

Fact Sheet – Test for Independence (12.2)

The **chi-square test for independence** is used to determine whether there is an association between a row variable and column variable in a contingency table constructed from sample data.

The null hypothesis is that the variables are not associated; in other words, they are independent.

The alternative hypothesis is that the variables are associated, or dependent.

Example: Test the claim that a person's burger preference (McDonald's, Burger King, Wendy's, In-N-Out) is independent of the person's gender (Male, Female).

Observed Frequencies

The observed frequencies, O_i , are the sample values in each cell of the contingency table.

Expected Frequencies

To compute the expected frequency for each cell in the table, use the formula

$$E_i = \frac{(\text{row total})(\text{column total})}{n}$$

Conditions

To perform a test for independence, the following two conditions must be met.

- Each expected frequency must be at least 1: $E_i \geq 1$ for each cell.
- No more than 20% of all expected frequencies are less than 5.

You can check the conditions by computing the test statistic/ P -value using StatCrunch.

The expected frequencies will be contained in the output.

If all of the expected frequencies are at least 5, then you can write up the test.

If any of the expected frequencies is lower than 5, you can still write up the test as long as they are all greater than or equal to 1 **and** no more than 20% of the expected frequencies are less than 5.

Hypothesis Test

Step 1

The null hypothesis will be of the form "*The row variable and column variable are independent.*"

This means that the two variables are not associated with each other.

H_1 will always be "*The row variable and column variable are dependent.*"

This means that the two variables are associated with each other.

Step 3

The test statistic is $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, where O_i is the observed sample frequency for category i and E_i is the computed expected frequency for category i .

Just write "Test for Independence", rather than writing the test statistic.

Step 4

To compute the test statistic and P-value using StatCrunch ...

1. Enter the row labels in the first column, and make the header for the first column the row variable.

Enter the column labels as headers of the next columns.

Enter the data from the contingency table in the correct rows under the column labels.

2. Select **Stat**, highlight **Tables**, select **Contingency**, then highlight **with summary**.

3. Select the column variable(s).

Then select the row variable.

Click Next>.

4. Select **Expected Count** and **Chi-Square Test**. Click Calculate.

Classroom Examples (12.2)

1) A random sample of 483 workers asked how they saved things – in piles, in files, or some combination of both. Here are the results, broken down by gender.

	Piles	Files	Both
Male	78	138	69
Female	30	110	58

At the 0.05 level, test the claim that the technique used is independent of gender.

(Step 1)

H_0 :

H_1 :

(Step 2) $\alpha =$ _____

(Step 3) Test for Independence

(Step 4) $\chi^2 =$ _____

P -value = _____

(Step 5) _____ H_0 .

There _____ sufficient evidence to support that technique and gender are dependent.

We _____ conclude that technique and gender are related.

2) A sample of 300 men and women were asked “How long should couples date before getting married?” Here are the results.

	Less Than 1 Year	1 Year	1 – 2 Years	2 – 3 Years	Longer Than 3 Years
Men	41	65	48	8	5
Women	19	23	51	24	16

At the 0.05 level of significance, test the claim that a person’s response to this question is independent of the person’s gender.

3) A random sample of 440 students were asked whether they had a computer at home. Here are the results, by family income.

	Computer	No Computer
Less than \$10,000	8	29
\$10,000-\$24,999	67	41
\$25,000-\$34,999	34	40
\$35,000-\$49,999	58	42
\$50,000 and over	97	24

At the 0.01 level of significance, test the claim that the proportion of students with a computer at home is the same for each income level.

4) A random sample of 400 Americans were asked whether they had health insurance. Here are the results, broken down by race.

	Asian	Black	Hispanic	White, Non-Hispanic
Insurance	16	52	40	207
No Insurance	5	16	22	42

At the 0.01 level of significance, test the claim that race is independent of health insurance status.