Case C→C: Finding the p-value

Learning Objective: In a given context, carry out the appropriate inferential method for comparing relationships and draw the appropriate conclusions.

Step 3: Finding the p-value

The p-value for the chi-square test for independence is the probability of getting counts like those observed, assuming that the two variables are not related (which is what is claimed by the null hypothesis). The smaller the p-value, the more surprising it would be to get counts like we did, if the null hypothesis were true.

Technically, the p-value is the probability of observing χ^2 at least as large as the one observed. Using statistical software, we find that the p-value for this test is 0.201.

Step 4: Stating the conclusion in context

As usual, we use the magnitude of the p-value to draw our conclusions. A small p-value indicates that the evidence provided by the data is strong enough to reject H_o and conclude (beyond a reasonable doubt) that the two variables are related. In particular, if a significance level of 0.05 is used, we will reject H_o if the p-value is less than 0.05.

Example

A p-value of 0.201 is not small at all. There is no compelling statistical evidence to reject H_o, and so we will continue to assume it may be true. Gender and drunk driving may be independent, and so the data suggest that a law that forbids sale of 3.2% beer to males and permits it to females is unwarranted. In fact, the Supreme Court, by a 7-2 majority, struck down the Oklahoma law as discriminatory and unjustified. In the majority opinion Justice Brennan wrote (http://www.law.umkc.edu/faculty/projects/ftrials/conlaw/craig.html):

"Clearly, the protection of public health and safety represents an important function of state and local governments. However, appellees' statistics in our view cannot support the conclusion that the gender-based distinction closely serves to achieve that objective and therefore the distinction cannot under [prior case law] withstand equal protection challenge."

Scenario: Alcohol Problems Among 9/11 First Responders

The purpose of this activity is to draw our conclusion regarding the relationship between participation in the 9/11 rescue and risk of alcohol problems among New York firefighters and first responders.

A chi-square test regarding the relationship between participation in the 9/11 rescue and risk of alcohol problems among New York firefighters and first responders produced the following results:

Cell format
Count Expected count

	No Risk	M to S Risk	Total
Yes	793 822.7	309 279.3	1102
No	441 411.3	110 139.7	551
Total	1234	419	1653

Chi-Square test:

Statistic	DF	Value	P-value
Chi-square	1	12.661749	0.0004

Learn By Doing

1/1 point (graded)

Based on the p-value results and a significance level of 0.05, which of the following is an appropriate conclusion?

- The data provided do not present enough evidence to reject the null hypothesis. Therefore, we can conclude that the risk of alcohol problems among New York firefighters **is** related to participation in the 9/11 rescue.

- The data provided do not present enough evidence to reject the null hypothesis. Therefore, we can conclude that the risk of alcohol problems among New York firefighters **is not** related to participation in the 9/11 rescue.
- The data provided present enough evidence to reject the null hypothesis. Therefore, we can conclude that the risk of alcohol problems among New York firefighters **is** related to participation in the 9/11 rescue. ✔
- The data provided present enough evidence to reject the null hypothesis. Therefore, we can conclude that the risk of alcohol problems among New York firefighters **is not** related to participation in the 9/11 rescue.

Answer
Correct:
In order to reject the null hypothesis of no relationship between two categorical variables, the p-value must be less than 0.05. The calculated p-value of 0.0004 is less than 0.05; therefore, the null hypothesis should be rejected and the conclusion drawn that there is a relationship between alcohol problems in New York firefighters and participation in the 9/11 rescue.

Submit

Comment

This is a good opportunity to illustrate an important idea that was discussed earlier in this unit: The larger the sample the results are based on, the more evidence they carry. Let's take the previous example and simply multiply each of the counts by 3:

Gender	Drank Alcohol in Last 2 Hours? Yes	No	Total
Male	231	1212	1443
Female	48	366	414
Total	279	1578	1857

and see what would have happened if these were the original data. Obviously, the conditional counts would remain the same:

Gender	Drank Alcohol in Last 2 Hours? Yes	No
Male	231/1443=16.0%	1212/1443=84.0%
Female	48/414=11.6%	366/414=88.4%

In other words, the sample provides the "same" results, but this time they are based on a much larger sample (1857 instead of 619). This is reflected by the chi-square test. In this case, software gives us a chi-square statistic of 4.910 and a p-value of 0.027.

As before, H_o states that gender and drunk driving are not related; H_a states that they are related. Since the observed counts are triple what they were before, the expected counts are also tripled. When done with software the original chi-square statistic was 1.637 since software doesn't round as much. The chi-square statistic when we tripled the data is 3 times 1.637, or 4.91 (which now is in the "large" range). Therefore, the p-value is smaller and is now .027.

Now, we do reject H_o, and we conclude that gender and drunk driving are related. In this case, the "largest contribution to chi-square" is large enough to provide evidence of a relationship. This is due to the fact that so few females drove drunk (48) compared to the number that would be expected (62.2, which is 414 * 279 / 1857) if the variables gender and drunk driving were not related. This contribution is $\frac{(48-62.2)^2}{62.2} = 3.242$.