Lesson 16:
Chapter 9 Section 1

Caleb Moxley

BSC Mathematics

2 November 15

§9.1 Inference for Two-Way Tables

So far, we've only conducted statistical inference tests to describe quantitative aspects of populations like the values of means, proportions, and standard deviations.

§9.1 Inference for Two-Way Tables

So far, we've only conducted statistical inference tests to describe quantitative aspects of populations like the values of means, proportions, and standard deviations. In Chapter 9, we'll conduct statistical inference tests which deal with the qualitative aspects of populations.

§9.1 Inference for Two-Way Tables

So far, we've only conducted statistical inference tests to describe quantitative aspects of populations like the values of means, proportions, and standard deviations. In Chapter 9, we'll conduct statistical inference tests which deal with the qualitative aspects of populations. We'll be able to answer questions like whether or not two populations are associated/dependent or if an underlying population has a particular distribution. The first question is discussed in this section, and it involves two-way tables.

§9.1 Inference for Two-Way Tables

So far, we've only conducted statistical inference tests to describe quantitative aspects of populations like the values of means, proportions, and standard deviations. In Chapter 9, we'll conduct statistical inference tests which deal with the qualitative aspects of populations. We'll be able to answer questions like whether or not two populations are associated/dependent or if an underlying population has a particular distribution. The first question is discussed in this section, and it involves two-way tables. Two-way tables are discussed in Chapter 2, Section 6 — which we skipped!

§9.1 Inference for Two-Way Tables

So far, we've only conducted statistical inference tests to describe quantitative aspects of populations like the values of means, proportions, and standard deviations. In Chapter 9, we'll conduct statistical inference tests which deal with the qualitative aspects of populations. We'll be able to answer questions like whether or not two populations are associated/dependent or if an underlying population has a particular distribution. The first question is discussed in this section, and it involves two-way tables. Two-way tables are discussed in Chapter 2, Section 6 — which we skipped! So, first, we'll discuss what a two-way table is.

§9.1 Inference for Two-Way Tables

Definition (two-way table)

A **two-way table** records the number of cases from a sample which fall into a category based on two variables.

§9.1 Inference for Two-Way Tables

Definition (two-way table)

A **two-way table** records the number of cases from a sample which fall into a category based on two variables.

Example (two-way table)

	M	F
< 30	12	30
≥ 30	14	10

§9.1 Inference for Two-Way Tables

Using two-way tables, we can easily compute **joint** and **marginal** distributions.

Example (two-way table: joint distribution)

	M	F	Total
< 30	12	30	42
≥ 30	14	10	24
Total	26	40	66

Using two-way tables, we can easily compute **joint** and **marginal** distributions.

Example (two-way table: joint and marginal distributions)

	M	F	Total
< 30	0.1818	0.4545	0.6363
≥ 30	0.2121	0.1515	0.3636
Total	0.3939	0.6061	1

In order to compute the test statistic for a test for independence using a two-way table, you must be able to compute the expected counts.

§9.1 Inference for Two-Way Tables

In order to compute the test statistic for a test for independence using a two-way table, you must be able to compute the expected counts.

Definition (expected count)

This is what we could expect to see in each of the cells of the two-way table if the two variables were independent.

In order to compute the test statistic for a test for independence using a two-way table, you must be able to compute the expected counts.

Definition (expected count)

This is what we could expect to see in each of the cells of the two-way table if the two variables were independent. The expected count is given below.

§9.1 Inference for Two-Way Tables

In order to compute the test statistic for a test for independence using a two-way table, you must be able to compute the expected counts.

Definition (expected count)

This is what we could expect to see in each of the cells of the two-way table if the two variables were independent. The expected count is given below.

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

In order to compute the test statistic for a test for independence using a two-way table, you must be able to compute the expected counts.

Definition (expected count)

This is what we could expect to see in each of the cells of the two-way table if the two variables were independent. The expected count is given below.

$$\text{expected count} = \frac{\text{row total} \times \text{column total}}{n}$$

Example

Our table for expected counts would look like this:

	M	F	Total
< 30	16.545	25.455	42
≥ 30	9.455	14.545	24
Total	26	40	66

§9.1 Inference for Two-Way Tables

Once you have the expected counts and the observed counts table, you compute your X^2 test statistic for the test for independence using the formula below.

Once you have the expected counts and the observed counts table, you compute your X^2 test statistic for the test for independence using the formula below.

Definition (X^2 test statistic for two-way table test of independence)

The X^2 **statistic** is a measure of how much the observed cell counts (from the sample) differ from the expected (computed) cell counts. It's calculated by

$$X^2 = \sum \frac{(\text{observed count} - \text{expected count})^2}{\text{expected count}},$$

where the sum is computed over all cells in the $r \times c$ two-way table.

Example (X^2 test statistic)

From our previous example:

	M	F	Total
< 30	16.545	25.455	42
≥ 30	9.455	14.545	24
Total	26	40	66

	M	F	Total
< 30	12	30	42
≥ 30	14	10	24
Total	26	40	66

Example (X^2 test statistic)

From our previous example:

	M	F	Total
< 30	16.545	25.455	42
≥ 30	9.455	14.545	24
Total	26	40	66

	M	F	Total
< 30	12	30	42
≥ 30	14	10	24
Total	26	40	66

So, $X^2 = \frac{(12-16.545)^2}{16.545} + \frac{(30-25.455)^2}{25.455} + \frac{(14-9.455)^2}{9.455} + \frac{(10-14.545)^2}{14.545} =$

§9.1 Inference for Two-Way Tables

Example (X^2 test statistic)

From our previous example:

	M	F	Total
< 30	16.545	25.455	42
≥ 30	9.455	14.545	24
Total	26	40	66

	M	F	Total
< 30	12	30	42
≥ 30	14	10	24
Total	26	40	66

So, $X^2 = \frac{(12-16.545)^2}{16.545} + \frac{(30-25.455)^2}{25.455} + \frac{(14-9.455)^2}{9.455} + \frac{(10-14.545)^2}{14.545} = 5.665$.

§9.1 Inference for Two-Way Tables

Once you have computed the test statistic, you can conduct the test.

§9.1 Inference for Two-Way Tables

Once you have computed the test statistic, you can conduct the test. **This is the first test we've encountered with null and alternative hypotheses that never change!** This test is designed only to test for independence, and thus the null and alternative hypotheses are fixed.

§9.1 Inference for Two-Way Tables

Once you have computed the test statistic, you can conduct the test. **This is the first test we've encountered with null and alternative hypotheses that never change!** This test is designed only to test for independence, and thus the null and alternative hypotheses are fixed.

Definition (X^2 test for independence using two-way tables)

To test H_0 : The row and column variables are not associated against H_a : The row and column variables are associated, compute the X^2 test statistic as before. The P-value for this test is

$$P(\chi^2 > X^2),$$

where χ^2 is the chi-squared distribution with degrees of freedom k given by $k = (r-1)(c-1)$.

Example (P-value for a X^2 test statistic)

So to continue our example, we see that the P-value is

$$P(\chi^2 > 5.665) = \text{between } 0.02 \text{ and } 0.01.$$

Example (P-value for a X^2 test statistic)

So to continue our example, we see that the P-value is

$$P(\chi^2 > 5.665) = \text{between } 0.02 \text{ and } 0.01.$$

Thus, if we were testing with $\alpha = 0.05$ to see if being male or female was not associated with being above or below 30 in the population from which our sample was drawn, we would reject H_0 and say that it looked like age and sex were not independent.

Example (P-value for a X^2 test statistic)

So to continue our example, we see that the P-value is

$$P(\chi^2 > 5.665) = \text{between } 0.02 \text{ and } 0.01.$$

Thus, if we were testing with $\alpha = 0.05$ to see if being male or female was not associated with being above or below 30 in the population from which our sample was drawn, we would reject H_0 and say that it looked like age and sex were not independent.

You can also use Minitab or a TI-83/84 calculator to get the exact P-value.

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$.

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$.

	90-110	110-130	130-150
Left	5	10	6
Right	11	60	8

§9.1 Inference for Two-Way Tables

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Answer: So, $X^2 = 6.947$, and the P-value is 0.031.

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Answer: So, $X^2 = 6.947$, and the P-value is 0.031. Thus, we fail to reject H_0 and support the notion that the IQ and handedness are independent among US college students.

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Answer: So, $X^2 = 6.947$, and the P-value is 0.031.

§9.1 Inference for Two-Way Tables

Example (handedness and IQ)

A sample of 100 college students in the US gave the following two-way table for IQ and handedness. Does it appear that handedness and IQ are associated? Use $\alpha = 0.02$. Let's look at the contribution of each cell to the test statistic.

	90-110	110-130	130-150
Left	0.800	1.503	3.185
Right	0.213	0.399	0.847

Answer: So, $X^2 = 6.947$, and the P-value is 0.031. Thus, we fail to reject H_0 and support the notion that the IQ and handedness are independent among US college students.

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

- If the table is more than a 2×2 table, then the average expected counts should be at least 5.

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

- If the table is more than a 2×2 table, then the average expected counts should be at least 5.
- If the table is a 2×2 table, then each cell should have an expected count of at least 5.

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

- If the table is more than a 2×2 table, then the average expected counts should be at least 5.
- If the table is a 2×2 table, then each cell should have an expected count of at least 5.

There are two paradigms for understanding this test: either the sample is a collection of samples from different populations or it is a single sample which "divides" a population into constituent parts.

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

- If the table is more than a 2×2 table, then the average expected counts should be at least 5.
- If the table is a 2×2 table, then each cell should have an expected count of at least 5.

There are two paradigms for understanding this test: either the sample is a collection of samples from different populations or it is a single sample which "divides" a population into constituent parts. The test is the same no matter what paradigm you use.

§9.1 Inference for Two-Way Tables

What are the requirements for conducting a two-way table test?

- If the table is more than a 2×2 table, then the average expected counts should be at least 5.
- If the table is a 2×2 table, then each cell should have an expected count of at least 5.

There are two paradigms for understanding this test: either the sample is a collection of samples from different populations or it is a single sample which "divides" a population into constituent parts. The test is the same no matter what paradigm you use. The text makes a big deal out of this distinction — I won't. I'd argue you shouldn't either.

Example (using Minitab)

Minitab will also conduct a two-way test for independence.

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

	Underweight	Normal	Overweight	Obese
Low Income	5	15	25	10
Medium Income	10	29	51	2
High Income	4	16	26	7

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

	Underweight	Normal	Overweight	Obese
Low Income	5	15	25	10
Medium Income	10	29	51	2
High Income	4	16	26	7

The *P*-value is 0.0698.

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

	Underweight	Normal	Overweight	Obese
Low Income	5	15	25	10
Medium Income	10	29	51	2
High Income	4	16	26	7

The P-value is 0.0698. The X^2 test statistic is 11.666.

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

	Underweight	Normal	Overweight	Obese
Low Income	5	15	25	10
Medium Income	10	29	51	2
High Income	4	16	26	7

The *P*-value is 0.0698. The X^2 test statistic is 11.666. Thus, we fail to reject the null hypothesis,

Example (using Minitab)

Minitab will also conduct a two-way test for independence. Use Minitab to test whether the following two-way table is such that its row and column variables are independent. Use 5% significance.

	Underweight	Normal	Overweight	Obese
Low Income	5	15	25	10
Medium Income	10	29	51	2
High Income	4	16	26	7

The P-value is 0.0698. The X^2 test statistic is 11.666. Thus, we fail to reject the null hypothesis, and it seems that income and weight are independent.

§9.1 Inference for Two-Way Tables

Note: If you use the χ^2 table (i.e. Table F) from the text, then you'll only be able to get an estimate for the P-value.

§9.1 Inference for Two-Way Tables

Note: If you use the χ^2 table (i.e. Table F) from the text, then you'll only be able to get an estimate for the P-value. Otherwise, you'll have to use technology to compute a P-value.

Note: If you use the χ^2 table (i.e. Table F) from the text, then you'll only be able to get an estimate for the P-value. Otherwise, you'll have to use technology to compute a P-value. The homework will expect you to compute the P-value in several instances.

§9.1 Inference for Two-Way Tables

Note: If you use the χ^2 table (i.e. Table F) from the text, then you'll only be able to get an estimate for the P-value. Otherwise, you'll have to use technology to compute a P-value. The homework will expect you to compute the P-value in several instances. Please make sure you have the capability to compute it exactly.