Chapter 12: Inference on Categories of Data

Source: M&M's

Did you know that M&M's® Milk Chocolate Candies are supposed to come in the following percentages: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown? (Note: These values are different from those that used to be available on the M&M's website, but have been confirmed by ScientificAmeriken.) Do they really? Could a quality control engineer test that? How far from those expected percentages is acceptable?

These are all questions we're going to answer in Section 12.1, using something called a **Goodness-of-Fit Test**.

In Chapter 4, we studied relationships between two variables. We learned that we could quantify the strength of the linear relationship between two quantitative variables with the correlation.

What about qualitative (categorical) variables, though? For example, suppose we consider a survey given to 82 students in a Basic Algebra course at ECC, with the following responses to the statement "I enjoy math."

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Men	9	13	5	2	1
Women	12	18	11	6	5

How do we study this relationship? Is there a way to tell if gender and whether the student enjoys math are related? In Section 4.4, we discussed construction conditional distributions and analyzing them, but can we be more precise? In fact, there is a way, and we'll study it in Sections 12.2.

If you're ready to begin, just click on the "start" link below, or one of the section links on the left.

Section 12.1: Goodness-of-Fit Test

Objectives

By the end of this lesson, you will be able to...

1. perform a goodness-of-fit test

In Example 4 from Section 6.2, we assumed that the number of free throws made out of 10 attempts followed the binomial distribution. But does it really? And how do we know? In that problem, we assumed that the free throws were independent, but is that something we could check?

Consider a standard 6-sided die. If we assume a die is fair, each side should be equally likely. Of course, out of 100 tosses, they won't show up an equal number of times (they can't, since 1/6 of 100 is about 16.7). But how far from an equal number is acceptable?

Did you know that M&M's® Milk Chocolate Candies are supposed to come in the following percentages: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown? (Note: These values are different from those that used to be available on the M&M's website, but have been confirmed by ScientificAmeriken.) Do they really? Could a quality control engineer test that? Like the dice, how far from those expected percentages is acceptable?

These are all questions we're going to answer in this section, using something called a **Goodness-of-Fit Test**. Before we do that, we need a little background.

Source: M&M's

Observed vs. Expected Values

Example 1

Consider the M&M's® example above. Suppose we purchase a standard bag of Milk Chocolate M&M's, and observe the following distribution:

	Blue	Orange	Green	Yellow	Red	Brown
Observed Frequency	13	9	12	8	7	7

If we follow the percentages above, we would expect 24% of the M&M's® to be blue. Since we had 56 total, we would expect 24% of 56, or about 13. Similarly, we could fill out the table for the rest of the colors as follows:

	Blue	Orange	Green	Yellow	Red	Brown
Observed Frequency	13	9	12	8	7	7
Expected Frequency	13	11	9	8	7	7

What we're not able to answer now is the severity of these differences - is it significant enough for us to say that the distribution is different from what the company claims?

Expected Counts

In general, the expected count for each category is the number of trials of the experiment, multiplied by the probability of that particular outcome.

$$E_i = n \cdot p_i$$

To test whether the observed values fit the stated distribution, we compare them with the expected, using the **Goodness-of-Fit Test**. Go to the next page to see the details and some examples.

You can also go to the video page for links to see videos in either Quicktime or iPod format.

The Goodness-of-Fit Test

The **Goodness-of-Fit Test** is used to test the distribution of a single variable. In essence it compares the observed values with what we would expect. Before we can begin, we need a new test statistic.

The Test Statistic for Goodness-of-Fit Tests

If we let O_i represent the observed counts for category i, and E_i represent the expected counts, with n independent trials and k categories, then the formula

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

approximately follows the chi-square distribution with k-1 degrees of freedom, provided that

1. all expected frequencies are greater than or equal to 1, and
2. no more than 20% of the expected frequencies are less than 5.

Note: If 1 or 2 fail, we can combine categories so they are satisfied.

Performing a Goodness-of-Fit Test

***Step 1*:** State the null and alternative hypotheses.

H_0: The random variable follows the claimed distribution.
H_1: The random variable does not follow the claimed distribution.

Note: The test is *always* a right-tailed test, since larger deviations from the expected values will result in larger X^2 values.

***Step 2*:** Decide on a level of significance, α.

***Step 3*:** Compute the test statistic, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$.

***Step 4*:** Determine the *P*-value.

***Step 5*:** Reject the null hypothesis if the *P*-value is less than the level of significance, α.

***Step 6*:** State the conclusion.

Example 2

Consider the M&M's®, from Example 1. Based on the observed counts in that bag, is there enough evidence at the 5% level of signficance to say that the distribution is different from what the company claims?

[reveal answer]

Chi-Square Goodness-of-Fit Test Using StatCrunch

1. You'll need to calculate the expected counts based on the assumed distribution.
2. Enter the observed counts in the first column, and the expected counts in the second column.
3. Choose **Stat** > **Goodness-of-fit** > **Chi-Square test**
4. Select the columns for the observed counts.
5. Select the columns for the expected counts.
6. Select **Calculate**.

The results should appear.

More information is available in the help file through StatCrunch.

Example 3

Consider the standard 6-sided die we mentioned earlier this section. If we assume a die is fair, each side should be equally likely. Suppose we roll a die 100 times and observe the results shown below. Is there enough evidence at the 5% level of significance to say that the die is not fair?

	1	2	3	4	5	6
Observed Frequency	22	12	17	13	10	26

[reveal answer]

Section 12.2: Tests for Independence and the Homogeneity of Proportions

Objectives

By the end of this lesson, you will be able to...

1. perform a test for independence
2. perform a test for homogeneity of proportions

In the previous section, we considered the relationship between a student's gender and whether he or she enjoys math. One question we might have as a result of this is whether we can determine whether there is a statistical test to determine if there is a relationship between the two variables.

Of course, we wouldn't be mentioning it if there wasn't! Before we discuss that test, we need a little background.

Determining Expected Counts

Let's assume that a student's gender and whether he or she enjoys math are independent. What frequencies would we expect in that case? Let's consider again the survey data from Example 2 in Section 4.4:

In that example, a survey was given to 82 students in a Basic Algebra course at ECC, with the following responses to the statement "I enjoy math."

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Men	9	13	5	2	1
Women	12	18	11	6	5

We then created a **relative frequency marginal distribution**, which was calculated by taking the row/column totals and dividing by the sample size of 82.

	SA	A	N	D	SD	Total
Men	9	13	5	2	1	30/82 ≈ 0.37
Women	12	18	11	6	5	52/82 ≈ 0.63
Total	21/82 ≈ 0.26	31/82 ≈ 0.39	16/82 ≈ 0.20	8/82 ≈ 0.10	6/82 = 0.07	1

Let's focus on the first cell - "Men" and "Strongly Agree". From the table, we can see that 30/82 or about 37% of the students were men, and 21/82 or about 26% of the students strongly agreed with the statement "I enjoy math." If they really are independent, we can use the Multiplication Rule for independent events, where P(E and F) = P(E)•P(F).

So if they are independent, the probability that a student is is both male and strongly agrees would be:

$$\text{P(male and strongly agrees)} = \frac{30}{82} \cdot \frac{21}{82} \approx 0.094$$

We can then use this probability to determine how many we would *expect* in that cell, if the two variables are actually independent. We just multiply the total number of individuals by the probability of being both male and strongly agreeing:

$$\text{Expected number of students who are male and strongly agree} = 82 \cdot \frac{30}{82} \cdot \frac{21}{82} = \frac{30 \cdot 21}{82} \approx 7.68$$

In general, we can find the expected values using this formula:

$$\text{Expected Frequency} = \frac{(\text{row total}) \cdot (\text{column total})}{\text{table total}}$$

Example 1

Use the table provided and find the expected frequency for each outcome.

[reveal answer]

Now that we have the expected frequencies for each outcome, we need a new hypothesis test to see if these expected counts are far enough from what we actually observed to say that the variables *aren't* independent.

The Test for Independence

The test we use to determine if there is an association between two qualitative variables is called the **chi-square test for independence**. In this test, the null hypothesis is always that the variables are not associated (independent), and the alternative is that they are associated (depedent).

The test works by comparing the observed counts with the expected counts if we assume the two variables are related. If those are far enough apart, we can say that we think there *is* a relationship. Here are the details:

The Test Statistic for the Test of Independence

If we let O_i represent the observed counts for the i^{th} cell, and E_i represent the expected counts, then

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

approximately follows the chi-square distribution with (r-1)(c-1) degrees of freedom, where r is the number of rows and c is the number of columns, provided that:

1. all expected frequencies are greater than or equal to 1, and
2. no more than 20% of the expected frequencies are less than 5.

Note: If 1 or 2 fail, we can combine categories so they are satisifed.

Performing a Chi-Square Test for Independence

***Step 1*:** State the null and alternative hypotheses.

H_0: The row and column variables are independent.
H_1: The row and column variables are dependent.

Note: Like the Goodness-of-Fit Test, this test is *always* right-tailed, since larger deviations from the expected values will result in larger X^2 values.

***Step 2*:** Decide on a level of significance, α.

***Step 3*:** Compute the test statistic, $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$.

***Step 4*:** Determine the *P*-value.

***Step 5*:** Reject the null hypothesis if the *P*-value is less than the level of significance, α.

***Step 6*:** State the conclusion.

Example 2

Use the data from earlier examples to determine if gender and whether a student enjoys math are related. Perform the test at the 5% level of significance.

From earlier, we have the observed counts:

	SA	A	N	D	SD
Men	9	13	5	2	1
Women	12	18	11	6	5

And from Example 1, we know the expected counts are:

	SA	A	N	D	SD
Men	7.68	11.34	5.85	2.93	2.20
Women	13.32	19.66	10.15	5.07	3.80

[reveal answer]

Chi-Square Test for Independence Using StatCrunch

1. You'll need to first enter the data, with row and column labels. (Leave the first column for row labels.)
2. Choose **Stat** > **Tables** > **Contingency** > **with summary**
3. Select the columns for the observed counts.
4. Select the column for the row variable.
5. Click **Next**.
6. Check "Expected Count" and select **Calculate**.

The results should appear.

More information is available in the help file through StatCrunch.

You can also go to the video page for links to see videos in either Quicktime or iPod format.

Example 3

Repeat the previous example using technology.

[reveal answer]

The Test for Homogeneity of Proportions

Suppose the Math Department at ECC would like to compare success rates in its College Algebra course based on how students placed into the class. There are currently three ways of placing into the course:

1. earning a C or better in the Mth098 - Intermediate Algebra; or
2. an appropriate placement test score; or
3. a Math ACT of 23 or better.

In this case, the department might want to analyze the proportion who are successful in College Algebra (i.e. earning a C or better). They wonder if the proportions are all the same, or if one is different. One way to answer this would be to do proportion tests with all of the possible pairs, but that would entail three separate tests like those we studies in Section 11.3.

Another option is a new test - the **chi-square test for homogeneity of proportions**. In a chi-square test for homogeneity of proportions, we test the claim that different populations have the same proportion of individuals with a certain characteristic.

> Interestingly, the procedures for performing a chi-square test for homogeneity of proportions is identical to that for the test of independence.

Example 4

In the Fall of 2005, the ECC math department asked the Institutional Research department to collect data from previous semesters to analyze. The table below shows the results for Fall 2004 and Spring 2005.

	Mth098	placement	ACT 23+
successful	132	94	163
not successful	140	52	62

Is there evidence to indicate that the proportion of students in each group who are successful is different at the α = 0.01 level of significance?

[reveal answer]