

Constructing a Chi Squared test for one categorical variable (goodness of fit)

Step 1: Set up your hypothesis based on the expectation that the distribution for the data collected will match the expected distribution.

H_0 = The observed proportions will match the expected proportions.

H_a = At least one of the expected proportions is incorrect.

Step 2: Make a table with the observed and expected frequency for each response type.

In the calculator, L1 = observed values and L2 = expected values

Step 3: Have you met the conditions for using a significance test?

Randomly selected samples – Look for the term SRS or randomized experiment. Without random selection, we lose the ability to make inferences about the population.

Large Sample Size – All values in the expected column must be at least 5.

Independent Observations – In order to use our standard deviation formula, we need the sample size to be less than 10% of the population size for BOTH data sets, so $10n < N$. This is especially true when sampling without replacement.

Step 4: Calculate the Chi Square for each response:
$$\chi^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

This value represents a standardized measure for the difference between the expected and observed value for each response. If the observed and expected values are close, then this value is small. If the observed and expected values are far apart, then this value is large.

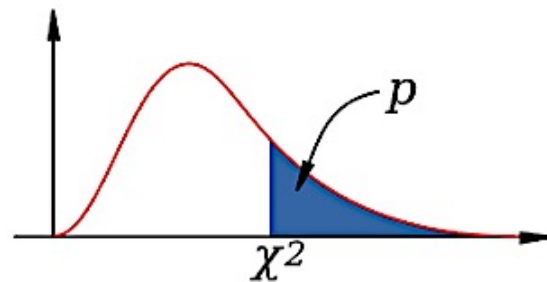
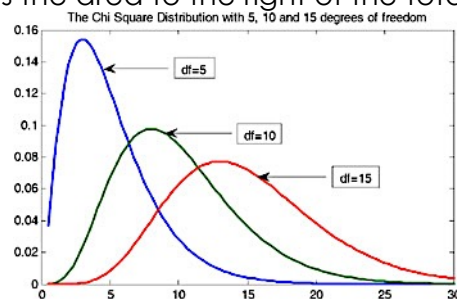
Step 5: Find the Sum of all Chi Square values. This is the total Chi Square amount for the test.

Step 6: Calculate the **p-value**: The probability of the observed and expected values being at least as far apart as found in our results. For this test, the Degrees of Freedom is one less than the number of response types. $DF = \text{types} - 1$

The Chi Square graph changes based on the Degrees of Freedom. See the graph below.

The x-axis for the Chi Square graph shows the total chi-square values.

The p-value is the area to the right of the total chi-square in our test. See the graph below.



Step 7: Compare the p-value to the significance level (called an α – level)

If the p-value is **below** the significance level:

- Then we reject the null hypothesis H_0 in favor of the alternate hypothesis H_a
- We say the results are statistically significant

If the p-value is **above** the significance level:

- Then we fail to reject the null hypothesis H_0 and reject the alternate hypothesis H_a
- We say the results are NOT statistically significant

Step 8: State your conclusion using AP Language.

If the observed and expected values of _____ truly match, then there is a _____
Describe the variable *p-value*

chance of obtaining the results found in our test. We conclude that at the _____ significance level the
 α -level
 claim [is rejected] or [fails to be rejected].

Now write a sentence giving the meaning of the conclusion tied to the purpose of the study.

Constructing a Chi Squared test for two categorical variables (association)

Step 1: Set up your hypothesis based on the expectation that there is not an association between the two variables (independence).

H_0 = An association between the variables does not exist

H_a = An association between the variables does exist

Step 2: Enter the table in as a matrix.

Press 2nd Matrix,
Cursor to EDIT
Press Enter

```
NAMES MATH [00]
1: [A]
2: [B]
3: [C]
4: [D]
5: [E]
6: [F]
7↓ [G]
```

Enter the Rows (2) and Columns (3)
Press Enter to move to next option
Enter the table information, Then Quit.

```
MATRIX[A] 2 × 3
[ 16    10    6 ]
[ 20    22    6 ]

2, 3=16
```

Step 3: Have you met the conditions for using a significance test?

Randomly selected samples – Look for the term SRS or randomized experiment. Without random selection, we lose the ability to make inferences about the population.

Large Sample Size – Large Sample Size – All values in the expected column should be at least 5

Independent Observations – In order to use our standard deviation formula, we need the sample size to be less than 10% of the population size for BOTH data sets, so $10n < N$. This is especially true when sampling without replacement.

Step 4: Run AP STAT Program → Chi-Square → Data in MTRX-A

Note the program asks you to check the table for the second check to perform a Chi-Square Test

Step 5: The program shows the expected values and chi-square for each location.

Formulas The expected count for each location = $\frac{\text{row total} \cdot \text{column total}}{\text{table total}}$

$$\chi^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$\text{Degrees of Freedom} = (\text{Number of rows} - 1) \cdot (\text{number of columns} - 1)$$

Step 6: Find the Sum of all Chi Square values. This is the total Chi Square amount for the test.

Step 7: Calculate the **p-value**: The probability of the observed and expected values being at least as far apart as found in our results. Compare the p-value to the significance level.

Step 8: State your conclusion using AP Language.

If no association exists between _____ and _____, then there is a _____

Variable 1 Variable 2 p-value

chance of obtaining the results found in our test. We conclude that at the _____ significance level the

α-level

claim [is rejected] or [fails to be rejected].

Now write a sentence giving the meaning of the conclusion tied to the purpose of the study.