Contents

Chapter 3

Types of Variables

3.1 Introduction

Statistics begins by defining and identifying a **population** which is to be examined. In Chapter 2, it was noted that the population is the set of individuals, objects or phenomena of interest to the researcher. The researcher then investigates various characteristics of the population. These characteristics of the population are called **variables**. Once these characteristics of the population are obtained, statistical analysis can be used to describe the population and analyze the relationship among the various variables which have been produced. Inferential statistics takes the variables which have been obtained from a sample or from an experiment, and uses these to make statements concerning these variables in the population as a whole.

In carrying out these various forms of statistical analysis, a researcher must first consider the nature of the variables which have been obtained. Various types of variables and the implications of having different types of variables is discussed in this chapter. Section 3.2 looks at types of measurement and presents the difference between nominal, ordinal, interval and ratio scales. Section 3.3 examines the difference between variables which are continuous and discrete.

3.2 Measurement

When organizing and analyzing data, one of the first steps to consider is the nature of the scale by which the variable is measured. Scales may be well established measurements such as kilograms to measure weight or centimetres to measure length. Or they may be scales such as strongly agree, agree, neutral, disagree or strongly disagree for measuring opinion. Each of these scales constitutes a different form of measurement, and the type of scale used to measure characteristics of population members has considerable effect on the manner in which this data is presented and analyzed.

In statistical analysis, variables are commonly measured on one of four possible scales. All variables classify population members into categories. This classification is called a **nominal** scale. If the categories into which the variables are classified can be ordered or ranked, then this ordering is considered to be the basis of an **ordinal** scale. When a scale of measurement allows the researcher to meaningfully measure distances between values or categories into which the variable has been classified, the scale is called an **interval** scale. Finally, if the ratio of the values of the categories into which the variable has been classified can be meaningfully determined, then the scale is a **ratio** scale.

The various types of scales are hierarchical, in that each successive level of measurement has the characteristics of the earlier level, but some additional characteristics. Each of these types of scale, along with examples, is discussed in the following sections.

3.2.1 Nominal Scale of Measurement.

A nominal scale is any classification or categorization of the members of the population.

Definition 3.2.1 A **nominal** scale is a scale which classifies the various values of the variable into categories. These categories are usually given names, or possibly numbers.

By definition, variables have different values which describe the various characteristics of a population. Thus every variable has at least a nominal scale of measurement. Some scales are no more than nominal, though, in that the values taken on by the variable differ, but one cannot meaningfully order these values or measure the distance between these values. That is, a scale which is nominal, but no more than nominal, is no more than a classification or categorization of these values into different categories.

Some examples of variables which are considered to be measured by nominal scales are sex, ethnicity, religion, nationality, political party preferred and province of residence. Sex can take on only two values, male and female, and these are different categories, but no more than that. One cannot meaningfully say that male exceeds female, or that female exceeds male, or measure the distance between these two values for sex. These two values, male and female, are merely different categories for the variable sex.

Ethnicity or nationality have considerably more categories into which people can be classified, but again the different ethnicities amount to no more than a classification of culture or nationality. For example, English, French, Aboriginal, Ukrainian, and Vietnamese are different categories into which people may be classified, and the differences between these cannot be measured in any meaningful manner. It may be that the income of respondents of English ethnicity exceeds the income of all these other ethnicities, but in this case, the variable becomes income, which can be ranked and measured in dollars, with ethnicity merely the classification of people. Religion is essentially the same as ethnicity or nationality, a classification of people into the various religious preferences. The same can be said of political preferences.

As can be noted in the above examples, a scale which is nominal usually has names associated with the different categories. The categories could alternatively be given numbers. For example, male could be given value 1 and female the value 2. These numbers may be used by data analysts when coding nominal scales for data entry to a computer. However, these codes are used only to keep track of the values male and female, and have no meaning as numerical values. The codes are purely arbitrary, in this example, female could instead have been coded 1 and male coded 2. When coding a nominal scale, one need only be consistent, and ensure that one has enough codes so that all possible values of the variable can be included. For example, in Canada, if ethnicity were coded with numbers, it might be necessary to have over 100 codes to take care of the great number of possible ethnicities that people have.

When working with statistics, one often wishes to have scales which are more than nominal. The nominal scale is the weakest or lowest level of scale in that one is only categorizing the values of the variable. The statistical methods appropriate for nominal scales are often relatively weak, and do not allow the researcher to carry out such detailed statistical analysis.

3.2.2 Ordinal Scale of Measurement

Frequently the values of the scale being used can be ranked or ordered so that one category exceeds or falls short of another category.

Definition 3.2.2 An **ordinal scale** is a scale which is nominal, and one in which each value of the variable can be ordered, or ranked, as more than, less than, or equal to any other value of the variable.

An ordinal scale is any ranking or ordering of the different possible values of a variable. If a scale is ordinal, but no more than ordinal, then the different values of the scale can be ranked or ordered, but the distances between the values cannot be meaningfully determined.

Some examples of ordinal scales are:

1. Order of finish in a race: first, second, third, etc. In this case, one may not know, or care about, the actual difference in time taken to complete a race. In the case of a race or contest where a gold is awarded for first, silver for second and bronze for third, all that matters is the ranking.

2. Attitudes are often measured on an ordinal scale. The attitudes may be ranked strongly agree, agree somewhat, disagree somewhat, and disagree strongly on some issue. In this case, the opinion of Person A can be compared with the opinion of Person B, and A's opinion is greater than, less than, or equal to B's opinion. The same situation may occur when respondents are asked to give their opinion on a scale from 1 to 7 where 1 means strongly disagree and 7 means strongly agree. On this scale, a respondent answering 4 is ranked as greater than someone else who rates himself or herself as 2.

3. Many scales used by psychologists and sociologists are ordinal since they rank respondents on some scale. In Canada, two commonly used scales of socioeconomic status are the Pineo-Porter scale and the Blishen scale. The Pineo-Porter scale is ordinal in that respondents were asked to rank occupations on the basis of social standing. Those occupations with a greater value on the Pineo-Porter scale are thus considered to have greater social status, social standing or prestige than those lower on the scale. The Blishen scale is based on a combination of income and education, with a higher Blishen score meaning the person has greater socioeconomic status than a person with a lower Blishen score.

   IQ is another ordinal scale, although there are some researchers who consider an interval scale. The case of IQ is discussed again when examining interval level scales.

If a variable can be measured on an ordinal scale, considerably more statistical analysis can be carried out on the variable, as compared with the situation where the variable has only a nominal scale. In the first place, the values can be ranked in order, and the value which has the middle rank determined. This measure is called the median, and will be discussed in detail in Section 5.3. The ranks of two variables which are each ordinal can be compared, to see how closely they match. For example, suppose a researcher expected that those of greater socioeconomic status tended to agree that Canadians who are better off should pay less in taxes, while those of lower status tend to disagree. The responses of those surveyed can be ranked on each of these variables (socioeconomic status and opinion concerning taxation) and the ranking compared. In Chapter 10, the comparison of these rankings is used to construct a correlation coefficient (Spearman's rho). (Or similar rankings of two different judges).

In summary, any variable where the different possible values of the variable can be meaningfully ranked is considered an ordinal scale. If one can rank these different possible values, then one can carry out a much wider range of analyses of the data than in the case of variable which are nominal only. For this reason, many social scientists attempt to create ordinal scale variables. (Other examples?)

3.2.3 Interval Scale of Measurement

It is often possible to measure a variable on a scale which indicates considerably more than whether one value ranks higher on the scale than does the other value. For example, if one person is taller than another person, the amount of difference in height can be measured. A person's height is certainly ordinal in nature, but with the additional characteristic that height differences can be meaningfully determined. The idea of a meaningfull difference is the basis for the definition of an interval level scale.

Definition 3.2.3 An **interval** scale of measurement is a scale in which equal numerical differences represent equal quantities or magnitudes of the characteristic.

To illustrate what this means, consider the heights of the four male undergraduates in Table 1. Each of the four heights can certainly be ranked as being greater than or less than any other height. But in addition, the heights can be measured on a numerical scale, and one where the numbers, and the differences in these numbers, are meaningful and of constant magnitude. That is, based on the metric scale, John is 3 cm. taller than James. In turn, Blaine is 3 cm. taller than Kyle. These two different values of 3 cm. each represent an equal magnitude, so that it makes sense and is correct to write $180 - 177 = 191 - 188$. Equal numerical differences of 3 cm. represent equal magnitudes of the characteristic *height*.

Name	Height	Attitude
James	177 cm.	3 (Agree Somewhat)
John	180 cm.	2 (Disagree Somewhat)
Kyle	188 cm.	1 (Strongly Disagree)
Blaine	191 cm.	2 (Disagree Somewhat)

Table 3.1: Hypothetical Heights and Attitudes of Four Male Undergraduates

In contrast, the same cannot be said concerning attitudes. Suppose attitude is measured on a 4 point scale from 4 as strongly agree to 1 as strongly disagree. While the numerical difference between James and John is $3-2 = 1$ unit of attitude, it may not make sense to say that this difference in attitude is equal to the same difference of $2 - 1 = 1$ between Blaine and Kyle. The numerical value is equal in each case. But the difference between 3 (Agree Somewhat) and 2 (Disagree Somewhat) may not represent the same quantity of difference in opinion that 2 (Disagree Somewhat) and 1 (Strongly Disagree) does. Attitude is an ordinal scale, and while differences in attitude certainly exist, it is very difficult to meaningfully measure these differences.

Definition 3.2.4 An **interval scale** is alternatively defined as a scale which has a well defined **unit of measure**.

If a well defined unit of measure exists, then this forms the basis for construction of an interval scale. This unit of measure must be defined in such a manner that different investigators can agree on its definition. The unit should also be invariant in time and place, so that when measuring some phenomenon, different researchers will produce measurements of the phenomenon which are more or less identical. In the case of height, the unit of measure is the centimetre (or the inch). This is a measure which has been adopted and agreed upon universally. A standard has been set and this standard does not vary from country to country or over time.

In the case of the measurement of attitudes, the difficulty is to define a unit of measure. If a researcher wishes to determine attitudes toward the view "There is too much difference between rich and poor in Canada," it is not really possible to define a unit of agreement or disagreement on an attitude scale with respect to this question. Even if a researcher did construct such a unit, it is likely that other researchers attempting to measure the same set of opinions would produce quite different units and measures. As a result, attitudes are usually considered to be measurable on an ordinal, but not on an interval level scale.

Once a unit of measure has been defined, it is possible to add or subtract values on an interval scale. Differences between values on an interval level scale have meaning in that they represent a certain number of units that form the basis for the scale. In the example of Table 3.1, Kyle is $188 - 177 = 11$ cm. taller than James, and this difference of 11 cm. in their heights would be the same, regardless of who measured their respective heights. It is, of course, possible that two measurements of a person's height will differ slightly. There are usually small errors of measurement, when measuring any phenomenon. But if the scale is interval, the unit is well defined and standardized, and errors of measurement merely represent the difficulty of applying an absolutely constant standard to any phenomenon.

Some Examples of Interval Scales.

1. Any variable that is measured in terms of time is an interval level scale. Age in years, the time taken to complete a Statistics assignment, and the number of weeks spent on vacation are all examples of interval scales. In each case the variable measures a certain amount of time elapsed. The unit of measure for *time* is seconds, minutes, hours or years. Each of these units is a well agreed upon standard which does not change. That is, a second is always a second, and each second that passes represents an equal extra amount of time that passes.

2. Any variable measured in dollars is measured on an interval scale. While the value of the dollar, and the amount of goods and services the dollar will purchase, changes over time, at any one point in time the value of a dollar is given. Certainly the unit of a dollar is well defined in Canada. Any loonie is exchangeable for any other loonie,

or two loonies are exactly equal to one two dollar bill in value. If my income is $3,000 per month and your income is $2,500 per month, the difference is $500, and this difference of $500 does not differ from the difference between an income of $1,500 and $1,000. In each case the gap is $500.

3. A variable that measures the number of people or the number of objects is usually an interval level variable. The number of letters on a page, the number of people taking this Statistics class, the number of restaurant meals sold weekly, are all interval level scales. In each case, the individual object or person is the unit of measure. In the case of people, if one is only considering how many there are, then each person counts as one, and one person is the unit. Each extra person counts as one more, so differences in the number of people represent constant magnitudes.

Some variables are treated as interval level, although they may not be so clearly considered as such. For example, when grading, instructors normally grade on a scale from 0 to 100, with each extra grade being equal in magnitude. While the unit of measurement here is 1 unit of grade, this is not so neatly defined as instructors might have students believe. There is often considerable discretion or judgment involved in grading, so that different instructors might grade the same question somewhat differently. Yet grades are added and averaged, producting a *grade point average*, and thus grades are certainly treated as if they were interval level scales. Even attitudes may be treated in this manner. Once a numerical scale from 1 to 4, or from 1 to 7, is attached to agreement or disagreement on some attitude question, then these attitude scales are often added and averaged. The implicit assumption the researcher makes here is that there may be some underlying unit appropriate to measuring attitudes.

Since variables measured on an interval level scale can be added or subtracted, differences determined, and averages calculated, researchers hope to be able to measure characteristics of populations with an interval level scale. Many more statistical procedures can be carried out with interval level scales than with scales that are only nominal or ordinal. Later in this textbook, it will be seen that many of the descriptive procedures and tests used in statistics require interval level scales. Commonly used t-tests and regression both assume that the scales used are interval. As a result, if the phenomenon being measured has a well defined unit of measure that makes construction of an interval level scale possible, it is preferable to use this level of measurement.

Different social science disciplines tend to have different types of scales of measurement. Economics, since it measures many of its variables in a monetary unit such as dollars, is generally able to use statistical methods that are appropriate for interval level scales. Sociology is often in the same position. The Blishen scale of socioeconomic status, or measures of income and wealth distributions, are based on interval level measurements. Some sociological variables such as social class, attitudes and opinions, are usually considered as only ordinal in nature. In Political Science and Psychology, the measures are often no more than ordinal. Often all that can be done in terms of measuring attitudes or mental processes is to rank values. In some areas of Political Science and of Sociology, variable may be no more than nominal. Political party preferred, ethnicity, region of the country, and sex are all nominal scales which play a major role in both Political Science and Sociology.

3.2.4 Ratio Scale of Measurement

The final commonly used scale of measurement is a scale for which one can legitimatley take the ratios of two values. This type of scale will at least be of the interval level, so that differences of values are meaningful, but with the additional characteristic that the ratio of two values will also be a meaningful magnitude.

Definition 3.2.5 A **ratio** scale of measurement is an interval scale with the additional property that equal ratios between two possible values of a variable represent equal magnitudes.

To illustrate a ratio scale, consider the variable *income*. A person with an income of $3,000 per month has an income that is 2 times the income of a person with a $1,500 monthly income. Similarly, an income of $1,000 per month is 2 times an income of $500 per month. In each case, the ratio is 2, or 2 to 1, representing the same ratio, or the fact that in each case the first income is double that of the second income.

Almost all interval level variables are also ratio level variables. For most purpose, one can assume that this is the case. However, there are a few interval level scales which are not ratio level. The most common is temperature, as measured on the Celsius or Fahrenheit scale. A temperature of

40° Celsius may seem double that of 20, but in what sense is this doubling equal to the doubling of 1 to 2? In the case of temperature on the Celsius scale, each extra degree represents an equal extra amount of heat, but ratios of two temperatures are not really meaningful. The problem with this scale is that the 0 point is arbitrary. On the Celsius scale, the 0 point is arbitrarily chosen as the freezing point of water at sea level and under normal atmospheric conditions.

Many social science scales also have an arbitray 0 point, and thus can be considered to be no more than interval level. The Blishen scale of socioeconomic status can be considered to be an interval level scale of measurement, in that it is constructed using a combination of income and education. Since the measures of each of the latter are also interval, the Blishen scale can be considered to be an interval level scale. However, it has then be further scaled so that occupations will ordinarily have values between 0 and 100. Again, the 0 point is arbitrary, and while differences in socioeconomic status may have meaning, the ratio of two different values of socioeconomic status would not appear to have meaning in the same way that the ratio of two incomes has meaning.

The same problem appears in the case of IQ, although it may be that IQ is really no more than ordinal.

3.2.5 Importance of Type of Scale

The type of scale used to measure the values of a variable is important in statistics for three main reasons. Much of this will become clear only later in this text. In what follows, some of the reasons for considering the different scales are given, but if this is the first time you have studied statistics, some of the procedures mentioned here will only become clear later in the semester.

First, the meanner in which the data is presented will often differ depending on whether it is measured on the basis of a nominal, ordinal, interval or ratio scale. Variables which can be measured on no more than a nominal scale often contain relatively few values and the whole range of these values can be presented. For example, sex, political party preferred or labour force status have only a few possible values, and data concerning the distribution of these is often presented. In addition, tables of relationships between pairs of nominal scale variables are one of the most common types of statistical data presented. For example, a cross-classification of sex by labour force status would show how many males and how many females there are in each of the categories of employed, unemployed or non labour force. Where the scale is ordinal, interval or ratio, there are often many more possible values, and the variable must then be grouped into categories. The creates the problem concerning how best to group the values of the variable. Examples of this will be presented in the next chapter.

Second, the nature of the summary statistics that will be used, depend largely on the type of scale. Averages, or measures of central tendency, as statisticians call these, which are appropriate for each type of scale differ. The mode, or most common value of the variable, can always be constructed, because it can be determined for even a nominal scale. For example, if there are more Protestants in the population than any other category of religion, then the modal religion is Protestant. Ordinal scales lend themselves to determining the median, or middle value, or a set of values. For variables which are measured on no higher a scale than a nominal or ordinal scale, a branch of statistics called *nonparametric statistics* is often most appropriate.

In contrast, for variables which can be measured on an interval or ratio scale, the mean can be calculated. This involves summing the values of the variables and dividing this total by the number of values. This is the *average* in the manner it is most commonly used, for example, as in *grade point average*. Since most interval scales are also ratio scales, for most interval level scales it is possible to use all of the ordinary arithmetic operations. Measures such as the mean and standard deviation are most commonly used in the case of interval and ratio scales, and the statistical analysis carried out using these scales is referred to as *parametric statistics*. Most of the statistical methods presented in this textbook are part of parametric statistics.

Third, quite different statistical procedures can be applied to each type of variable. In the case of nominal variables, there are a number of procedures such as the chi square test, and measures of association such as Cramer's V, lambda or phi which may be used. These are discussed later in the text. When the variables have been measured at the ordinal level, other measures such as tau, rho and gamma are more appropriate in order to determine the nature of the relationship between two or more variables. Finally, when the variables are measured at the interval or ratio level, the full range of statistical methods can be used. Means and standard deviations can be meaningfully calculated, and tests such as t-tests along with regression and correlation methods, all can legitimately be used. For example, if a researcher is interested in estimating the effect of extra years of schooling on the potential income later in life, this can be done using the regression method. This will allow the researcher to determine how many extra dollars of income a typical member of the labour force receives as a result of completing another year of schooling. In this way, a rate of return to investment in education can be calculated.

Since these latter methods associated with interval or ratio scales tend to be more powerful, and can yield stronger conclusions, these methods are to be preferred. Often researchers will use these methods even with ordinal scales. While this has some problems associated with it, if carefully done, and if the researcher is aware of the potential errors associated with doing this, it may be appropriate. Some examples of this, and the cautions that should be taken, are given later in the text.

3.3 Continuous and Discrete

In the last section, it was seen that there are several different ways in which variable may be measured. Another type of distinction concerning the type of variables, when working with quantitative data, is a distinction between discrete and continuous. While the distinction between discrete and continuous is less important than the scales of measurement noted above, whether a variable is discrete or continous may have considerable impact on the presentation of the data.

Definition 3.3.1 A **discrete variable** is a variable which can take on only a countable number of values.

Examples of discrete variables are the number of people in a family, the number of books in a library, the political parties in a country, and the number of children in poverty. In each case, the set of all the possible values of the variable can be counted, 1, 2, 3, 4, ... and thus the variables are discrete.

Any variable which takes on only a finite number of possible values is discrete. For example, the number of chairs in a room could only be a finite number, so this represents a discrete variable. This set of possible values may be very large, in fact it could be infinite. But so long as the set of possible values of the variable can be matched up with the integers 1, 2, 3, 4, 5, ... this process could continue forever. The number of stars in the universe is extemely large, perhaps even infinite, but each star represents 1 extra star, so that the set of possible stars is, in principle, countable.

Another way in which one can think of discrete variables is to consider whether the objects of concepts involved are separate and discrete entities. For example, the variable *religious preference* might be measured as support for one or other religious denomination. Each religious denomination is a separate and discrete entity, having a set of doctrines and an organization associated with it. Similarly, political parties, or ethnicities, each represent separate and distinct organizations or groups so they would also be considered discrete.

In contrast, some characteristics of members of a population can potentially take on a set of values which cannot be counted. If this is the case, then the variable is said to be continous in nature.

Definition 3.3.2 A **continuous variable** is a variable whose values can assume any possible value along some line interval.

The reason this definition represents continuity is that mathematicians can prove that one cannot count all the possible values along some line interval. For example, variables such as height, weight, time, age, and temperature are all continuous variables. All can vary along some interval, and the set of all possible values that can potentially be taken on by these variables cannot possibly be counted. Some examples are as follows.

1. **Height**. The height of a person varies continuously from a very short height at birth to the full height one reaches as an adult. If the interval from 0 cm. to 300 cm. (about 10 feet) is considered, this should cover all the heights that a human can possibly take on. The height of any individual at any point in time is a particular value somewhere on this interval. The set of all such possible points is uncountable, or continuous. Alternatively, one can consider the growth process to be a continuous process. As children, we do not grow in jumps, but grow continuously until the full height of an adult is reached. Much the same considerations also apply to weight, and the process of continual change in weight, even as adults, is something that happens every day.

2. **Time or Age**. Any variable that is measured in terms of time is a continuous variables, because time varies continuously. Time does not pass in discrete jumps from one one instant to another, but time moves along a continuous spectrum. A person's age may be rounded off to years of age at last birthday, say age 20 or 21 or 22. However, these are merely conveniences of reporting. Each person ages in a continuous

manner, but age is usually reported as age at last birthday, or age to the nearest year.

3. **Temperature** is much the same. The line interval from -100 degrees to plus 100 degrees should cover the set of all possible temperatures that could occur at any point on the earth. Within this range, the temperature could take on any possible value, although only one distinct value is taken on at any one place and time. The thermometer shows this process of continuity, the alcohol or mercury moves continuously up or down the stem of the thermometer.

4. **Money**. Any variable measured in terms of dollars and cents, or some other monetary unit, can be considered to be continuous. While these monetary values may be rounded to the nearest dollar, or the nearest cent, in theory a sum of money can be reported to many decimal places, and the calculations of monetary values can be considered to vary continously.

5. **Attitudes**. Measurement of attitudes and opinions is less clear cut. Suppose that a researcher is attempting to measure the degree of support of Canadians for the federal Liberal party. The researcher could ask respondents in a survey whether they would vote for this party if a federal election were to be held. If this question were asked, the responses would be either *yes*, *no*, or *uncertain.* This is a discrete set of categories, and the variable measuring *party for which the respondent would vote* is a discrete variable.

   In contrast, if the researcher were to ask respondent how strongly she or he supports the Liberal party, then the variable might be considered continuous in nature, at least in principle. Support could be considered to vary continuously from outright opposition to this party, to complete support. An individual could be anywhere along the line joining these two points. In practice though, it would be difficult to measure opinions along this line, and different researchers would likely produce different values concerning the level of support. This is because of the earlier problem of constructing a means of measuring attitudes.

   In practice, the researcher would be more likely to use a 4, 5 or 7 point scale, from strong support to strong opposition, or strongly favour to strongly oppose. As a result, the *level of support* for the federal Liberal party is likely to be measured as a discrete variable, and for most purposes it is considered to be a discrete variable. But in principle, the underlying attitudes or opinions may be considered to vary continuously.

The last example illustrates the difference between the manner in which variables may be reported and their underlying nature. Any member of a population takes on a specific value for each variable which is obtained concerning that person. These values are also rounded to the nearest integer, or the nearest tenth of a point, etc. when they are reported. This makes it appear as if all variables are discrete, they take on only specific values in actual reporting. But this is only a requirement of reporting the data. Some variables such as number of children, number of trees, ethnicities or religions, etc. are inherently discrete in nature, because each value is distinct from any other value, and the set of all possible values can be counted. For other variables such as liquid measures, time, distance, area, monetary values, etc., the potential values can take on any value within some interval, and thus are continuous.

The main reason for distinguishing discrete and continuous is in terms of how the data is presented. Each value of a discrete variable, especially if the number of categories is relative small, can be separately presented. For a continuous variable, the set of values must be grouped, so the data is presented in an interval form. Then the researcher must decide what the appropriate intervals are, so that the data can be accurately and clearly represented. In the following section concerning the diagrams of frequency distributions, and histograms, this difference in presentation will be seen.

In Chapter 6, in the normal approximation to the binomial distribution, this distinction is again important. There a continuous variable is used to approximate the probabilities for a discrete variable. In doing this, the researcher has to decide which parts of the line interval most closely match the discrete values of the binomial probability distribution. A similar situation later occurs in the case of the chi square tests. These are fully applicable only when the sample size is sufficiently large, mainly because a continuous distribution is again being used to approximate a discrete distribution, and this is legitimate only in the case when a certain sample size is reached.

3.4 Conclusion

This chapter has discussed the various types of variables which are commonly used in statistical analysis. When working with a variable, the researcher should be aware of what type of variable it is, and which statistical measures and methods are most appropriate for that type of variable. Each variable can be either nominal, ordinal, interval or ratio. These levels are hierarchical in that all variables are nominal, but some such as sex or ethnicity are no more than nominal. Others allow ranking of the potential values and are considered ordinal. If the differences between these different values can be meaningfully measured, then the scale is interval. Finally, if the ratios between values are meaningful, then the variable is ratio. In addition, each variable is either discrete or continuous. One could have a continuous ordinal scale, a discrete interval scale, a continuous nominal scale, etc. That is, the scale of measurement and the discrete-continuous classifications are two separate and distinct types of classification for a variable.

In the next chapter, various forms of presenting data in tables, charts, and diagrams are presented. The manner in which the different types of variables affect the nature of this presentation is also discussed. In Chapter 5, summary measures of the data, measures of central tendency and of variation are presented. Again, the measures most appropriate for each of the types of variable is discussed.