6.6: Choosing Statistical Tests

Inferential statistics are used to test hypotheses and build upon and beyond descriptive statistics. Descriptive statistics can be used to summarize things such as counts, means, and standard deviations, yet they are unable on their own to tell you whether differences in any of those are likely true beyond the samples whose data were being summarized. Inferential techniques take descriptive statistics such as these into further calculations to estimate whether differences or patterns in them are strong enough to indicate a likely difference or pattern in the broader population. Therefore, inferential statistics are not fully distinct from descriptive statistics; instead, they subsume various descriptive statistics into more complex calculations to establish the probabilities that they are indicative of population truths.

Breaking Down a Hypothesis for Testing

There are many different inferential techniques that can be used to test hypotheses. Each inferential test fits specific kinds of hypotheses and their corresponding data. Inferential testing starts with the statement of a hypothesis. Thus, a hypothesis must be stated before an appropriate inferential test can be chosen to test it. Once stated, a hypothesis can be broken down to identify:

1. How many variables are included
2. How many of those variables are quantitative (such as scores) or qualitative (such as group names)
3. Which of those variables, if any, are considered independent variables and which, if any, are considered dependent variables
4. The proposed nature of those variables in regards to one another

Let's take a look at a hypothesis and walk through each of these. Suppose that it is hypothesized that:

> Cholesterol levels will be lower for those who ate oatmeal daily for six weeks compared to those who did not.

Number of Variables

First, identify how many variables are in the hypothesis. Remember, a variable is anything that is measured and is not always the same. There are two proposed variables in this hypothesis:

1. Cholesterol Level and
2. Oatmeal Consumption.

Types of Variables

Second, identify whether each of the variables is quantitative or qualitative. Cholesterol levels are quantifications of how much cholesterol is circulating in the bloodstream. Thus, cholesterol level is a quantitative variable. Oatmeal consumption is qualitative in this example. This is because members of the sample are being grouped based on whether they did or did not consume the oatmeal for the specified timeframe. Comparisons of one group to the other are going to be made. However, you may ask yourself whether oatmeal consumption can be quantitative, and rightly so because it can be in other hypotheses. It is possible to measure the *amount* of oatmeal each person consumed rather than to identify simply whether or not they had consumed it for a specified time frame. However, what matters when identifying variables to choose an inferential tests is not how a variable *could* be measured but, instead, how it is being identified for measurement in the current hypothesis. In the aforementioned research hypothesis, it is clear that oatmeal consumption is being measured qualitatively. Therefore, for this hypothesis, there is one quantitative variable (cholesterol level) that is continuous in nature and one qualitative variable (oatmeal consumption).

Third, differentiate between the independent variable(s) and the dependent variable(s). Recall that an independent variable is the theorized or hypothesized cause and the dependent variable is the theorized or hypothesized affected in a cause-effect relationship (for a review, see Chapter 1). Not all studies have variables which are true independent variables and dependent variables. Instead, some hypotheses are about whether a pattern or difference exists, without positioning variables as causes and effects in relation to one another. However, it is necessary to distinguish between them when applicable. In addition, sometimes a theory on which a hypothesis was generated specifies cause-effect even when the hypothesis does not do so overtly or the data collected cannot be used to deduce cause-effect. We will return to these caveats and expand on them in later chapters where they are most applicable. For now, let's continue breaking down the components of our example hypothesis.

Though causal language is not used in the hypothesis, it may be implied by the structure of the hypothesis. Oatmeal consumption is occurring first, over the course of six weeks, before cholesterol levels are measured in the two groups. Thus, oatmeal consumption is being positioned as though it is an independent variable and cholesterol is being positioned as though it is a dependent variable.

Thus, if the method of data collection is in keeping with this, oatmeal consumption is a qualitative, independent variable and cholesterol is a quantitative, dependent variable.

Proposed Relationship between Variables

Finally, we must identify the nature of the connection, pattern, or distinction between the variables which is being proposed in the hypothesis. Though there are many forms of relationships that can be proposed and tested, some of which can be quite complex, the focus of this section will be limited to the forms of tests covered in this book.

Common relationships that can be proposed and tested are:

Group Differences

1. That the counts of groups are different
2. That the means of groups are different

Patterned Connections

3. That variables trend, or form a pattern, together
4. That established trends/patterns can be used to predict connections in new data

Let's first decide which category best fits the proposed relationship in the aforementioned hypothesis. We have a grouping independent variable so we turn to the Group Differences options. Now we need to decide whether our other variable quantifies how many cases have cholesterol (option 1) or whether cholesterol is being measured in each case allowing a mean cholesterol level to be computed for each group (option 2). Amount of cholesterol is being measured for each case. As we saw in Chapter 3, quantitative variables are often summarized using means. Thus, mean cholesterol scores would be computed for each group. Together this means that option 2 (That group means are different) is the best categorization for the proposed hypothesis.

It is often useful to create a summary as you progress through any processes in research and statistics to keep track of all the necessary information. In keeping, here is a summary of the hypothesis and how we have broken it down:

Hypothesis: Cholesterol levels will be lower for those who ate oatmeal daily for six weeks compared to those who do not.

Variables	Variable Type	Variable Position (if any)
1. Cholesterol Level	Quantitative (continuous)	Potential DV
2. Oatmeal Consumption	Qualitative	Potential IV

Proposed Relationship: The means of groups are different.

> **Note**
>
> Hypotheses are often presented in future tense before data are collected and in past tense after data have been collected and analyzed.

Summary of Inferential Tests

Once a hypothesis has been broken down, we can review the options for inferential testing and choose the one that best fits the hypothesis.

Group Differences

Each of the group differences tests requires that the data set and hypothesis include at least one grouping (qualitative) variable.

1. That the counts of groups are different
   a. To compare the counts of cases or objects in different groups either to each other or to other specified counts: Chi-Squared Goodness of Fit Test
   b. To test whether counts in one group appear to be dependent upon counts in another group: Chi-Squared Test of Independence
2. That the means of groups are different
   a. To compare a sample mean to the known or hypothesized population mean: One-Sample *t*-Test

b. To compare the means of two different groups to each other: Independent Samples *t*-Test
c. To compare the means of three or more different groups to one another: One Way ANOVA (also known as Simple ANOVA or Independent Groups ANOVA)
d. To compare one group to itself at two time points: Dependent Samples *t*-Test (also known as a Paired Samples of Repeated Measures *t*-test
e. To compare one group to itself at three or more time points: Repeated Measures ANOVA

Patterned Connections

Each of the patterned connections tests requires that the data set and hypothesis include two quantitative variables.

3. That variables trend, or form a pattern, together
   a. To test the relationship between two quantitative variables: Bivariate Correlation
4. That established trends/patterns can be used to predict connections in new data
   a. To test whether one quantitative variable is useful for predicting another quantitative variable: Simple Linear Regression

Though there are many variations beyond these tests, these are the foundational inferential statistics used by many social and behavioral researchers and, thus, will be the ones covered in this book. Each of these tests is covered in its own chapter with the exception of the two Chi-Squared test which are covered together in the final chapter of this book.

Reading Review 6.5

Complete each table using the hypotheses provided above it.

1. **Hypothesis:** People will have greater mean hours of sleep after they stop consuming caffeine compared to before.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		

2. **Hypothesis:** Level of happiness will be related to income.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		

3. **Hypothesis:** Students who are given study guides will have higher mean test scores compared to those who are not given study guides.

Variable Name	Variable Type	Variable Position (if any)
Which inferential test is the best fit for this hypothesis?		