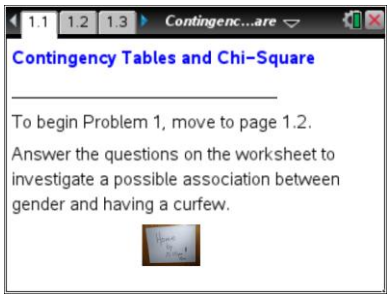
Contingency Tables and Chi-Square
Student Activity

Name _____

Class _____

Open the TI-Nspire™ document *Contingency_Tables_and_Chi_Square.tns.*



Is it more likely that a female high school student will have a curfew than a male high school student? A statistic called a chi-square is calculated from a random sample of 100 students and interpreted to determine whether or not gender and having a curfew are associated.

Move to pages 1.2 and 1.3.

Press ctrl ▶ and ctrl ◀ to navigate through the lesson.

This activity involves generating a number of random samples from a population. In order to avoid having your results be identical to those for another student in the room, it is necessary to "seed" the random number generator. Read the instructions on Page 1.2 for seeding your random number generator, then carry out that seeding on Page 1.3.

In the activity entitled **Two-way Tables and Association**, a randomly selected sample of 100 students from Rufus King High School was asked the question, "Do you have a curfew?" The results from the 100 students are summarized in the following table:

	Had a curfew	Did not have curfew	Total
Females	33	35	68
Males	20	12	32
Total	53	47	100

Move to page 1.4.

1. Based on the conditional probabilities displayed in this table, determine the following:
   a. What is the probability that if a male is selected from the sample, he has a curfew?

   b. What is the probability that if a female is selected from the sample, she has a curfew?

   c. Do you think this sample provides us with enough evidence to conclude that males at King are more likely to have a curfew? Explain your reasoning.

The sample of King students represents the **observed values of one sample**. If gender and having a curfew were **independent variables**, then the probability of males having a curfew would be the same as the probability of females having a curfew. If, however, the probability of males with a curfew was not the same as the probability of females with a curfew, then the variables would be **associated** and not independent.

2. If a student's gender and having a curfew were independent, what percent of males and what percent of females would have a curfew? Explain your reasoning.

3. The highlighted cell in the table on page 1.4 represents the number of males with a curfew. Enter values in the highlighted cell to estimate the number of males and the number of females that result in approximately the same proportion of students with a curfew.
   a. How many of the males would have a curfew? How many of the females?

   b. Fill in the following table with the exact values that would indicate the proportion of males and females with a curfew are the same. (Note that the number of males and females cannot be represented by decimals values in reality.) These values, however, are the **expected values** as they represent the values you would expect if curfew and gender were independent.

Expected values

	Had a curfew	Did not have curfew	Total
Females			68
Males			32
Total	53	47	100

A statistic, called a chi-square, is derived from the following formula:

$$\chi^2 = \sum \frac{(Expected - Observed)^2}{Expected}$$

4. a. Based on the formula, explain what you think a chi-square statistic represents.

   b. Why do you think the square of the difference is divided by the expected value rather than summing the squared differences and calculating the average?

Move to page 1.5.

Assume the following key for the cells in the table displayed on Page 1.5:

- MC represents males with a curfew.
- FC represents females with a curfew.
- M_NC represents males with no curfew.
- F_NC represents females with no curfew.

5. Refer back to the original set of observations. The observed value for the number of males from King who had a curfew was 20. The table on Page 1.5 begins with a sample in which 20 males indicated they had a curfew and 33 females indicated they had a curfew. What is the value of the chi-square statistic for 20 males and 33 females with a curfew?

The question to think about is whether a chi-square value indicates that the differences between the observed and expected values are extreme or not. The following questions are designed to figure out the answer to that question by assuming it is possible to change the number of males with a curfew.

6. The cell highlighted on Page 1.5 in the row labeled as MC represents the number of males with a curfew.
   a. Change the number of observed males with a curfew to determine the following chi-square values.

Number of males with a curfew	Chi-square
27 males	
22 males	
20 males	
17 males	
12 males	
5 males	

   b. What was the smallest chi-square value in your answer to part a? Explain why this number of males with a curfew results in a small chi-square value.

   c. Find the largest chi-square value. Explain why this number of males with a curfew results in such a large chi-square value.

7. Use the table on Page 1.5 to determine the value of the chi-square in the following situations:
   a. Consider the case where none of the males had a curfew. What would be the value of the chi-square for this sample?

   b. Consider the case where all of the males had a curfew. What would be the value of the chi-square for this sample?

   c. What conjecture might you make about the association between gender and curfew if you observed either of the chi-squares values in parts a and b?

Given your answer to part c, the question to be addressed is how large would a chi-square value have to be to conjecture that curfew and gender are associated.

Move to page 1.6.

Page 1.6 will generate simulated samples drawn from a population in which the variables gender and curfew are not associated. Each sample has 32 males and 68 females. 53 of the students are randomly selected from the 100 students and are considered the students who have a curfew (recall that 53% of the students have a curfew). The value of interest is the number of males selected from the 53 students with a curfew.

8. Use the arrow on the left to select a sample.
   a. How many males have a curfew if 53 students were randomly selected from this sample of 32 males and 68 females?

   b. If this sample were compared to the expected values in which gender and curfew are independent, what is the value of the chi-square for this sample?

9. Generate at least five more samples. Note that clicking on a dot in either the top or bottom dot plot will highlight the corresponding dot in the other dot plot. Move the cursor over a dot to display its value. (Note: Click in an empty region in either plot to deselect the dots.)

   Use the dot plots to find the observed number of males that resulted in the smallest chi-square value and the observed number of males that resulted in the largest chi-square value. Compare your answer to others in the class.

10. Continue to use the arrow to obtain a total of 100 samples. As you generate each sample, note the relationship between the observed number of males and the location of the dot that represents the chi-square value.
    a. How does this relationship seem to fit with your answer to 7c?

    b. Describe the simulated distribution for the number of males with a curfew.

    c. Describe the simulated distribution representing the chi-squares values for the 100 samples.

A sample that results in a large chi-square value is a noticeably different sample than the one identified with the expected values. A sample resulting in a large chi-square value indicates the variables are not independent. The question remains, however, what value of a chi-square is considered sufficiently large to suggest that gender and having a curfew are not likely to be independent? If we can identify a large chi-square value, then any sample resulting in this chi-square value or greater would be used to conjecture that there is an association between the variables in our population. Statisticians often identify “large” chi-squares as the largest chi-square values representing 5% or less of the chi-square distribution.

11. Consider the chi-square distribution for the 100 samples:
    a. Estimate the five largest chi-square values in the simulated sampling distribution. Click each point, and estimate the corresponding number of males with a curfew identified in the simulated distribution of males with a curfew. (There might be a cluster of points around some of the chi-square values. Estimate the largest chi-square values in the cluster and the connections to the number of males with a curfew.)

    b. Explain why some of the large chi-square values were a result of more males than expected having a curfew and other large chi-square values were a result of fewer males than expected having a curfew.

12. Recall the original sample obtained from King at the beginning of this activity.
    a. What was the number of males with a curfew in the original sample? Click on one of the points representing that number and estimate the corresponding chi-square value in the simulated chi-square distribution.

    b. Look at the distribution in the lower panel on Page 1.6. Approximately how many of the 100 chi-square values were greater than or equal to 1.7?

13. Based on the general standard for evaluating large chi-square values, does the observed number of 20 males with a curfew indicate there is an association between curfew and gender at King? Explain your answer.

14. Henry was overheard complaining that it was unfair that he had curfew, while the females in his school were less likely to have a curfew. Based on the random sample that started this activity, is Henry's claim that females are less likely to have a curfew at King an accurate statement? Explain your answer.