CHAPTER 3

DATA ANALYSIS: DESCRIBING DATA

In the analysis process, the researcher tries to evaluate the data collected both from written documents and from other sources such as questionnaires, interviews, and experiments. A scientific method of analysis involves

- describing and explaining phenomena which incorporates the principles of empirical verification,
- defining the operational terms,
- explaining the procedures of the controlled observation,
- arriving at statistical generalizations, and
- providing empirical confirmation for the findings in order to make inference to the whole population.

The first three items are related to describing the data. For this reason, the procedures are discussed under the term "descriptive statistics." These first issues will be discussed in this chapter. Chapter 4 contains different types of graphic illustrations with which the descriptive data are represented. In Chapter 5, measures of relationship will be introduced for the purpose of making generalizations in relation to the given variables. Chapter 6 will be devoted to the application and computation of tests confirming the significance of the differences observed between groups.

The analysis of the data has to be done according to the questions or tentative statements formulated in the hypotheses (see Vol. 1, Chapter 3). The variables mentioned in each hypothesis have to be explicitly stated to inform the readers about the content of the study. The hypothesis of the study serves as a guidance as to what steps to take and what to focus on regarding the process of analysis. For instance, this is very well illustrated in the following example by Glidden (1964, p. 128). The study is on the airworthiness condition of an aircraft engine that can take two different directions depending on the way the hypothesis is formulated:

H_1 "Airworthiness can be determined by finding out the parts needed according to the given classification (pistons, connecting-rod, etc.) and evaluation (...etching, calipering, etc.)."

H_2 "Airworthiness of an engine depends on tolerance limits of oil consumption, compression and several similar functions."

When the two hypotheses are carefully examined, the difference on the procedure of analysis to be applied for each of these studies can easily be recognized. The first hypothesis requires the analysis of each part of the engine. The second hypothesis, on the other hand, is conceptual; therefore, it requires an experimental approach to gather reliable data.

Computers are excellent for processing mass data in a very short time. However, in order to make use of computers for analysis, it is essential to know what type of scale is represented by the collected data because different scales of measurement require the application of different statistical evaluation.

Types of Measurement Scales

Different kinds of scales permit different kinds of descriptive statements. There are four types of measurement scales:

1. Nominal Scale
2. Ordinal Scale
3. Interval Scale
4. Ratio Scale

Nominal Scale

The nominal scale refers to naming variables. In the nominal scale, the existing variable is indicated by letters or numbers. Presence or absence of the variable is indicated by signs such as (+/-) or (1/0).

In a nominal scale, numerals or letters are used to represent categories or levels of a variable. For instance, the letters M and F are used to name the characteristics of the subjects as male and female. In such a case, the findings of the research are analyzed according to these categories of sex. Such a scale classifies subjects into two or more categories. Under such categorization, each subject is a member of one category only depending on the given characteristics.

Ordinal Scale

The ordinal scale indicates the rank order, and it is used to measure how much of a variable is present within the sample population when the intervals between these ranks are not equal. An ordinal scale ranks the subjects depending on the degree of values of the variable they possess. For instance, subjects can be ranked depending on which year they are in at the high school they are attending, or at which mental age they are, or what score they have received from an examination. The ordinal ranking of these variables are expressed with terms indicating an order as in the following expressions:

- 1st, 2nd, 3rd, 4th, 5th etc.
- poor - fair - good - excellent

Interval Scale

Interval scale tells us how much of that variable is present. This type of measurement is used to study differences among levels of a variable. In an interval scale, each number represents a different level, and the distance or intervals between these levels are equal. If we take the years A. D. as a variable, we can say that 1980 is 20 equal units greater than 1960, but there is no zero point because we do not know the starting point of the years although the birth of Christ is given as a selected zero point.

The following is an example of indicating the results of a test both in ordinal and interval scales:

ORDINAL SCALE (ranking)	INTERVAL SCALE (test scores)
First	85
Second	80
Third	70
Fourth	60
Fifth	40

<u>Ratio Scale</u>

The ratio scale is the most precise type of measurement. It carries all the characteristics of the first three scales. Moreover, it has the distinguishing characteristic of an absolute (true) zero point. This distinct property of the ratio scale enables researchers to make comparisons in terms of ratios.

Most physical measures represent ratio scales, but psychological ones do not. For instance, mathematical operations are expressed in ratio scale because they represent quantity. Variables such as time, weight, distance, and height are expressed in ratio scale because they have a true zero starting point and equal intervals to operate on. Most of the psychological tests, on the other hand, do not have true equal intervals. The progress observed in subjects are expressed in terms of a movement from less refined to a more refined situation without mentioning any equal intervals between the initial and the final stage of the development.

In the application of statistics, one important issue that needs to be taken into consideration is that "a statistic appropriate for a lower level of measurement may be applied to data representing a higher level of measurement; the reverse is not true" (Gay 1987:342). Moore (1983) demonstrates the application of the variable height (normally a ratio scale) to other scales:

1. Ratio .... absolute height of a student; measured from zero inches to maximum height in class
2. Interval ... height of a student measured from a table top; requires an arbitrary zero point and equal intervals
3. Ordinal ..... relative height of students measured from tallest to shortest; requires the value of the tallest position in the series
4. Nominal ... the height of students categorized into three groups: tall, average, short; requires the value of 3, 2, and 1 for each group

If the variables are sex and the habit of smoking or nonsmoking, other scales cannot be applied because, in the nominal order, there is no ranking, nor any interval or ratio scales. Statisticians mention four properties in relation to these scales. The absence or presence of these properties are indicated in Table 3.1.

TABLE 3.1 Properties of Measurement Scales

Property	Nominal	Ordinal	Interval	Ratio
Distinctiveness	+	+	+	+
Ordering	_	+	+	+
Equal intervals	_	_	+	+
Absolute zero point	_	_	_	+

Descriptive Statistics

Descriptive statistics is used to describe the basic features of the data in a study. It provides simple summaries about the sample and the measures. Together with simple graphics analysis, it serves as the basis of virtually every quantitative analysis of data. For instance, the degree of success of each class on a specific examination can be described by giving the average score (mean) for each class. The mean of each class on that examination can be illustrated by means of tables or graphs (see also Chapter 4).

Descriptive statistics is used to present quantitative descriptions. Each descriptive statistical measure condenses the data into a comprehensible summary as in the case of a simple number used to summarize how well a specific group of students have done on a specific test. This single number may simply be the total number of scores divided by the number of students in the class.

Types of Descriptive Statistics

Descriptive statistics can be discussed under four main headings:

1. Basic measures
    - frequency
    - proportion
    - percentage
2. Measures of central tendency
    - mean
    - median
    - mode
3. Measures of variability
    - variance
    - standard deviation
    - range
4. Measurement of relative position
    - percentile rank
    - standard scores (z-score, T-score)

Basic Measures

After the scores are tallied on a sheet of paper or on cards, the frequency, the proportion, or the percentage can easily be calculated. Frequency is the number of subjects in a particular category. Proportion is the ratio of subgroups to the total group. It is expressed in decimal values. Percentage is the proportion of a subgroup to the total group, and is expressed in percentages.

Moore (1983, p. 243) gives the following example to demonstrate the distinctions among these types of measurements:

In a class of 200, there are 80 males and 120 females.

1. What is the *frequency* (f) of women in the class?

   frequency = headcount; so f = 120

2. What is the *proportion* (p) of women in the class?

   $$\text{proportion of women} = \frac{\text{subgroup}}{\text{total group}}$$

   so p = 120/200 = 0.60

3. What is the *percentage* (%) of women in the class?

   percentage of women = proportion x 100 ; so
   percentage of women = .60 x 100 = 60%

The graphic representations of these obtained results can be represented in graphs (see Chapter 4).

Measures of Central Tendency

Measures of central tendency help the researcher to describe a set of data using raw scores. As a result of the computation of these scores, measure of central tendency is obtained. The common measures of central tendency are mean, median, and mode.

Mean. The mean, which is symbolized as X (X bar) in statistics, is the same as the arithmetical average. It is used with interval or ratio scale variables. The formula to calculate the mean is as follows:

$$X = \frac{\sum x}{N}$$

1. X is the mean score.
2. ∑ is a summation sign.
3. ∑X means the sum of all the X scores.
4. N is the total number of scores.

Suppose we have the following scores for a class of 10:

Students	Grades	Calculation

1	52	
2	60	
3	67	
4	69	
5	69	700 - 10 = 70
6	71	Mean = 70
7	73	
8	76	
9	80	
10	83	
N =10	∑X= 700	

In order to calculate the mean of these scores, first, we add up all the scores to find the sum of the scores, which turns out to be 700. Then we divide this number by the total number of scores, which is 10. Thus 700 divided by 10 is 70.

Median. The median is defined as "the middle point in a distribution, or that point below which 50 percent of the scores fall, and logically above which 50 percent fall" (Brown, 1990, p. 67). It is used with ordinal scale as well as interval and ratio scales. In distribution with odd number scores, the median is the middle score. If we take the same list of scores given above, we see that the number of scores in the list is even; so we take the two scores in the middle (69 and 71). Then we take the average of these two scores, and in that case, the median is calculated as 70. Suppose there are 11 students in the class, and the eleventh student got 75. His score, when placed on the list, would take the 7th position; thus, 71 (the score indicated by the asterisk in the second column) would move upward to the middle of the list. Therefore, the median would be 71.

With even number of scores		With odd number of scores			
	52				52
	60				60
	67				67
	69	two middle scores		69	
69]	69 +71	= 140	69		
	71]	140÷ 2 = 70*		71*	
	73				73
	76				**75**
	80				76
	83				80
					83

The median is less reliable as compared to the mean because it is not sensitive to extreme scores. If the extreme scores are very high or low, it is not reflected in the median in a positive or a negative way. For instance, if the first two scores in the list were 20 and 35 instead of 52 and 60 respectively, the mean would go down to 65.3 while the median would still remain as 70 (indicated by the asterisk in the calculation for the scores listed in the first column).

Mode. The mode is the score that is attained by more than one subject in the group. In other words, the score that is *most frequently encountered* in a distribution of scores is the mode. If we refer to our list of scores, we see that 69 has been scored twice, and there is no other score that has been scored more than that. Thus, the mode is 69. The mode is used with nominal variables as well as with the other three scales. The mode is the least stable since it cannot use mathematical operations. For instance, in our group, there could have easily been two subjects who have scored 73. In that case, we would have claimed our mode to be 73.

Measures of Variability

We interpret the typical behavior of a group in the form of central tendency, and we try to investigate how individuals in the group vary from that typical behavior by measures of variability. Variance, standard deviation, and range are the common indicators utilized for this purpose.

Variance. Suppose a student who has scored 75 in the second achievement test wants to know how close his/her score is to the average for that test. If we consider the mean of the achievement test to be 73, then the student's score would be 2.0 points from the mean. Let us take another student with a score of 69. In this case, his/her score would be - 4.0 points from the mean. The minus sign is the indication of the score being below the average. This type of calculation only indicates the deviation of one score from the mean of the whole test. If we take the distribution of all scores into account, we would be measuring the variance. In statistical terms, the variance of this kind is indicated as s^2. This means that the variance is "the sum of the squared deviations about the mean divided by the number of scores" (Moore, 1983, p. 253). In calculating the variance, first, we subtract the mean from each score in order to find out how far each score is from the mean; then we square each difference. This application is indicated as $(X - X)$ in statistical terms where X represents the score of the individual and X is the mean. Then we take the sum of the squared difference. Let us apply this procedure to the scores in our original list:

Score	Mean	Difference	Difference square
(X)	(X)	(X - X)	$(X - X)^2$
52	70	- 18	324
60	70	- 10	100
67	70	- 3	9
69	70	- 1	1
69	70	- 1	1
71	70	+ 1	1
73	70	+ 3	9
76	70	+ 6	36
80	70	+10	100
83	70	+13	169

$\sum(X - X)^2$ = 750

In the second step, we divide the sum of the squared difference by the number of students. This is called the variance.

$$750 \div 10 = 75.0$$

Standard deviation. Standard deviation indicates, in standard units, how scores deviate from the mean. A large standard deviation suggests a large amount of variability of scores below and above the mean. A small standard deviation, on the other hand, indicates little variation.

In the calculation of the standard deviation, we apply the same procedures to calculate the variance, and in the third step, we take the square root of the obtained result (75.0) to arrive at the desired result:

Standard deviation (SD) = 8.66 = 8.7

If we have to formulate the procedures we have applied so far, we would see that the formula for the standard deviation is:

$$SD = V\sqrt{\frac{\sum (X - X)^2}{N}}$$

With new technology, standard deviation can easily be calculated by the use of computers or calculators when the raw data are entered.

We need two sets of scores to compare the development between the groups. Suppose the class we did the calculation for is our experimental group (Group A); and let us assume that we have the following results for our control group (Group B):

Central Tendency				Variability			
Groups	X	Mode	Median	Low	High	Range	SD
A	69	70	71	52	83	31	8.7
B	60	59	58	55	68	14	2.0

Small standard deviation, as in Group B, indicates that the scores are close together. If the mean and the standard deviation of the scores in a given group are known, it is easy to estimate what the possible low and high grades could be. This is because the mean always corresponds to the 50th percentile, and 99 percent of the scores fall into the plus or minus 3 standard deviation (- /+3 SD). **If scores are normally distributed** (see Figure 3.1.), the following generalizations can be made about the scores in Group A:

1. Since the SD for Group A is 9, we multiply 9 by three (8.7 x 3 = 26), and subtract this number from the mean to find the possible lowest score (70 - 26 = 34), and add the same number to the mean to find the possible highest score (70 + 26 = 96). Since the mean is 70 for this group, we can say that approximately 99 percent of all the scores in that group should be between 34 and 96. In our case, we see that 100 percent of the scores in Group A fall between 34 and 96.

The mean (70) is larger than the mode (69). Therefore, the distribution is negatively skewed. In other words, the extreme scores are at the lower end of the distribution. That is, most of the subjects did well but a few did very poorly. In a positively skewed distribution, where the mean is smaller than the mode and the median, the extreme scores will be at the upper or higher end of the distribution:

negatively skewed: mean > median > mode
positively skewed: mean < median < mode

2. Again, in a normal distribution, about 95 percent of the scores fall into -/+ 2 SD. Accordingly, 95 percent of the scores in Group A should fall between 52.6 and 87.4. Since our sample is too small, and the scores are not distributed within a normal curve, it does not reflect the details. The 100 percent of the attained scores still fall within -/+ 2SD.
3. In a normal distribution, approximately 68 percent of the scores fall into -/+ 1 SD. With our scores, when we add and subtract 8.7 from the mean, we get 61.3, and when we add 8.7 to 70, we get 78.7. As a result of this calculation, we expect 68 percent of our scores to fall between 61.3 and 78.7. When we look at our scores, we see that 60 percent of the scores fall into this slot. In spite of the sample being small, the obtained results are close to the normal distribution.

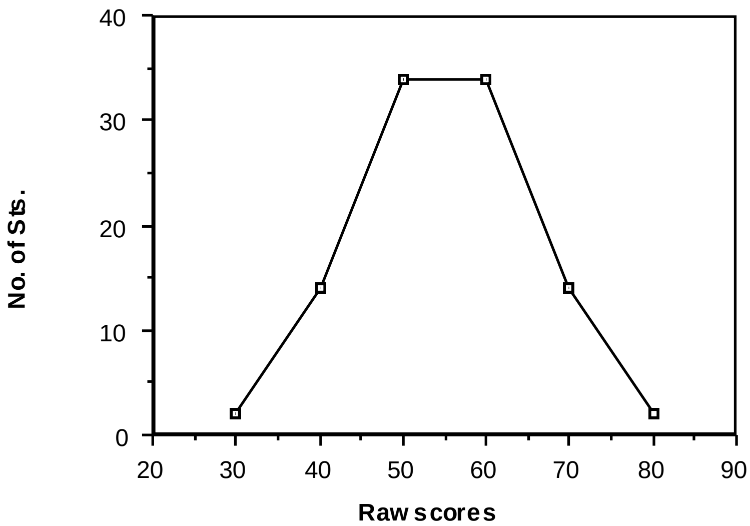

Figure 3.1 An example of a normal curve

Range: Range is "the most unstable and crudest measure of variability of scores within the distribution" Moore, 1983, p. 254). Range is the difference between the highest and the lowest score. In calculating the range, we subtract the lowest score from the highest score. For instance, let us suppose that there are five students in a class and they have the following scores: 40, 57, 78, 86, 90. By subtracting 40 from 90, we find the range as 50. Range does not allow us to make interpretation of a score within a distribution. In other words, the range does not enable us to make judgments regarding where a single score stands in terms of other obtained scores. It only allows us to see how skewed or wide spread the scores are within the distribution. If all the scores are very close, the range is narrow. This means that there is little variability among the scores. If the scores are widespread within the distribution, the range is said to be wide. This is an indication of a high variability among the scores.

Measures of Relative Position

These measures indicate the place of a score within a set of scores when it is compared to all other scores in the set. They can be used to compare the performance of the subject on two or more tests. We will mention two types of derived score measures here: percentile ranks and standard scores.

Percentile ranks. Percentiles are used to indicate the percentage of scores that fall below a given score. They are frequently applied on interval data but they are appropriate for computing data representing an ordinal scale. For instance, it has already been calculated that 45 percent of the students in a specific group have scored below 60 on a specific test. If the same result is to be expressed in terms of percentiles, it would be the 45th percentile. This means that 45 percent of the students scored equal to or below a given score (in our case it is 60).

In some international tests, the percentile of the score of the test taker is provided. If a person scores in the 76th percentile on TOEFL, this means that the score of this individual is equal to or higher than 75 percent of the people who have taken the test but scored lower than 25 percent of the total group of test takers.

Standard scores. Standard scores are appropriate for data representing an interval or ratio scale. T score and z score are the commonly used ones. "The z score expresses raw score performance in terms of the number of standard deviation units above or below the mean. It is standardized on X = 0 and s [SD] = 1" (Moore, 1983, p. 221). For instance, 40 in our set of scores falls within 3 SD below the mean and 71 falls 1 SD above the mean.

z score:

If the mean and the standard deviation of a distribution are known, the score of an individual can be located in the distribution by using z scores. The formula for z score is:

$$z = \frac{X - X}{SD}$$

According to this formula, the z score is calculated for any score (X) by subtracting the mean from the score and dividing the obtained result by the standard deviation of the same distribution. If we try to calculate the z score for a person whose score is 86 in Group A, we subtract 70 from 86, and we get 16. We divide 16 by 4 (which is the SD for the scores of this group), and we find the z score for that individual, which is calculated as +4. This means that the score of this individual is 4 standard units (which is the highest unit in the distribution) above the mean.

If we want to find out the probability level of this score (86) for any paper we pick up from the set belonging to this group (Group A), we make use of the normal distribution table. We know that in a normal distribution, 68 percent of the scores fall 1 standard deviation below or above the mean. The probability of the scores to fall between 1 and 2 standard deviations below or above the mean is 98 percent. Since the z score for the score we are concerned with (86) is +4, the probability of coming across a score this high (since it is positive) is very low. When we check this z score value on the table of distribution (Hatch & Lazaraton, 1991, p. 594; see Appendix C, Table C6), we see that the area between the mean and the z score is .49997 and the area beyond is .00004. This means that within the given set of scores, the change to come across another score higher than the given one (84) is .00004, which is almost impossible.

T score:

Application of the T score eliminates the minus and plus signs and the decimals used in the z score:

$$T = 10z + 50$$

If we take the score 60 from our list again, we calculate the z score as -2.5. In order to eliminate the negative sign and the decimals, we can apply the T score :

$$T = 10 \times (-2.5) + 50$$
$$T = -25 + 50$$
$$T = 25$$

The use of z scores and T scores has advantages over using raw scores in the analysis of the results. However, both are only appropriate if the scores are distributed normally; or else it is difficult to interpret these scores.

In this chapter, we have tried to discuss the simple measures which serve as the base to descriptive statistics. We have demonstrated how different statistics are used to compute or assess values.

EXERCISES

A. 1. Compare and contrast nominal, ordinal, and interval scales.
2. Name the scale the following variables are likely to belong to.
Variables: Turkish, second, intermediate, has experience abroad, female, 11 years old, score on a test.

B. In a longitudinal study, a Turkish child is observed for 10 weeks and the sequence of appearance of seven case markers (CM) were indicated on a chart noting the presence (1) or absence (0) of these case markers.

1. Study the chart and pose a natural hypothesis for the acquisition of these case markers.
2. Which case marker would you claim to be the easiest and which one to be the most difficult?
3. How would you design the chart if you had been observing several children in a cross-sectional study.

Week	CM 1	CM 2	CM 3	CM 4	CM 5	CM 6	CM 7
1	0	1	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	1	0	0	0	0	0
4	0	1	1	0	0	0	0
5	0	1	1	0	1	0	0
6	0	1	1	1	1	0	0
7	1	1	1	1	1	0	0
8	1	1	1	1	1	0	0
9	1	1	1	1	1	1	0
10	1	1	1	1	1	1	1

C. The following are writing course final test scores.
5, 7, 11, 20, 35, 47, 49, 50, 50, 52, 55, 65, 68, 70, 70, 70, 74, 75, 87,92

Calculate:
a) mean
b) mode
c) median
d) variance
e) standard deviation
f) z score for 20, 52, 74, and 87
g) T score for 20, 52, 74, and 87

D. Suppose you gave a reading test to 30 students. The mean for the group was 70 and the standard deviation was 5.

1) What is the probability of scoring below the mean?
2) What is the probability of scoring below 30?
3) What is the probability of scoring above 65?

4) Suppose that the passing grade is 60.
   a) What is the probability of not passing this particular test?
   b) What is the probability of passing this test?
5) Suppose any score above 80 is considered "A"
   a) Draw a distribution curve.
   b) Label the mean and standard deviation points.
   c) Shadow the areas for scores below the "A "point.
   d) Tell what proportion of scores fall in this area.
   e) Find out the probability of getting an "A" on this test.
6) Estimate the probability of scoring for that specific test:
   20, 52, 74, 87
7) Find out the percentile ranking for the following scores:
   20, 52, 74, 87

C. Below are the statistical results obtained from different observers regarding three different teaching materials. These materials have been rated from 1 to 5 with 5 being the highest score.

Material I	Mean	SD
Six observers		
Presentation	3.8	1.1
Format	3.7	1.4
Teachableness	4.4	.81
Student participation	3.8	1.2

Material II	Mean	SD
Five observers		
Presentation	4.2	.87
Format	4.3	1.0
Teachableness	4.6	.68
Student participation	3.9	1.1

Material III	Mean	SD
Seven observers		
Presentation	3.7	1.3
Format	3.3	1.4
Teachableness	4.1	1.1
Student participation	3.7	1.4

1. In which ratings have the observers most closely agreed?
2. In which cases do their ratings vary the most?
3. Can you think of any other way of displaying the given data ?

(Adopted from Hatch and Lazaraton, 1991, p. 185)