if we can generalize findings derived from a sample to the population from which the sample was drawn. In other words, knowing what we know about a particular sample, can we infer that the rest of the population is similar to the sample that we have studied?

Before you read farther, we encourage you to reread Chapter 11 on levels of measurement. It's extremely important for you to have a sound understanding of the four levels of measurement before you dive into statistics.

Entering Data into Computers

The use of computers has revolutionized the analysis of quantitative and qualitative data. Where previous generations of researchers had to rely on hand-cranked adding machines to calculate every small step in a data analysis, today we can enter raw scores into a personal computer and with few complications direct the computer program to execute just about any statistical test imaginable. Seconds later, the results are available.

While the process is truly miraculous, the risk is that, even though we have conducted the correct statistical analysis, we may not understand what the results mean, a factor that will almost certainly affect how we interpret the data. We can code data from all four levels of measurement into a computer for any given data analysis (see chapter 9). The coding of nominal data is perhaps the most complex because we have to create categories that correspond to certain possible responses for a variable.

One type of nominal-level data that is often gathered from research participants is *place of birth*. If, for the purposes of our study, we were interested in whether our research participants were born in either Canada or the United States, we would assign only three categories to *place of birth*:

1 Canada
2 United States
9 Other

The *other* category (9) appears routinely at the end of lists of categories and acts as a catchall to cover any category that may have been omitted.

When entering nominal-level data into a computer, because we do not want to enter *Canada* every time the response on the questionnaire is *Canada*, we may assign it the code number 1 so that all we have to enter is 1. Similarly, the *United States* may be assigned the number 2, and "other" may be assigned the number 9.

These numbers have no mathematical meaning. We are not saying that Canada is better than the United States because it comes first, or that the United States is twice as good as Canada because the number assigned to it is twice as high. We are merely using numbers as a shorthand device to record *qualitative* differences: differences in *kind*, not in *amount*.

Most coding for ordinal-, interval-, and ratio-level data is simply a matter of entering the final score, or number, from the measuring instrument that was used to measure the variable directly into the computer. If a person scored a 45 on a standardized measuring instrument, for example, the number 45 would be entered into the computer. Although almost all data entered into computers are in the form of numbers, we need to know at what level of measurement the data exist so that we can choose the appropriate statistic(s) to describe and compare the variables.

As you know from Chapter 11 about the four different measurement levels, let's turn to the first group of statistics that can be helpful for the analyses of data—descriptive statistics. This would is an excellent time to review Chapter 11 as this chapter assumes you know Chapter 11's content inside and out—and don't forget backwards.

Descriptive Statistics

Descriptive statistics are commonly used in most quantitative and qualitative research studies. They describe and summarize a variable(s) of interest and portray how that particular variable is distributed in the sample, or population. Before looking at descriptive statistics, however, let's examine a social work research example that will be used throughout this chapter.

Thea Black is a social worker who works in a treatment foster care program. Her program focuses on children who have behavioral problems who are placed with "treatment" foster care parents. These parents are supposed to have parenting skills that will help them provide for the children's special needs.

Thus, Thea's program also teaches parenting skills to these treatment foster care parents. She assumes that newly recruited foster parents are not likely to know much about parenting children who have special needs. Therefore, she believes that they would benefit from a training program that teaches these skills in order to help them deal effectively with the special needs of these children who will soon be living with them.

Thea hopes that her parenting skills training program will increase the knowledge about parental management skills for the parents who attend. She assumes that with such training the foster parents will be in a better position to support and provide clear limits for their foster children.

Descriptive statistics describe and summarize a variable and portray how that particular variable is distributed in the sample, or population.

After offering the training program for several months, Thea became curious about whether the foster care providers who attended the program were, indeed, lacking in knowledge of parental management skills as she first believed (her tentative hypothesis).

She was fortunate to find a valid and reliable standardized instrument that measures the knowledge of such parenting skills, the *Parenting Skills Scale* (*PSS*). Thea decided to find out for herself how much the newly recruited parents knew about parenting skills—clearly a descriptive research question.

At the beginning of one of her training sessions (before they were exposed to her skills training program), she handed out the *PSS*, asking the twenty individuals in attendance to complete it and also to include data about their gender, years of education, and whether they had ever participated in a parenting skills training program before. All of these variables could be potentially extraneous ones that might influence the level of knowledge of parenting skills of the twenty participants.

For each foster care parent, Thea calculated the *PSS* score, called a *raw score* because it has not been sorted or analyzed in any way. The total score possible on the *PSS* is 100, with higher scores indicating greater knowledge of parenting skills. The scores for the PSS scale, as well as the other data collected from the twenty parents, are listed in Table 18.1 on the following page.

At this point, Thea stopped to consider how she could best utilize the data that she had collected. She had data at three different levels of measurement. At the nominal level, Thea had collected data on gender (3rd column), and whether the parents had any previous parenting skills training (4th column). Each of these variables can be categorized into two responses.

The scores on the *PSS* (2nd column) are ordinal because, although the data are sequenced from highest to lowest, the differences between units cannot be placed on an equally spaced continuum. Nevertheless, many measures in the social sciences are treated as if they are at an interval level, even though equal distances between scale points cannot be proved. This assumption is important because it allows for the use of inferential statistics on such data.

Finally, the data on years of formal education (5th column) that were collected by Thea are clearly at the ratio level of measurement because there are equally distributed points and the scale has an absolute zero.

In sum, it seemed to Thea that the data could be used in at least two ways. First, the data collected about each variable could be described to provide a picture of the characteristics of the group of foster care parents. This would call for descriptive statistics. Second, she might look for relationships between some of the variables about which she had collected data, procedures that would utilize inferential statistics. For now let us begin by looking at how the first type of descriptive statistic can be used with Thea's data set.

On a general level, descriptive statistics can be divided into three subgroups of statistics (Weinbach & Grinnell, 2017):

1. Frequency distributions
2. Measures of central tendency
3. Measures of variability

Table 18.1

Data Collection for Four Variables from Foster Care Providers

Case Number	*PSS* Score	Gender	Previous Training	Years of Education
01	95	Male	No	12
02	93	Female	Yes	12
03	93	Male	No	08
04	93	Female	No	12
05	90	Male	Yes	12
06	90	Female	No	12
07	84	Male	No	14
08	84	Female	No	18
09	82	Male	No	10
10	82	Female	No	12
11	80	Male	No	12
12	80	Female	No	11
13	79	Male	No	12
14	79	Female	Yes	12
15	79	Female	No	16
16	79	Male	No	12
17	79	Female	No	11
18	72	Female	No	14
19	71	Male	No	15
20	55	Female	Yes	12

FREQUENCY DISTRIBUTIONS

One of the simplest procedures that Thea can employ is to develop a frequency distribution of her data. Constructing a frequency distribution involves counting the occurrences of each value, or category, of the variable and ordering them in some fashion. This *absolute* or *simple frequency distribution* allows us to see quickly how certain values of a variable are distributed in our sample.

The *mode,* or the most commonly occurring score, can be easily spotted in a simple frequency distribution (see table 18.2 on the right). In this example, the mode is 79, a score obtained by five parents on the *PSS* scale. The highest and the lowest scores are also quickly identifiable. The top score was 95, while the foster care parent who performed the least well on the *PSS* scored 55.

Table 18.2

Frequency Distribution (from table 18.1 above)

PSS Score	Absolute Frequency
95	1
93	3
90	2
84	2
82	2
80	2
79	5
72	1
71	1
55	1

Table 18.3

Cumulative Frequency and Percentage Distribution of Parental Skill Scores (from table 18.1)

PSS Score	Absolute	Cumulative	Percentage Distribution
95	1	1	5
93	3	4	15
90	2	6	10
84	2	8	10
82	2	10	10
80	2	12	10
79	5	17	25
72	1	18	5
71	1	19	5
55	1	20	5
Totals . . .	20		100

There are several other ways to present frequency data. A commonly used method that can be easily integrated into a simple frequency distribution table is the *cumulative frequency distribution*, shown in Table 18.3 above.

In Thea's data set, the highest *PSS* score, 95, was obtained by only one individual. The group of individuals who scored 93 or above on the *PSS* measure includes four foster care parents. If we want to know how many scored 80 or above, if we look at the number across from 80 in the cumulative frequency column, we can quickly see that twelve of the parents scored 80 or better.

Other tables use percentages rather than frequencies, sometimes referred to as *percentage distributions*, shown in the far-right column in Table 18.3. Each of these numbers represents the percentage of participants who obtained each *PSS* value. Five individuals, for example, scored 79 on the *PSS*. Since there were a total of 20 foster care parents, five out of the twenty, or one-quarter of the total, obtained a score of 79. This corresponds to 25 percent of the participants.

Table 18.4

Grouped Frequency Distribution of Parental Skill Scores (from table 18.1)

PSS Score	Absolute	Cumulative	Absolute Percentage
90–100	6	6	30
80–89	6	12	30
70–79	7	19	35
60–69	0	19	0
50–59	1	20	5
Totals . . .	20		100

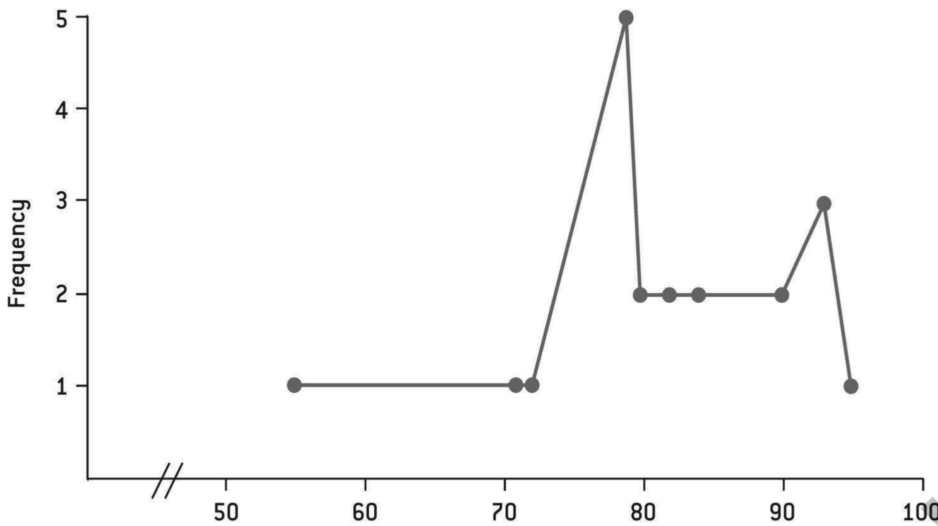


Figure 18.1

Frequency Polygon of Parental Scores (from table 18.2)

Finally, *grouped frequency distributions* are used to simplify a table by grouping the variable into equal-sized ranges, as shown in Table 18.4 on the previous page. Both absolute and cumulative frequencies and percentages can also be displayed using this format. Each is calculated in the same way that was previously described for nongrouped data, and the interpretation is identical.

Looking at the absolute frequency column, for example, we can quickly identify the fact that seven of the foster care parents scored in the 70–79 range on the *PSS.* By looking at the cumulative frequency column, we can see that twelve of twenty parents scored 80 or better on the *PSS*. Further, from the absolute percentage column, it's clear that 30 percent of the foster parents scored in the 80–89 range on the parenting skills scale.

Note that each of the other variables in Thea's data set could also be displayed in frequency distributions. Displaying years of education in a frequency distribution, for example, would provide a snapshot of how this variable is distributed in Thea's sample of foster care parents. However, with two category nominal variables, such as gender (male, female) and previous parenting skills training (yes, no), cumulative frequencies become less meaningful, and the data are better described as percentages.

Thea noted that 55 percent of the foster care parents who attended the training workshop were women (obviously the other 45 percent were men) and that 20 percent of the parents had already received some form of parenting skills training (while a further 80 percent had not been trained).

MEASURES OF CENTRAL TENDENCY

We can also display the values obtained on the *PSS* in the form of a graph. A frequency polygon is one of the simplest ways of charting frequencies. The graph in Figure 18.1 displays the data that we had previously put in Table 18.2. The *PSS* score is plotted in terms of how many of the foster care parents obtained each score. As can be seen from Figure 18.1 above, most of the scores fall between 79 and 93. The one extremely low score of 55 is also quickly noticeable in such a graph because it's so far removed from the rest of the values.

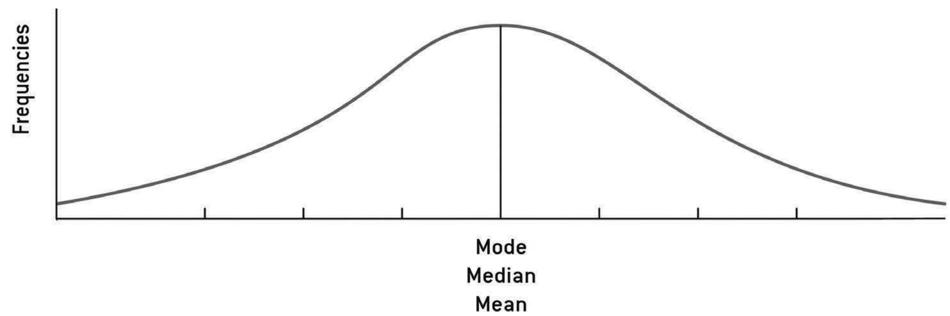


Figure 18.2

The Normal Distribution

A frequency polygon allows us to make a quick analysis of how closely the distribution fits the shape of a normal curve. A normal curve, also known as a *bell-shaped distribution* or a *normal distribution,* is a frequency polygon in which the greatest number of responses fall in the middle of the distribution and fewer scores appear at the extremes of either very high or very low scores (see figure 18.2 above).

Many variables in the social sciences are assumed to be distributed in the shape of a normal curve. Low intelligence, for example, is thought to be relatively rare as compared to the number of individuals with average intelligence. On the other end of the continuum, extremely gifted individuals are also relatively uncommon. Of course, not all variables are distributed in the shape of a normal curve. Some are such that a large number of people do very well (as Thea found in her sample of foster care parents and their parenting skill levels). Other variables, such as juggling ability, for example, would be charted showing a fairly substantial number of people performing poorly.

Frequency distributions of still other variables would show that some people do well, and some people do poorly, but not many fall in between. What is important to remember about distributions is that, although all different sorts are possible, most statistical procedures assume that there is a normal distribution of the variable in question in the population.

When looking at how variables are distributed in samples and populations it's common to use measures of *central tendency*, such as,

1. Mode
2. Median
3. Mean

which help us to identify where the typical or the average score can be found. These measures are used so often because not only do they provide a useful summary of the data, they also provide a common denominator for comparing groups to each other.

MODE

As mentioned earlier, the mode is the score, or value, that occurs the most often—the value with the highest frequency. In Thea's data set of parental skills scores the mode is 79, with five foster care parents obtaining this value. The mode is particularly useful for nominal level data. Knowing what score occurred the most often, however, provides little information about the other scores and how they are distributed in the sample or population.

Because the mode is the least precise of all the measures of central tendency, the median and the mean are better descriptors of ordinal level data and above. We now turn our attention to the second measure of central tendency, the median.

Median

The median is the score that divides a distribution into two equal parts or portions. In order to do this, we must rank-order the scores, so at least an ordinal level of measurement is required. In Thea's sample of twenty *PSS* scores, the median would be the score above which the top ten scores lie and below which the bottom ten fall. As can be seen in Table 18.2, the top ten scores finish at 82, and the bottom ten scores start at 80. In this example, the median is 81, since it falls between 82 and 80.

Mean

The mean is the most sophisticated measure of central tendency and is useful for interval or ratio levels of measurement. It's also one of the most commonly utilized statistics. A mean is calculated by summing the individual values and dividing by the total number of values. The mean of Thea's sample is 95 + 93 + 93 + 93 + 90 + 90 + ... 72 + 71 + 55/20 = 81.95. In this example, the obtained mean of 82 (we rounded off for the sake of clarity) is larger than the mode of 79 or the median of 81.

The mean is one of the previously mentioned statistical procedures that assumes that a variable will be distributed normally throughout a population. If this is not an accurate assumption, then the median might be a better descriptor. The mean is also best used with relatively large sample sizes where extreme scores (such as the lowest score of 55 in Thea's sample) have less influence.

Measures of Variability

While measures of central tendency provide valuable information about a set of scores, we are also interested in knowing how the scores scatter themselves around the center. A mean does not give a sense of how widely distributed the scores may be. This is provided by measures of variability such as the range and the standard deviation. There are many types of variability. We will only discuss two:

1. Range
2. Standard deviation

Range

The range is the distance between the minimum and the maximum score. The larger the range, the greater the amount of variation of scores in the distribution. The range is calculated by subtracting the lowest score from the highest. In Thea's sample, the range is 40 (95 − 55). The range does not assume equal interval data. It is, like the mean, sensitive to deviant values because it depends on only the two extreme scores.

We could have a group of four scores ranging from 10 to 20: 10, 14, 19, and 20, for example. The range of this sample would be 10 (20 − 10). If one additional score that was substantially different from the first set of four scores was included, this would change the range dramatically. In this example, if a fifth score of 45 was added, the range of the sample would become 35 (45 – 10), a number that would suggest quite a different picture of the variability of the scores.

Standard Deviation

The standard deviation is the most utilized indicator of dispersion. It provides a picture of how the scores distribute themselves around the mean. Used in combination with the

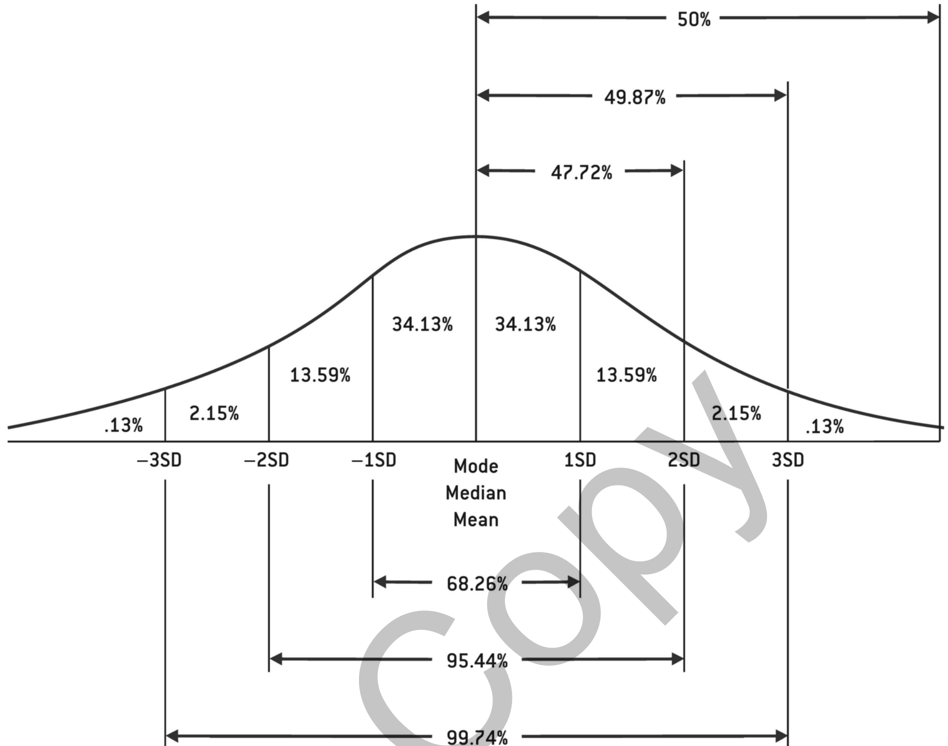


Figure 18.3

Proportions of the Normal Curve

mean, the standard deviation provides a great deal of information about the sample or population without our ever needing to see the raw scores. In a normal distribution of scores (as described previously) there are six standard deviations: three below the mean and three above, as is shown in Figure 18.3 above.

In this perfect model we always know that 34.13 percent of the scores of the sample fall within 1 standard deviation above the mean, and another 34.13 percent fall within 1 standard deviation below the mean. Thus, a total of 68.26 percent, or about two-thirds of the scores, is between +1 standard deviation and −1 standard deviation from the mean. This leaves almost one-third of the scores to fall farther away from the mean, with 15.87 percent (50 to 34.13 percent) above +1 standard deviation, and 15.87 percent (50 to 34.13 percent) below 1 standard deviation.

In total, when we look at the proportion of scores that fall between +2 and −2 standard deviations, 95.44 percent of scores can be expected to be found within these parameters. Furthermore, 99.74 percent of the scores fall between +3 standard deviations and −3 standard deviations about the mean. Thus, finding scores that fall beyond 3 standard deviations above and below the mean should be a rare occurrence.

The standard deviation has the advantage, like the mean, of taking all values into consideration in its computation. Also similar to the mean, it's utilized with interval or ratio levels of measurement and assumes a normal distribution of scores.

Several different samples of scores could have the same mean, but the variation around the mean, as provided by the standard deviation, could be quite different, as is shown in Figure 18.4a. Two different distributions could have unequal means and equal standard deviations, as in Figure 18.4b, or unequal means and unequal standard deviations, as in Figure 18.4c.

The standard deviation of the scores of Thea's foster care parents was calculated to be 10. Again, assuming that the variable of knowledge about parenting skills is normally distributed in the population of foster care parents, the results of the *PSS* scores from the sample of parents about whom we are making inferences can be shown in a distribution like Figure 18.5 on the following page.

As can also be seen in Figure 18.5, the score that would include 2 standard deviations, 102, is beyond the total possible score of 100 on the test. This is because the distribution of the scores in Thea's sample of parents does not entirely fit a normal distribution. The one extremely low score of 55 (see table 18.1) obtained by one foster care parent would have affected the mean as well as the standard deviation.

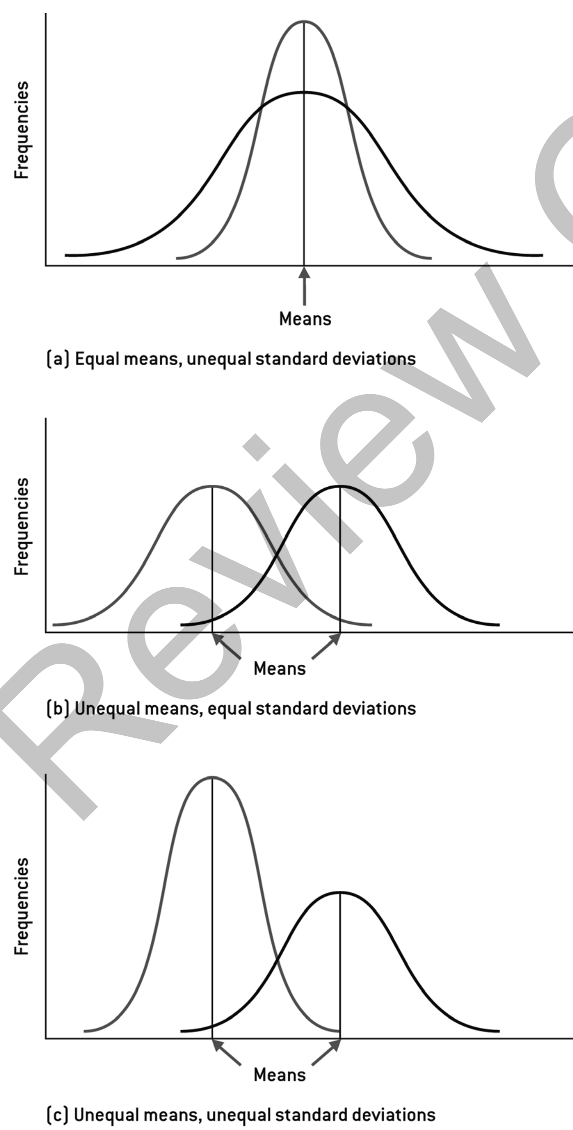


Figure 18.4

Variations in the Normal Distribution

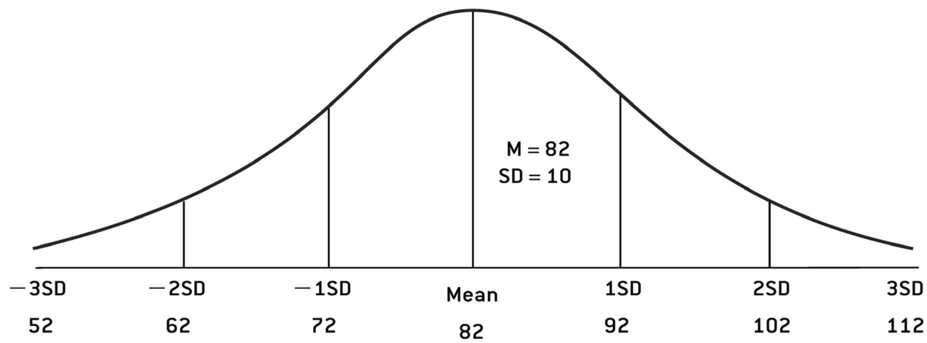


Figure 18.5

Distribution of Parental Skills Scores

INFERENTIAL STATISTICS

The goal of inferential statistical tests is to rule out chance as the explanation for finding either associations between variables or differences between variables in our samples. Because we can rarely study an entire population, we are almost always dealing with samples drawn from that population. The danger is that we might make conclusions about a particular population based on a sample that is uncharacteristic of the population it's supposed to represent.

For example, perhaps the group of foster parents in Thea's training session happened to have an unusually high level of knowledge of parenting skills. If she assumed that all the rest of the foster parents that she might train in the future were as knowledgeable, she would be overestimating their knowledge, a factor that could have a negative impact on the way she conducts her training program.

Inferential statistics try to rule out chance as the explanation for finding either associations between variables or differences between variables in our samples.

To counteract the possibility that the sample is uncharacteristic of the general population, statistical tests take a conservative position as to whether we can conclude that there are relationships between the variables within our sample. The guidelines to indicate the likelihood that we have indeed found a relationship or difference that fits the population of interest are called *probability levels.*

The convention in most social science research is that variables are significantly associated or groups are significantly different if we are relatively certain that in nineteen samples out of twenty (or 95 times out of 100) from a particular population, we would find the same relationship. This corresponds to a probability level of .05, written as: $p < .05$.

Probability levels are usually provided along with the results of the statistical test to demonstrate how confident we are that the results actually indicate statistically significant differences. If a probability level is greater than .05 (e.g., .06, .10), this indicates that we did not find a statistically significant difference.

STATISTICS THAT DETERMINE ASSOCIATIONS

There are many statistics that can determine whether there is an association between two variables. We will briefly discuss two:

1. Chi-square
2. Correlation

Table 18.5

Frequencies (and Percentages) of Gender by Previous Training (from table 18.1)

	Previous Training?		
Gender	Yes	No	Total
Male	1 (11)	8 (89)	9
Female	3 (27)	8 (73)	11
Totals . . .	4 (20)	16 (80)	20 (100)

CHI-SQUARE

The chi-square test requires measurements of variables at only the nominal or ordinal level. Thus, it's very useful because much data in social work are gathered at these two levels of measurement. In general, the chi-square test looks at whether specific values of one variable tend to be associated with specific values of another.

In short, we use it to determine whether two variables are related. It cannot be used to determine whether one variable caused another, however. In thinking about the foster care parents who were in her training program, Thea was aware that women are more typically responsible for caring for their own children than men. Even if they are not mothers themselves, they are often in professions such as teaching and social work where they are caretakers.

Thus, she wondered whether there might be a relationship between gender and previous training in parenting skills, such that women were less likely to have taken such training because they already felt confident in their knowledge of parenting skills. As a result, her one-tailed hypothesis was that fewer women than men would have previously taken parenting skills training courses. Thea could examine this possibility with her twenty foster care parents using a chi-square test. In terms of gender, Thea had data from the nine (45 percent) men and eleven (55 percent) women. Of the total group, four (20 percent) had previous training in foster care training, while sixteen (80 percent) had not.

As shown in Table 18.5, the first task was for Thea to count the number of men and women who had previous training and the number of men and women who did not have previous training. She put these data in one of the four categories in Table 18.5. The actual numbers are called *observed frequencies*. It's helpful to transform these raw data into (percentages), making comparisons between categories much easier.

We can, however, still not tell simply by looking at the observed frequencies whether there is a statistically significant relationship between gender (male or female) and previous training (yes or no). To do this, the next step is to look at how much the observed frequencies differ from what we would expect to see if, in fact, if there was no relationship. These are called *expected frequencies*. Without going through all the calculations, the chi-square table would now look like Table 18.6 for Thea's data set.

Because the probability level of the obtained chi-square value in Table 18.6 is greater than .05, Thea did not find any statistical relationship between gender and previous training in parenting skills. Thus, statistically speaking, men were no more likely than women to have received previous training in parenting skills; her research hypothesis was not supported by the data.

Table 18.6

Chi-Square Table for Gender by Previous Training (from table 18.5)

	Previous Training?	
Gender	Yes	No
Male		
Observed	1.0	8.0
Expected	1.8	7.2
Female		
Observed	3.0	8.0
Expected	2.2	8.8

Note: $\chi^2 = 0.8$; $df = 1$; $p > .05$

CORRELATION

Tests of correlation investigate the strength of the relationship between two variables. As with the chi-square test, correlation cannot be used to imply causation, only association. Correlation is applicable to data at the interval and ratio levels of measurement. Correlational values are always decimalized numbers, never exceeding ±1.00.

The size of the obtained correlation value indicates the strength of the association, or relationship, between the two variables. The closer a correlation is to zero, the less likely it is that a relationship exists between the two variables. The plus and minus signs indicate the direction of the relationship. Either high positive (close to +1.00) or high negative numbers (close −1.00) signify strong relationships.

In most positive correlations, though, the scores vary similarly, either increasing or decreasing. Thus, as parenting skills increase, so does self-esteem, for example. A negative correlation, in contrast, simply means that as one variable increases the other decreases. An example would be that as parenting skills increase the stresses experienced by foster parents decrease.

Thea may wonder whether there is a relationship between the foster parents' years of education and score on the *PSS* knowledge test. She might reason that the more years of education completed, the more likely the parents would have greater knowledge about parenting skills.

To investigate the one-tailed hypothesis that years of education are positively related to knowledge of parenting skills, Thea can correlate the *PSS* scores with each person's number of years of formal education using one of the most common correlational tests, Pearson's *r*. The obtained correlation between *PSS* score and years of education in this example is $r = -.10$, $p > .05$. It was in the opposite direction of what she predicted. This negative correlation is close to zero, and its probability level is greater than .05. Thus, in Thea's sample, the parents' *PSS* scores are not related to their educational levels.

If the resulting correlation coefficient (*r*) had been positive and statistically significant ($p < .05$), it would have indicated that as the knowledge levels of the parents increased so would their years of formal education. If the correlation coefficient had been statistically significant but negative, this would be interpreted as showing that as years of formal education increased, knowledge scores decreased.

If a correlational analysis is misinterpreted, it's likely to be the case that the researcher implied causation rather than simply identifying an association between the two variables. If Thea were to have found a statistically significant positive correlation between knowledge and education levels and had explained this to mean that the high knowledge scores were a result of higher education levels, she would have interpreted the statistic incorrectly.

STATISTICS THAT DETERMINE DIFFERENCES

There are three commonly used statistical procedures to determine whether two or more groups are statistically different from one another.

1 Dependent *t*-tests
2 Independent *t*-tests
3 Analysis of variance (ANOVA)

T-tests are used with only two groups of scores, whereas ANOVA is used when there are more than two groups. All are characterized by having a dependent variable at the interval or ratio level of measurement, and an independent, or grouping, variable at either the nominal or ordinal level of measurement. Several assumptions underlie the use of both *t*-tests and ANOVA.

First, it's assumed that the dependent variable is normally distributed in the population from which the samples were drawn. Second, it's assumed that the variance of the scores of the dependent variable in the different groups is roughly the same. This assumption is called *homogeneity of variance*. Third, it's assumed that the samples are randomly drawn from the population. Nevertheless, as mentioned in Chapter 15 on group research designs, it's a common occurrence in social work that we can neither randomly select from a population nor randomly assign individuals to either the experimental or the control group. In many cases this is because we are dealing with already preformed groups, such as Thea's foster care parents.

Breaking the assumption of randomization, however, presents a serious drawback to the interpretation of the research findings, which must be noted in the limitations and the interpretations section of the final research report. One possible difficulty that might result from nonrandomization is that the sample may be uncharacteristic of the larger population in some manner.

It's important, therefore, that the results not be used inferentially; that is, the findings must not be generalized to the general population. The design of the research study is thus reduced to a descriptive level, being relevant to only those individuals included in the sample.

DEPENDENT *T*-TESTS

Dependent *t*-tests are used to compare two groups of scores from the same individuals. The most frequent example in social work research is looking at how a group of individuals change from before they receive a social work intervention (pre) to afterward (post). Thea may have decided that although she knew the knowledge levels of the foster care parents before receiving training was interesting, it did not give her any idea about whether her program helped the parents to improve their skill levels. In other words her research question became: After being involved in the program, did parents know more about parenting skills than before they started? Her hypothesis was that knowledge of parenting skills would improve after participation in her program.

Thea managed to contact all the foster care parents in the original group (group A) one week after they had graduated from the program and asked them to fill out the *PSS*

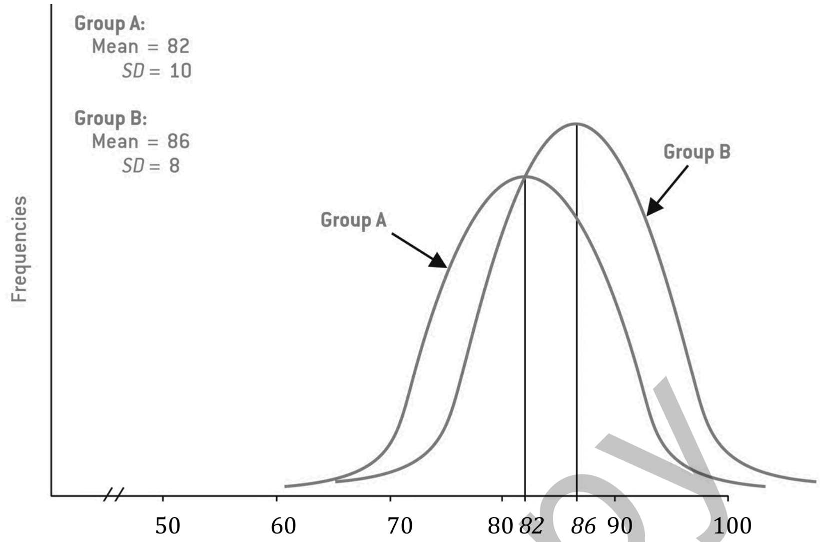


Figure 18.6

Frequency Distributions of *PSS* Scores from Two Groups of Foster Care Providers

knowledge questionnaire once again. Because it was the same group of people who were responding twice to the same questionnaire, the dependent *t*-test was appropriate.

Using the same set of scores collected by Thea previously as the pretest, the mean *PSS* was 82, with a standard deviation of 10. The mean score of the foster care parents after they completed the program was calculated as 86, with a standard deviation of 8.

A *t*-value of 3.9 was obtained, significant at the .05 level, indicating that the levels of parenting skills significantly increased after the foster care parents participated in the skills training program. The results suggest that the average parenting skills of this particular group of foster care parents significantly improved (from 82 to 86) after they had participated in Thea's program.

INDEPENDENT *T*-TESTS

Independent *t*-tests are used for two groups of scores that have no relationship to each other. If Thea had *PSS* scores from one group of foster care parents and then collected more *PSS* scores from a second group of foster care parents, for example, these two groups would be considered independent, and the independent *t*-test would be the appropriate statistical analysis to determine whether there was a statistically significant difference between the means of the two groups' *PSS* scores.

Thea decided to compare the average *PSS* score for the first group of foster care parents (group A) with the average *PSS* score of parents in her next training program (group B). This would allow her to see whether the first group (group A) had been unusually talented or conversely were less well versed in parenting skills than the second group (group B). Her hypothesis was that there would be no differences in the level of knowledge of parenting skills between the two groups.

Because Thea had *PSS* scores from two different groups of participants (groups A and B), the correct statistical test to identify whether there are any statistical differences between the means of the two groups is the independent *t*-test. Let's use the same set of numbers that we previously used in the example of the dependent *t*-test in this analysis, this time considering the posttest *PSS* scores as the scores of the second group of foster care parents.

As can be seen from Figure 18.6, the mean *PSS* of Group A was 82 and the standard deviation was 10. Group B scored an average of 86 on the *PSS*, with a standard deviation of 8. Although the means of the two groups are four points apart, the standard deviations in the distribution of each are fairly large, so there is considerable overlap between the two groups. This would suggest that statistically significant differences will not be found.

The obtained *t*-value to establish whether this four-point difference (86 − 82) between the means for two groups is statistically significant was calculated to be $t = 1.6$ with a $p > .05$. The two groups were thus not statistically different from one another, and Thea's hypothesis was supported.

Note, however, that Thea's foster care parents were not randomly assigned to each group, thus breaking one of the assumptions of the *t*-test. As discussed earlier, this is a serious limitation to the interpretation of the study's results. We must be especially careful not to generalize the findings beyond the groups included in the study.

Also note that in the previous example, when using the same set of numbers but a dependent *t*-test, we found a statistically significant difference. This is because the dependent *t*-test analysis is more robust than the independent *t*-test, because having the same participant fill out the questionnaire twice, under two different conditions, controls for many extraneous variables, such as individual differences, that could negatively influence an analysis of independent samples.

ANALYSIS OF VARIANCE

A one-way ANOVA is the extension of an independent *t*-test that uses three or more groups. Each set of scores is from a different group of participants. For example, Thea might use the scores on the *PSS* test from the first group of foster care parents from whom she collected data before they participated in her program, but she might also collect data from a second and a third group of parents before they received the training. The test for significance of an ANOVA is called an *F*-test.

We could actually use an ANOVA procedure on only two groups, and the result would be identical to the *t*-test. Unlike the *t*-test, however, obtaining a significant F-value in a one-way ANOVA does not complete the analysis. Because ANOVA looks at differences between three or more groups, a significant *F*-value only tells us that there is a statistically significant difference among the groups. It does not tell us between which ones. To identify this, we need to do a post hoc test.

A variety is available, such as Duncan's multiple range, Tukey's honestly significant difference test, and Newman-Keuls, and they are provided automatically by most computer statistics programs. But one caution applies: a post hoc test should be used only after finding a significant *F*-value, because some of the post hoc tests are more sensitive than the *F*-test and so might find significance when the *F*-test does not. Generally, we should use the most conservative test first, in this case the *F*-test.

In the example of Thea's program, let's say that she collected data on a total of three different groups of foster care parents. The first group of foster care parents scored an average of 82 on the *PSS* (standard deviation 10). The second group scored an average of 86 (standard deviation 8), and the mean score of the third group was 88 with a standard deviation of 7.

The obtained *F*-value for the one-way ANOVA is 2.63, with a $p > .05$. Thus, we must conclude that there are no statistically significant differences between the means of the groups (i.e., 82, 86, and 88). Because the *F*-value was not significant, we would not conduct any post hoc tests. This finding would be interesting to Thea because it suggests that all three groups of foster care parents started out with approximately the same knowledge levels, on average, before receiving training.

EVALUATING A QUANTITATIVE ANALYSIS

Assessing the overall credibility of a quantitative data analysis can be a daunting task at the best of times. Nevertheless, you can use Appendix A and ask a series of questions about a particular data analysis (e.g., questions 217–238 on pages 490–492).

Two people can assess the same quantitative data analysis and come up with two different opinions in respect to its overall credibility and usefulness.

How to evaluate the overall creditability of a quantitative data analysis is totally up to you—and only you. You are the one who must ask and answer questions. You just don't want to ask questions without knowing why you're asking them. More important, you need to address two extremely difficult tasks:

1 You must have a rationale for why you're asking each question in the first place. Appendix A only lists potential questions you can ask. Can you guess why each one was listed and what importance your answer to each question has for your overall evaluation of the research study?

2 You also must have some kind of idea of what you're going to do with your answer to each question. For example, question 227 on page 491 asks, "Apart from statistically significant changes, was the clinical significance of any improvements discussed (if applicable)?" Let's say your answer was "no". What are you going to do with your answer? That is, how are you going to use it in your overall evaluation of the study's creditability? Now what about if your answer was "yes"?

CHAPTER RECAP

- On a simple level there are two types of statistics: descriptive and inferential.
- Descriptive statistics describe and summarize a variable(s) of interest and portray how that particular variable is distributed in the sample or population.
- Inferential statistics try to rule out chance as the explanation for finding either associations between variables or differences between variables in samples drawn from a population.
- Descriptive statistics can be broken down into three categories: frequency distributions, measures of central tendency, and measures of variability.
- Measures of central tendency contain three groups of statistics: mode, median, and mean.
- Measures of variability contain two groups of statistics: range and standard deviation.
- Inferential statistics can be broken down into two categories: statistics that determine associations (i.e., chi-square, correlation) and statistics that determine differences (i.e., dependent *t*-tests, independent *t*-tests, one-way analysis of variance).

REVIEW EXERCISES

1 Before you entered your social work program, and before you read this chapter, how knowledgeable were you about the methods of analyzing quantitative data? Discuss fully.

2 What are *quantitative data*? Provide social work examples throughout your discussion to illustrate your main points.

3 What are the four levels of measurement for quantitative data? Provide an example of each and discuss how it can be used in a social work research situation and a social work practice situation.

4 Discuss the purposes of using descriptive statistics to analyze quantitative data. Provide social work examples throughout your discussion to illustrate your main points.

5 What are *frequency distributions*? Provide social work examples throughout your discussion to illustrate your main points.

6 What are *measures of central tendency*? Provide social work examples throughout your discussion to illustrate your main points.

7 What are *modes*? Provide social work examples throughout your discussion to illustrate your main points.

8 What are *medians*? Provide social work examples throughout your discussion to illustrate your main points.

9 What are *means*? Provide social work examples throughout your discussion to illustrate your main points.

10 Locate a quantitatively oriented journal article that used one or more of the three measures of central tendency. Were the statistics meaningful to you? If so, why? If not, why not?

11 What are *measures of variability*? Provide social work examples throughout your discussion to illustrate your main points.

12 What are *ranges*? Provide social work examples throughout your discussion to illustrate your main points.

13 What are *standard deviations*? Provide social work examples throughout your discussion to illustrate your main points.

14 Locate a quantitatively oriented journal article that used standard deviations in its data analysis. Was it meaningful to you? If so, why? If not, why not?

15 Discuss the main purpose of using inferential statistics to analyze quantitative data. Provide social work examples throughout your discussion to illustrate your main points.

16 Discuss how the chi-square test can be used to determine if there is an association between two variables.

17 Locate a quantitatively oriented journal article that used the chi-square test in its data analysis. Was it meaningful to you? If so, why? If not, why not?

18 Discuss how tests of correlation can be used in social work research studies.

19 Discuss how the dependent *t*-test can be used in social work studies.

20 Discuss how the independent *t*-test can be used in social work research studies.

21 What's the main difference between an independent *t*-test and an analysis of variance? Provide social work examples throughout your discussion to illustrate your main points.

22 Discuss how a one-way analysis of variance can be used to determine if there are any significant differences between the means of three or more groups. Provide social work examples throughout your discussion to illustrate your main point.