Chi-Squared Hypothesis Test for Homogeneity

H_0 : The true category proportions are the same for all populations.

H_a : The true category proportions are NOT the same for all populations.

Note #1: Use on CATEGORICAL data

Note #2: df = (rows - 1)(columns -1)

Chi-Squared Hypothesis Test for Independence

H_0 : The two variables are independent.

H_a : The two variables are NOT independent.

Note #1: Use on CATEGORICAL data

Note #2: df = (rows - 1)(columns -1)

Note #3: Same mechanics as homogeneity test

Note #4: When you see a chart with more than one column or row of data...use your matrices

Chi-Squared Hypothesis Test for Homogeneity

Steps in Chi-Squared Homogeneity Hypothesis Testing

2. H_0 : The true category proportions are the same for all populations.

3. H_a : The true category proportions are NOT the same for all populations.

4. State α.

5. <u>Assumptions:</u>
   1. Random **Independent** Samples
   2. Expected Counts ≥ 5

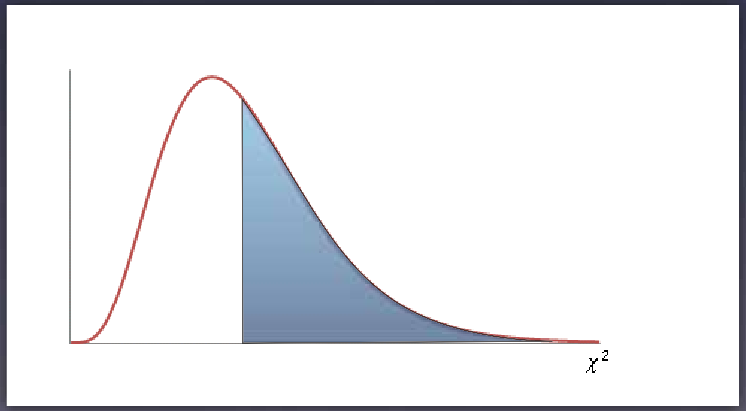
$$\text{Expected Count} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

6. χ^2 Test for Homogeneity

7. df = (rows - 1)(columns -1)

8/9. $\chi^2 = \sum \frac{(O-E)^2}{E} = \#$

10. $P-value = P\left(\chi^2 > \#\right) = \chi^2 cdf\left(\#, 1E99, df\right)$

11. 

12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0 .
    2. Summarize in context discussing H_a .

Several AP Stats students noticed a difference in air circulation and temperature between Cowell and McCullough Gymnasiums. They wondered if the difference in environments affect AP test results so they randomly sample 100 AP results from SI students who took their exam in Cowell and 100 results from students who took their exam in McCullough. The results are tabulated below:

Test Grade	5	4	3	2	1
Cowell	10	14	40	24	12
McCullough	15	21	39	15	10

To answer this, we will need to conduct a Chi Square Test for Homogeneity because we are drawing students from two different populations. First let's state our hypotheses:

H_0: Students will test at the same success rate for both gyms

H_a: Students will test at different success rates for each gym

Significance level: $\alpha = 0.05$ …but first, let's check some assumptions:

1. The data consist of *independently chosen random samples.*
2. The *sample size is large*. The sample size is large enough for the chi-square test to be appropriate as long as every expected cell count is at least 5.

Several AP Stats students noticed a difference in air circulation and temperature between Cowell and McCullough Gymnasiums. They wondered if the difference in environments affect AP test results so they randomly sample 100 AP results from SI students who took their exam in Cowell and 100 results from students who took their exam in McCullough. The results are tabulated below:

Test Grade	5	4	3	2	1
Cowell	10	14	40	24	12
McCullough	15	21	39	15	10

In order to confirm that each expected count is at least 5, we have to divide the product of the row total and the column total by the total number of students. For example, in the AP score of 5 column, both expected counts will be (70*25)/140 = 12.5

Expected Counts	**5**	**4**	**3**	**2**	**1**
Cowell	12.5	17.5	39.5	19.5	11
McCullough	12.5	17.5	39.5	19.5	11

Several AP Stats students noticed a difference in air circulation and temperature between Cowell and McCullough Gymnasiums. They wondered if the difference in environments affect AP test results so they randomly sample 100 AP results from SI students who took their exam in Cowell and 100 results from students who took their exam in McCullough. The results are tabulated below:

Test Grade	5	4	3	2	1
Cowell	10	14	40	24	12
McCullough	15	21	39	15	10

Now we can find our test statistic:

$$\chi^2 = \frac{(10-12.5)^2}{12.5} + \frac{(14-17.5)^2}{17.5} + \frac{(40-39.5)^2}{39.5} + \frac{(24-19.5)^2}{19.5} \ldots = 4.6714$$

The number of rows is 2 and columns is 5 so our $df = (5 - 1)(2 - 1) = 4$

$P(\chi^2 > 121.25) = \chi^2 cdf(4.6714, 1E99, 4) = 0.3227$ so we don't even come close to rejecting the null hypothesis. This data is not sufficient evidence of any effect on student test achievement by the gym in which they take the AP exam.

Before we move on, let's see how this can all be done on the calculator using matrices:

Chi-Squared Hypothesis Test for Homogeneity

Steps in Chi-Squared Homogeneity Hypothesis Testing

2. H_0 : The true category proportions are the same for all populations.

3. H_a : The true category proportions are NOT the same for all populations.

4. State α.

5. <u>Assumptions:</u>
   1. Random **Independent** Samples
   2. Expected Counts ≥ 5

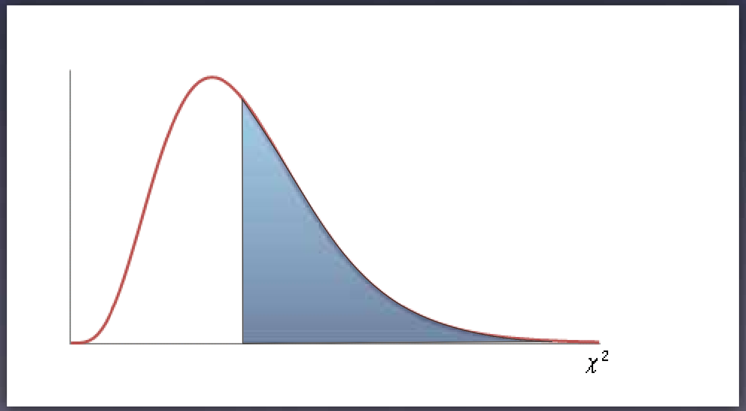
$$\text{Expected Count} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

6. χ^2 Test for Homogeneity

7. df = (rows - 1)(columns -1)

8/9. $\chi^2 = \sum \frac{(O-E)^2}{E} = \#$

10. $P-value = P\left(\chi^2 > \#\right) = \chi^2 cdf\left(\#, 1E99, df\right)$

11. 

12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0 .
    2. Summarize in context discussing H_a .

Chi-Squared Hypothesis Test for Independence

Steps in Chi-Squared Homogeneity Hypothesis Testing

2. H_0 : The two variables are independent.

3. H_a : The two variables are NOT independent.

4. State α.

$$\text{Expected Count} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

5. Assumptions:
   1. Random Samples
   2. Expected Counts ≥ 5

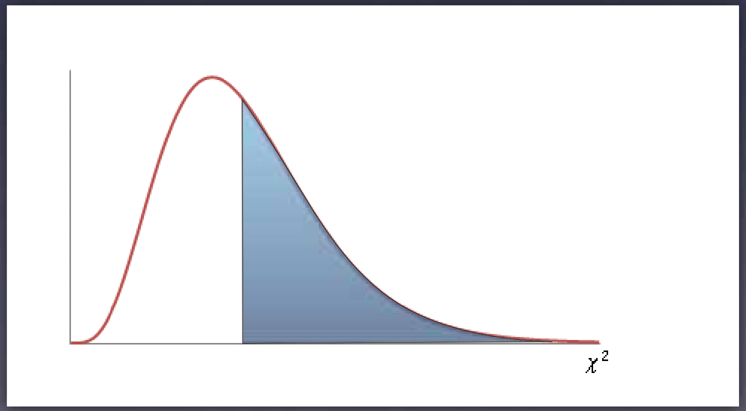
6. χ^2 Test for Independence

7. $df = (\text{rows} - 1)(\text{columns} - 1)$

8/9. $\chi^2 = \sum \frac{(O-E)^2}{E} = \#$

10. $P - value = P\left(\chi^2 > \#\right) = \chi^2 cdf\left(\#, 1E99, df\right)$

11.



12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0 .
    2. Summarize in context discussing H_a .

Below is a table showing who survived the sinking of the Titanic based on whether they were crew members, or passengers booked in first-, second-, or third-class staterooms:

Test Grade	Crew	First	Second	Third	Total
Survived	212	202	118	178	710
Died	673	123	167	528	1491
Total	885	325	285	706	2201

If an individual is drawn at random, what is the probability that we will draw a member of the crew?

What's the probability of randomly selecting a third-class passenger who survived?

If someone's chances of surviving were the same regardless of their status on the ship, how many members of the crew would you have expected to survive?

Below is a table showing who survived the sinking of the Titanic based on whether they were crew members, or passengers booked in first-, second-, or third-class staterooms:

Test Grade	Crew	First	Second	Third	Total
Survived	212	202	118	178	710
Died	673	123	167	528	1491
Total	885	325	285	706	2201

Is there a relationship between a person’s status on the Titanic and whether or not they survived the sinking? Perform the appropriate test.

Chi-Squared Hypothesis Test for Independence

Steps in Chi-Squared Homogeneity Hypothesis Testing

2. H_0 : The two variables are independent.

3. H_a : The two variables are NOT independent.

4. State α.

$$\text{Expected Count} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

5. Assumptions:

   1. Random Samples

   2. Expected Counts ≥ 5

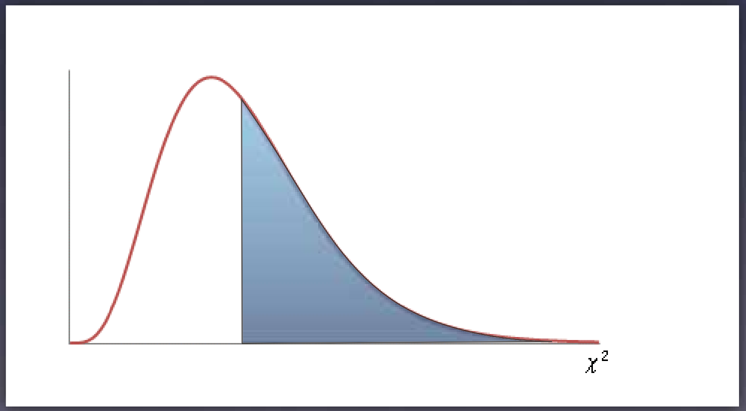
6. χ^2 Test for Independence

7. $df = (\text{rows} - 1)(\text{columns} - 1)$

8/9. $\chi^2 = \sum \frac{(O-E)^2}{E} = \#$

10. $P - value = P\left(\chi^2 > \#\right) = \chi^2 cdf\left(\#, 1E99, df\right)$

11.



12. State the conclusion in two sentences -
    1. Summarize in theory discussing H_0 .
    2. Summarize in context discussing H_a .