# Lesson 10, part 1
# Inferences for Categorical-Categorical Association

**Outline of the Lesson**[1]



As we have discussed for several lessons now, *inferential statistics* has two important features:

- Information is obtained from a *sample*.
- The information from the sample is used to draw a conclusion (an *inference*) about the entire *population* from which the sample was drawn.

We have learned methods of statistical inference for a categorical variable measured using a single sample. For example, suppose we want to know what percent of the population of Pennsylvania Republicans plans to vote in the upcoming election. We would obtain a random sample of Pennsylvania Republicans, and measure what percent of the sample plans to vote. Using this figure, we use either confidence intervals or hypothesis tests to make statements about the entire population consisting of all Pennsylvania Republicans.

[1] This "part 1" document contains sections 10.1 through 10.6. The "part 2" document contains the remaining Section 10.7. Some, but not all, courses which use these lessons will study part 2 as well as part 1 of the lesson.

In this lesson, the focus shifts slightly. Rather than asking questions about a single population, we want to ask about multiple populations. When we do this, we are almost always interested in these questions:

- Is there a difference between the different populations?
- If so, how big is that difference?

For example, is there a difference between Democrats and Republicans, with respect to whether or not they plan to vote in the upcoming elections? In fact, we are not limited to two groups; we could ask the same question for three different populations: Democrats, Republicans, and Independents.

The methods of this lesson combine ideas on hypothesis testing, covered in Lesson 8, with ideas you first encountered way back in Lesson 3. We begin with a very quick review of these two major concepts.

## 10.1 – Review: Association Between Categorical Variables

In Lesson 3, we studied the association between two variables. For example, we might ask these questions.

- Is there an association between whether a person smokes and whether or not that person has cancer?
- Is there an association between high school GPAs (grade point averages) and college GPAs?
- Is there an association between gender and SAT scores?

An **association** (or **relationship**, or **connection**, or **correlation**) exists between two variables if a particular value for one variable is more likely to occur with certain values of the other variable. If no association exists, we say the variables are **independent**.

In this lesson, both variables will be categorical variables. As we learned in Lesson 3, we can conveniently represent the data we gather in our survey or experiment using a **contingency table**. Once the data is represented in this form, we can easily answer questions about various proportions or probabilities seen in the data.

**Example:** A researcher surveyed 1765 U.S. households, asking whether paying bills was a major concern. Some of the people had children, and others did not. There are two categorical variables:

- Whether or not the household has children.
- Whether or not the household view paying bills as a major concern.

Notice that you can think of each variable as corresponding to a question posed by the researcher. (“Do you have children?” and “Do you view bill-paying as a major concern?”)

This table summarizes the results of the survey:

| | Major concern | Not a major concern |
|---|---|---|
| Have children | 515 | 340 |
| No children | 473 | 437 |

Calculate the following proportions as decimal fractions rounded to 4 places (for example, 0.6134).

a. What proportion (of the entire sample) have children in the household?
b. What proportion view bill-paying as a major concern?
c. What proportion of the households with children are households that view bill-paying as a major concern?
d. What proportion of the households that view bill-paying as a major concern are households with children?

**Solution:** Answering these questions will be easier if we first write down totals for each row of the table, for each column of the table, and the grand total for the entire survey, as shown below. For example, in the "Have children" row 515 household said bill-paying was a major concern, and 340 said it was not, for a total of 515 + 340 = 855.

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

a. Out of the 1765 people in the sample, 855 have children in the household, so the proportion is $\frac{855}{1765} = 0.4844 = 48.44\%$
b. The question as stated does not include the words "of the entire sample" but those words are implied. Out of the 1765 people in the sample, 988 view bill-paying as a major concern, so the proportion is $\frac{988}{1765} = 0.5598 = 55.98\%$
c. The question asks for a "proportion of the households with children," so this is a question only about the households with children. There are 855 of these households. Of these 855 households, 515 of them view bill-paying as a major concern. The proportion is therefore $\frac{515}{855} = 0.6023 = 60.23\%$
d. Of the 988 households that view bill-paying as a major concern, 515 have children, leading to the proportion $\frac{515}{988} = 0.5213 = 52.13\%$

You may wish to use the following applet, first introduced in Lesson 3, to practice the calculations illustrated by this example:

Proportions for contingency tables

In this example, there were two categorical variables. One variable, the **explanatory** variable, we think of as establishing the groups we wish to study. In this study, the explanatory variable was whether or not the household had children. The other variable, the **response** variable, is the outcome from our survey or experiment. In this study, the response variable was whether or not the household viewed bill-paying as a major concern.

Here is another example. Suppose we want to know if there a difference between Democrats, Republicans, and Independents, with respect to whether or not they plan to vote in the upcoming elections. In this case, political affiliation would be the explanatory variable and whether they plan to vote the response variable. As in the previous example, we can use a contingency table to present the data we obtain from our survey or experiment. It is convenient, but not mandatory, to make the rows of the table correspond to the values of the explanatory variable, with the columns representing the values for the response variable. This would lead to a table in this form:

| | Plans to vote | Does not plan to vote |
|---|---|---|
| Democrat | | |
| Republican | | |
| Independent | | |

In general, there are a number of ways to ask the primary question we are interested in; all these English-language questions mean the same thing.

- Is there an association (correlation, relationship, connection) between the two variables?
- Are the groups different?
- Does the response variable depend on the explanatory variable?

The answer to the question could be either "yes" or "no." This table summarizes the different ways we can express these two possible answers, applied specifically to our situation involving political affiliation and voting plans.

| **yes** | **no** |
|---|---|
| There *is* an association (connection, relationship, correlation) between political affiliation and whether or not you plan to vote. | There is *no* association (connection, relationship, correlation) between political affiliation and whether or not you plan to vote. |
| There *is* a difference between the different groups (Democrat, Republican, Independent), with respect to whether or not they plan to vote. | There is *no* difference between the different groups (Democrat, Republican, Independent), with respect to whether or not they plan to vote. |
| Whether you plan to vote *depends on* political affiliation. | Whether you plan to vote *is independent of* political affiliation. |

What the table presents is not mathematics; it is English-language usage. However, it is very important for you to become comfortable with this. An important note is this:

- To say that there *is an association* between the explanatory variable and the response variable says that the groups identified by the explanatory variable are *different* from one another, for the response variable you are studying. The value for the response variable can be expected to be different, depending on the value for the explanatory variable.

**Exercise 1[2]:** Here again is the data from our earlier example.

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

a. To get a preliminary feel for possible association, we can calculate, *for each value of the explanatory variable*, the corresponding proportion who do and do not view bill-paying as a major concern. Fill in this table, writing all proportions as percents rounded to the nearest percent (for example, 14%). The sum of each row should be 100%, except for possible rounding.

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | | | 100% |
| No children | | | 100% |

b. Based on these percentage calculations, does it appear that, *for the people in the sample*, having children made a difference in the opinion about whether bill-paying is a major concern? Circle your choice:
   - No, it made no difference.
   - It appears that people with children were more likely to view it as a major concern.
   - It appears that people with children were less likely to view it as a major concern.

**Exercise 2:** In Exercise 1, we concluded that, for the people in the survey, it appears that people with children were more likely to view bill-paying as a major concern. Use this conclusion to answer the following questions, for the people in the survey.

a. Is there an association between the two variables (having children, viewing paying bills as a major concern)?
b. Are the two variables independent?
c. Does the opinion on paying bills depend on whether or not you have children?
d. Is there a difference between the two groups (those with and those without children), relative to the issue of viewing bill-paying as a major concern?

## <u>10.2 – Review: One-sample Proportion Hypothesis Testing</u>

In a single short paragraph, here is the logic of two-tail hypothesis testing for proportions, for a single sample drawn from a single population. The null hypothesis states that the proportion within the entire population is equal to some specific fixed value, typically written $p_0$. The alternative hypothesis simply states that the null hypothesis is incorrect. To judge whether one should or should not reject the null hypothesis, the researcher takes a sample from the population. If the sample has a proportion that is "close to" the claimed proportion $p_0$, this would seem to support the null hypothesis. On the other hand, if the sample's proportion is "far away from" what the null hypothesis claims, this is evidence against the null hypothesis. If the evidence is strong enough, the researcher rejects the null hypothesis. If we *do*

[2] Solutions to the exercises may be found at the end of the lesson.

reject the null hypothesis, this means that we believe the *alternative* hypothesis is true. Here are a few details:

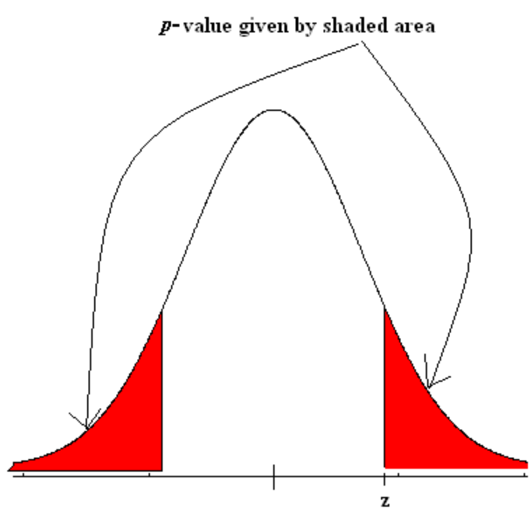
- The null and alternative hypotheses take the form
  $H_0 : p = p_0$
  $H_a : p \neq p_0$
  The null hypothesis claims that, in the entire population, the proportion $p$ is equal to the specific value $p_0$. The alternative hypothesis states that it is not equal to that value.

- The proportion obtained in the sample is identified by the variable $\hat{p}$. If the null hypothesis is correct, the proportion in the sample should be close to $p_0$. The only reason it is not exactly equal to $p_0$ is that there is random variability when samples are taken. The larger the sample size, the closer we expect these two numbers to be, since there is less variability when samples are larger.

- The question, "Should we reject the null hypothesis?" boils down to this related question: "How far is the sample proportion from what it should have been, based on the null hypothesis?" To answer this question, we do two calculations. First, we calculate a *test statistic*. For this type of hypothesis test the test statistic we use is the *z*-score, which measures the distance between the two, in terms of the standard error:

  $$z = \frac{\hat{p} - p_0}{SE}, \text{ where } SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$

  Next, we calculate a *p*-value which measures how far this *z*-score lies out in the tail of the normal distribution, as pictured here:




  We can do these calculations "by hand," or we can use technology such as a statistical calculator or statistical software to assist in the calculations (refer to Lesson 9) [3].

[3] We are putting the phrase *by hand* in quotes for at least two reasons. First, we certainly use a standard calculator for most of the calculations. Second, we frequently use statistics-based technology to calculate the *p*-value (the alternative would be to use a table such as Table A). So, in terms of the calculator supplied with these lessons, "by hand" simply means without using the ***Intervals*** or ***Tests*** menu options. We will continue to follow this convention in the remainder of the lessons.

- If the *p*-value is small, this indicates the sample proportion is a long way from what the null hypothesis would lead us to expect, so we reject the null hypothesis.

- We judge "small" in terms of being smaller than the significance level $\alpha$. Thus, if *p*-value < $\alpha$, we reject the null hypothesis; otherwise, we fail to reject the null hypothesis.

- Two frequently-used values for the significance level are $\alpha = 0.05$ and $\alpha = 0.01$. If we use $\alpha = 0.05$, the likelihood of a Type 1 error is 0.05 or 5%; similarly, using a significance level of 0.01 limits the likelihood of a Type 1 error to 1%. Rejecting a true null hypothesis is a Type 1 error.

- To state our final conclusion, we ask the question: "Was there enough evidence to support the alternative hypothesis?" The sentence we write expresses the answer, using appropriate words to express the meaning of the alternative hypothesis:

    - If we reject the null hypothesis, we write, "There *was* evidence to support the alternative hypothesis."
    - If we fail to reject the null hypothesis, we write, "There *was not* enough evidence to support the alternative hypothesis."

## 10.3 – The Logic of Hypothesis Testing for Categorical-Categorical Association

In this section we describe the logic of hypothesis testing for multiple groups, for the situation where we have two categorical variables. (The actual calculations are covered in later sections.) The explanatory variable identifies the groups we are interested in, and the response variable is what we want to study relative to those groups. The overall logic can be expressed in a short paragraph which is almost identical to the short paragraph on page 5 describing one-sample proportion hypothesis testing.

> The null hypothesis states that there is no difference between the groups identified by the explanatory variable. The alternative hypothesis simply states that the null hypothesis is incorrect. To judge whether one should or should not reject the null hypothesis, the researcher takes a sample from the population. If the sample has data in which the groups are "close to" identical, this would seem to support the null hypothesis. On the other hand, if the sample's data is "far away from" what the null hypothesis claims, this is evidence against the null hypothesis. If the evidence is strong enough, the researcher rejects the null hypothesis. If we *do* reject the null hypothesis, this means that we believe the *alternative* hypothesis is true.

We illustrate the details using the example from Section 10.1. The question we are studying is whether there is an association between having children in a household, and considering bill-paying a major concern. Put another way, is there any difference between the two groups (those with children, those with no children)?

**Null and alternative hypotheses**

The null hypothesis asserts that there is no difference between the groups. As we discussed in Lesson 3, and reviewed briefly in Section 10.1, there are several ways to state a claim that there is no difference between the groups. Speaking in general, we can state:

- There is no difference between the groups identified by the explanatory variable.
- The explanatory variable and the response variable are independent.
- The response variable does not depend on the explanatory variable.
- There is no association (relationship, connection, correlation) between the explanatory variable and the response variable.

One common way to state the null hypothesis is to use the word *independent* (or, equivalently, *does not depend*). Another common usage involves the word *association* (or *relationship*, *connection*, *correlation*). The alternative hypothesis simply states that the null hypothesis is false.

For our continuing example, the explanatory variable is "whether or not you have children" and the response variable is "whether or not you view bill-paying as a major concern." Here is one way to state the hypotheses:

$H_0$ : whether or not you have children, and whether or not you view bill-paying as a major concern, are *independent*
$H_a$ : whether you view bill-paying as a major concern *depends on* whether you have children

Here is another:

$H_0$ : there is *no* association between whether or not you have children, and whether or not you view bill-paying as a major concern
$H_a$ : there *is* an association between whether or not you have children, and whether or not you view bill-paying as a major concern

**Example – looking at the data**

To try to answer the question of association, the researchers took a random sample of 1765 households, with the results shown in Exercise 1 and reproduced here:

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

To further analyze the results, in Exercise 1 we calculated the percentages for each value of the explanatory variable, shown here:

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 60% | 40% | 100% |
| No children | 52% | 48% | 100% |

Based on these percentages, we made a very informal decision that, for the 1765 households that were surveyed, it looked like the two groups are different – the households with children seemed to be more likely to view bill-paying as a major concern.

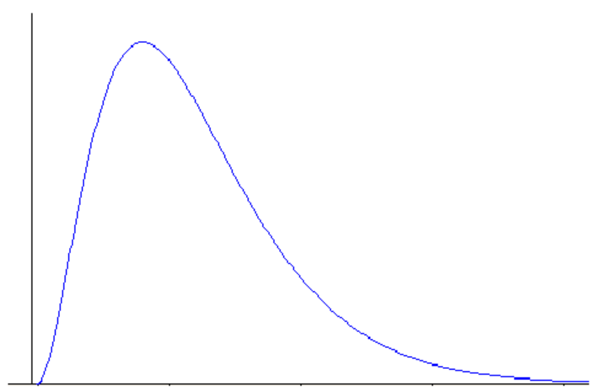
However, just as for the one-sample-proportion problems of Lesson 8, taking samples introduces variability. The question we have to ask is this: ***Are the differences we see in the sample large enough to suggest a difference in the entire population, or are they just the result of sampling variability?*** The next subsection describes the steps statisticians take to answer this question.

**Should we reject the null hypothesis?**

With a few minor differences, the logic used to make this decision is identical to the logic, summarized in Section 10.2, for the one-sample proportion situation.

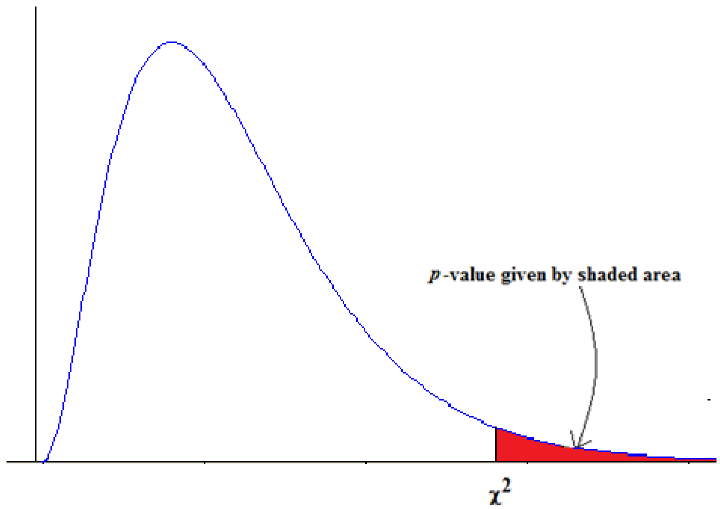
- The question, "Should we reject the null hypothesis?" boils down to this related question: "How far is the sample data from what it should have been, based on the null hypothesis?" To answer this question, we do two calculations.

    1. First, we calculate a test statistic which measures the difference between the sample and the null hypothesis claim. For this type of hypothesis test, the test statistic is known as a **chi-square** score, and the hypothesis test is referred to as a **chi-square hypothesis test**. The test statistic is written as $\chi^2$, where $\chi$ is the Greek letter "chi." The $\chi^2$-score is analogous to the *z*-score of the previous lesson, but the calculations are of course different.
       Statisticians know what should happen when many samples are taken from a population in which the null hypothesis is true, and the resulting $\chi^2$-scores are plotted in a histogram. The resulting *sampling distribution* is called a **chi-square distribution**, and its general shape is illustrated here:

       The results are always positive, and results near zero support the null hypothesis. Just as for one-sample proportion tests, we want to reduce the likelihood of making a Type 1 error (rejecting a true null hypothesis). So our logic is this: if we get a large $\chi^2$-score (out in the right tail of the chi-square distribution), we will reject the null hypothesis.

    2. So, we calculate a *p*-value which measures how far this $\chi^2$-score lies out in the tail of the chi-square distribution, as pictured below. The *p*-value calculates the probability of obtaining a $\chi^2$-score this large, or larger, assuming a true null hypothesis.




Section 10.4 describes how to calculate the $\chi^2$-score and the *p*-value using technology, namely the calculator provided by the author of these lessons. Your instructor may provide information on other forms of technology. It is also possible (not difficult, but a bit tedious) to do the calculations entirely "by hand." This topic is addressed in Section 10.5.

- If the *p*-value is small, this indicates that the sample contains data that is very different from what we would expect if the null hypothesis were true. Therefore, we reject the null hypothesis (and accept the alternative hypothesis).

- We judge "small" in terms of being smaller than the significance level α. Thus, if *p*-value < $\alpha$, we reject the null hypothesis; otherwise, we fail to reject the null hypothesis.

- As always, there is a connection between the significance level and the occurrence of Type 1 error. We can control the likelihood of committing a Type 1 error (rejecting a true null hypothesis) by our choice of significance level.

- If the *p*-value is not less than α, we fail to reject the null hypothesis. We acknowledge that the null hypothesis could be true. Notice that this does not imply that the null hypothesis *is* true, only that it *might be* true.

**Stating conclusions**

To state our final conclusion, we ask the question: "Was there enough evidence to reject the null hypothesis?" Put another way, we ask: "Was there enough evidence to support the alternative hypothesis?" The sentence we write expresses the answer, using appropriate words to express the meaning of the alternative hypothesis.

- If we reject the null hypothesis, we write a sentence stating that there *was* evidence to support the alternative hypothesis.
- If we fail to reject the null hypothesis, we write a sentence stating that there *was not* enough evidence to support the alternative hypothesis.

**Note:** As was true for one-sample proportion hypothesis tests, the final conclusion is written in terms of what the alternative hypothesis states. We either state that there *was* enough evidence to support the alternative hypothesis, or that there *was not* enough evidence to support the null hypothesis. In either case, the sentence will include a description of what the alternative hypothesis states.

For our example involving children and bill-paying, we can use the methods described in the following sections to calculate the test statistic ($\chi^2$-score) and $p$-value, obtaining these results:

$\chi^2 = 12.1928$
$p$-value = 0.0005

Using either $\alpha = 0.05$ or $\alpha = 0.01$, we reach the same conclusion: reject the null hypothesis. We therefore need to come up with a sentence describing the fact that we have found evidence supporting the alternative hypothesis. For our current example, we wrote the alternative hypothesis in a couple of different ways:

$H_a$ : whether you view bill-paying as a major concern depends on whether you have children

$H_a$ : there *is* an association between whether or not you have children, and whether or not you view bill-paying as a major concern

Corresponding to these different ways to write the alternative hypothesis, here are two ways to write our conclusion (using significance level 0.01):

There was evidence, at $\alpha = 0.01$, to indicate that whether you view bill-paying as a major concern depends on whether you have children.

There was evidence, at $\alpha = 0.01$, to suggest an association between having children and viewing bill-paying as a major concern.

As we learned in Lesson 8, another common way to express the conclusion uses the word **significant** or **statistically significant**. Use of this word indicates that the researchers rejected the null hypothesis. So, since we did reject the null hypothesis, we might write something similar to these examples:

Researchers studied whether or not bill-paying was a major concern, for households with and without children. They found a significant difference between the two groups.

Researchers found a significant connection between having children and finding bill-paying a major concern.

**Comment.** Remember that the word "significant" in this context does not necessarily imply "large." As used by statisticians, it means exactly this:

1. We found a difference in the sample we surveyed.
2. That difference was large enough that we believe it reflects a difference in the entire population from which the sample was drawn.
3. The researcher herself will no doubt have reported either a significance level $\alpha$ or the $p$-value obtained in the study. But that part of the report almost certainly will not be reported in the news media. In fact, the news media frequently omits the word "significant" in its report.

If we do not reject the null hypothesis, we simply negate the sentence we would write if we rejected the null hypothesis. Here are the resulting sentences, corresponding to the sentences written above – differences are highlighted in bold face.

> There was **not enough** evidence, at $\alpha = 0.01$, to indicate that whether you view bill-paying as a major concern depends on whether you have children.
>
> There was **not enough** evidence, at $\alpha = 0.01$, to suggest an association between having children and viewing bill-paying as a major concern.
>
> Researchers studied whether or not bill-paying was a major concern, for households with and without children. They found **no** significant difference between the two groups.
>
> Researchers found **no** significant connection between having children and finding bill-paying a major concern.

**Example[4].** Is one party or the other more likely to vote in the upcoming election – Democrats, Republicans, or perhaps Independents? To study this, a researcher randomly sampled 1423 registered voters, asking two questions: 1. What is your party affiliation (Democrat, Republican, Independent)? 2. Do you plan to vote in the upcoming election?

Of the 449 Democrats surveyed, 205 indicated they planned to vote; for Republicans the results were 145 out of 387; for Independents 220 out of 592.

Describe a suitable hypothesis test to answer the question posed. Report the results using $\alpha = 0.05$; using $\alpha = 0.01$.

**Solution.** First of all, a chi-square test is appropriate because we are asking about the association between two categorical variables, party affiliation (Democrat, Republican, or Independent) and voting plans in the upcoming election (will vote, will not vote). The explanatory variable is party affiliation, with voting plans as the response variable. One possible first step would be to organize the reported results in a contingency table.

| | Plans to vote | Does not plan to vote |
|---|---|---|
| Democrat | 205 | 449 – 205 = 244 |
| Republican | 145 | 387 – 145 = 242 |
| Independent | 220 | 592 – 220 = 372 |

We can do some preliminary analysis as follows: for Democrats the proportion who plan to vote is $\frac{205}{449} = 45.66\%$; for Republicans, $\frac{145}{387} = 37.47\%$; and for Independents, $\frac{220}{592} = 37.16\%$. Based on these calculations, we can see that, ***in the sample***, the proportions are different. But it requires a chi-square test to see if the difference observed in the sample is large enough to suggest a difference in the entire population.

**Hypotheses**. As usual, the null hypothesis states that the groups identified by the explanatory variable are all the same. Using the word *association*, we write:

---

[4] Reminder: Many examples in these lessons, including the current example, involve studies of the type which professional statisticians typically carry out, but with data created by the author of the lessons to illustrate the statistical concepts involved.

$H_0$: there is *no* association between party affiliation and voting plans in the upcoming election

The alternative hypothesis states that the null hypothesis is false:

$H_a$: there *is* an association between party affiliation and voting plans in the upcoming election

**Calculations.** Using one of the techniques described in the following two sections, we calculate a test statistic and a *p*-value. Here are the results we obtain:

$\chi^2 = 9.0097$
*p*-value = 0.0111

**Conclusions using $\alpha = 0.05$.** Because 0.0111 is less than 0.05, we reject the null hypothesis. We have enough evidence to support the alternative hypothesis. In the context of this problem, we write: *We found evidence, at $\alpha = 0.05$, to suggest that there is an association between party affiliation and voting plans in the upcoming election.*

**Conclusions using $\alpha = 0.01$.** Because 0.0111 is **not** less than 0.01, we **do not** reject the null hypothesis. We **do not** have enough evidence to support the alternative hypothesis. In the context of this problem, we write: *There was not enough evidence, at $\alpha = 0.05$, to suggest that there is an association between party affiliation and voting plans in the upcoming election.*

Notice that the conclusion you reach can depend upon your choice of significance level. As always, the choice of significance level is related to the likelihood of rejecting a true null hypothesis, that is the likelihood of a Type I error.

**Exercise 3:** Researchers are studying the connection between smoking status and highest level of education. In a randomly chosen sample, they obtain this data:

| | Nonsmoker | Smoker |
|---|---|---|
| Primary school | 45 | 126 |
| Secondary school | 35 | 87 |
| University | 43 | 68 |

a. Which of the following are correct ways to write the null hypothesis?
   - There is a connection between smoking status and highest level of education.
   - Highest level of education and smoking status are independent.
   - Smoking status depends on highest level of education.
   - There is no correlation between smoking status and highest level of education.

b. Using the methods you will learn in later sections, we obtain a test statistic of $\chi^2 = 5.1602$, with a corresponding *p*-value of 0.0758. At significance level $\alpha = 0.01$, does the evidence establish a connection between smoking status and highest level of education?

c. Write the conclusion, at $\alpha = 0.05$, using the word *significant*.

The applets at the following links provide additional practice in formulating hypotheses and drawing conclusions for categorical-categorical association.

Hypotheses and conclusions for chi-square tests

Chi-square interpretation

## 10.4 – Calculations for $\chi^2$ Hypothesis Tests: Using Technology

In this section we will show you how to use technology – specifically, the online calculator[5] provided by the author of these lessons – to do the calculations involved in a $\chi^2$ hypothesis test. Here again is a link to that calculator:

Statistical calculator

We will use the data from Exercise 1, reproduced below, to illustrate the process. We have labeled this the table of ***observed counts*** – it contains the counts that were observed in the sample.

| **Observed counts** | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

Here are the steps you use to carry out the chi-square hypothesis test using the calculator.

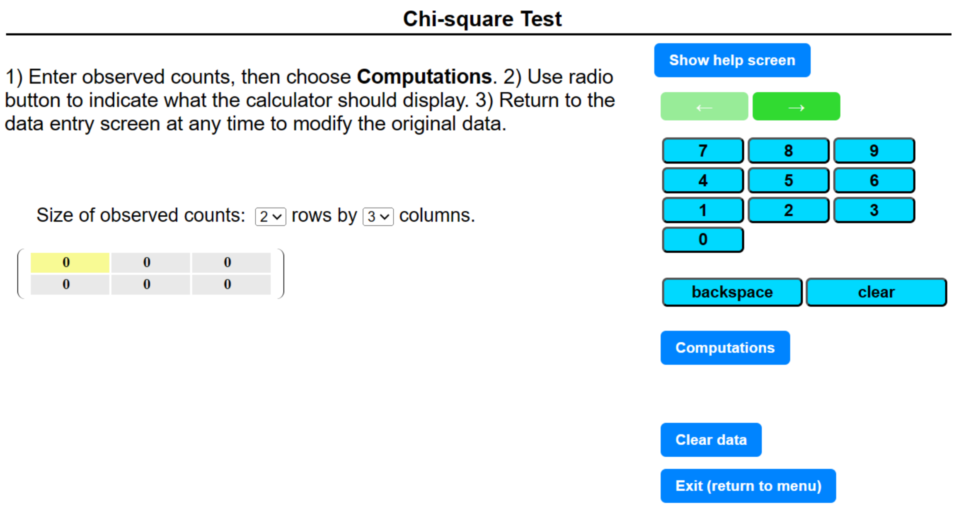
**Note:** The table as shown above contains a row of totals and a column of totals. However, these values will not be entered into your calculator. You will instead enter only the actual counts observed by the researchers, as shown in this version of the contingency table:

| **Observed counts** | Major concern | Not a major concern |
|---|---|---|
| Have children | 515 | 340 |
| No children | 473 | 437 |

**1.** Begin by entering the observed counts, as follows.

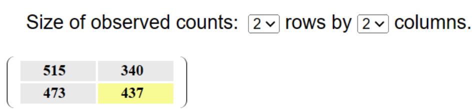
- Choose menu option *Tests* (this is a hypothesis test), then submenu option *Chi square*, to obtain this screen:

[5] In this section we describe the calculator designed for solving problems such as those typically encountered in an introductory statistics class. In Section 10.6 we will use the datafile-based calculator to analyze categorical variables in a data file.

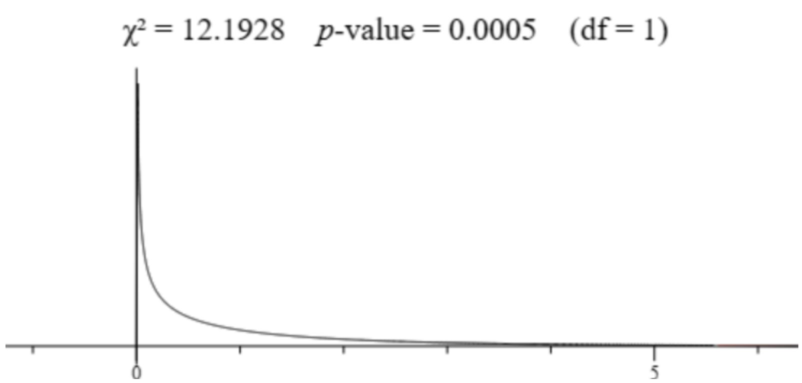



- Use the pull-down lists to set the number of rows and the number of columns.
  **Reminder: Do not include the "totals" row and column! This table has two rows and two columns.**

- Input the observed counts exactly as they appear in the table, as shown here:




**2.** Run the test by clicking the *Computations* button, with these results:




Using technology, the test statistic is $\chi^2$ = 12.1928, with a corresponding *p*-value of 0.0005. At either significance level (0.05 or 0.01) we reach the same conclusion: reject the null hypothesis.

**Comments**:

- The meaning of the "df = 1" result will be explained in Section 10.5 on doing the calculations "by hand."
- By default the calculator displays the results and a corresponding graph. The radio button can be used to display details of the "by hand" calculations.
- Generally speaking, the graph displays a visualization of the *p*-value. For this particular example the $\chi^2$ value of 12.1928 is so large it is past the end of the scale for the graph.

**Exercise 4:** Is there an association between gender and political party preference? Test the following:

$H_0$ : gender and political party preference are independent
$H_a$ : there is an association between gender and political party preference

Data from the 2002 GSS gave the following results:

| | **Democrat** | **Independent** | **Republican** | **Total** |
|---|---|---|---|---|
| **Female** | 567 | 534 | 395 | 1496 |
| **Male** | 356 | 460 | 369 | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

Is this enough evidence to refute the null hypothesis? Give the test statistic, the *p*-value, and your conclusion using $\alpha = 0.01$. Report your conclusion four ways:

- Do you reject the null hypothesis? (YES / NO)
- There __________ (is / is not) enough evidence to conclude that there _________ (is / is not) an association between gender and political party preference.
- Researchers _________________ (did / did not) find a significant difference between males and females when it comes to political party preference.
- There ________________ (is / is not) enough evidence to conclude that the political party preference ____________________ (depends on / is independent of) gender.

The applet at the following link provides additional practice in formulating hypotheses, doing calculations, and drawing conclusions for chi-square tests.

Chi-square calculations

In addition, you may use this applet for more practice in drawing conclusions about the association between two categorical variables.

Chi-square interpretation

## 10.5 – Calculations for $\chi^2$ Hypothesis Tests: "by hand"

We will use the example in Exercise 1 to illustrate how the $\chi^2$ score is calculated "by hand." Before we begin the explanation, it will be useful to review some additional ideas from Lesson 3. Specifically, the null hypothesis states that there is no association between the variables. What patterns would we expect to find in a contingency table if the null hypothesis is true? To answer this, we revisit two exercises you first worked in Lesson 3.

**Exercise 5:** Consider the survey described in Exercise 1. In another town, the results were somewhat different, as shown here:

| | Bills a major concern? | | |
|---|---|---|---|
| Have children? | Yes | No | Totals |
| Yes | 492 | 328 | 820 |
| No | 459 | 306 | 765 |
| Totals | 951 | 634 | 1585 |

Using "do you have children" as the explanatory variable, add the conditional proportions; the sum of the rows should be 100%.

If you worked this exercise, you may have noticed the following. The percentage answering "yes" to the question about bills was the same for both groups (60%). As a result, the percentage for the entire town was also 60%. This is a pattern which always holds:

*If the percentages for each possible value of the explanatory variable are the same, the percentages for the entire set of data will be the same also. Put another way, if there is no association between the variables, the percentages for each group will match the percentages for the entire set of data.*

This pattern is also true for situations where the variables have multiple values, as illustrated in the following exercise.

**Exercise 6:** Suppose a survey question has three possible answers (call them choice (a), choice (b), and choice (c)). Of the 1200 males surveyed, 52% choose (a), 37% (b), and the rest (c). For the 900 females, the percentages are the same. Fill in the counts in this contingency table. Then use the results to calculate the percentages for the entire set of data.

| | (a) | (b) | (c) | Totals |
|---|---|---|---|---|
| Males | | | | 1200 |
| Females | | | | 900 |
| Totals | | | | 2100 |

**Expected counts**

We now return to the data from Exercise 1, reproduced here. We have labeled this the table of ***observed counts*** – it contains the counts that were observed in the sample.

| **Observed counts** | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

The first step is to calculate what "should have happened" if the null hypothesis is true. The null hypothesis states that the two groups are the same: viewing bill-paying as a major concern does not depend on whether you have children. This means that the percent answering yes should be the same in both groups. As we have noted, this further means that the percent answering yes in each group should match the percent for the entire sample. Specifically:

- 988 out of 1765 (approximately 56%) answered "major concern" in the entire sample. Therefore, if the null hypothesis is true, 56% of the families with children should have answered "major concern", and 56% of the families with no children should have answered "major concern."

- 777 out of 1765 (approximately44%) answered "not a major concern" in the entire sample. Therefore, if the null hypothesis is true, 44% of the families with children should have answered "not a major concern", and 44% of the families with no children should have answered "not a major concern."

We use this to build a table of "expected counts" as described below.

We can determine *how many* households with children should have answered "major concern," 56% of the 855 households. For more accuracy, we use $\frac{988}{1765}$ rather than 56% or 0.56 in our calculation:

$$855 * \left(\frac{988}{1765}\right) = 478.61$$

Similarly, we can determine these "expected counts" for each of the other three cells, using these calculations:

44% , or more precisely $\frac{777}{1765}$, of the 855 with children should have said "not a major concern":

$$855 * \left(\frac{777}{1765}\right) = 376.39$$

56% of the 910 with no children should have said "major concern":

$$910 * \left(\frac{988}{1765}\right) = 509.39$$

44% of the 910 with no children should have said "not a major concern":

$$910 * \left(\frac{777}{1765}\right) = 400.61$$

This gives the following table of expected counts.

| **Expected counts** | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 478.61 | 376.39 | 855 |
| No children | 509.39 | 400.61 | 910 |
| Totals | 988 | 777 | 1765 |

**Note:** Once we calculated the 478.61 figure, we could use the totals to calculate all the other expected count. For example, 855 – 478.61 = 376.39 for the "have children / not a major concern" entry. We say this table has ***degrees of freedom*** = 1. Once one cell has been calculated using the formula, all the others are determined (forced, in fact) by the totals in the rows and columns. In general, the degrees of freedom is given by the formula

(# rows – 1) * (# columns – 1)

In this case (2 – 1) * (2 – 1) = 1 * 1 = 1. Note that only the data rows and columns are counted (not the "totals" row and column).

**Exercise 7:** Calculate the expected counts for the data shown below, as follows:

- Do the first two entries in the first row.
- Use subtraction from the totals to fill in the rest of the table.
- What is the value of the degrees of freedom for this table?

| **Observed** | **Democrat** | **Independent** | **Republican** | **Total** |
|---|---|---|---|---|
| **Female** | 567 | 534 | 395 | 1496 |
| **Male** | 356 | 460 | 369 | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

| **Expected** | **Democrat** | **Independent** | **Republican** | **Total** |
|---|---|---|---|---|
| **Female** | | | | 1496 |
| **Male** | | | | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

## The $\chi^2$ statistic

The $\chi^2$ statistic gives a measure of how far the observed counts are from the expected counts. The obvious thing to do is to calculate "*observed* minus *expected*" for each cell, but this would give some positive numbers and some negative numbers, which would cancel each other out. Similar to the calculation for standard deviation, the solution to this problem is to square the results. We also need to adjust for the size of the sample. (A difference of 10 when the expected count was 3 is huge; a difference of 10 when the expected count was 23,000 is small.) This adjustment is done by dividing by the expected count. To summarize, we do this calculation for each cell:

$$\frac{(observed - expected)^2}{expected}$$

For example, for the first cell:

$$\frac{(515 - 478.61)^2}{478.61} = 2.7668$$

Here are the complete results:

| **$\chi^2$ contribution** | Major concern | Not a major concern |
|---|---|---|
| Have children | 2.7668 | 3.5182 |
| No children | 2.5996 | 3.3055 |

The $\chi^2$ statistic is the sum of these four numbers, $\chi^2 = 12.1901$.

**Exercise 8:** Use the results from Exercise 7 to fill in this table. Then calculate the $\chi^2$ statistic.

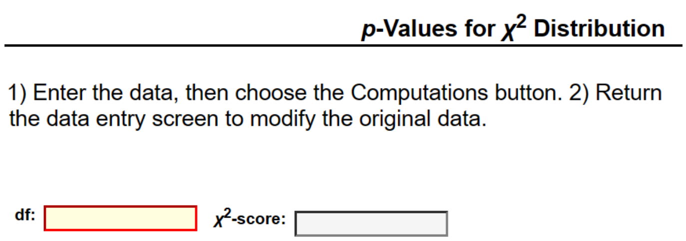
| **$\chi^2$ contribution** | **Democrat** | **Independent** | **Republican** |
|---|---|---|---|
| **Female** | | | |
| **Male** | | | |

**The *p*-value**

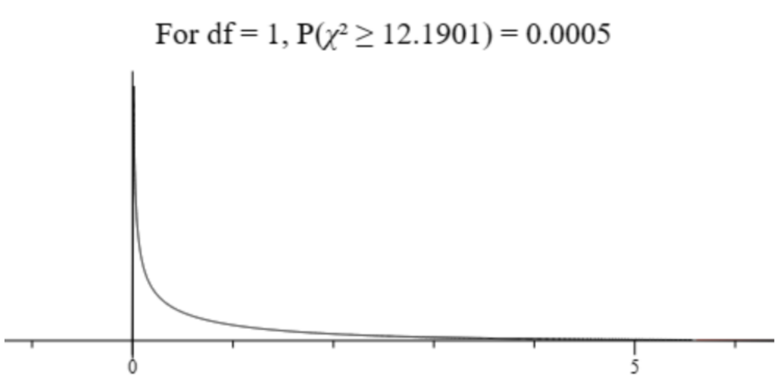
Once you have the $\chi^2$ value, you can use tables to get a range of associated *p*-values. (This is the approach taken by many textbooks.) Preferably, you can use technology to do the calculation. One option is the calculator provided with these lessons, at this link:

Statistical calculator

Use option *Distributions,* submenu option *$\chi^2$ distribution p-value* to obtain this screen:




Fill in the degrees of freedom (1) and the $\chi^2$-score (12.1901). When you click *Computations* you obtain this result:



Because the *p*-value is very small, we reject the null hypothesis. There *is* strong evidence that whether you view bill-paying as a major concern *does* depend on whether there are children in the household.

**Exercise 9:** Calculate the p-value based on the results of Exercise 8.

**Comment on expected counts.** When you do the calculations "by hand," the first step is to find the table of expected counts. When you perform the calculations using technology, it is also possible to determine the expected counts. For the calculator provided with these lessons, the output concludes these choices for what should be displayed:

Display:
- (●) Results and graph
- ( ) Observed counts
- ( ) Expected counts
- ( ) Contribution to chi-square

If you choose the third option, you obtain this output – compare it to what we just calculated.

| | |
|---|---|
| 478.6062 | 376.3938 |
| 509.3938 | 400.6062 |

Choosing the fourth option displays each cell's contribution to the $\chi^2$ score. Again, compare it to what we just calculated – the small difference are because we rounded at each step of our calculations.

| | |
|---|---|
| 2.7674 | 3.5189 |
| 2.6002 | 3.3063 |

The applet at the following link provides additional practice in formulating hypotheses, doing calculations, and drawing conclusions for chi-square tests.

[Chi-square calculations]

In addition, you may use this applet for more practice in drawing conclusions about the association between two categorical variables.

[Chi-square interpretation]

## 10.6 – $\chi^2$ Calculations for Data Files

In Lessons 2, 3, and 9 we have examined the use of the second calculator provided with these lessons to analyze the data in a data file. In this lesson we will see that this same calculator can be used to carry out a $\chi^2$ test to study the association between two categorical variables present in a data file. The data file we use is the same used in those earlier lessons. You should have it saved to your own computing device, but in any case here again is a link to that file:

[First day survey]

As a reminder, the file contains student responses to a first day survey containing these questions:

1. What is your gender? (M) Male (F) Female
2. What is your class year? (FR) Freshman (SO) Sophomore (JR) Junior (SR) Senior
3. How many states have you visited?
4. Do you currently smoke? (Y) Yes (N) No
5. How tall are you (in inches)?
6. How many days per week do you read a newspaper?

To begin, open the data file calculator using the following link, then use the *Load vertical file* button to load the data file containing the student responses.

Data file calculator

**Example. A**nalyze the association between a student's class year and whether or not they smoke, by running an appropriate $\chi^2$ test, with the class year will be the explanatory variable.

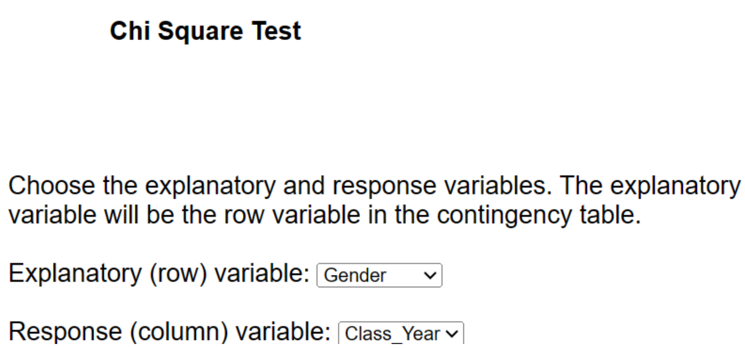
**Solution.** First of all, notice that both variables are categorical, so a $\chi^2$ test is the proper tool for analyzing the association. These are the pertinent hypotheses written in terms of the word "association":

$H_0$ : there is **no** association between a student's class and whether or not they smoke
$H_a$ : there **is** an association between a student's class and whether or not they smoke

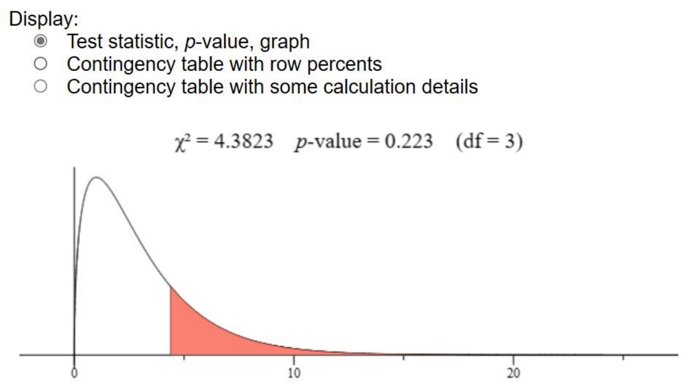
We can also write the hypotheses in terms of independence:

$H_0$ : whether or not a student smokes is independent of their class
$H_0$ : whether or not a student smokes depends on their class

To carry out the test, we use menu option *Tests*, submenu option *Chi Square Test*, obtaining this screen:




Change the explanatory variable to *Class_Year* and the response variable to *Smoke*, and choose *Computations*, to obtain this result:




At either α = 0.01 or α = 0.05 the result is the same. There is not enough evidence to conclude that whether a student smokes depends on the class year. Notice that we can use the radio buttons to obtain additional information about the corresponding calculations.

**Comment.** In Lesson 3 we examined this same data using a contingency table and side by side bar charts. At that time we had this conclusion:

> *Both the contingency table and the graph make it clear that the percentage of smokers for freshmen is more than that for the other classes.*

Now in this lesson we have concluded:

> *There is not enough evidence to conclude that whether a student smokes depends on the class year.*

On the surface these conclusions may seem to be contradictory. However, we must remember that in Lesson 3 we were drawing a conclusion about the particular set of students that had been surveyed. In that collection (sample) of students there was a difference between the different classes. However, in this lesson we found that that difference in the sample was not large enough to be able to conclude that there is a difference in the entire population of all students who take that course.

**Exercise 10:** Carry out a chi square test to analyze the association between gender and class year. Do female students tend to take the course in a different year than their male counterparts?

**Exercise 11:** In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Use that data file to carry out a chi square test to analyze the association between the *Political party* variable and the *Politics* variable.

## Solutions to Exercises

**1:** Here again is the data from our earlier example.

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | 515 | 340 | 855 |
| No children | 473 | 437 | 910 |
| Totals | 988 | 777 | 1765 |

a. To get a preliminary feel for possible association, we can calculate, *for each value of the explanatory variable*, the corresponding proportion who do and do not view bill-paying as a major concern. Fill in this table, writing all proportions as percents rounded to the nearest percent (for example, 14%). The sum of each row should be 100%, except for possible rounding.

| | Major concern | Not a major concern | Totals |
|---|---|---|---|
| Have children | $\frac{515}{855}$ =60% | $\frac{340}{855}$ =40% | 100% |
| No children | $\frac{473}{910}$ =52% | $\frac{437}{910}$ =48% | 100% |

b. Based on these percentage calculations, does it appear that *for these people* having children made a difference in the opinion about whether bill-paying is a major concern? Circle your choice:

- No, it made no difference.
- It appears that people with children were more likely to view it as a major concern.
- It appears that people with children were less likely to view it as a major concern.

It appears that people with children were more likely to view it as a major concern (60% answered yes, compared to only 52% for the other group).

**2:** In exercise 1, we concluded that, for the people in the survey, it appears that people with children were more likely to view bill-paying as a major concern. Use this conclusion to answer the following questions, for the people in the survey.

a. Is there an association between the two variables (having children, viewing paying bills as a major concern)? YES
b. Are the two variables independent? NO
c. Does the opinion on paying bills depend on whether or not you have children? YES
d. Is there a difference between the two groups (those with and those without children), relative to the issue of viewing bill-paying as a major concern? YES

**3:** Researchers are studying the connection between smoking status and highest level of education. In a randomly chosen sample, they obtain this data:

| | Nonsmoker | Smoker |
|---|---|---|
| Primary school | 45 | 126 |
| Secondary school | 35 | 87 |
| University | 43 | 68 |

a. Which of the following are correct ways to write the null hypothesis?
   - There is a connection between smoking status and highest level of education. Not correct (this would be the alternative hypothesis)
   - Highest level of education and smoking status are independent. Correct
   - Smoking status depends on highest level of education. Not correct (this would be the alternative hypothesis)
   - There is no correlation between smoking status and highest level of education. Correct

b. Using the methods you will learn in later sections, we obtain a test statistic of $\chi^2 = 5.1602$, with a corresponding $p$-value of 0.0758. At significance level $\alpha = 0.01$, does the evidence establish a connection between smoking status and highest level of education? No

c. Write the conclusion using the word *significant*. Researchers found no significant connection, at $\alpha = 0.01$, between smoking status and highest level of education.

**4:** Is there an association between gender and political party preference? Test the following:

$H_0$ : gender and political party preference are independent
$H_a$ : there is an association between gender and political party preference

Data from the 2002 GSS gave the following results:

| | **Democrat** | **Independent** | **Republican** | **Total** |
|---|---|---|---|---|
| **Female** | 567 | 534 | 395 | 1496 |
| **Male** | 356 | 460 | 369 | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

Is this enough evidence to refute the null hypothesis? Give the table of expected counts, the test statistic, the *p*-value, and your conclusion using $\alpha = 0.01$. Report your conclusion four ways:

- Do you reject the null hypothesis? (YES / NO)
- There __________ (is / is not) enough evidence to conclude that there _________ (is / is not) an association between gender and political party preference.
- Researchers _________________ (did / did not) find a significant difference between males and females when it comes to political party preference.
- There ________________ (is / is not) enough evidence to conclude that the political party preference ____________________ (depends on / is independent of) gender.

The $\chi^2$-test statistic value is 18.8056 and the $p$-value is 0.0001.

Since the *p*-value is small we **do reject** the null hypothesis, and report that:

- There **is** enough evidence to conclude that there **is** an association between gender and political party preference.
- Researchers **did** find a significant difference between males and females when it comes to political party preference.
- There **is** enough evidence to conclude that the political party preference **depends on** gender.

**5:** Consider the survey described in Exercise 1. In another town, the results were somewhat different, as shown here:

| | Bills a major concern? | | |
|---|---|---|---|
| Have children? | Yes | No | Totals |
| Yes | 492<br>60% | 328<br>40% | 820 |
| No | 459<br>60% | 306<br>40% | 765 |
| Totals | 951<br>60% | 634<br>40% | 1585 |

Using "do you have children" as the explanatory variable, add the conditional proportions; the sum of the rows should be 100%. See the table above.

**6:** Suppose a survey question has three possible answers . Call them choice (a), choice (b), and choice (c). Of the 1200 males surveyed, 52% choose (a), 37% (b), and the rest (c). For the 900 females, the percentages are the same. Fill in the counts in this contingency table. Then use the results to calculate the percentages for the entire set of data.

| | (a) | (b) | (c) | Totals |
|---|---|---|---|---|
| Males | 624 | 444 | 132 | 1200 |
| Females | 468 | 333 | 99 | 900 |
| Totals | 1092<br>52% | 777<br>37% | 231<br>11% | 2100 |

**7:** Calculate the expected counts for the data shown below, as follows:

- Do the first two entries in the first row.
- Use subtraction from the totals to fill in the rest of the table.
- What is the value of the degrees of freedom for this table?

| Observed | Democrat | Independent | Republican | Total |
|---|---|---|---|---|
| **Female** | 567 | 534 | 395 | 1496 |
| **Male** | 356 | 460 | 369 | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

| Expected | Democrat | Independent | Republican | Total |
|---|---|---|---|---|
| **Female** | 515.03 | 554.65 | 426.32 | 1496 |
| **Male** | 407.97 | 439.35 | 337.68 | 1185 |
| **Total** | 923 | 994 | 764 | 2681 |

Here are the calculations:

$1496\left(\frac{923}{2681}\right) = 515.03$ $1496\left(\frac{994}{2681}\right) = 554.65$ 1496 – 515.03 – 554.65 = 426.32

923 – 515.03 = 407.97 994 – 554.65 = 439.35 764 – 426.32 = 337.68

The degrees of freedom (df) is 2. We know this because once we calculated the first two values, all the rest were determined by simply subtracting from the totals. Alternatively, since there are 2 rows and 3 columns of data, we can calculate using the formula (rows – 1)(columns – 1) = (1)(2) = 2.

**8:** Use the results from Exercise 7 to fill in this table. Then calculate the $\chi^2$ statistic.

| $\chi^2$ contribution | Democrat | Independent | Republican |
|---|---|---|---|
| **Female** | 5.2441 | 0.7688 | 2.3010 |
| **Male** | 6.6203 | 0.9706 | 2.9049 |

Here are the calculations:

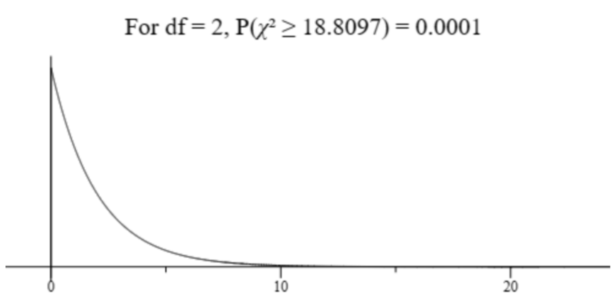
$$\frac{(567-515.03)^2}{515.03} = 5.2441 \quad \frac{(534-554.65)^2}{554.65} = 0.7688 \quad \frac{(395-426.32)^2}{426.32} = 2.3010$$

$$\frac{(356-407.97)^2}{407.97} = 6.6203 \quad \frac{(460-439.35)^2}{439.35} = 0.9706 \quad \frac{(369-337.68)^2}{337.68} = 2.9049$$

$\chi^2$ = 5.2441 + 0.7688 + 2.3010 + 6.6203 + 0.9706 + 2.9049 = 18.8097. Notice that more than half of the total comes from the entries in the first column.

**9:** Calculate the p-value based on the results of Exercise 8.

Use, option *Distributions,* submenu option *$\chi^2$ distribution p-value.* Fill in the values 2 for the df and 18.8097 for the $\chi^2$-score. Click *Computations* to obtain this result:



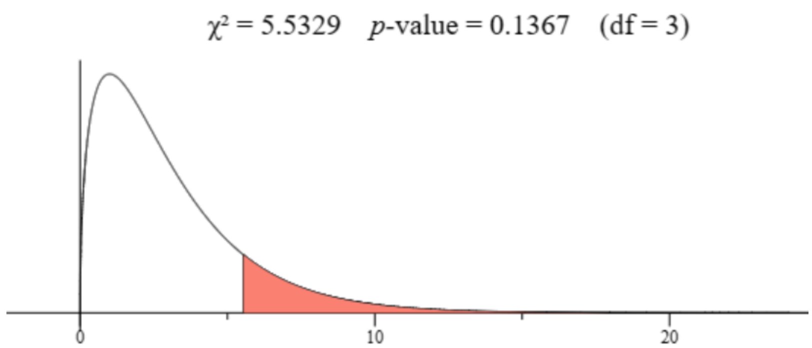
**10:** Carry out a chi square test to analyze the association between gender and class year.

These are the hypotheses:

$H_0$ : there is **no** association between gender and class year

$H_a$ : there **is** an association between gender and class year

With menu option *Tests*, submenu *Chi square*, we choose *Gender* as the explanatory variable and *Class_Year* as response variable. Clicking *Computations* gives this result:




The *p*-value is larger than 0.05, so we conclude: There is not enough evidence to support an association between gender and class year in the population of all students who take this course.

Do female students tend to take the course in a different year than their male counterparts? NO

**Comment.** Here is the contingency table.

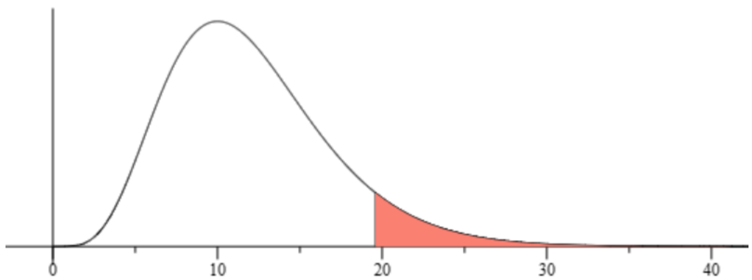
Rows: *Gender*
Columns: *Class_Year*

| | FR | JR | SO | SR | Totals |
|---|---|---|---|---|---|
| F | 0<br>0.00% | 9<br>37.50% | 13<br>54.17% | 2<br>8.33% | 24<br>100.00% |
| M | 7<br>17.95% | 14<br>35.90% | 14<br>35.90% | 4<br>10.26% | 39<br>100.00% |
| Totals | 7<br>11.11% | 23<br>36.51% | 27<br>42.86% | 6<br>9.52% | 63<br>100.00% |

In the **sample** of students surveyed, the female student are more likely to take the class as a sophomore and less likely to take it as a freshman, with the percentages for junior and senior approximately the same for both genders. **But the difference is not large enough to imply a difference in the entire population.**

**11:** In Exercise 16 of Lesson 2 you created a data file containing the data presented originally in Lesson 1. This data was collected in a statistics course at a public university. Use that data file to carry out a chi square test to analyze the association between the *Political party* variable and the *Politics* variable.

$\chi^2 = 19.584$  $p$-value = 0.0754  (df = 12)




At $\alpha = 0.05$ there is not enough evidence ($p = 0.0754$) to establish an association between the variables.
**Note:** Below is the corresponding contingency table, which definitely does show an association in the group of students surveyed. In particular, all the Democrats' responses to the *Politics* question ranged from 1 to 4, with the Republicans' responses ranging from 3 to 7. But the association we see in the sample is not enough to establish a corresponding association in the entire population.

Rows: *Political Party*
Columns: *Politics*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Totals |
|---|---|---|---|---|---|---|---|---|
| D | 1<br>8.33% | 6<br>50.00% | 1<br>8.33% | 4<br>33.33% | 0<br>0.00% | 0<br>0.00% | 0<br>0.00% | 12<br>100.00 |
| I | 0<br>0.00% | 0<br>0.00% | 1<br>33.33% | 0<br>0.00% | 1<br>33.33% | 1<br>33.33% | 0<br>0.00% | 3<br>100.00 |
| R | 0<br>0.00% | 0<br>0.00% | 1<br>7.69% | 5<br>38.46% | 3<br>23.08% | 3<br>23.08% | 1<br>7.69% | 13<br>100.00 |
| Totals | 1<br>3.57% | 6<br>21.43% | 3<br>10.71% | 9<br>32.14% | 4<br>14.29% | 4<br>14.29% | 1<br>3.57% | 28<br>100.00 |