

MAT117: Introduction to Chi-square (Sections 11.1/11.2)

Question: Is there a connection between gender and opinion on the legalization of marijuana?

- If the answer is yes, we would expect that the proportion who favor legalization would be different for males and females.
- If the answer is no, then we would expect both male and female proportions to be the same.

Some terms:

- Independent, dependent
- Association, relationship (connection)

Hypotheses:

H_0 : gender and opinion on marijuana legalization are *independent*

H_a : there is *some association between* gender and opinion on marijuana legalization

The approach we develop here can be generalized to multiple groups, or to categorical question with more than two answers, or both

Example 1. Researchers sample 1000 people, and 230 of them favor the legalization of marijuana.

- This is _____ % of the people sampled. If there is no difference between male and female opinion, then _____ % of the males and _____ % of the females would have said yes.
- The sample had 400 males. If the null hypothesis is true, how many of them would you expect to have said yes?
- Fill in this *hypothetical* contingency table of “Expected Counts,” assuming H_0 were true.

		Do you favor legalization?		
		Yes	No	Totals
Male				400
Female				600
Totals		230	770	1000

Now let’s see what did happen when they did the survey. Of the 400 males, 116 said yes; of the 600 females, 114 said yes. Fill in this table of “Observed Counts” – what actually happened in the survey.

		OBSERVED		
		Do you favor legalization?		
		Yes	No	Totals
Male				400
Female				600
Totals		230	770	1000

In a TI calculator:

- Enter the OBSERVED counts into Matrix [A] after setting it to be 2 by 2.
- Choose “Chi-squared test” from the STAT > TESTS menu. (χ^2 -Test) and hit calculate.

What is the test statistic $\chi^2 =$ _____? What is the P -value = _____?

Do you conclude that the null hypothesis is probably false?

How do you write the final English conclusion?

Example 2: In an experiment on aspirin use and heart attack, 293 of the 22,071 subjects had a heart attack. In the study, 11,034 of the subjects were given a placebo and 11,037 were given an aspirin.

a. Write the hypotheses:

H_0 : _____ and _____ are *independent*

H_a : there is *some association between* _____ and _____

b. Fill in this *hypothetical* contingency table of “Expected Counts,” assuming H_0 were true. Round to two places after the decimal.

EXPECTED			
Heart attack			
	Yes	No	Totals
Placebo			11034
Aspirin			11037
Totals	293	21778	22071

c. Here is the table of observed counts from the study (from the textbook).

OBSERVED			
Heart attack			
	Yes	No	Totals
Placebo	189	10845	11034
Aspirin	104	10933	11037
Totals	293	21778	22071

d. Use a TI calculator to find the χ^2 test statistics and the P -value.

e. Write your conclusions, first about rejecting the null hypothesis and then the conclusion in English.

Exercises. For each of the following situations, state the hypotheses, then run the chi-squared test from the data, report your test statistic and P-value, and then make your conclusions.

1. Is there a statistically significant association between income level and marital happiness?

Income	Marital happiness		
	Not	Pretty	Very
Above	6	43	75
Average	6	113	178
Below	6	57	117

2. Is there statistically significant evidence of an association between education and smoking status for French men?

Education	Smoking Status	
	Nonsmoker	Smoker
Primary school	56	131
Secondary school	37	102
University	53	80

3. Subjects from the 2002 GSS indicated whether they identified more strongly with the Democratic Party, the Republican Party or with Independents. Is there a gender gap in political beliefs in the United States?

	DEM	IND	REP	TOTAL
Female	567	534	395	1496
Males	356	460	369	1185
TOTAL	923	994	764	2681

4. In an experiment on aspirin use and heart attack, 11201 patients were given aspirin, 10590 patients were given Tylenol and 11994 patients were given neither. Over the five years in the study, 112 of the aspirin patients had a heart attack, 125 of the Tylenol patients had a heart attack and 201 of the other patients had a heart attack. Is there enough evidence in support of an association between pain reliever use and whether or not one has a heart attack?

ANSWERS

Example 1. Assuming the null hypothesis is true, this is the expected contingency table:

EXPECTED			
Do you favor legalization?			
	Yes	No	Totals
Male	92	308	400
Female	138	462	600
Totals	230	770	1000

From a calculator, we find that the test statistic $\chi^2 = 13.55$, and the P -value = 0.00023.

Since the P -value is small, we reject H_0 , and we conclude, "We have statistically significant evidence that there is an association between gender and opinion on marijuana legalization."

Example 2.

H_0 : Taking aspirin and having a heart attack are **independent**

H_a : there is **some association between** between taking aspirin and having a heart attack.

Assuming the null hypothesis is true, this is the expected contingency table:

EXPECTED			
Heart attack			
	Yes	No	Totals
Placebo	146.5	10887.5	11034
Aspirin	146.5	10890.5	11037
Totals	293	21778	22071

From a calculator, we find that the test statistic $\chi^2 = 25.01$, and the P -value = 0.00000057.

Since the P -value is small, we reject H_0 , and we conclude, "We have statistically significant evidence that there is an association between taking aspirin and having a heart attack."

Exercises

1. H_0 : Income level and marital happiness are **independent**

H_a : there is **some association between** income level and marital happiness

From a calculator, we find that the test statistic $\chi^2 = 4.23$, and the P -value = 0.376.

Since the P -value is **not** small, we **do not** reject H_0 , and we conclude, "We **do not** have statistically significant evidence that there is an association between income level and marital happiness."

2. H_0 : Smoking level and educational level are **independent**

H_a : there is **some association between** smoking level and educational level

From a calculator, we find that the test statistic $\chi^2 = 5.99$, and the P -value = 0.05.

Since the P -value is **not** small, we **do not** reject H_0 , and we conclude, "We **do not** have statistically significant evidence that there is an association between smoking level and educational level."

3. H_0 : Gender and political affiliation are **independent**

H_a : there is **some association between** gender and political affiliation

From a calculator, we find that the test statistic $\chi^2 = 18.8$, and the P -value = 0.000082.

Since the P -value is small, we reject H_0 , and we conclude, "We have statistically significant evidence that there is an association between gender and political affiliation."

4. H_0 : Pain reliever taken and having a heart attack are **independent**

H_a : there is **some association between** pain reliever taken and having a heart attack

From a calculator, we find that the test statistic $\chi^2 = 22.3$, and the P -value = 0.000014

Since the P -value is small, we reject H_0 , and we conclude, "We have statistically significant evidence that there is an association between pain reliever taken and having a heart attack."