

10.3: Testing for Independence

In this section, we continue with the χ^2 (chi-square) tests in a special test for independence.

Two-Way Tables

The data used in this problem are based on the article, Attitudes about Marijuana and Political Views²². The two-way table below summarizes data on marijuana use and political views from a random sample of 270 adults. Each frequency in an interior cell is a count of the number of adults in the sample that have two specific characteristics. The two-way table is missing several values. We will investigate whether the variables political views and smoking frequency are independent of one-another.

Political Views	Never Smoke	Rarely Smoke	Frequently Smoke	Totals
Liberal	96	35		155
Conservative	43	9		55
Other				60
Totals	173	53	44	270

1. What is the explanatory variable in this study?
2. What is the response variable?
3. Do you think these variables are independent, or dependent? Explain your answer.
4. Enter the missing values into the table above.
5. Compute the following conditional probabilities. Write the probability as a decimal rounded to 2 decimal places.
 - a. $P(\text{Conservative} \mid \text{Never Smoke}) =$
 - b. $P(\text{Conservative} \mid \text{Frequently Smoke}) =$
 - c. Given your previous answer, would you consider marijuana smoking frequency and political views as independent or dependent variables? Explain your answer.

We will now look at the number of degrees of freedom involved in this problem. Below is the two-way table from the start of this section, before values in the “Frequently Smoke” column and “Other” row were entered.

Political Views	Never Smoke	Rarely Smoke	Frequently Smoke	Totals
Liberal	96	35	24	155
Conservative	43	9	3	55
Other	34	9	17	60
Totals	173	53	44	270

6. With the values originally given in the table, are the values you entered free, random values, or dependent on other values? Explain.

7. If the degrees of freedom in the observed frequencies (O) are the number of free, independent observations, how many degrees of freedom are there among the observed frequencies in the table above?

8. Suppose a two-way table has r rows and c columns (don't count the totals). Make a rule for the degrees of freedom among the observed frequencies in the table. Express this rule as a formula for the degrees of freedom.

$df =$ _____

Hypothesis Testing Process

Now that we understand the degrees of freedom in a two-way table, it is time to think about a hypothesis test. This test is similar to the goodness-of-fit test already discussed, but the degrees of freedom are different (as discussed) and the expected frequencies have a special formula.

We are conducting a test for independence. In the prior statistical study, the question is, “Are political views independent of marijuana smoking frequency?” The data in the two-way table are from a random sample. By examining conditional probabilities in the sample data, we saw that the variables appear to not be independent. We will now perform a chi-square goodness-of-fit test to examine whether the variables are independent across the entire population.

To conduct a test for independence, we construct expected frequencies based on the assumption that these variables are independent (this will be our null hypothesis).

If events A and B are independent then $P(A \cap B) = P(A) \cdot P(B)$. Suppose A = liberal, and B = never smoke. Refer back to your completed two-way table.

9. Suppose we randomly pick an adult from the sample. If “being liberal” and “never smoking” are independent events, what is the probability that an adult is liberal and never smokes? Round the answer to four decimal places.

$$P(\text{liberal} \cap \text{never smoke}) = P(\text{liberal}) \cdot P(\text{never smoke}) = \underline{\hspace{2cm}}$$

This probability is the proportion of people who should be in the liberal and never smoke category if the events are independent. If this is p , the population proportion of people in this category, then the expected frequency (the number of people in this sample expected to be in this category) is $E = np$. In this formula, n is the sample size (grand total of the two-way table).

10. Find E . Round to two decimal places. $E = np = \underline{\hspace{2cm}}$

The expected frequency is actually quite close to what was observed, $O = 96$. For this event, the null hypothesis of independence seems to lead to reliable predictions of what actually occurred. Keep in mind, this expected frequency is computed based on our null hypothesis that assumes that our variables are independent. Look once more as we backtrack through the computation which we performed to compute the expected frequency.

$$E = np = 270 \cdot 0.3678 = 270 \cdot \frac{155}{270} \cdot \frac{173}{270} = \frac{155 \cdot 173}{270} = \frac{\text{row total} \cdot \text{column total}}{\text{grand total}}$$

The key here is that the expected frequency of an event can be found directly from the row, column and grand totals. In general, To compute an expected frequency for an observation in a given row and column of a two-way table, use the formula,

$$E = \frac{\text{row total} \cdot \text{column total}}{\text{grand total}}$$

Step 1: Determine the Hypotheses

We are now ready to conduct a test for independence using a two-way table. We want to test if political views are independent of marijuana smoking frequency at the 1% significance level.

The hypotheses for this test are:

H_0 : The explanatory and response variables are independent.

H_a : The explanatory and response variables are dependent.

11. Write the null and alternative hypotheses in the context of the current problem (naming the explanatory and response variables).

Step 2: Collect the Data

The test for independence between bivariate categorical variables requires that data be summarized in a two way table. Row and column totals should be computed so that expected frequencies can be computed. The expected frequencies are computed with the formula

$$E = \frac{\text{row total} \cdot \text{column total}}{\text{grand total}}$$

As with the previous goodness-of-fit test, we require that each expected frequency is at least 5.

12. Compute the expected frequencies for the cells in the first row. Round the values to two decimal places. The expected frequencies of cells in the 2nd and 3rd rows are provided.

Political Views	Never Smoke	Rarely Smoke	Frequently Smoke	Totals
Liberal	96 $E = \frac{155 \cdot 173}{270} = 99.31$	35	24	155
Conservative	43 $E = \frac{55 \cdot 173}{270} = 35.24$	9 $E = \frac{55 \cdot 53}{270} = 10.8$	3 $E = \frac{55 \cdot 44}{270} = 8.96$	55
Other	34 $E = \frac{60 \cdot 173}{270} = 38.44$	9 $E = \frac{60 \cdot 53}{270} = 11.78$	17 $E = \frac{60 \cdot 44}{270} = 9.78$	60
Totals	173	53	44	270

13. Are the criteria satisfied for the approximate χ^2 distribution for us to perform a test for independence? Explain.

Step 3: Assess the Evidence

The test statistic for a χ^2 test for independence is: $\chi^2 = \sum \frac{(O-E)^2}{E}$ This is approximately distributed according to the χ^2 distribution with degrees of freedom equal to: $df = (r - 1) \cdot (c - 1)$

Here, r is the number of rows in the table and c is the number of columns. Note – these do not include the total row or column. As before, this test for independence is always a right-tailed test.

14. Enter the expected frequencies next to the corresponding observed frequencies in the table below. For each pair, compute the contribution to the χ^2 statistic, $\frac{(O-E)^2}{E}$ and total these values.

O = Observed Frequency	E = Expected Frequency	$\frac{(O-E)^2}{E}$
96	99.31	$\frac{(96-99.31)^2}{99.31} = 0.1103$
43		
34		
35		
9		
9		
24		
3		
17		
Total (χ^2 test statistic rounded to two decimal places)		

15. What are the degrees of freedom for this test statistic? (Don't count the total row or column)
16. Use this desmos graph <https://www.desmos.com/calculator/bjohldwaym> to find the P-value for this right-tailed test. Round the value to 4 decimal places.

Step 4: Make a Decision and State a Conclusion

17. What conclusion do you make regarding the null and alternative hypotheses? Why? Recall the level of significance is 1%.
18. Write a brief conclusion in the context of this problem.

Reference

²² *Psychological Reports*, 1973, pp. 1051 to 1054

This page titled [10.3: Testing for Independence](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Hannah Seidler-Wright](#).