Chapter 11

Describing Bivariate Relationships

In Chapter 8 we looked at a number of ways in which we can describe the univariate data (data obtained from a single variable) drawn from a sample or population. In that chapter we pre sented visual methods of describing data, such as pie charts, bar graphs and frequency poly-gons, but we were much more extensive in our discussion of mathematical ways of describing data, particularly the measures of central tendency and the measures of dispersion. In this chapter, we will use these univariate descriptive statistics to build descriptive measures of relationships be-tween pairs of variables. By coupling these descriptions of relationships to the ideas of sampling distributions and probabilities (which we discussed in the preceding chapters), we will have the necessary tools to test hypotheses, as we'll see in the next chapter.

The purpose of this chapter is to outline the basic techniques for describing the relationship between two variables. Not all the statistics that can be used to measure relationships will be pre-sented here. However, it is important that we cover several necessary steps before discussing formal hypothesis testing.

First, we want to introduce the general strategy for detecting relationships. This strategy is based on the discrepancy between what we actually observe in the data, versus what we would expect to observe if there were no relationship between two variables. The logical method for de-tecting a relationship between variables is by contradiction: if we find a result that should not be there if no relationship existed, we conclude that there is probable cause to conclude that the rela-tionship really exists.

This conceptual strategy brings the Research and Null hypotheses discussed in Chapter 3 squarely back on center stage. The Null hypothesis states that no relationship exists between two variables. If we can visualize what data would look like if the Null hypothesis were true, then we can interpret deviations from this picture as evidence supporting the Research hypothesis that a relationship between the two variables does in fact exist.

There are two steps in determining if a relationship exists. The first step is to describe the magnitude of the relationship; the second is to determine if this magnitude is large enough to be considered unlikely to have happened by chance alone. This chapter focuses only on the first step; on outlining the various general ways in which we go about describing relationships. The next chapter will outline the second step of the process, which is the formal decision making process for testing whether the Null hypothesis or the Research hypothesis has the better probability of being correct. And in Chapter 19 we will present a general guide to methods for testing for the presence of relationships under different conditions of measurement and in different research settings.

Second, we will introduce the fact that the level of measurement determines our choice between two fundamentally different methods that we use to describe relationships between variables: contrasting groups versus measuring relationships directly.

Relationships and Levels of Measurement

If you remember, in Chapter 3 we presented an extensive discussion about the nature of relationships between variables. A relationship between two variables can be diagramed as shown in Figure 11-1.

(a) Relationship Hypothesis

Variable A
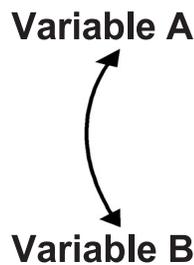
Variable B

(b) Comparative Hypothesis

$$\overline{B} \text{ for Level 1 of A} \quad \neq \quad \overline{B} \text{ for Level 2 of A}$$

FIGURE 11-1   **Diagramming a Relationship**

Both notations really represent the same relationship: different levels of Variable *B* will be observed whenever different levels of Variable *A* are observed, and the levels of Variable *A* and Variable *B* will be associated in some nonrandom fashion (that is, they will covary).

In order to say that a relationship between some variable *A* and another variable *B* indeed exists, the following criteria need to be met:

1.   Variable B must be observed under at least TWO levels of variable A; and

2.   The observed levels of variable B must be coordinated with the different levels of variable A, so that the levels of Variable A can be used to predict the levels of Variable B (and possibly vice versa). Criterion (2) is another way of stating that the variables must covary in a nonrandom fashion.

The level of measurement forces us to choose between these two different ways of stating the nature of the relationship. As you recall from Chapter 3, one way in which we can distinguish types of hypotheses is by separating them into the following two groups: (1) relationship hypotheses, those hypotheses that use expressions of the amount of covariance between the independent variables and the dependent variable as their basis, and (2) comparative hypotheses, which use statements of dependent variable differences among two or more groups which are known to differ on the independent variable to describe the relationship. These two types of hypotheses are presented in Figure 11-1 as (a) and (b), respectively.

Whenever we formulate a hypothesis about the relationship between two variables, we will need to consider the level of measurement of the two variables. Table 11-1 shows the various permutations of levels of measurement that might occur when we are examining the relationship between any two variables. From this figure, we can see that for some combinations of levels of measurement, only comparative hypotheses can be used to describe the predicted relationship between the two variables. For other combinations, however, relationship hypotheses are appropriate. And as is illustrated in Table 11-1, whenever relationship hypotheses are appropriately used, comparative hypotheses can also be stated.

Table 11-1   Levels of Measurement and Types of Hypotheses

		Independent Variable		
		Nominal	*Ordinal*	*Interval/Ratio*
Dependent Variable	*Nominal*	Comparative	Comparative	Comparative
	Ordinal	Comparative	Relationship Comparative	Relationship Comparative
	Interval/ Ratio	Comparative	Relationship Comparative	Relationship Comparative

In looking at Table 11-1 it is apparent that a comparative hypothesis is the only possible alternative whenever either the independent variable or the dependent variable or both variables are measured as nominal variables.

The reason is quite simple. Nominal measurement does not possess an underlying dimension that implies "more" or "less", while the other three levels of measurement do have this characteristic. The classes or categories of a nominal factor merely state that the observations which are assigned to the different categories are "different". In contrast, the classes in ordinal, interval and ratio variables add to this distinction the idea that the theoretical concept may be present to some degree, i.e., that measurement may show the presence of more or less of the quantity indicated by the concept.

An important consequence of this distinction is that a relationship involving at least one nominal variable cannot be characterized as, for instance, a positive or negative relationship. For example, Gender is a nominal factor, which means that we cannot say that Gender is positively related to Frequency of Initiation of Intimate Conversations. We can't say it because it doesn't make any sense. Although the Frequency of Initiation of Intimate Conversations variable clearly has an underlying dimension which allows us to translate "positive" to changes in the "more" direction, the variable Gender certainly does not have this characteristic. The nominal factor Gender is made up of two nominal variables: Male and Female. Which one should we call "positive" or "more", and which one "negative" or "less"? Neither should be.

Consequently, the only option we have is to state relationships involving nominal factors in

terms of the groups that are defined by the various classes which make up the nominal factor. For instance, we may represent the relationship between Gender and Frequency of Initiation of Intimate Conversations with some statement like "Males will initiate such conversations less often than will Females". To describe the magnitude of this hypothetical relationship, we will contrast Males (one class of the independent variable) and Females (the other class of the independent variable) on their performance for Frequency of Initiation of Intimate Conversations (the interval-level dependent variable). This contrast method forms the basis for a number of techniques for describing the degree of relationship between a nominal independent variable and other variables at whatever level of measurement.

Describing Relationships by Comparing Nominal Groups

A hypothesis involving a nominal independent variable will have some statement about how two (or more) groups, representing two (or more) classes of that independent variable, are expected to differ from one another on some dependent variable. Such a difference is described as a contrast between an appropriate univariate descriptive statistic for one group and the equivalent descriptive statistic for the other group or groups.

The kind of group contrast and the descriptive statistics involved are determined by the level of measurement of the dependent variable and by the number of groups to be contrasted.

Table 11-2   Levels of Measurement and Descriptive Statistics

		Descriptive Statistics			
		Proportion	*Mode*	*Median*	*Mean*
	Nominal	yes	yes	no	no
Measurement	*Ordinal*	yes	yes	yes	no
	Interval/ Ratio	yes	yes	yes	yes
yes = statistic is appropriately applied at this level					
no = statistic is inappropriate					

Comparisons and the Level of Measurement

In Chapter 8 we discussed the appropriateness of different univariate descriptive measures of central tendency and dispersion for different levels of measurement. Our discussion is summarized in Table 11-2.

If a relationship is to be described by contrasting two or more groups on some dependent variable, then such a contrast should be done with a statistic which is appropriate to the level of measurement of the dependent variable.

For instance, nominal measurement will allow us to contrast two groups in terms of proportions observed in the various classes of the dependent variable. Interval and ratio variables are appropriately represented by the proportion, too. However, we have argued before (and we will continue to remind the reader regularly) that important information will be "thrown away" if we treat interval/ratio measurement as if it was nominal or ordinal measurement. Such underuse of available information would occur if we were to contrast groups in terms of their proportions or medians if the level of dependent variable measurement was appropriate for contrasts in terms of means. The information contained in a measurement should be exploited to the fullest extent possible, by

using descriptive statistics at the highest level warranted by the data. We'll later give some examples to illustrate how contrasts between groups can be carried out at the highest level appropriate for each level of measurement of the dependent variable.

The Number of Groups to be Contrasted

The second factor we must consider when choosing an appropriate way to conduct such a contrast is the number of groups to be compared. This is determined by the number of classes or categories that constitute the nominal variable. The variable "Newspaper Subscriber" has only two classes: "Yes" and "No", and the test of a relationship between this variable and, for instance, knowledge of local political affairs, would require comparing the knowledge of subscribers and non-subscribers. But nominal factors may have more than two classes. For instance, we may wish to describe a relationship between Political Party Affiliation and the Frequency of Exposure to News Coverage of Politics. To describe such a relationship we will have to make contrasts of the Frequency of Exposure among groups of Republicans, Democrats, Communists, Socialists, Unaffiliateds and Others. Again, in Chapter 19 we will provide a guide to statistical methods for comparing multiple groups, regardless of level of measurement. We'll try to keep it as simple as possible in this chapter, but the importance of the number of classes to the correct choice of descriptive statistic should not be lost.

Nominal Group Contrasts with a Nominal Dependent Variable

When the criterion or dependent variable (the variable on which the groups are to be compared) is a nominal variable, the appropriate statistic is the proportion. We can multiply this by 100 (i.e., move the decimal point two places to the right) to convert this to the familiar statistic called

---

Exhibit 11-1 Testing a Comparitive Hypothesis: Group Contrast with Nominal Dependent Variable

An experiment is carried out to determine whether there is a relationship between the gender of a speaker and audience members' evaluations of that speaker.

Two students, one a female, the other a male, are audio taped while reading identical passages from a book. The two tapes are played to two groups of 50 male students each, where the first group listens to the tape of the male reader, and the second group listens to the tape of the female reader.

After listening, the participants are asked to fill out a questionnaire about the presentation. Among other questions, they are asked whether they thought that the person whose voice they just heard came from an "Urban" or a "Rural" background.

"percentage". The size of a relationship between the independent and the dependent variable is detected by looking at the proportions of cases assigned to different classes of the dependent variable, for each of the groups which represent the different classes of the nominal independent variable. We'll illustrate this with the example described in Exhibit 11-1.

In this exhibit the Independent Variable is Gender (of the speaker), which is a nominal factor with the classes "Male" and "Female". The Dependent Variable, Origin, is also a nominal factor which consists of the classes "Urban" and "Rural". The hypothesis of No Relationship can be stated like this:

H_0: The proportion of listeners who think the Male Speaker is "Urban" is the same as the proportion of listeners who think that the Female Speaker is "Urban".

Note that this H_0 implies that the proportion of Males who think that the speaker is rural should also equal the proportion of Females who think so. The corresponding nondirectional research hypothesis is:

H_R: Proportion of listeners who think the Male Speaker is "Urban" is NOT the same as the proportion of listeners who think that the Female Speaker is "Urban".

If there is no relationship between gender and place of origin, the female speaker and the male speaker should be equally likely to be identified as being from an "Urban" background. If there are differences in the proportion of listeners to the male and female speakers who identify the speaker

Table 11-3   Group Contrast with Nominal Dependent Variable Observed Frequencies in Cells

		Speaker		N
		Male	*Female*	
Origin	*Urban*	25	35	60
	Rural	25	15	40
		N=50	*N*=50	Total *N*=100

as "Urban", then we would be more likely to believe that a relationship exists.

We will have to do a formal test of this belief, to establish our confidence in its truth, but this will be discussed in the next chapter. From what we see in Table 11-3 it appears that the Female Speaker is somewhat more likely to be identified as "Urban" than is the Male Speaker.

How can we describe the relationship more accurately? Essentially, we must look at the pattern of proportions predicted by the null hypothesis, and contrast that with the pattern of proportions that we actually observe. But how do we determine the pattern of proportions that we would observe if the hypothesis of No Relationship were true?

We know from our results (Table 11-3) that of the 100 people who participated in the experiment, 60 (a proportion of .60, or 60%) identified the speaker as "Urban" and 40 (or .40, 40%) did not. If the Gender of the speaker did not matter (that is, the Null hypothesis is true), then we will expect that 60% will call the speaker "Urban" and 40% will call the speaker "Rural", regardless of whether the speaker is Male or Female. Since 50 people listened to the Male speaker, we would expect that .60 of these 50 listeners (or 30 listeners) will characterize the Male Speaker as "Urban".

Likewise, the proportion of subjects who characterize the Male Speaker as "Rural" will be the remaining .40 of these 50, which would be 20 individuals. Since the Null Hypothesis states that Gender does not matter, an identical pattern should hold for the Female Speaker. For the Males as well as the Females, the proportion who'd call the speaker Urban and the proportion who'd call the speaker Rural should be .60 and .40 respectively. The frequencies associated with these proportions, the 30 who would call the speaker Urban and the 20 who would call the speaker Rural, are called the expected frequencies. These are the frequencies that are expected, if the Null Hypothesis is true. We can then contrast the expected frequencies with the actual observed frequencies, as can be seen in Table 11-4.

The statistical test we'd apply would focus on the difference between these two frequencies. Generally speaking, the greater the difference between the two sets of frequencies, the greater the magnitude of the relationship and the lower the probability that the difference was due to a chance occurrence.

Table 11-4   Group Contrast with Nominal Dependent Variable, Observed and Expected Frequencies for Two Nominal Variables

		Speaker		N
		Male	*Female*	
Origin	*Urban*	(30)* 25	(30) 35	60
	Rural	(20) 25	(20) 15	40
		N=50	*N*=50	Total *N*=100
* Expected frequencies are in parentheses				

Nominal Group Contrasts with a Discrete Ordinal Dependent Variable

When the dependent or criterion variable (the variable on which the groups are to be compared) is an ordinal variable, a number of descriptive statistics are appropriate for contrasting the groups representing the different classes of the independent variable.

Because ordinal measurement can take two forms, two different statistics can be used for making contrasts. The statistic encountered in the previous example, the proportion or percentage, can also be used here, as we'll see. The other statistic appropriately used with data at this level of measurement is the median. And whether we would use the proportion or the median as the statistic of choice depends on the type of ordinal measurement being used. Let us continue with our running example to explain this further.

The level of measurement associated with the dependent variable is ordinal; there is an underlying dimension (amount of liking) associated with the variable, but the dimension is divided into a set of five discrete categories which the respondents were requested to use. Discrete refers to the fact that there is no possibility of finer distinctions in categorization. Respondents have to either chose "Dislike" or "Dislike very much"; they cannot chose a location in between these two.

The results, presented in Table 11-5, indicate that, of all those who listened to the Male Speaker,

Table 11-5   Group Contrast with Discrete Ordinal Dependent Variable

		Speaker		N
		Male	*Female*	
Liking	*Like Very Much*	2 (10)*	18 (10)	20
	Like	10 (15)	20 (15)	30
	Neither Like nor Dislike	10 (10)	10 (10)	20
	Dislike	18 (10)	2 (10)	20
	Dislike Very Much	10 (5)	0 (5)	10
		N=50	*N*=50	Total *N*=100
*Expected Frequencies are in parentheses.				

Exhibit 11-2 Testing a Comparitive Hypothesis: Group Contrast with Discrete Ordinal Dependent Variable

After the research participants listened to the either the tape of the Male or of the Female speaker, they were asked to use a scale to indicate their liking of the speaker. The scale used was the following:

The research participants were requested to place a check mark in the space on the scale which most closely represented their feelings about the speaker.

Like very much	Like	Neither Like nor Dislike	Dislike	Dislike very much
|	|	|	|	|

---

2 indicated that they liked him very much, 10 said that they liked him, etc. The analysis necessary to determine whether the Gender and Liking variables are related is analogous to the procedure outlined above for contrasting groups on nominal dependent variables.

The H_0 to be tested is the following:

$H_{0:}$ The respondents who listened to the Male speaker and the respondents who listened to the Female speaker will use the response categories in the same proportions.

Associated with the H_0 is the following nondirectional H_R:

$H_{0:}$ The respondents who listened to the Male speaker and the respondents who listened to the Female speaker will use the response categories in different proportions.

Again, from our results we know that of the 100 people who participated in the experiment, 20 (or .20) liked the speaker "Very Much", 30% "Liked" the speaker, etc. By applying the same procedures as in the previous example, we can calculate the expected frequencies for each cell in the table. Once again, these frequencies (those expected under the Null hypothesis) are those that are presented between parentheses in the cells of the table above. Note that the discrete ordinal dependent variable in this example was treated as if it was a nominal variable.

Nominal Group Contrasts with a Continuous Ordinal Dependent Variable

Ordinal measurement, however, does not necessarily result in a discrete variable. Another way in which an ordinal variable can appear is as a continuous measure. This is a variable which potentially has an infinitely large number of classes or categories (even if all these classes do not appear in the operational definition of the variable).

In Exhibit 11-3 we present another variation on our current example, but this time with a continuous ordinal variable as the dependent variable.

The data that we have obtained represent continuous ordinal data. The system of measurement is ordinal because the numbering system merely represents more or less liking. We are not assuming that the measure is interval, as there is no standard amount that separates a single unit of "Liking". The amount of distance between 1 and 2 may be much smaller than the amount between 9 and 10. But the measurement is continuous in that any one of a potentially infinitely large number of values between 1 and 10 could have been selected by a respondent. Respondents could reply "1.5" or even "2.32145".

Exhibit 11-3 Testing a Comparitive Hypothesis: Group Contrast with Continuous Ordnial Dependent Variable

Imagine that the research participants in our study had been requested to use a different scale to evaluate the speaker. Assume that they had been asked to do the following:

"Please rate the speaker you have just heard on a scale from 1 to 10, where 10 means that you liked the speaker very much, and where 1 means that you did not like the speaker at all".

Every listener will use this scale to rate the speaker, and the result of this process will be two frequency distributions, one for each of the speakers:

	Speaker		
Male		Female	
Score	Freq	Score	Freq
1	6	1	0
2	4	2	0
3	8	3	1
4	10	4	1
5	5	5	4
6	5	6	6
7	6	7	8
8	4	8	12
9	2	9	10
10	0	10	8
	N=50		*N=50*

Data at this level of measures opens some options in the computation of the descriptive statistics. By the logic laid out in Table 11-2, the frequencies obtained here can be readily converted to proportions, and we could revert to treating the ordinal responses as if they were nominal. A better alternative, however, is to compute the median and use this statistic to contrast the two groups.

The logic behind using the median as a descriptive statistic for contrasting groups is this. If the independent variable "Gender" does not have an effect on the speaker evaluation, then the frequency distributions of ratings for each gender should have the same value for the median and this median should be equal to the median of the combined frequency distributions.

As discussed in Chapter 8, the value of the median is the one which divides a distribution in half, so if the Null Hypothesis is true, .50 of all cases in each distribution should be above the median of the two groups combined, and the remaining .50 of cases should be below that median.

How the median for the combined frequency distribution can be computed is illustrated below in Table 11-6.

We can use this median to test the H_0:

H_0: The Proportion of scores above the combined median in the frequency distribution associated with Male Speaker should be equal to the proportion of scores above the combined median for the frequency distribution associated with the Female Speaker and should equal .50;

which is simply a very explicit restatement of:

$\text{Median}_{\text{Male}} = \text{Median}_{\text{Female}}$.

The associated H_R is:

H_R: The Proportion of scores above the combined median in the frequency distribution of the Male Speaker should NOT be equal to the proportion of scores above the combined median for the frequency distribution of the Female Speaker and should NOT equal .50.

which could be restated simply as:

$\text{Median}_{\text{Male}} \neq \text{Median}_{\text{Female}}$.

Table 11-6   Group Contrast with Continuous Ordinal Dependent Variable

Combined Frequency Distribution		
Score	*Frequency*	*Cumulative Frequency*
1	6	6
2	4	10
3	9	19
4	11	30
5	9	39
6	11	50
7	14	64
8	16	80
9	12	92
10	8	100
N = 100		

Using the formula for the median from Chapter 8:

$$\text{Median} = \text{TLL} + \frac{N/2 - cf}{F}(I)$$

$$= 5.5 + \frac{100/2 - 39}{11}(1)$$

$$= 5.5 + 1 = 6.5$$

Since the median of the combined frequency distributions equals 6.5, we will observe 50 of the 100 Liking scores to be above 6.5, and the remaining 50 to be below 6.5. Likewise, under a true H_0, we would expect 25 of the 50 ratings (.50) of the Male Speaker to be above 6.5, and 25 to be below that value. The same pattern of responses would be expected for the Female Speaker's ratings, as can be seen below in Table 11-7. In that table the frequencies expected under the true H_0 (seen there in parentheses) can then be contrasted with the frequencies of scores actually observed to be above and below the combined frequency distribution's median of 6.5 in each of the frequency distributions for the Male and Female speaker. We simply go through the two frequency distributions and find the number of scores greater than, and those less than, 6.5. Again, the same logic for finding evidence of a relationship prevails. The greater the discrepancy between the values expected given a true Null Hypothesis and those actually observed the stronger the relationship between the independent and dependent variables.

Nominal Group Contrasts with an Interval/Ratio Dependent Variable

When the criterion variable is an interval or ratio variable, the descriptive statistic that is most appropriate for contrasting the groups which represent the different classes of the independent variable is the mean. This statistic exploits all the measurement characteristics in these levels of

---

EXHIBIT 11-4   Testing a Comparative Hypothesis: Group Contrast with Interval/Ratio Dependent Variable

After the research participants rated the speakers in terms of liking, they also filled out a brief questionnaire to measure their recall of factual information provided in the book passage which was read to them by the Male and Female Speakers.

A series of 15 questions were prepared and each listener was assigned a recall score reflecting the number of questions he or she could correctly answer. The following results were obtained:

Speaker			
Male		Female	
# Correct	Freq	# Correct	Freq
6	3		
7	4		
8	8	8	3
9	10	9	1
10	5	10	4
11	5	1	7
12	6	12	7
13	4	13	11
14	4	14	9
15	1	15	8
	$N=50$		$N=50$

$$\bar{X} \, \text{Recall}_{\text{Male Speaker}} \qquad \bar{X} \, \text{Recall}_{\text{Female Speaker}}$$

$$= \frac{\sum f \cdot x_i}{N} \qquad\qquad = \frac{\sum f \cdot x_i}{N}$$

$$= \frac{500}{50} = 10.0 \qquad\qquad = \frac{623}{50} = 12.46$$

---

measurement, in which the classes of the dependent variable carry not only the idea of "more" and "less", but also have constant and unvarying units of measurement which separate the classes.

We'll return again to our running example.

This recall data is ratio (and thus interval) in nature, which allows for the computation of the Mean as the descriptive statistic. We can use the mean to contrast groups who heard the Male and Female speakers. Although contrasts using proportions or medians are also possible with interval/ratio measurement, such practices are not recommended because they are less sensitive and therefore likely to be less accurate.

The H_0 associated with this problem is the following:

H_0: The Mean recall of information of subjects who listened to the Male speaker is equal to the Mean recall of information of subjects who listened to the Female speaker, or:

$\bar{X}$ Recall Male speaker = $\bar{X}$ Recall Female speaker

and the associated nondirectional H_R:

H_R: The Mean recall of information of subjects who listened to the Male speaker is NOT equal to the Mean recall of information of subjects who listened to the Female speaker, or:

$\bar{X}$ Recall Male speaker > or < $\bar{X}$ Recall Female speaker

The logic for determining what contrast should be observed if the H_0 were true is very simple indeed. If the Gender of the speaker is not related to recall of information provided by that speaker,

then the average level of recall should be the same for both groups of listeners. Consequently, a true H_0 ought to yield the following:

$\overline{X}$ Recall Male speaker – $\overline{X}$ Recall Female speaker = 0

The greater the difference between the two means, the more likely it is that a relationship between the two variables exists, and the less likely that such a difference was due to sampling error.

Describing Relationships by Comparing Groups Created from Ordinal or Interval Variables

An independent variable does not necessarily have to be a discrete nominal or ordinal variable to be used in a comparative hypothesis. Any continuous independent variable, either ordinal, interval or ratio, also lends itself to this type of hypothesis. However, such continuous variables must be reduced to a set of discrete classes, before the techniques described above can be applied. The general strategies described above merely require that the independent variable consist of at least two classes, each represented by one groups. Whether these groups have an underlying dimension is quite immaterial to the analysis. We can choose to ignore the additional information supplied by the underlying dimension, if we wish.

The process of defining groups by collapsing ranges of a continuous independent variable is not recommended, since doing so forces a very large number of potential differences to fit in a relatively small number of nominal classes. As we've explained previously, this will cause us to lose some of the information contained in the data.

We might accept this loss and choose to categorize ordinal or interval independent variables, however, if the dependent variable is a nominal variable and we want to use proportional comparison procedures. But if the dependent variable is also a continuous ordinal or interval variable, then establishing whether a relationship exists can be much more efficiently handled by techniques that measure relationships directly without ignoring available information in the data. These techniques measure the degree to which there is simultaneous variation of both continuous variables and will be introduced in the next part of this chapter.

Describing Relationships by Measures of Association

The methods for directly assessing a relationship between variables assume that both the independent and the dependent variable are ordinal, interval, or ratio in nature. In this section of this chapter we will outline, in general terms, the various ways of determining directly whether a relationship exists. We will present two variations on a conceptual approach to the measurement of the degree and the nature of association between variables which are at least ordinal. Once again we will see that the choice of procedures is dependent upon the level of measurement of the independent and dependent variables. The specific computations required to use these statistics can be found in any statistics text and we'll defer descriptions of the computations until later in this book. Here we'll focus on the logic behind the procedures.

Measures of Association and Relationship Hypotheses

Direct measures of relationships are appropriate when we have formulated a relationship hypothesis. When two nonnominal variables are thought to be related to one another, changes in the magnitude of one of the variables should be associated with changes in magnitude of the other variable.

There are two general strategies for determining whether a relationship exists, and the choice between them depends on the level of measurement of the two variables thought to be covarying.

The presence of an underlying dimension is shared by both ordinal measurement and interval/ratio measurement. Both categories of measurement convey a sense of magnitude. Interval and ratio measurement, in addition, are characterized by equal intervals between classes or categories. This characteristic is critical to this discussion, because of the constraints it imposes on the use of measures of association between variables that do not share the same level of measurement. Let us first turn our discussion to a measure of relationship between ordinal variables, which do not have equal intervals between classes.

Relationships between Ordinal Variables

The general strategy for measuring relationships between two ordinal variables is based on agreement/disagreement in observations' rankings of both the independent and the dependent variable. The ordinal nature of measurement will allow the N observations of the independent variable to be arranged in order of increasing or decreasing magnitude.

To these ordered observations we can then assign the rank numbers 1 through N (or N through 1). Observations' scores on the dependent variable can be ordered and have ranks assigned in the same way as they were for the independent variable. Let us look at an example.

If the ranks largely agree with one another, that is, we see that each organization has similar ranks on both variables, we infer that a positive relationship exists between Degree of Change and Reliance on Hierarchical Communication. In fact, if the rank numbers for both the X and the Y variable are identical, we have a perfect positive relationship. Conversely, in the event we observe totally dissimilar rank pairs we draw the conclusion that a perfect negative relationship exists. In this case high ranks on one variable would always be associated with low ranks on the other variable. Inconsistent results (for instance, high ranks on one variable are observed to be associated with both high and low ranks on the other variable) are indicative of a Null relationship.

The directional H_R regarding the relationship between Degree of Change and Reliance on Hierarchical Communication is the following:

Exhibit 11-5   Testing a Relationship Hypothesis: Ordinal Measurement for Both Variables

A researcher is interested in establishing the relationship between the "Degree of Change" confronting an organization and that organization's "Reliance on Hierarchical Communication". "Degree of Change" is measured by presenting key informants in the organization with a number of statements about change in organizations (e.g. "There is something different to do every day") The informants rate how true these statements are for their organization on a seven-point scale ranging from 1 (not at all true) to 7 (completely true). "Reliance on Hierar-

chical Communication" is measured using a set of similar statements (e.g. "I have to ask my boss before I do almost anything") and the same rating scale.

For every organization a total "Degree of Change" score is computed by adding all the scales, as is a total "Reliance on Hierarchical Communication" score. However, the researcher is not willing to assume that the measurement scales have equal intervals, so the resulting scores are to be considered as ordinal data. The following results were obtained:

	Scores		**Rank**		
Org.	*Degree of change (X)*	*Reliance on hierarchical communication (Y)*	*X*	*Y*	*Difference in rank*
A	53	27	1	7	-6
B	50	25	2	8	-6
D	47	32	3	6	-3
G	43	27	4	4	0
F	36	45	5	2	3
H	35	53	6	1	5
E	30	42	7	3	4
C	23	35	8	5	3

H_R: Degree of change $\rightarrow$ Reliance on hierarchical communication

This hypothesis reflects the reasoning that increased change requires that decisions be made at lower levels in the organization, without communicating with higher-ups. The associated H_0 is:

H_0: Degree of change   Reliance on hierarchical communication

The first step in determining whether the variables are related is the conversion of the variable raw scores to ranks, as is shown above in Exhibit 11-5. The next step is to determine, for any given organization, the discrepancy or difference in rank numbers between its rank on the X variable and its rank on the Y variable. We then look at the pattern of these discrepancies across the whole set of observations. In Exhibit 11-5 we see that the pattern of differences between these ranks in our data is quite revealing: high ranks on the Degree of Change variable tend to be associated with low ranks on the Hierarchy of Communication variable. This indicates some support for our research hypoth-

Table 11-8   Differences in Ranks of Variables for a Null Relationship

Observation	Variable *X* Rank	Variable *Y* Rank	*D*
A	1	7	-6
B	2	1	+1
C	3	6	-3
D	4	5	-1
E	5	4	+1
F	6	3	+3
G	7	8	-1
H	8	2	+6

esis. It is instructive to compare these results with the kind of results that we would see if the Null Hypothesis was correct. Table 11-8 shows a set of ranks and their differences typical of an absence of a relationship.

Notice that the observations which have a high rank on the X variable (e.g. *A* and *B*) have both a high rank (*B*) and a low rank (*A*) on the Y variable. The same can be said for all observations with low ranks on the X variable. They are associated with both high and low Y ranks. The ranking of Y cannot be predicted by knowing the rank on X. This lack of predictive power indicates that the two variables are not related.

We focus on the extent to which a pattern of rank differences deviates from a random pattern to tell us the extent to which the variables are related (the patterning of ranks) and the direction of the relationship.

Relationships between Interval/Ratio Variables

Interval and Ratio level variables share the characteristic of equal intervals between classes in the variables. This characteristic has two significant consequences. First, equal intervals in measurement allow for the computation and inspection of the mean as the key descriptive statistic. Second, we can conceptualize covariation differently from the way we conceptualize covariation among ordinal variables. With ordinal variables, we are restricted to talking about variation in ranks, without considering the distances between adjacent ranks (as these may be quite different, for different

rank numbers). With interval level data, these distances become part of the definition of the dispersion of the variable, as we saw in Chapter 8. We can then develop another conceptualization of covariation.

In the context of interval/ratio variables covariation is defined as the coordinated, simultaneous deviation of an observation from the means of two distributions. Positive covariation occurs when either (1) observations that are above the mean for the *X* variable will also be above the mean for the *Y* variable by similar distances, or (2) observations for which the *X* value falls below the mean also will have *Y* values that fall similar distances below the mean. This pattern of similar deviations defines a positive relationship. The reverse is true for negative relationships: *X* values above the mean are associated with *Y* values below the mean, and vice versa.

In this conceptualization of covariation we retain the idea that two variables can be related in either a positive or negative direction. This idea was introduced with ordinal variables, but add to it the following: that observations' distances from their respective means for the *X* and *Y* variables must also be coordinated, in order to define a relationship. This is a more stringent requirement, and it gives us a better description of the relationship between the variables.

When the direction of the deviation for one variable is observed in conjunction with deviations of the other variable in both positive and negative directions, or if large deviations on one variable are seen in conjunction with both large and small deviations of the other variable, then we will conclude that there is no relationship between the variables. In interval/ratio covariance both the direction and the magnitude of the deviation from the means must be coordinated before we conclude that there is a relationship. Exhibit 11-6 shows an example of a relationship which is

---

Exhibit 11-6   Testing a Comparative Hypothesis: Interval/Ratio
                          Measurement for Both Variables

Let us look again at the "Degree of Change" occurring in an organization and let us see this time how this variable might be related to the degree of Communicated Formalization in the organization. Communicated Formalization refers to the extent to which job descriptions and functions have been explicitly made available for employees.

For the same eight organizations used in the previous example, we have defined the "Degree of Change" by measuring the number of new products that have been introduced by the organization in the past year. "Communicated Formalization" has been defined as the percentage of positions in the organization for which written job descriptions exist.

The following results were obtained:

			Deviations from Means	
Organization	# of New Products	% with Job Descriptions	New Prod.	Job Descr.
A	6	20	+1.625	-18.12
B	5	25	+ .625	-13.12
C	3	35	-1.375	-3.12
D	5	32	+ .625	-6.12
E	3	50	-1.375	+11.88
F	4	45	- .375	+6.88
G	4	40	- .375	+1.88
H	5	60	+ .625	+21.88
	Σ = 35	Σ = 307		

$$\bar{X}_{\text{No. of new products}} = 35/8 = 4.375$$

$$\bar{Y}_{\text{Percent with job descriptions}} = 307/8 = 38.12$$

---

being investigated using variables at the interval/ratio measurement level.

We will propose the following directional Research Hypothesis about the relationship between these variables:

H_0:   # of New Products $\rightarrow$ Positions with Job Descriptions

based on the theoretical linkage that the greater the change occurring in an organization, the less sense it makes to define jobs, as they are bound to undergo change. The associated H_0 is:

H_0:   No. of new products   Positions with Job Descriptions (%)

In Exhibit 11-6 we see the extent to which the data reveal a pattern of "simultaneous deviation" from the means of the two data distributions. Marking the individual organizations' locations on these two distributions with + signs for positive deviations and - signs for negative deviations reveals a fairly consistent pattern of negative association between the two variables. Positive deviations on the New Products variable are associated with negative deviations on the Job Description variable, and vice versa. Furthermore, the larger deviations in the X variable tend to be associated to some degree with larger deviations in the Y variable. The logic for distinguishing between a non-Null and a Null relationship follows exactly the same logic as was used above for ordinal variables. But it adds the idea of the magnitude of the deviations to the simpler inspection of the direction of the deviations.

Mixed-Measurement Relationships

Mixed measurement relationships occur whenever the variables we expect to be related are measured at different levels.

An example of such a situation could easily occur in the last few examples we have used. Imagine, for instance, that we are interested in exploring the relationship between "Reliance on Hierarchical Communication" and "Communicated Formalization".

Recall that the measurement we used for "Reliance on Hierarchical Communication" was ordinal, while the "Communicated Formalization" measurement was interval. To determine the degree of association between these two variables we will have to use the measure of association that is appropriate for describing both variables. This will always be the measure of association that is appropriate for the variable characterized by the lowest level of measurement.

Since the variable "Reliance on Hierarchical Communication" was operationalized at the ordinal level, it does not allow the computation of means. So, for example, using a method of establishing relationships based on deviations from means (like that described in the last example) would be inappropriate. The only option that remains is to treat the interval/ratio variable as if it were merely an ordinal variable and to determine its relationship with the true ordinal variable (Reliance) by using the method based on agreement-disagreement between ranks. This will work, but we will lose the magnitude information contained in the interval variable. Once again, this points out the importance of creating operational definitions that are at the highest level of measurement possible.

Summary

In this chapter we have outlined the interdependency that exists between measurement, descriptive statistics for characterizing the size of relationships, and the basic procedures used for detecting relationships between variables. Students (and some communication researchers) sometimes tend to draw a line between communication theory construction and research design on the one hand, and data analysis on the other, considering them to be almost unrelated activities. We hope we have shown in this and in previous chapters that there is an uninterrupted path from the theoretical definition of concepts to the operational definition, with its associated level of measurement, to the detection and testing of hypothetical relationships. The way in which an independent variable is operationalized will dictate whether comparative hypotheses or relationship hypotheses will be appropriate. Decisions that are made when the dependent variable is operationalized will determine on what descriptive statistic the contrast between the groups will be made. Decisions

regarding measurement should always be made with intended analysis procedures explicitly considered.

We have outlined very basic strategies for describing relationships at the measurement level. The fundamental strategy consists of comparing the observed pattern of some descriptive statistic with the pattern which we would expect to find if no relationship existed. Deviations from the No Relationship pattern are taken as evidence for the existence of a relationship. The greater the deviation, the greater the presumed strength of the relationship.

The kind of contrast that we can make is determined by the lowest level of measurement of either the independent or the dependent variable. We can always reduce a higher level of measurement to a lower one and use contrast procedures appropriate for the lower level. But when we do so, we lose information and statistical sensitivity. In describing relationships and testing hypotheses, we should always strive to operate at the highest level of measurement possible, as this will give us the most powerful way of scrutinizing our theoretical statements.

In future chapters we will explore in greater detail some variations on the basic strategies for describing and testing relationships that we have outlined here.

References and Additional Readings

Buchanan, W. (1984). *Nominal and ordinal bivariate statistics: The practitioner's view*. In H. Asher, H.F. Weisberg, J. H. Kessel & W. Phillips Shively (Eds.), Theory building and data analysis in the social sciences (pp. 189210). Knoxville: University of Tennessee Press.

Loether, H.J. & McTavish, D.G. (1974). *Descriptive statistics for sociologists: An introduction*. Boston: Allyn & Bacon. (Part III, "Descriptive Statistics: Two Variables").

Weisberg, H. F. (1984). *Fundamentals of data analysis*. In H. Asher, H.F. Weisberg, J. H. Kessel & W. Phillips Shively (Eds.), Theory building and data analysis in the social sciences (pp. 151-188). Knoxville: University of Tennessee Press.