Chapter 5

Cross-tabulated Frequency Distributions: Chi-square

This chapter focuses on a very common data analysis situation. Suppose you observe students who are sitting in the student center and discover that 35% of them are reading whereas 65% are doing other things such as talking or eating. You then wonder whether there is a sex difference – were more females reading, were more males reading, or were they equal? In other words, is the reading behavior you observed in the student center dependent upon gender? You can address this question with a cross-tabulated frequency distribution; statistical significance will be determined by applying the chi-square test, a test statistic that measures relationships or associations across variables.

Introduction to Chi-square

There are various ways you could examine the relationship between gender and reading behavior. In the student center, it would be easy for you to note whether each person you observe is a male or female and whether the person is reading or engaging in another activity. You could make similar observations at a beach during the summer (for example, noting whether someone is male or female, and if they are sun-bathing or engaged in another activity). You could distribute a questionnaire that includes one or more questions about reading behavior along with a question about the person's sex.

Our study examines the relationship between two variables: gender and reading behavior. Each variable has two levels. We will be able to see whether females are more likely to be reading that males. Some data from such a study are presented below.

To study the relationship between gender and reading, we will use the chi-square test statistic. Notice that as we discuss this test, we use the term *test statistic* as opposed to the term

test. The reason, as we saw earlier in Chapter 4 on hypothesis testing, is that the formal test of significance is based on the conditional probability that the resulting test statistic exceeds some predetermined probability level. Therefore, we are trying to make the point that test statistics (such as chi-square, *t*, and *F*) are not tests, but are stand-alone values derived from our obtained date. The obtained values are then evaluated as to whether they are consistent with what is expected by the null hypothesis.

There are a number of chi-square measures, but the main test statistic we will be covering is the chi-square test of independence (also known as the Pearson chi-square; later we will cover another chi-square test, called the chi-square "goodness of fit" test). When we say "test of independence" we are assessing whether our variables are independent of one another (in other words, neither has an influence on the other), or whether they are dependent (the variables influence each other). The chi-square test of independence is applied primarily to categorical or nominal level data (recall from Chapter 2 that this is data summarized into categories, including many demographic variables such as gender or race/ethnicity, variables that address the presence or absence of something, the exposure to something, or variables reflecting group membership), although ordinal applications are also made (e.g., socio-economic status). Categorical variables have levels that represent different categories – the scale describing the levels of the variable is called nominal because we name the levels but do not make quantitative distinctions. The variables in our gender-reading example are nominal or categorical. The measure for sex is the observer's indication of whether the person is male or female – males and females are different. The reading measure is nominal because it only categorizes people as either currently reading or not reading. A nominal scale provides no information about how much the person reads – e.g., number of books or time spent reading.

Now let's take a look at some hypothetical data from this study. Table 1 displays the results in what is called a 2 x 2 contingency table. The *2 x 2* refers to the fact that there are two levels of gender and two levels of reading. The term *contingency* is used to indicate that we are looking at whether one variable is contingent upon the other – is reading contingent on gender? The 2 x 2 contingency table is often called a 2 x 2 cross-tabulated frequency distribution. This is because the data in the table are frequencies or counts for the two variables. The table shows frequencies for each variable and also all the cells that are created when the two variables are cross-tabulated.

Begin by looking at the outside column and row labeled *Total,* The totals represent frequency information for each variable separately (called "Marginal Totals"). First, you can see that there were a total of 40 participants in the study. Of these, 15 were females and 25 were males. Also, 25 of the 40 participants were not reading and 15 were reading.

Table 1: Observed counts of reading behavior by gender

Gender

Reading?		Females	Males	Total
	No	**5**	**20**	25
	Yes	**10**	**5**	15
	Total	15	25	40

To see if the two variables are related, we need to focus on the cross-tabulated cells (these are bolded in Table 1). We can ask if the two variables are independent (not related) or dependent. Of the 15 females, 5 were not reading and 10 were reading. However, of the 25 males, only 5 were reading.

Based on the information in Table 1, it appears that females are more likely to be reading. However, it is not always easy to see patterns when looking at the frequencies. It is usually easier to visualize the data in a contingency table if we convert the frequencies to percentages. Graphing the data is also helpful. Table 2 shows the percent of females who were reading along with the percent of males who were reading. Now we see that 66.7% of females were reading (10 of 15); in contrast, 20.0% (5 of 25) of the males were reading.

We can also depict data in a bar graph – see Figure 1. The graph allows you to visualize the pattern of results. The graph illustrates that there are more females among the readers and more males among the students who were not reading.

Table 2: Observed frequency of reading behavior by gender

		Gender	
		Females	Males
Reading?	No	**33.3%**	**80.0%**
	Yes	**66.7%**	**20.0%**
	Total	100.0%	100.0%

A simple inspection of the data displayed in the tables and figure leads to the conclusion that the variables are related. However, this conclusion is based on an informal interpretation of the data. There is a more formal method for deciding whether the variables are related: this is the job of statistical significance testing. For our example data, we will apply the chi-square test statistic.

How Chi-square Works

The chi-square test statistic works by comparing the observed cell frequencies to the frequencies that would be expected if the null hypothesis is true. For example, Table 2 contains the observed frequencies of 33.3%, 80.0%, 66.7%, and 20%, and these will be compared to a

series of expected cell frequencies (which will be calculated shortly). The null hypothesis is that the variables are not related. In our example, the null hypothesis is that gender has nothing to do with reading. This means that we have our actual, obtained data in each cell of the table that we can compare to what is expected if the null hypothesis is true.

The chi-square statistic is calculated by making these comparisons in all table cells. We calculate the difference between the obtained and expected frequencies (these are called *residuals*). The differences are subsequently summed to generate the chi-square test statistic. Eventually, the chi-square test statistic is evaluated for statistical significance.

In a moment, we will walk you through how this is formally calculated. But first, let's take a look at the chi-square formula (Formula 1), and then apply it to our gender-reading example.

Formula 1: Chi-square Test of Independence

$$\chi^2 = \sum_j \sum_k \frac{\left(fo_{jk} - fe_{jk}\right)^2}{fe_{jk}}$$

Take a moment and look at the formula. The "χ^2" is the Greek letter chi that is our test statistic. The "Σ" is the Greek letter sigma, and here it stands for summation of terms (recall the discussion of summing values in Chapter 2 when calculating the variance). The "*fo*" part refers to the observed cell frequency (the numbers 5, 20, 10, 5 from Table 1), and "fe" refers to the expected cell frequency (yet to be calculated). The subscripts "*j*" and "*k*" refer to row and column numbers. Recall that our working example of gender and reading is a 2 x 2 contingency table, so to refer to the cell containing "20," we would say *j* is 1 and *k* is 2 (row 1, column 2). In essence, the formula specifies that we are going to take all the observed counts and subtract them from expected

counts (squaring the outcomes and dividing by a term), then sum all of these deviations across rows and columns.

Let's apply the formula to our example. To make the calculation, we need:

- observed cell counts (that we have already)

- expected cell counts (that we need to calculate)

How do we calculate the expected cell counts? One thing we will need is the overall count or the Total N. Here, that value is 40. Another thing we need is the product of the marginal counts (i.e., the row and column marginal totals). To get the expected cell count for each cell, we apply this formula:

Formula 2: Expected Cell Frequency for Contingency Tables

$$E_{jk} = \frac{R_j C_k}{N}$$

The expected count is the product of the marginal cell counts divided by the Total N (the subscripts j and k refer to row and column numbers). Thus, the expected count for the first cell in Table 1 is:

$$E_{11} = \frac{15 \times 25}{40} = 9.375$$

So, the expected cell frequency for the cell *Male and Not Reading* is 9.375. Let's take the next cell over (*Female and Not Reading*) with the observed count of 20. The marginal total for *No Reading* is 25, and the marginal total for *Females* is 25. Their product is 625 (25 x 25 = 625). Dividing this number by the Total N (40), we get an expected cell frequency of 15.625 (625 / 40 = 15.625). After completing the remaining two cells, the 2 x 2 contingency table may be populated with both the observed and expected counts as follows:

Table 3: Observed and expected counts of reading behavior by gender

			Females	Males	Total
			Gender		
			Females	Males	Total
Reading?	No	Observed	5.000	20.000	25
		Expected	*9.375*	*15.625*	
	Yes	Observed	10.000	5.000	15
		Expected	*5.625*	*9.375*	
		Total	15.000	25.000	40

A check to see whether the correct expected counts were calculated is that they should sum to 40. And indeed, this is the case (9.375 + 15.625 + 5.625 + 9.375 = 40). You may be wondering why these are the expected counts; that is, how are these counts related to what is expected if the null hypothesis is true. The answer lies in the marginal counts (i.e., the row and column marginal totals). Note that 15 out of 40 participants in the study were reading; this represents 37.5% of the participants. Similarly, 25 of 40 participants were not reading (62.5%). Now note that 25 participants are male. The null hypothesis expectation is that these same percentages will apply to the 25 males.

Thus, 37.5% of the 25 should be reading and 62.5% should not be reading. So, we can see that 25 x .375 = 9.375 and 25 x .625 = 15.625. The same logic can be applied to the 15 females. The main point is that the null hypothesis expectation is that the percentages we found in the margins should also be reflected in the individual cells.

Now that we have both the observed and expected counts, we can calculate the chi-square test statistic. As previously mentioned, for the chi-square test statistic, we take the observed cells and subtract them from the expected. The result provides us with what is called a "residual" value. Residuals can be either positive in value or negative depending on whether the observed value

exceeds the expected. Are you getting a sense yet as to what the chi-square test is actually doing? The test essentially derives differences, called residuals, between the observed and the expected values. And these residuals are eventually added to derive a test statistic. Later, we will see whether the observed counts differ from the expected values beyond what is expected by the null hypothesis. Hence, deriving these residuals is the first step toward that assessment.

Let's focus on calculating our first difference using that same upper-left cell "Female / Not Reading." The observed cell count is 5, and the expected cell count is 9.375. Subtracting the observed from the expected produces a value of –4.375 (5 – 9.375 = -4.375). Moving to the "Male / Not Reading" cell, we see an observed cell count of 20, and an expected count of 15.625. Subtracting these yields a value of 4.375 (20 – 15.625 = 4.375). The remaining cell deviations are 4.375 for "Female / Reading" and –4.375 for "Male / Reading." Hence, the following differences are derived: -4.375, 4.375, 4.375, -4.375. Notice that all the cells have equal residuals from their expected values; in other words, no particular cell has a residual greater than any other cell.

You have probably noticed that if these residuals are added, they sum to zero (-4.375 + 4.375 + 4.375 + -4.375 = 0), which is of little use to us. One thing you will get used to seeing throughout this text is that when residuals are calculated and used in a formula like chi-square, they typically are squared. Why? Because residuals, when added, typically total zero. Yet, we still need to capture these residuals – what do we do? Since the chi-square test statistic at this stage of the calculations requires non-negative residuals, we square the residuals to get rid of the negative signs, thus retaining the residual value (albeit squared).[1] The residuals squared are 19.141, 19.141, 19.141, 19.141 (e.g., both -4.375^2 and $4.375^2 = 19.141$). It might make sense to sum these now, but in fact the values are out of comparative proportion – they have all been

---

[1] Recall that when we calculated sample variances in Chapter 2, we also dealt with "residuals."

squared, and must be reduced. Further, the values will not fit the chi-square distribution (more on

that later). The original residuals were -4.375, 4.375, 4.375, -4.375. When squared they became

19.141, 19.141, 19.141, 19.141. We next must reduce these squared residuals so they reflect the

original context from which they came.

The value chosen to reduce the squared residuals is the "expected" cell frequency. This

will provide a "contextual" adjustment accounting for expected cell frequencies. Thus, each

squared residual is divided by its expected cell frequency, and these results are then summed to

give us the chi-square test statistic. Here's a layout of the process so far:

Table 4: Observed and expected counts, residuals (squared and adjusted) for
reading behavior by gender

			Gender		
			Females	Males	Total
Reading?	No	Observed	5.000	20.000	25
		Expected	*9.375*	*15.625*	
		Residual	*-4.375*	*4.375*	
		Squared Residual	*19.141*	*19.141*	
		Squared Residual/Expected	*2.042*	*1.225*	
	Yes	Observed	10.000	5.000	15
		Expected	*5.625*	*9.375*	
		Residual	*4.375*	*-4.375*	
		Squared Residual	*19.141*	*19.141*	
		Squared Residual/Expected	*3.403*	*2.042*	
		Total	15.000	25.000	40

From the table, we now have all the necessary information to calculate the chi-square test statistic.

Summing these adjusted residuals (2.042 + 1.225 + 3.403 + 2.042), we finally get our chi-square

value: 8.71. In other words, the resulting chi-square test statistic for our data is 8.71. By the way,

notice the largest adjusted residual is the cell with the females who were reading (3.403). This cell is contributing the most to the chi-square statistic.

What does the 8.71 chi-square statistic mean? It is best understood by examining the value in the context of the chi-square distribution. This is the distribution of all possible chi-square values that might be obtained if the null hypothesis is correct. Recall that the chi-square value is derived from your data so it represents the outcome of your study. The chi-square distribution allows you to see how your study compares to the outcomes that are possible if the null hypothesis is true. Figure 2 is a display of the distribution of all possible chi-square values when you have a 2x2 contingency table and the null hypothesis is true – the variables are in fact not related. This represents 100% of possible values that you might calculate from your data when the null hypothesis is true. Note that most of the chi-square values are quite small. If the null hypothesis is true:

- The observed and expected counts should be the same (or nearly the same)

- Thus, the residuals that we calculate should be very small.

- The obtained chi-square value will be small

Recall from the chapter on hypothesis testing (Chapter 1) that we need to derive the probability of getting a chi-square value of 8.71 given the null hypothesis is true. You see, every value in the distribution has a probability of occurrence – low values are highly probable. Our value of 8.71 fits into the conditional probability formula as follows: $p(8.71|H_o) = ?$. In other words, what is the probability of getting a chi-square value of 8.71 given that the null hypothesis is true? It isn't easy to calculate this exact probability without the help of a computer programs such as SPSS or SAS. However, we *can* turn this problem around and figure out the test value (D) at that we would fail to reject the null hypothesis at a given probability level: $p(D|H_o) = .05$.

The term "fail to reject" is wordy, but it stems from the arguable belief that we never "accept" the null hypothesis as true. We either "reject" the null hypothesis or "fail to reject" the null hypothesis.

Figure 2 illustrates the portions of the chi-square distribution for a 2 x 2 table that characterize the two decisions we have available: Fail to reject the null hypothesis or reject the null hypothesis. The large portion constitutes 95% of all possible chi-square values when the null hypothesis is true. If the obtained chi-square value is in this area, your decision is to fail to reject the null hypothesis. The remaining 5% of possible values are in the dark area of the distribution. If the obtained chi-square value is in this area, the decision is to reject the null hypothesis at the .05 level.

You can see in Figure 2 that our obtained chi-square test value of 8.71 falls well within the dark "reject the null hypothesis" area. The chi-square value that falls exactly at the .05 level is 3.84. This means that 95% of all the values are smaller than 3.84; the remaining 5% are larger than 3.84. If your obtained chi-square is exactly 3.84, there is exactly a .05 probability of your results occurring, if the null hypothesis is true.

This figure is not available for every situation. In actuality, we use one of two methods to decide whether to reject or fail to reject the null hypothesis. In either method, we first have to adopt an alpha level. The alpha level is also known as the Type I error rate. Typically in the social-behavioral sciences a .05 level (two-tailed) is used, which indicates that we are willing to accept a 5% or less chance (i.e., 5% or smaller chance) that our test results could have occurred by chance (see Chapter 4 for a discussion on probability applied to hypothesis testing). Second, we need to derive something called the degrees of freedom.

The *degrees of freedom* refers to the number of values that are *free to vary* when we are calculating a statistic. Degrees of freedom will be noted throughout the remainder of the text for most of our inferential test statistics. Therefore, we probably should spend a little time on the concept. For a 2 x 2 contingency table, degrees of freedom are calculated by taking the number of rows minus 1 ("r - 1"), and the number of columns minus 1 ("c – 1"), and then multiplying these two values together. In our example, we have 2 rows (2 – 1 = 1), and 2 columns (2 – 1 = 1). Multiplied together, we have 1 degree of freedom (1 x 1 = 1). This means that of the four cells in the table that we use to calculate the chi-square statistic, only one cell is free to vary. Once you have the frequencies for each of the two variables, the data in one cell determines the data for all other cells. Note that when you place the value of 5 in the first cell of Table 1, the other cells are automatically filled in because the marginal totals cannot change.

Why do we care about the degrees of freedom? By knowing how many degrees of freedom one has, an appropriate null hypothesis test statistic can be derived. We can accurately compare our obtained data to the null hypothesis expectation. Thus, the distribution of chi-square values is different depending on the degrees of freedom. You need to know the degrees of freedom when you evaluate your obtained data.

We now have all the information needed to derive the null hypothesis test statistic. Our alpha level is .05 (i.e., we are willing to accept a 5% or less chance that our test results could have occurred by chance), and we have a table with 1 degree of freedom. Appendix C contains a listing of null hypothesis test statistic values at various alpha levels, and Table 5 (below) is an excerpt from that appendix. Looking at Table 5, at the .05 alpha level with 1 degrees of freedom, we see the value of 3.84 that we encountered above. Applying the conditional probability formula, we get: $p(3.84|H_o) = .05$. Read this as "The probability of obtaining a chi-square value of 3.84 if the

null hypothesis is true equals .05." Because 8.71 is greater than 3.84, we can conclude that our

obtained data are very unlikely, given the null hypothesis expectation. We therefore decide to

reject the null hypothesis using the .05 level of significance. Note that Table 5 also shows us test

values for .01 and .001 levels of significance. Now we see that $p(6.65|H_o) = .01$ and $p(10.83|H_o)$

$= .001$. If we had used the .001 level as a decision rule to reject the null hypothesis, we would fail

to reject the null hypothesis with our data in the gender and reading study.

A second way to proceed is to analyze the data using a computer program such as SPSS

or SAS (see the "Applied Examples" section at the end of the chapter). These programs calculate

the chi-square value that is derived from the obtained data and they provide an exact probability.

With our data, the computer program tells us that $p(8.71|H_o) = .003$. This tells us that the

probability of getting our obtained data if the null hypothesis is true is exactly .003. Such a low

probability tells us that it is very unlikely that we could obtain our findings if the null hypothesis is

true. This is why we now reject the null hypothesis – it seems like a much better decision to

conclude that the alternative hypothesis is true. Therefore, our conclusion is that the variables are

in fact related. Gender is related to reading behavior.

Table 5: Chi-square critical value table (partial – from Appendix C)

| | Level of Significance | | |
Df	0.05	0.01	0.001
1	3.84	6.63	10.83
2	5.99	9.21	13.82
3	7.81	11.34	16.27
4	9.49	13.28	18.47
5	11.07	15.09	20.51
6	12.59	16.81	22.46
7	14.07	18.48	24.32
8	15.51	20.09	26.12

<u>Chi-Square Example Write-up: Gender and reading:</u>

A study was conducted to examine the association between gender and reading –

do males and females differ in their reading in public places? A total of 40

individuals were observed in the student center. There were 15 females and 25

males seated alone. Of the 15 females, 10 were reading (67%). Only 5 of the 25

males were reading (20%). A chi-square test of independence revealed an

association beyond the .05 significance level, $\chi^2(2, N = 40) = 8.82$, $p < .05$. The

data indicate that females are more likely to be reading than are males.

Chi-square Assumptions

For inferential tests, assumptions are made about the underlying distributions that the data

form. Tests such as chi-square are carried out on data assuming that certain data properties are in

place. First, we need to insure that our *data are counts or frequency data*, not interval or ratio-

level scores. If you have such score-level data (sometimes interval and ratio data are called

"score-level" data), consider other tests covered in this textbook such as correlation, or a t-test

depending on whether your data are grouped. Second, your nominal variable levels or categories should be *mutually exclusive*. As noted in the chapter on probability, the possible results of a coin-flip are mutually exclusive (heads or tails). Third, the observations or cases in the study should be independent. This means that cases should be unrelated to the other cases – if they are related in some way, this can affect the underlying association to be evaluated by the chi-square test statistic. In our current study, "if" we had prior knowledge that some individuals in the study were married, then the *independence of observations* assumption would be violated.

Fourth is the assumption of *normality*. Normality for chi-square focuses on the expected cell counts; do the expected cell counts resemble the normal distribution of a bell-shape curve? This may seem a little confusing at first, but recall that normality on count data was previously demonstrated when we applied the binomial distribution to test hypotheses. The distribution of probabilities became more and more normal as the number of observations increased. This may be applied to the chi-square test; as the number of expected cell counts increases, normality within each cell is more likely to be observed. The general cutoff for normality to be assumed within chi-square is having an expected cell count of 5 or more. Therefore, as you evaluate your cells in a contingency table, make sure all your expected cell counts are 5 or more.

What if normality is violated? First, realize that when the assumption is violated, your power to detect an association is limited (small samples indicate less power), and in some cases Type I error may increase. Typically in larger contingency tables (when your df > 1), you are allowed at least 1 cell where the expected cell count can be less than 5, yet when you are working with a 2 x 2 contingency table (df = 1), all expected counts should be 5 or greater. When violated, one thing you can do is collapse cells together "if" it makes sense per your research design. This will not increase your overall sample size, but will increase your marginal counts, thus leading to

bigger expected cell counts. For example, say you had a 2 x 5 contingency table, and the variable with 5 levels was race/ethnicity (White, African-American, Asian-American, Latino, and Mixed/Other). In looking at your cells, it appears that the Latino cell and the Mixed/Other cell have low expected counts. In this situation, you may be able to combine those cells, yielding a 2 x 4 contingency table and a new category simply labeled "Other" race/ethnicity. This may done "assuming" you didn't have any direct hypotheses regarding Latinos or Mixed/Other race. If your overall goal was to see if there was an association between race/ethnicity and the other variable, then collapsing the cells would make sense.

What if you violate normality and you have a 2 x 2 contingency table? Obviously, collapsing cells would be problematic. In this case, there are several rules that may be applied to help steer the type of adjustment that needs to be made. And, in some cases, the adjustment is actually utilizing a different measure of association. First, anytime your sample size is 20 or fewer cases in a 2 x 2 contingency table, instead of using the chi-square test statistic, you should use Fisher's exact test (that is presented later in this chapter). Second, if your sample size is between 20 and 40, and at least one of your expected cell counts is less than 5, then you would utilize an alternative test of association called "Fisher's Exact Test" (covered later in this chapter). Third, if your sample size is greater than 40 and you violate the normality assumption, use a continuity correction (that involves making fine adjustments to the numerator of the chi-square formula and may be invoked through most computer programs).

Other Chi-square Applications

Thus far, chi-square has been applied to evaluate the association between two categorical variables with 2-levels each (a 2 x 2 contingency table). In the following section, chi-square will

also be applied to situations with only one categorical variable, and when we have more complex

tables (e.g., a 3 x 3 design).

One categorical variable only. Chi-square may be applied when we have only one

categorical variable. In this case, we are interested in whether the observed counts match the

expected counts, or whether they differ. Applied in this manner, chi-square is referred to as a

"goodness of fit" test because we are assessing the "fit" between the observed and expected

frequencies within a single categorical variable. The formula is almost exactly the same as the

"test of independence" except that there is only one summation sign due to only having one

variable:

Formula 3: Chi-square Goodness of Fit Test

$$\chi^2 = \sum_j \frac{\left(fo_j - fe_j\right)^2}{fe_j}$$

You might now be asking, "One variable? Then what are we doing?" With only one variable, we

will be able to see whether the resulting counts (recall chi-square focuses on counts) deviate from

what would be expected by chance per the null hypothesis. In other words, do the observed

counts match what is expected, or do they differ?

Say that you have data reflecting drink preference at a coffee house. You hypothesize that

individuals will have a preference for coffee-based drinks with caffeine over other possible

choices. After observing drink choice for an hour, you have collected the following data:

Table 6: Observed counts of coffee drink choice

	Coffee with Caffeine	Coffee without Caffeine	Non-Coffee Drink
Observed Coffee-House Order	20	5	5

The data in Table 6 clearly show a preference for coffee beverages with caffeine, with 20 people ordering coffee beverages with caffeine, and only 5 individuals ordering coffee beverage without caffeine, or a non-coffee drink. There indeed does appear to be a pattern in beverage preference. However, is this preference for coffee with caffeine beyond what is expected by the null hypothesis?

The next step is to calculate the chi-square test statistic. The calculations are almost exactly the same as the one used for the 2 x 2 chi-square. The observed counts (20, 5, and 5) are subtracted from the expected counts, squared, reduced by the expected count, then summed. Expected counts here are calculated a little differently since we only have one variable. To obtain the expected counts, the total sample size (30) is divided by the total number of categories (3), yielding expected cell counts of "10" for each cell. Next, taking each observed count and subtracting it from the expected, we get the following values: 10, -5, -5. As with the 2 x 2 example, these values are squared to remove the negative signs, resulting in squared differences of 100, 25, and 25. These values are now divided by the expected cell count (e.g., 100/10 = 10; 25/10 = 2.5; 25/10 = 2.5), and finally summed to provide the chi-square test statistic (10 + 2.5 + 2.5 = 15 = χ^2). Below is breakdown of the data:

Table 7: Observed and expected counts, residuals (squared and adjusted) for coffee drink choice

	Coffee Beverage with Caffeine	Coffee Beverage without Caffeine	Non-Coffee Drink
Observed Counts	**20**	**5**	**5**
Expected Counts	10	10	10
Residuals	10	-5	-5
Residuals Squared	100	25	25
Residuals Squared divided by Expected	10	2.5	2.5

$$\chi^2 = 15$$

For a one-dimensional chi-square table, degrees of freedom are derived by taking the number of categories and subtracting by 1 ("c - 1"). In our example, we have 3 categories, so the degrees of freedom are: 3 - 1 = 2.

We now have all the information needed to derive the null hypothesis test statistic. Our alpha level is .05 (a 5% or less chance that our test results could occur by chance), with 2 degrees of freedom (3 categories – 1 = 2). From Table 5 (and Appendix C), at the .05 alpha level with 2 degrees of freedom, a critical value of 5.99 is found. Using this number, we now know the exact probability at that we fail to reject the null hypothesis. Applying the conditional probability formula, we have: $p(5.99|H_o) = .05$. Our test statistic of 15 exceeds 5.99, and we thus can reject the null hypothesis. The general interpretation of this finding is that individuals prefer coffee-based drinks over the other beverages.

Two categorical variables with 3 levels each. Chi-square may be applied to any situation where categorical data is used regardless of the number of levels. The 2 x 2 formula (presented above) is simply expanded to research in which one or both variables has 3 or more categories. Consider the following study in which both variables have three categories (based on Lachin,

1977). In this example, an experimental design was used. The independent variable was drug

dosage (placebo, low dose, or high dose). The dependent variable was disease recovery (complete

recovery, partial recovery, or no recovery). In the experiment, a total of 75 participants were

assigned to the three dosage conditions (25 in each condition). Is there an association between

these two variables? Example data are displayed in Table 8.

Table 8: Observed counts of drug dosage by disease recovery status

		Disease Recovery			
		None	Partial	Complete	Total
Drug Dosage	Placebo	15	5	5	25
	Low Dose	5	5	15	25
	High Dose	5	10	10	25
	Total	25	20	30	75

$$\chi^2 = 15.5$$

Calculation of the chi-square value is done in the typical way, as are degrees of freedom (*c -1* x *r-1*), yielding a final chi-square value of 15.5 with 4 degrees of freedom. Try these calculations

yourself, and see if can derive (or come close) to this chi-square test value (be careful with your

rounding). Our critical value at a .05 level of significance from Table 5 is 9.45. Since our value of

15.5 exceeds 9.45, we can reject the null hypothesis, and assert that there is an association

between Drug Dosage and Disease Recovery.

In these more complex tables, however, discerning exactly what is driving the association

is a challenge. Certainly we can say that association exists, and by looking at the observed and

expected counts, can make an educated guess that low dosage appears to have the best results.

However, to formally test that a low dosage is superior to a placebo or a high dosage, you need to

break this larger 3 x 3 table into a series of 2 x 2 tables. Space does not permit that demonstration here, but if curious, the technique is outlined in the book *Nonparametric Statistics for the Behavioral Sciences* (Siegel & Castellan, 1988).

Other Measures of Association for Categorical Data

So far, the chi-square statistic has been applied to one or two categorical variable examples. There are a number of alternative tests, though, that in some instances may be more appropriate to use, or evaluate association in a slightly different manner. This section addresses two such tests; Fisher's Exact test and the Odds Ratio.

Fisher's exact test. Fisher's exact test as a measure of association allows us to calculate an exact probability value instead of having to generate a test statistic as with chi-square. When might one use this test instead of chi-square? First, it may only be applied to 2 x 2 contingency tables. Second, Fisher's exact test should be used when sample sizes are extremely small (N of 20 or less), and/or when normality has been violated with samples of 40 or less. Third, note that the formula presented here is a 1-tail probability (not our typical 2-tail probability). Your hypothesis assessment should take this into account. Ways to alter the formula for 2-tail assessments are beyond the scope of this text (see Hays, 1994, for adjustments). Therefore, if you use these hand calculations for Fisher's exact test in research, your hypotheses should clearly be directional (i.e., clearly state the expected direction of the association). For example, in our gender-reading example, stating that females would be expected to be reading more than males would be a directional hypothesis, and therefore may be evaluated using a 1-tail probability level; see Chapter 4 regarding directional hypotheses). Both SPSS and SAS will provide 2-tailed probabilities, and if available those results should be reported.

Let's walk through a numerical example. Say you were curious about the association

between gender and eating behaviors at the University, hypothesizing that males will tend to eat

unhealthy meals compared to females. You stake-out your College dining area, and note the

gender of individuals purchasing food, and whether they purchase healthy or unhealthy foods.

You collect data on only seven individuals, and will be assessing whether there is an association

between gender and eating healthy/unhealthy foods:

Table 11: Data on Gender and Eating Behaviors

Gender

		Females	Males
Food Type	Unhealthy	2	1
	Healthy	1	3

To apply Fisher's exact test, the cells within a 2 x 2 table take the form of the following:

A	B
C	D

Applied to our data, cell "A" has 2, cell "B" has 1, cell "C" has 1, and cell "D" has 3.

2	1
1	3

The formula for Fisher's exact test utilizes the cell labels as follows:

Formula 4: Fisher's Exact Test

$$p = \frac{(A + B)! \, (C + D)! \, (A + C)! \, (B + D)!}{N! \, A! \, B! \, C! \, D!}$$

The formula for Fisher's exact test takes into account the observed cell counts only – the marginal

counts are not needed, nor are expected cell counts. The exact probability (*p*) is a function of the

various cells first added together, then the factorial of each is taken, followed by the product of

these results that, taken together, form the numerator. The denominator is simply the total sample

size and all cell counts, multiplied together after the factorial of each has been taken. Applying the

formula to our data:

Applied to our data:

$$p = \frac{(2 + 1)!\ (1 + 3)!\ (2 + 1)!\ (1 + 3)!}{7!\ 2!\ 1!\ 1!\ 3!}$$

Further working the problem, we get:

$$p = \frac{3!\ 4!\ 3!\ 4!}{7!\ 2!\ 1!\ 1!\ 3!} \quad = \quad \frac{6*24*6*24}{5040*2*1*1*6} \quad = \quad \frac{20736}{60480} \quad = \quad .34$$

The final probability obtained here is *p* = .34 (one-tail test). By evaluating our original hypothesis

with this data, and looking at the observed cell counts, we fail to reject the null hypothesis and

conclude that gender is not associated with eating behaviors.

Odds ratio. Chi-square and Fisher's exact test provide us with measures of association

between two variables in a 2 x 2 contingency table, but fail to tell us anything about the "degree"

of association. The odds ratio provides a degree of association between variables. Never heard of

an odds ratio? Many news reports on health-related issues commonly utilize odds ratios since this

measure of association provides a "sound bite" for various media outlets. For example, a New

York Times article (Wilkinson, 2005) on anxiety disorders stated "…women have twice the risk

[for anxiety disorders] as men." In another example, an article addressing pregnancy in Ethiopia (Lacey, 2004) read, "Pregnant girls under 15 are 5 times more likely to die from childbirth than women in their 20's." Although the values mentioned above are all above 1.0 (and therefore make excellent sound bites), odds ratios may also range between 0 and 1. Even in that form (0 to 1), they actually have the same meaning, but can be easily converted into something worthy of the evening news.

An odds ratio is a statistic where a ratio is formed between the odds of an event occurring in one group or category, and dividing it by the odds of an event occurring in another group or category. Let's focus on an example. Say you completed a study looking at association between coffee consumption and performance on a statistics test. The following data were collected:

Table 10: Observed counts of test performance by coffee consumption

	Consumed One Cup of Coffee?	
	No (0)	Yes (1)
Test Performance Failed (0)	10	5
Passed (1)	5	10

The odds of failing a test and "not" consuming any coffee is 10/5 = 2. The odds of passing a test and "not" consuming any coffee is 5/10 = .5. It is these odds that will be used as the basis for the odds ratio.

The formula for a 2 x 2 table uses the same A, B, C, D labels as seen in Fisher's exact test:

Formula 5: Odds Ratio

$$O.R. = \frac{(A)(D)}{(B)(C)}$$

Appling the odds ratio to the data above:

$$O.R. = \frac{(10)(10)}{(5)(5)} \quad = \quad \frac{100}{25} \quad = \quad 4$$

The final odds ratio is 4 (100/25 = 4). This may also be calculated by taking the calculated odds

("2" and ".5") and creating a ratio: 2/.5 = 4. However, we suggest using the odds ratio formula

– therefore one does not have to calculate the individual odds for each group or category. What

does an odds ratio of 4 mean? Based on the observed counts, those who passed the test were 4

times more likely to have had a cup of coffee than those who failed the test. There are ways to

formally assess the significance of odds ratios, but those are beyond the scope of this text. Yet,

even without significance tests, the odds ratio may be used as a measure of the degree of

association between variables – higher odds ratios indicate a higher degree of association.

Interesting Web-Links:

http://www.stat.sc.edu/~west/applets/contable.html {2 x 3 table example applet}

http://www.stat.sc.edu/~west/applets/chisqdemo.html {tail distribution applet}

http://web.soc.psu.edu/courses/oliver/mcr/topics/chisq/chiscript.html {2 x 2 table example}

References

Hays, W.L. (1994). Statistics (5th ed.). NY: Holt, Rinehart, & Winston.

Lachin, J.M. (1977). Sample size determinants for r x c comparative trials. *Biometrics, 33*,

    315-324.

Lacey, M. (2004, December 12). For Africa's poor, pregnancy is often life threatening. *The*

    New York Times, Section 1, p. 22.

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd

    ed.). NY: McGraw-Hill.

Wilkinson, D. (2005, June 5). Anxiety disorders: Living on a thin line. *The New York*

    Times, Section 15, p. 9.

Applied Examples Using SPSS and SAS

In SPSS, chi-square analyses may be performed in different ways. The "Crosstabs" procedure is used with two-way tables (e.g., 2 x 2, 2 x 3, etc.) and is demonstrated below. For goodness-of-fit tests (one categorical variable only), the "Nonparametric" procedure is used. For either approach, you can write syntax or use the pull-down menus (here, we demonstrate the syntax approach). Also, raw data may be utilized, or cell count data may be used with a weight applied.

In SAS, chi-square is performed using the procedure "Proc Freq" – either raw data or weighted cell data may be used. Syntax code is typically required (although there are now analyses you may perform in SAS using pull-down menus).

SPSS Chi-Square Example: Reading behavior and gender (2 x 2)

Table 11: Data in "weighted cell" format for reading behavior and gender example

Reading	Gender	Weight
1	1	5
1	2	20
2	1	10
2	2	5

This example is the data from the reading and gender example from Table 1. To perform a chi-square test in SPSS, follow the steps below.

Step 1: Enter this data directly into SPSS "Data View." Only enter the data (do not enter the variable names in Data View; they are entered in Variable View).

Step 2: Once you have entered the 3 columns of data, go to "Variable View" and enter the three variable names (Reading, Gender, and Weight). You can also fill-in the variables labels (for example, you could enter "Reading?" for Reading, and "Gender" for Gender). Here, because we

have our pivot table output set to provide both variable names and labels, we have opted "not" to

enter variable labels for output clarity. Value labels for Reading are 1 = "No," and 2 = "Yes."

Value labels for Gender are 1 = "Females," and 2 = "Males."

Step 3: Next, you must WEIGHT the data. No doubt you have been wondering about the

"weight' variable – now you get to see it in action. The values in the "weight" column are 5, 20,

10, and 5, and match exactly the cell counts from Table 1. Hence, the weight variable contains the

cell counts. The values for Reading and Gender (1 and 2) above are cell categories (not counts).

Reading (1) and Gender (1) has a cell count (using "weight") of 5. Reading (1) and Gender (2)

has a cell count of 20.

To weight data, using the pull-down menus, go to DATA, then WEIGHT CASES. From

this box, click WEIGHT CASES BY, and then enter the variable "weight," and click okay. Going

back to the "Data View" window, you should now see a little box in the lower right corner of the

screen that reads "Weight On." This indicates the data are now weighted. Using this technique,

you can run any chi-square analysis just by having the observed cell frequencies!

Step 4: Assuming you have weighted the data, you can now proceed to try our example

from Table 11. The syntax approach to running this data is illustrated below, as is the output

(edited).

By the way, if you are using syntax, first enter your data in SPSS and then weight the

data. Then, open a syntax window (go to "File", then "New", then "Syntax"). You may directly

enter the "Crosstabs" commands in this window. Then, highlight all the syntax, and press the

"Play" icon (which looks like a "play" button from a VCR or a DVD player). Your output should

resemble what is illustrated on the next page.

```
CROSSTABS
  /TABLES=Reading  BY Gender
  /FORMAT= AVALUE TABLES
  /STATISTIC=CHISQ
  /CELLS= COUNT EXPECTED
  /COUNT ASIS .
```

Crosstabs

Reading * Gender Crosstabulation

			Gender		Total
			1.00 Females	2.00 Males	
Reading	1.00 No	Count	5.000	20.000	25.000
		Expected Count	9.4	15.6	25.0
	2.00 Yes	Count	10.000	5.000	15.000
		Expected Count	5.6	9.4	15.0
Total		Count	15.000	25.000	40.000
		Expected Count	15.0	25.0	40.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.711[b]	1	.003		
Continuity Correction[a]	6.834	1	.009		
Likelihood Ratio	8.810	1	.003		
Fisher's Exact Test				.006	.004
Linear-by-Linear Association	8.493	1	.004		
N of Valid Cases	40				

[a.] Computed only for a 2x2 table

[b.] 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.63.

SAS Chi-Square Example: Reading behavior and gender (2 x 2)

Here, we use the exact same data setup noted in Table 11. We have dispensed with the value

label codes (you can investigate these for yourself through the *Proc Format* procedure) for ease

of application. Unlike the SPSS example, we have provide the full data setup and output.

Step 1: Once you open SAS, simply go to your *Program Editor* window, and type the

following syntax:

```
data example;
input reading gender wt;
datalines;
1.00      1.00    5.00
1.00      2.00   20.00
2.00      1.00   10.00
2.00      2.00    5.00
;
proc freq;
weight wt;
tables reading*gender / expected chisq;
run;
```

Step 2: Next, highlight all of this syntax in the *Program Editor*, then press the "Run" icon

(its looks like a person running) at the top of your screen. The output produced will be similar to

the output presented on the next page. Notice that all the relevant information from the SPSS run

is also provided within SAS.

The SAS System   14:45
Wednesday, May 17, 2006   8

The FREQ Procedure

Table of reading by gender

```
reading      gender

Frequency|
Expected |
Percent  |
Row Pct  |
Col Pct  |       1|       2|  Total
         ────────┼────────┼
       1 |      5 |     20 |     25
         |  9.375 | 15.625 |
         |  12.50 |  50.00 |  62.50
         |  20.00 |  80.00 |
         |  33.33 |  80.00 |
         ────────┼────────┼
       2 |     10 |      5 |     15
         |  5.625 |  9.375 |
         |  25.00 |  12.50 |  37.50
         |  66.67 |  33.33 |
         |  66.67 |  20.00 |
         ────────┼────────┼
Total          15       25       40
            37.50    62.50   100.00
```

Statistics for Table of reading by gender

Statistic	DF	Value	Prob
Chi-Square	1	8.7111	0.0032
Likelihood Ratio Chi-Square	1	8.8095	0.0030
Continuity Adj. Chi-Square	1	6.8338	0.0089
Mantel-Haenszel Chi-Square	1	8.4933	0.0036
Phi Coefficient		-0.4667	
Contingency Coefficient		0.4229	
Cramer's V		-0.4667	

Fisher's Exact Test
────────────────────────────────
Cell (1,1) Frequency (F)	5
Left-sided Pr <= F	0.0044
Right-sided Pr >= F	0.9995
Table Probability (P)	0.0040
Two-sided Pr <= P	0.0062

Sample Size = 40

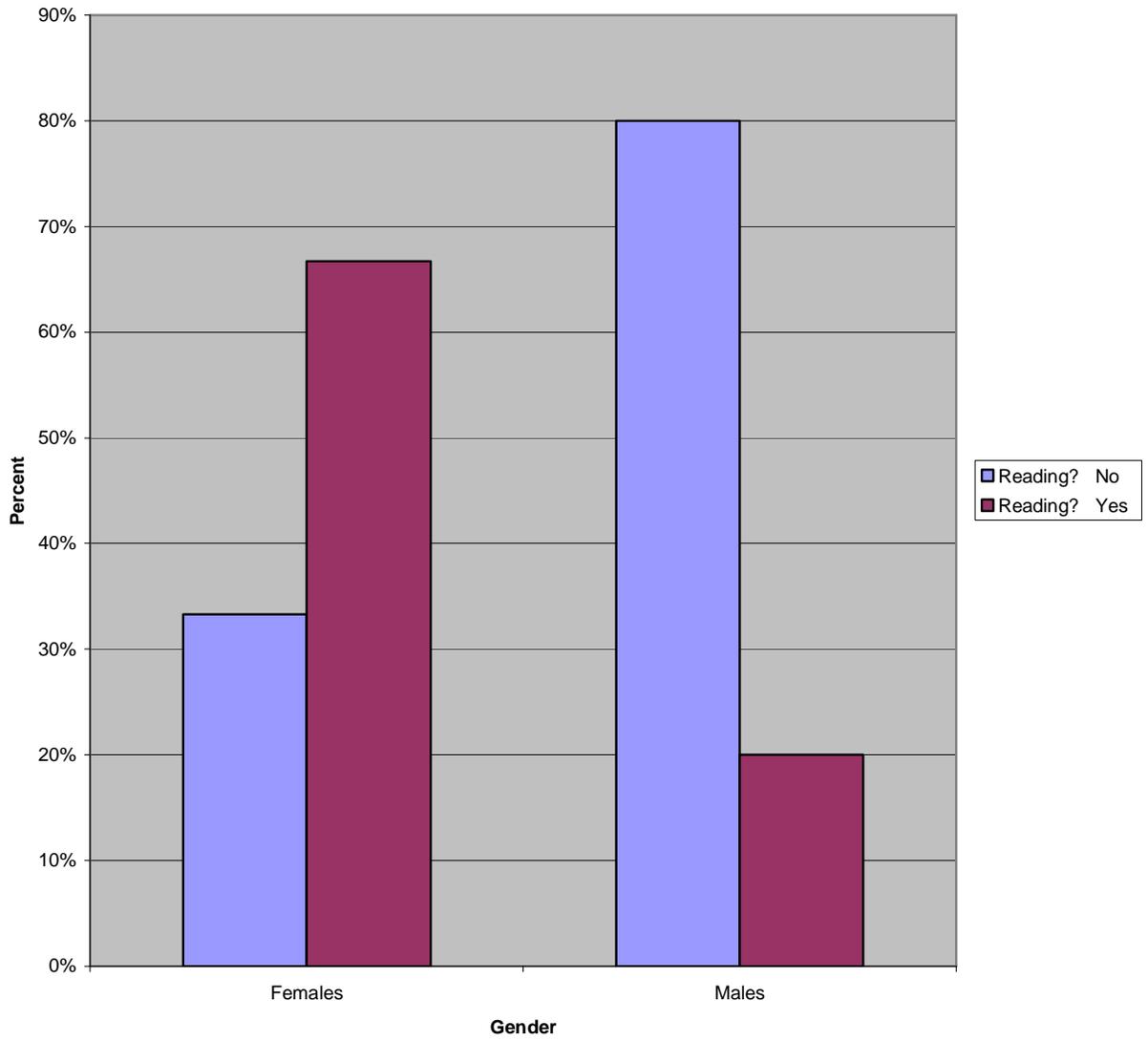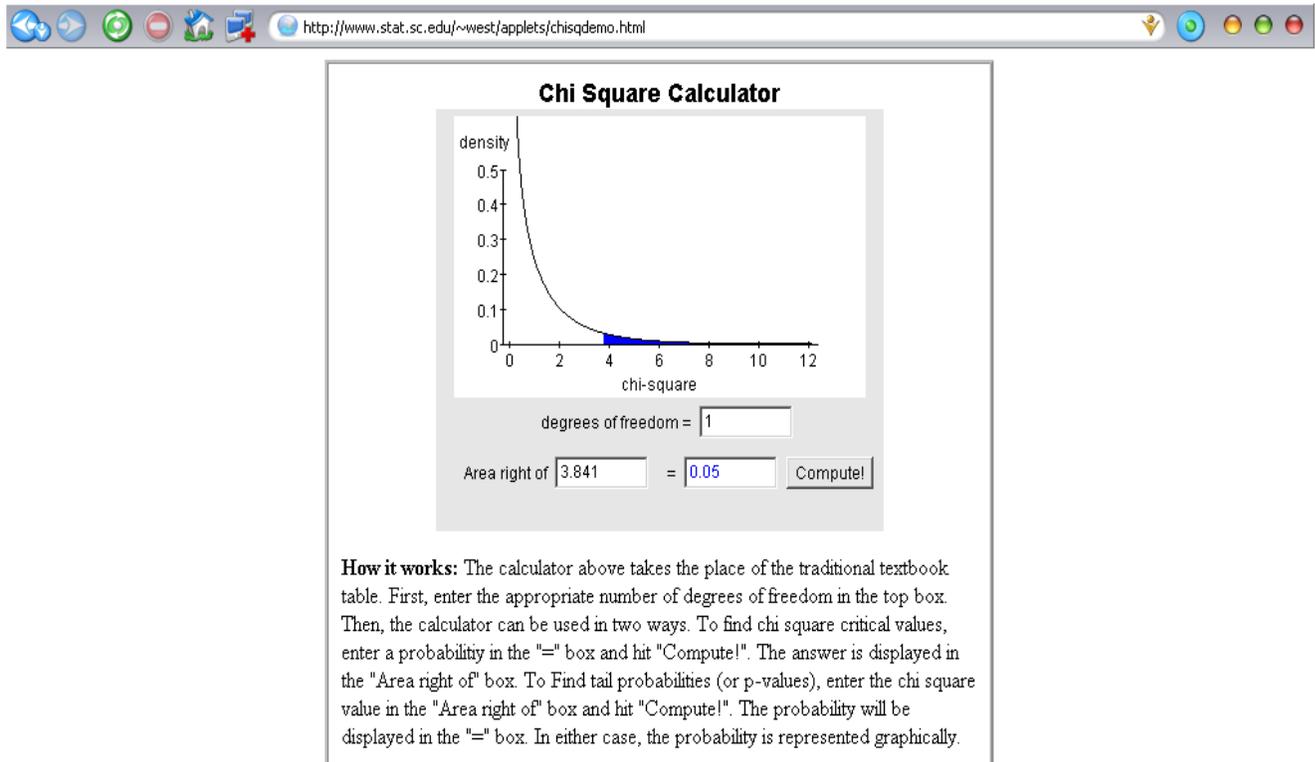*Figure 1: Observed frequencies of reading behavior by gender*

Figure 2: Chi-square distribution at p = .05 approximation with df=1

How it works: The calculator above takes the place of the traditional textbook table. First, enter the appropriate number of degrees of freedom in the top box. Then, the calculator can be used in two ways. To find chi square critical values, enter a probabilitiy in the "=" box and hit "Compute!". The answer is displayed in the "Area right of" box. To Find tail probabilities (or p-values), enter the chi square value in the "Area right of" box and hit "Compute!". The probability will be displayed in the "=" box. In either case, the probability is represented graphically.

west@stat.sc.edu.

Sample Exercises

1.   Using a sample of California Drivers, we have been questioned by the press about the representativeness of our sample. In particular, there have been claims that our sample has too many old drivers as opposed to young drivers. Our age categories are 16-28, 29-41, 42-54, 55-67, and 68+. Your data (in order) are as follows: 100, 120, 180, 100, 200. Do our obtained frequencies for age differ from what would be expected by chance?

(a)   Write the null and alternative hypothesis based on the information above.

(b)   Plot the observed counts and make a brief interpretation based on the counts.

(c)   What are your overall findings based on chi-square (at a .05 level)?

(d)   Interpret your findings (e.g., what do these findings mean)?

2.   HIV infection remains problematic in the U.S. You have completed a survey of HIV seroconversion (i.e., going from being HIV negative to positive) and safe-sex behavior (condom use). You surveyed a population of gay males, getting a sample of 300. You have obtained information on their seroconversion in the past year, and frequency of condom use (two levels – rarely vs. almost always). You found in your data that of the 100 seroconverting cases, 80 rarely used a condom. For the 200 non-seroconverters, 60 rarely used a condom. Are those who rarely use condoms more likely to seroconvert than those who almost always use condoms?

(a)   Write the null and alternative hypothesis based on the information above.

(b)   Plot the observed counts for both variables and interpret based on the counts.

(c)   Evaluate the assumptions of chi-square. Did you meet the assumptions? Explain.

(d)   What are your overall findings based on chi-square (at a .05 level)?

(e)   Interpret your findings appropriately.

3.   Using the same data as #2 above, provide an odds-ratio appropriate to this problem and interpret your findings.

4.   Using the data below, perform the following evaluations to assess the association between gender and history of a heart attack:

(a)   Write a null and a *directional* hypothesis regarding gender and history of heart attack.

(b)   Plot the observed counts for both variables and interpret based on the counts.

(c)   Apply Fisher's Exact test to measure the association between gender and history of a heart attack (use a 1-tail evaluation). Justify why you would use Fisher's test over the chi-square test of independence by evaluating the chi-square assumptions.

(d)   What is the odds ratio of men vs. women on history of heart attack?

(e)   Interpret your findings (both Fisher's Exact test and the odds ratio).

Sub#	Gender	Heart Attack
1.00	Male	Yes
2.00	Male	Yes
3.00	Male	Yes
4.00	Male	No
5.00	Male	Yes
6.00	Male	Yes
7.00	Male	Yes
8.00	Female	Yes
9.00	Female	No
10.00	Female	No