Stat 201: Statistics I
Chapter 11

Metropolitan State University

April 9, 2018

Chapter 11
Goodness-of-Fit and Contingency Tables

Section 11.1
Goodness-of-Fit

Frequency distributions

Recall, a **frequency distribution** is a listing of counts, or number of occurrences, of data within a class (quantitative data) or category (categorical data).

Frequency distributions

Recall, a **frequency distribution** is a listing of counts, or number of occurrences, of data within a class (quantitative data) or category (categorical data).

A frequency table is a list of the distribution of a sample drawn from a population.

Frequency distributions

Recall, a **frequency distribution** is a listing of counts, or number of occurrences, of data within a class (quantitative data) or category (categorical data).

A frequency table is a list of the distribution of a sample drawn from a population.

A test can be conducted to see if the population a sample is drawn from has an expected distribution.

Frequency distributions, example

Example

- A six-sided die is "fair" if the frequencies of each possible result of a roll (1 through 6) are equal.

  Given a sample of results from a number rolls of a particular die, a test could be conducted to test whether the die is "fair".

Frequency distributions, example

Example

- A six-sided die is “fair” if the frequencies of each possible result of a roll (1 through 6) are equal.

  Given a sample of results from a number rolls of a particular die, a test could be conducted to test whether the die is “fair”.

- M&Ms should have the following distribution of colors:

Color	Blue	Brown	Green	Orange	Red	Yellow
Percent	24%	14%	15%	20%	13%	14%

  Given a sample a M&Ms, a test could be conducted to test whether M&Ms really do have that distribution of colors.

Goodness-of-fit tests

A **goodness-of-fit test** is an hypothesis test which tests whether an observed frequency distribution matches, or fits, an expected distribution.

Goodness-of-fit tests

A **goodness-of-fit test** is an hypothesis test which tests whether an observed frequency distribution matches, or fits, an expected distribution.

- H_0 : The frequency counts agree with the expected distribution.

Goodness-of-fit tests

A **goodness-of-fit test** is an hypothesis test which tests whether an observed frequency distribution matches, or fits, an expected distribution.

- H_0 : The frequency counts agree with the expected distribution.
- H_a : The frequency counts do not agree with the expected distribution.

Goodness-of-fit tests

A **goodness-of-fit test** is an hypothesis test which tests whether an observed frequency distribution matches, or fits, an expected distribution.

- H_0 : The frequency counts agree with the expected distribution.
- H_a : The frequency counts do not agree with the expected distribution.
- Test statistic follows a χ^2 (chi-squared) distribution with $k-1$ degrees of freedom

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

  where
  - k is the number of classes or categories
  - O is the observed count for each class or category, from sample
  - E is the expected count for each class or category if the expected distribution is true

Expected counts

The expected count for each class or category can be calculated by

$$E = P(c) \times n$$

where

- $P(c)$ is the probability of class or category c
- n is the sample size

Expected counts

The expected count for each class or category can be calculated by

$$E = P(c) \times n$$

where

- $P(c)$ is the probability of class or category c
- n is the sample size

For expected uniform distributions, since $P(c) = 1/k$ where k is the number of classes or categories, the expected count for each class is

$$E = \frac{1}{k} \times n = \frac{n}{k}$$

Expected counts, example

Example

- Since a "fair" die has a uniform frequency distribution, the expected counts for each result for a sample of 100 die rolls is

$$E = \frac{n}{k} = \frac{100}{6} = 16.67$$

Expected counts, example

Example

- Since a "fair" die has a uniform frequency distribution, the expected counts for each result for a sample of 100 die rolls is

$$E = \frac{n}{k} = \frac{100}{6} = 16.67$$

- Blue M&Ms are expected to have a frequency of 24%. Thus, out of a sample of 150 M&Ms, the expected count for blue is

$$E = P(c) \times n = 0.24 \times 150 = 36$$

Decisions for goodness-of-fit tests

Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

Decisions for goodness-of-fit tests

Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.

Decisions for goodness-of-fit tests

Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.
- Critical value: If the calculated test statistic is greater than the critical value for the significance level and degrees of freedom ($\chi^2 > \chi^2_{\alpha,k-1}$), then reject the null hypothesis. Critical values can be found in Table A-4.

Decisions for goodness-of-fit tests

Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.
- Critical value: If the calculated test statistic is greater than the critical value for the significance level and degrees of freedom ($\chi^2 > \chi^2_{\alpha,k-1}$), then reject the null hypothesis. Critical values can be found in Table A-4.

For both methods, if conditions are not met to reject, fail to reject the null hypothesis.

Decisions for goodness-of-fit tests

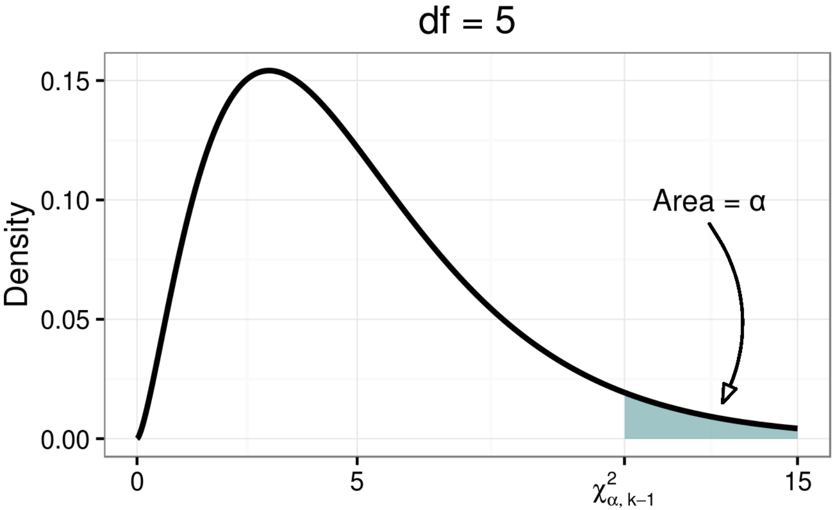
Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.
- Critical value: If the calculated test statistic is greater than the critical value for the significance level and degrees of freedom ($\chi^2 > \chi^2_{\alpha,k-1}$), then reject the null hypothesis. Critical values can be found in Table A-4.

For both methods, if conditions are not met to reject, fail to reject the null hypothesis.

Note: Chi-square test statistics are always positive and chi-square tests are always one-sided. Large values of χ^2 cause rejection of the null.

Chi-square distribution



Requirements for goodness-of-fit tests

- The sample is a simple random sample

Requirements for goodness-of-fit tests

- The sample is a simple random sample
- The sample data consists of frequency counts for each of the different categories

Requirements for goodness-of-fit tests

- The sample is a simple random sample
- The sample data consists of frequency counts for each of the different categories
- For each class or category, the expected count is at least 5

Goodness-of-fit tests in StatCrunch

- Stat $\rightarrow$ Goodness-of-fit $\rightarrow$ Chi-Square Test
- Select column that contains observed data
- Specify expected distribution:
  - For uniform distributions, select "All cells in equal proportion"
  - For non-uniform distributions, select the column which contains expected frequencies
- Leave default value of "Expected" for display
- Click "Compute!"
- The test statistic and p-value are found in "Chi-Square" and "P-value"

Uniform goodness-of-fit test, example

Example

To determine if there is evidence is is not a die is “fair”, roll the die 40 times and perform a goodness-of-fit test on the results.

A “fair” die will have a uniform frequency distribution, so each result has a probability of $1/6$ (16.67%).

Uniform goodness-of-fit test, example

Example

To determine if there is evidence is is not a die is "fair", roll the die 40 times and perform a goodness-of-fit test on the results.

A "fair" die will have a uniform frequency distribution, so each result has a probability of $1/6$ (16.67%).

- H_0 : The frequency distribution of rolls fits a uniform distribution
  H_a : The frequency count of at least one result differs from the others

Uniform goodness-of-fit test, example

Example

To determine if there is evidence is is not a die is "fair", roll the die 40 times and perform a goodness-of-fit test on the results.

A "fair" die will have a uniform frequency distribution, so each result has a probability of $1/6$ (16.67%).

- H_0 : The frequency distribution of rolls fits a uniform distribution
  H_a : The frequency count of at least one result differs from the others
- Requirements: The expected count for each result is
  $E = \frac{1}{6} \times 40 = 6.667 > 5$

Uniform goodness-of-fit test, example

Example

To determine if there is evidence is is not a die is "fair", roll the die 40 times and perform a goodness-of-fit test on the results.

A "fair" die will have a uniform frequency distribution, so each result has a probability of $1/6$ (16.67%).

- H_0 : The frequency distribution of rolls fits a uniform distribution
  H_a : The frequency count of at least one result differs from the others
- Requirements: The expected count for each result is $E = \frac{1}{6} \times 40 = 6.667 > 5$
- Find test statistic χ^2, p-value and report decision

Group work

- Complete question 1.

Section 11.2
Contingency Tables

Contingency tables

Recall, a **contingency table** is a two dimensional table (rows and columns) displaying frequency counts of classes or categories of two factors for a single sample.

Contingency tables, example

Example

Recall the cancer screening example. A sample of 1000 randomly selected people where given a new screening test for a particular kind of of cancer. Each subject either has cancer or doesn't, and either tested positive or tested negative.

	Test Result	
Diagnosis	Positive	Negative
Cancer	74	13
No cancer	26	887

Contingency tables, example

Example

Recall the example of the school distract attempting to reduce the rate of teen drivers who text or email. The school distract created an educational program that was attend by about half the students. Afterwards, a survey was taken of a sample of teen drivers. Each teen driver either attended the program or didn't, and either texted or emailed while driving or didn't.

	Texted or emailed?	
Attended program?	Yes	No
Yes	62	150
No	59	114

Independence in contingency tables

An important question that can be asked about data in contingency tables is whether the two factors are independent.

Independence in contingency tables

An important question that can be asked about data in contingency tables is whether the two factors are independent.

Factors are independent if the value of one factor does not impact the value of the other factor. In other words, if the probability of being in a category of one factor does not change depending on the category of the second factor, for all categories of both factors, then the factors are independent

Independence in contingency tables

Example

- For the cancer screening example, if the probability of testing positive is the same regardless of whether the subject has cancer or not, then the test results and cancer status are independent.

Independence in contingency tables

Example

- For the cancer screening example, if the probability of testing positive is the same regardless of whether the subject has cancer or not, then the test results and cancer status are independent.
- For the teen driver example, if the probability of a teen driver texting or emailing is the same regardless of whether they attended the educational program or not, then texting or emailing and program attendance are independent.

Test for independence

A **test for independence** is an hypothesis test which tests whether data contained in a contingency table represents two factors that are independent.

Test for independence

A **test for independence** is an hypothesis test which tests whether data contained in a contingency table represents two factors that are independent.

- H_0 : The two factors are independent

Test for independence

A **test for independence** is an hypothesis test which tests whether data contained in a contingency table represents two factors that are independent.

- H_0 : The two factors are independent
- H_a : The two factors are dependent

Test for independence

A **test for independence** is an hypothesis test which tests whether data contained in a contingency table represents two factors that are independent.

- H_0 : The two factors are independent
- H_a : The two factors are dependent
- Test statistic follows a χ^2 (chi-squared) distribution with $(r-1)\times(c-1)$ degrees of freedom

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where

- r is the number of rows and c is the number of columns
- O is the observed count for each table cell, from sample
- E is the expected count for each table cell if the factors are independent

Expected counts

Like with goodness-of-fit tests, the expected count for each cell is the probability for that cell under the null hypothesis times the sample size.

$$E = P(c) \times n$$

Expected counts

Like with goodness-of-fit tests, the expected count for each cell is the probability for that cell under the null hypothesis times the sample size.

$$E = P(c) \times n$$

Recall, if events are independent that the probability of both being true is the product of the probabilities of both.

$$P(A \text{ and } B) = P(A) \times P(B)$$

Expected counts

Like with goodness-of-fit tests, the expected count for each cell is the probability for that cell under the null hypothesis times the sample size.

$$E = P(c) \times n$$

Recall, if events are independent that the probability of both being true is the product of the probabilities of both.

$$P(A \text{ and } B) = P(A) \times P(B)$$

Thus, if A is an event of one factor and B is an event of the other factor, the expected count for the cell of A and B is

$$E = P(A) \times P(B) \times n$$

Expected counts, cont.

The probability of an event of one factor is the marginal probability, the total count for the row or column divided by the total sample size.

	Factor 2		
Factor 1	B	∼B	Total
A	# (A and B)	# (A and ∼B)	# A
∼A	# (∼A and B)	# (∼A and ∼B)	# ∼A
Total	# B	# ∼B	n

$$P(A) = \frac{\#A}{n} \qquad P(B) = \frac{\#B}{n}$$

Expected counts, cont.

Thus, the expected count for the cell of A and B is

$$E_{A,B} = P(A) \times P(B) \times n = \frac{\#A}{n} \times \frac{\#B}{n} \times n$$

Expected counts, cont.

Thus, the expected count for the cell of A and B is

$$E_{A,B} = P(A) \times P(B) \times n = \frac{\#A}{n} \times \frac{\#B}{n} \times n$$

After some algebra, a simpler formula for expected count is

$$E_{A,B} = \frac{\#A \times \#B}{n}$$

Expected counts, example

Example

	Test Result		
Diagnosis	Positive	Negative	Total
Cancer	74	13	87
No cancer	26	887	913
Total	100	900	1000

Expected counts, example

Example

	Test Result		
Diagnosis	Positive	Negative	Total
Cancer	74	13	87
No cancer	26	887	913
Total	100	900	1000

- $E_{+,\text{cancer}} = \dfrac{100 \times 87}{1000} = 8.7$

Expected counts, example

Example

	Test Result		
Diagnosis	Positive	Negative	Total
Cancer	74	13	87
No cancer	26	887	913
Total	100	900	1000

- $E_{+,\text{cancer}} = \dfrac{100 \times 87}{1000} = 8.7$
- $E_{-,\text{cancer}} = \dfrac{900 \times 87}{1000} = 78.3$

Expected counts, example

Example

	Test Result		
Diagnosis	Positive	Negative	Total
Cancer	74	13	87
	(8.7)	(78.3)	
No cancer	26	887	913
	(91.3)	(821.7)	
Total	100	900	1000

Decisions for tests for independence

Once a test statistic χ^2 is found, the decision process for a test for independence is identical to the process for a goodness-of-fit test:

Decisions for tests for independence

Once a test statistic χ^2 is found, the decision process for a test for independence is identical to the process for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.

Decisions for tests for independence

Once a test statistic χ^2 is found, the decision process for a test for independence is identical to the process for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.
- Critical value: If the calculated test statistic is greater than the critical value for the significance level and degrees of freedom ($\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$), then reject the null hypothesis. Critical values can be found in Table A-4.

Decisions for tests for independence

Once a test statistic χ^2 is found, the decision process for a test for independence is identical to the process for a goodness-of-fit test:

- P-value: If the calculated p-value is less than the significance level ($p < \alpha$), then reject the null hypothesis.
- Critical value: If the calculated test statistic is greater than the critical value for the significance level and degrees of freedom ($\chi^2 > \chi^2_{\alpha,(r-1)(c-1)}$), then reject the null hypothesis. Critical values can be found in Table A-4.

For both methods, if conditions are not met to reject, fail to reject the null hypothesis.

Requirements for tests for independence

- The sample is a simple random sample

Requirements for tests for independence

- The sample is a simple random sample
- The sample data consists of frequency counts for every cell of a contingency table

Requirements for tests for independence

- The sample is a simple random sample
- The sample data consists of frequency counts for every cell of a contingency table
- For every cell, the expected count is at least 5

Tests for independence in StatCrunch

- Stat $\rightarrow$ Tables $\rightarrow$ Contingency $\rightarrow$ With Summary
- Select the columns that contain observed data
- Select the column that contains row labels
- If desired, select calculated values to be displayed ("Expected count" can be useful)
- Leave "Hypothesis tests" on default value of "Chi-Square test for independence"
- Click "Compute!"
- In the "Chi-Square test" table, the test statistic and p-value are found in "Value" and "P-value"

Test for independence, example

Example

Recall the cancer screening data:

	Test Result		
Diagnosis	Positive	Negative	Total
Cancer	74	13	87
	(8.7)	(78.3)	
No cancer	26	887	913
	(91.3)	(821.7)	
Total	100	900	1000

Test whether cancer diagnosis has an effect on the screening test result at $\alpha = 0.01$ level of significance.

Test for independence, example

Example

Test for independence, example

Example

- H_0 : Test result and cancer diagnosis are independent
  H_a : Test result and cancer diagnosis are dependent (test result is associated with cancer diagnosis)

Test for independence, example

Example

- H_0 : Test result and cancer diagnosis are independent
  H_a : Test result and cancer diagnosis are dependent (test result is associated with cancer diagnosis)
- Requirements: The smallest expected count, for cancer and positive test, is 8.7 which is larger than 5

Test for independence, example

Example

- H_0 : Test result and cancer diagnosis are independent
  H_a : Test result and cancer diagnosis are dependent (test result is associated with cancer diagnosis)
- Requirements: The smallest expected count, for cancer and positive test, is 8.7 which is larger than 5
- $\chi^2 = 596.47717$
  $p < 0.0001$

Test for independence, example

Example

- H_0 : Test result and cancer diagnosis are independent
  H_a : Test result and cancer diagnosis are dependent (test result is associated with cancer diagnosis)
- Requirements: The smallest expected count, for cancer and positive test, is 8.7 which is larger than 5
- $\chi^2 = 596.47717$
  $p < 0.0001$
- $p < 0.0001 < 0.01 = \alpha$. Reject null hypothesis.

Test for independence, example

Example

- H_0 : Test result and cancer diagnosis are independent
  H_a : Test result and cancer diagnosis are dependent (test result is associated with cancer diagnosis)
- Requirements: The smallest expected count, for cancer and positive test, is 8.7 which is larger than 5
- $\chi^2 = 596.47717$
  $p < 0.0001$
- $p < 0.0001 < 0.01 = \alpha$. Reject null hypothesis.
- There is evidence that test results and cancer diagnosis are associated.

Test for independence, example

Example

Recall the teen driver data:

	Texted or emailed?	
Attended program?	Yes	No
Yes	62	150
No	59	114

Test whether texting or emailing while driving is associated with program attendance at $\alpha = 0.05$ level of significance.

Test for independence, example

Example

Test for independence, example

Example

- H_0 : Texting or emailing and program attendance are independent
  H_a : Texting or emailing and program attendance are dependent (texting or emailing is associated with program attendance)

Test for independence, example

Example

- H_0 : Texting or emailing and program attendance are independent
  H_a : Texting or emailing and program attendance are dependent (texting or emailing is associated with program attendance)
- Requirements: We don't have expected counts yet, but the sample size is in the hundreds and no outcome appears very unlikely. We can confirm when we have expected counts.

Test for independence, example

Example

- H_0 : Texting or emailing and program attendance are independent
  H_a : Texting or emailing and program attendance are dependent (texting or emailing is associated with program attendance)
- Requirements: We don't have expected counts yet, but the sample size is in the hundreds and no outcome appears very unlikely. We can confirm when we have expected counts.
- $\chi^2 = 1.0435298$
  $p = 0.307$

Test for independence, example

Example

- H_0 : Texting or emailing and program attendance are independent
  H_a : Texting or emailing and program attendance are dependent (texting or emailing is associated with program attendance)
- Requirements: We don't have expected counts yet, but the sample size is in the hundreds and no outcome appears very unlikely. We can confirm when we have expected counts.
- $\chi^2 = 1.0435298$
  $p = 0.307$
- $p = 0.307 > 0.05 = \alpha$. Fail to reject null hypothesis.

Test for independence, example

Example

- H_0 : Texting or emailing and program attendance are independent
  H_a : Texting or emailing and program attendance are dependent (texting or emailing is associated with program attendance)
- Requirements: We don't have expected counts yet, but the sample size is in the hundreds and no outcome appears very unlikely. We can confirm when we have expected counts.
- $\chi^2 = 1.0435298$
  $p = 0.307$
- $p = 0.307 > 0.05 = \alpha$. Fail to reject null hypothesis.
- There is no evidence that texting or emailing while driving and program attendence are associated.

Test for independence vs. proportion test

Recall, we performed a test with this data in Chapter 9. Testing the alternative hypothesis that the proportion of teen drivers who texted or emailed while driving was lower than for teens who attended the educational program than those who did not.

Test for independence vs. proportion test

Recall, we performed a test with this data in Chapter 9. Testing the alternative hypothesis that the proportion of teen drivers who texted or emailed while driving was lower than for teens who attended the educational program than those who did not.

- The results were $z = -1.0215331$, $p = 0.1535$.

Test for independence vs. proportion test

Recall, we performed a test with this data in Chapter 9. Testing the alternative hypothesis that the proportion of teen drivers who texted or emailed while driving was lower than for teens who attended the educational program than those who did not.

- The results were $z = -1.0215331$, $p = 0.1535$.
- Remember, the p-value for a two-sided test with the same data is twice that of a one-sided test. So, a two-sided proportion test would have a p-value of $p = 0.1535 \times 2 = 0.307$.

Test for independence vs. proportion test

Recall, we performed a test with this data in Chapter 9. Testing the alternative hypothesis that the proportion of teen drivers who texted or emailed while driving was lower than for teens who attended the educational program than those who did not.

- The results were $z = -1.0215331$, $p = 0.1535$.
- Remember, the p-value for a two-sided test with the same data is twice that of a one-sided test. So, a two-sided proportion test would have a p-value of $p = 0.1535 \times 2 = 0.307$.
- The test for independence, while conducted as a one-sided test, is actually a two-sided test in that is does not distinguish between observed values that are lower or higher than expected values.

Test for independence vs. proportion test

Recall, we performed a test with this data in Chapter 9. Testing the alternative hypothesis that the proportion of teen drivers who texted or emailed while driving was lower than for teens who attended the educational program than those who did not.

- The results were $z = -1.0215331$, $p = 0.1535$.
- Remember, the p-value for a two-sided test with the same data is twice that of a one-sided test. So, a two-sided proportion test would have a p-value of $p = 0.1535 \times 2 = 0.307$.
- The test for independence, while conducted as a one-sided test, is actually a two-sided test in that is does not distinguish between observed values that are lower or higher than expected values.
- Thus, the test for independence gave us identical results as the equivalent proportion test.

Group work

- Complete questions 2 and 3.