Stat 201: Statistics I
Week 10

Metropolitan State University

Week 10
Inference for Categorical Data

Section 10.1
Hypothesis tests for proportions

Testing claims about population proportions

Claims about a population proportion p are tested using a sample proportion $\hat{p}$.

Recall, sampling distributions of proportion data follow a binomial distribution which, under certain conditions, approximates a normal distribution.

Test statistic for proportion tests

Tests for population proportions use a standard normal sampling distribution. The test statistic is a z-score.

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}}$$

where $\hat{p}$ is the sample proportion, p_0 is the population proportion under the null hypothesis, $q_0 = 1 - p_0$ and n is the sample size.

P-values for proportion tests

Given a z test statistic, the p-value is the probability of getting z-scores more extreme on a standard normal distribution.

- For two-sided test:
  $H_a : p \neq p_0$, p-value $= P(Z < -z) + P(Z > z) = 2 \times P(Z > z)$
- For one-sided test:
  $H_a : p < p_0$, p-value $= P(Z < z)$
  $H_a : p > p_0$, p-value $= P(Z > z)$

Requirements for a proportion test

- Like all hypothesis tests, sample must be a simple random sample.
- Samples of proportion data follow a binomial distribution and must satisfy the binomial requirements:
  - Fixed number of trials
  - Trials are independent
  - Two possible outcomes (success/failure)
  - Probability of each outcome is constant
- For the binomial to approximate a normal distribution, np and nq must both be at least 5.

Steps for a proportion hypothesis test

1. Identify null and alternative hypotheses from research question
   $H_0 : p = p_0$
   $H_a : p \neq p_0,\ p < p_0,\ p > p_0$
2. Check the requirements for using normal distribution
3. Calculate z test statistic
4. Calculate p-value
5. Compare p-value to significance level α and report decision
6. State conclusion in terms of original research question

Hypothesis test for a proportion, example

Example

The proportion of teen drivers who text or email while driving is 40%. A school district adds an education program for all high school students in hopes of lowering that proportion. Two months after the program, the district surveys a random sample of teen drivers. The survey finds that 62 out of 212 teens surveyed had texted or emailed while driving during the previous month. Test whether the program was successful at a 0.05 level of significance.

1. Identify null and alternative hypotheses from research question
   $H_0 : p = 0.4$
   $H_a : p < 0.4$
   Population: teen drivers who had attended the education program

Hypothesis test for a proportion, example

Example

2. Check the requirements for using normal distribution
   - Simple random sample
   - Requirements of a binomial satisfied (fixed sample size, two outcomes, independent, constant probability of success)
   - $p = 0.4$, $q = 0.6$, $n = 212$
     $np > 5$, $nq > 5$
3. Calculate z test statistic
   $z = -3.1964013$
4. Calculate p-value
   $p = 0.0007$

Hypothesis test for a proportion, example

Example

5. Compare p-value to significance level α and report decision
   $p = 0.0007 < \alpha = 0.05$. Reject null hypothesis.
6. State conclusion in terms of original research question
   There is evidence that teen drivers who attended the program text and email while driving at a lower rate than 40%.

Two sample hypothesis tests for proportions

A two sample hypothesis test can be conducted to compare the proportions of two populations.

Notationally, p_1 is the population proportion from the first population and p_2 is the population proportion from the second population. Similarly, $\hat{p}_1 = x_1/n_1$, the sample proportion of the sample from the first population is the number of successes in sample 1 divided by the sample size of sample 1, etc.

It is usually not important which of the two populations is "first", but it is very important that which ever designation is used, it is used consistently through the hypothesis test process.

Requirements

- Both samples must meet the requirements of a binomial distribution (simple random samples, consistent probability of success for each trial, etc.)
- For each sample, the number of successes and the number of failures must both be 5 or greater.

Null hypotheses

The null hypothesis for a two sample proportion test is *always* proportion 1 is equal to proportion 2, or

$$H_0 : p_1 = p_2$$

However, the null hypothesis can be rewritten, without changing it's meaning, as

$$H_0 : p_1 - p_2 = 0$$

This second representation more accurately depicts how the hypothesis is conducted. In most circumstances, either can be used.

Thus, under the null hypothesis, the two population proportions are the same and the common proportion can be estimated by

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Alternative hypotheses

The alternative hypothesis for a two sample proportion test is that proportion 1 differs from proportion 2. Like one sample tests, two sample alternative hypotheses can be one-sided or two-sided. And like the null hypothesis for two sample tests, each alternative hypothesis can be written in two ways.

Less than:	$H_a : p_1 < p_2$	$H_a : p_1 - p_2 < 0$
Greater than:	$H_a : p_1 > p_2$	$H_a : p_1 - p_2 > 0$
Not equal:	$H_a : p_1 \neq p_2$	$H_a : p_1 - p_2 \neq 0$

Test statistic

Like with one sample proportion tests, two sample proportion tests use the standard normal sampling distribution. As always, z-scores are calculated by

$$z = \frac{x - \mu}{\sigma}$$

In this case,

- x is the difference of the sample proportions, or $\hat{p}_1 - \hat{p}_2$
- the expected value is the difference of the population proportions, or $p_1 - p_2$, which under the null is 0
- σ (standard deviation) is... complicated...

$$\sigma = \sqrt{(\bar{p}\bar{q})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Test statistic, cont.

The test statistic for two sample proportion tests is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{(\bar{p}\bar{q})\left(1/n_1 + 1/n_2\right)}}$$

As before, z-scores and p-values can be found using technology.

Hypothesis test for two proportions, example

Example

A school district in an effort to reduce the number of teen drivers who text or email while driving creates an education program that half the high school students in the district attend. Two months after the program, the district surveys a random sample of teen drivers. The survey finds that 62 out of 212 teens surveyed who had attended the program (population 1) had texted or emailed while driving during the previous month, while among those that did not attend the program (population 2), 59 out of 173 did so.

Test whether the program had an effect at 0.05 level of significance.

1. Identify null and alternative hypotheses from research question
   $H_0 : p_1 = p_2$ or $H_0 : p_1 - p_2 = 0$
   $H_a : p_1 \neq p_2$ or $H_a : p_1 - p_2 \neq 0$

Hypothesis test for two proportions, example

Example

2. Check the requirements for using normal distribution
   - Simple random sample
   - Requirements of a binomial satisfied (fixed sample size, two outcomes, independent, constant probability of success)
   - Successes and failures from both samples exceed 5
3. Calculate z test statistic
   $z = -1.0215331$
4. Calculate p-value
   $p = 0.307$

Hypothesis test for a proportion, example

Example

5. Compare p-value to significance level α and report decision
   $p = 0.307 > \alpha = 0.05$. Fail to reject null hypothesis.
6. State conclusion in terms of original research question
   There is not evidence that teen drivers who attended the program text and email while driving at a different rate than those who did not attend.

   or

   There is not evidence that the education program changed the rates of teen drivers texting or emailing while driving.

Section 10.2
Goodness-of-Fit Tests

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".

- 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.

- 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

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

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

- 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

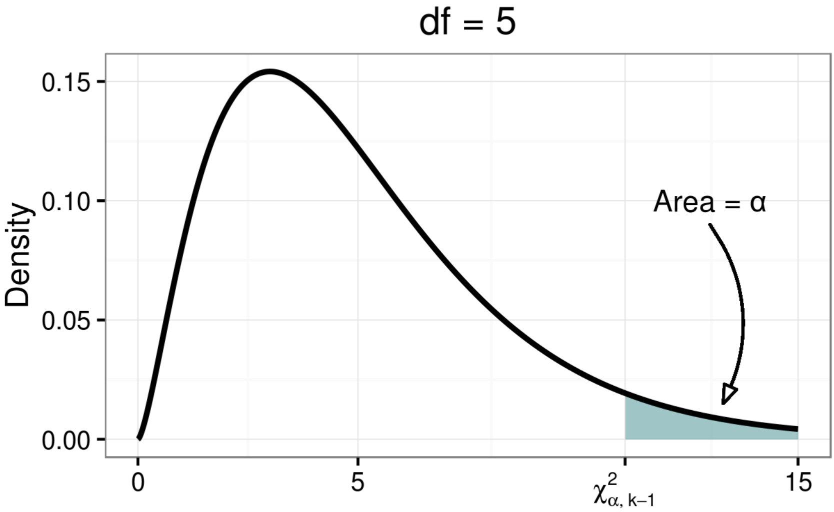
Like most hypothesis tests, there are two ways to make a decision for a goodness-of-fit test:

- 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.

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
- 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

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

Section 10.3
Tests for Independence

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.

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.
- 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.

- 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.

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

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

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

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

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

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:

- 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.

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

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

- 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

- 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

We performed a proportion test with this data earlier. 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 $z = -1.02$, $p = 0.307$.
- 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.