Chi-Square Test

Reporter: **VANESSA O. NACAR**

χ^2

Objectives:

- Define and describe chi-square test.
- Describe uses of chi-square test.
- Perform chi-square goodness of fit test.
- Perform chi-square test of independence and test of homogeneity.

Chi-Square Tests

01

Goodness-of-Fit Test

typically used to determine if data fits a particular distribution.

02

Test of Independence

use of a contingency table to determine the independence of two factors

03

Test of Homogeneity

determines whether two populations come from the same distribution, even if this distribution is unknown

What is Chi-Square Test?

- Common statistical test used for nominal data
- Non-parametric or distribution-free test
- Developed by Karl Pearson in 1900

A. Nominal Variables

- variables that have two or more categories, but which do not have an intrinsic order.

Example:

- ❑ Type of property into distinct categories such as houses, condos, co-ops or bungalows.
- ❑ Classifying where people live in the USA by states

B. Dichotomous Variables

- nominal variables which have only two categories or levels.

Example:

- Gender: somebody as either "male" or "female"
- Do you own a mobile phone? Ownership as either "Yes" or "No"
- Type of property had been classified as either residential or commercial.

When should you use a chi-square test?

The chi-square test is appropriate when the following conditions are met:

- the sampling method is **simple random sampling**
- the variable under study is **categorical**
- the expected value of the number of sample observations in each level of the variable is **at least 5.**

01

GOODNESS-OF-FIT TEST

GOODNESS-OF-FIT TEST

- used to test whether a frequency distribution fits an expected distribution

- to calculate the test statistic for the chi-square goodness-of-fit test, the **observed frequencies** and the **expected frequencies** are used.

- The **observed frequency (*O*)** of a category is the frequency for the category observed in the sample data.

- The **expected frequency (*E*)** of a category is the *calculated* frequency for the category. Expected frequencies are obtained assuming the specified (or hypothesized) distribution. The expected frequency for the *i*th category is

$$E_i = np_i$$

where n is the number of trials (the sample size) and p_i is the assumed probability of the *i*th category.

Example

200 teenagers are randomly selected and asked what their favorite pizza topping is. The results are shown below.

Find the observed frequencies and the expected frequencies.

$$E_i = np_i$$

Topping	Results ($n = 200$)	% of teenagers
Cheese	78	41%
Pepperoni	52	25%
Sausage	30	15%
Mushrooms	25	10%
Onions	15	9%

Observed Frequency	Expected Frequency
78	$200(0.41) = 82$
52	$200(0.25) = 50$
30	$200(0.15) = 30$
25	$200(0.10) = 20$
15	$200(0.09) = 18$

FORMULA

$$x^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

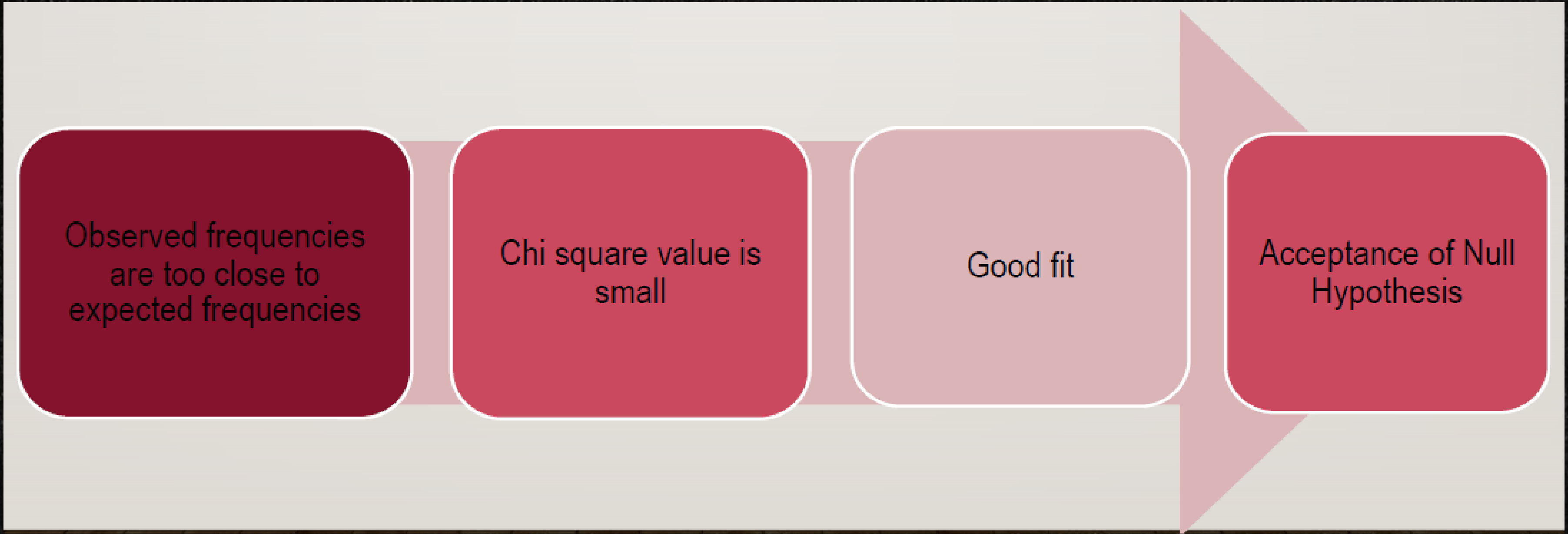
where,

O_i is the observed frequency in i^{th} category;

E_i is the expected frequency in the i^{th} category

Goodness-of-Fit is typically used to see if the population is uniform (all outcomes occur with equal frequency), the population is normal, or the population is the same as another population with a known distribution. The null and alternative hypotheses are:

H_o: The population **fits** the given distribution.
H_a:The population **does not fit** the given distribution.


Observed frequencies are too close to expected frequencies
Chi square value is small
Good fit
Acceptance of Null Hypothesis

Example

In the 2010 Philippine Census, the ages of individuals in a small town were found to be the following:

Less than 18	18-40	Greater than 40
20%	30%	50%

In 2020, ages of n = 1000 individuals were sampled. Below are the results:

Less than 18	18-40	Greater than 40
288	571	141

Using $\alpha = 0.05$, conclude that the population distribution of ages has changed in the last 10 years.

$n = 1000$

$E_i = np_i$

	Less than 18	18-40	Greater than 40
Expected	20%	30%	50%

	Less than 18	18-40	Greater than 40
Observed	288	571	141
Expected	1000(0.20)=**200**	1000(0.30)=**300**	1000(0.50)=**500**

Step 1. Formulate the hypotheses

H_o: The data meet the expected distribution.

H_a: The data do not meet the expected distribution.

Step 2: Set the level of significance or alpha level

$$\alpha = 0.05$$

Step 3: Find the degrees of freedom

df = k-1
df =3-1
df = 2

$\alpha = 0.05$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 5.99

Step 4: Set up the decision rule

Decision Rule: Reject H_0 if chi-square statistic is greater than the critical value. Otherwise, fail to reject H_0. (If x^2 is greater than 5.99, reject H_0)

Step 5: Compute the test statistic

	Less than 18	18-40	Greater than 40
Observed Values	288	571	141
Expected Values	200	300	500

$$X^2 = \Sigma \frac{(O-E)^2}{E}$$

$$X^2 = \frac{(288-200)^2}{200} + \frac{(571-300)^2}{300} + \frac{(141-500)^2}{500}$$

$$X^2 = \frac{(88)^2}{200} + \frac{(271)^2}{300} + \frac{(-359)^2}{500}$$

$$X^2 = \frac{7744}{200} + \frac{73441}{300} + \frac{128881}{500} = 541.29$$

Step 6: Interpret the results

Since, the chi-square statistic value (541.29) is greater than 5.99, we reject the null hypothesis (H_0). Therefore, at 5% level of significance, there is sufficient data that the observed distribution do not meet the expected distribution.

Try This!

Employers want to know which days of the week employees are absent in a five-day work week. Most employers would like to believe that employees are absent equally during the week. Suppose a random sample of 60 managers were asked on which day of the week they had the highest number of employee absences. The results were distributed as in the table below. For the population of employees, do the days for the highest number of absences occur with equal frequencies during a five-day work week? Test at a 5% significance level.

	Monday	Tuesday	Wednesday	Thursday	Friday
Number of Absences	15	12	9	9	15

Step 1. Formulate the hypotheses

H_0: The absent days occur with equal frequencies, that is, they fit a uniform distribution.

H_a: The absent days occur with unequal frequencies, that is, they do not fit a uniform distribution.

Step 2: Set the level of significance or alpha level

$$\alpha = 0.05$$

Step 3: Find the degrees of freedom

$\alpha = 0.05$

df = k-1
df =5-1
df = 4

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 9.49

Step 4: Set up the decision rule

Decision Rule: Reject H_0 if chi-square statistic is greater than the critical value. Otherwise, fail to reject H_0. (If x^2 is greater than 9.49, reject H_0)

Step 5: Compute the test statistic

Number of Absences	Monday	Tuesday	Wednesday	Thursday	Friday
Observed Values	15	12	9	9	15
Expected Values	12	12	12	12	12

If the absent days occur with **equal frequencies**, then, out of 60 absent days (the total in the sample: 15 + 12 + 9 + 9 + 15 = 60), there would be 12 absences on Monday, 12 on Tuesday, 12 on Wednesday, 12 on Thursday, and 12 on Friday.

Number of Absences	Monday	Tuesday	Wednesday	Thursday	Friday
Observed Values	15	12	9	9	15
Expected Values	12	12	12	12	12

$$X^2 = \Sigma \frac{(O-E)^2}{E}$$

$$X^2 = \frac{(15-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(15-12)^2}{12}$$

$$X^2 = \frac{(3)^2}{12} + \frac{(0)^2}{12} + \frac{(-3)^2}{12} + \frac{(-3)^2}{12} + \frac{(3)^2}{12}$$

$$X^2 = 3$$

Tabular chi-square value is 9.49

Step 6: Interpret the results

Since, the chi-square statistic value (3) is less than 9.49, we fail to reject the null hypothesis (H_0). Therefore, at 5% level of significance, from the sample data, there is sufficient evidence to conclude that the absent days occur with equal frequencies.

CONTINGENCY TABLE

- A contingency table is a type in a matrix format that displays the frequency distribution of the variables.
- They provide a basic picture of the interrelation between two variables and can help find interaction between them.

	Column 1	Column 2	Totals
Row 1	A	B	R1
Row 2	C	D	R2
Totals	C1	C2	N

The chi-square statistic compares the observed count in each table cell to the count which would be expected under the assumption of no association between the row and column classifications.

Tests using Contingency Table

The **test of independence** is used to determine whether two variables are independent of or related to each other when a single sample is selected.

The **test of homogeneity** is used to determine whether the proportions for a variable are equal when several samples are selected from different populations.

02

TEST OF INDEPENDENCE

Test of Independence

A **chi-square independence test** is used to test the independence of two variables. Using a chi-square test, you can determine whether the occurrence of one variable affects the probability of the occurrence of the other variable.

Test of Independence

To test whether two categorical variable are associated with each other, the formula employed is:

$$x^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

rows columns

where,

O_{ij} is the observed frequency in i^{th} and j^{th} category;

E_{ij} is the expected frequency in the i^{th} and j^{th} category

For a contingency table that has r rows and c columns, the Chi-square test can be generalized as a test of independence. Thus, as a test of independence, hypotheses are as follows:

H_o: There is no relationship between two categorical variables. (The two variables are independent.)

H_a: There is a relationship between two categorical variables. (The two variables are not independent.)

Decision Rule

- Reject H_o if $x^2 \geq x^2_{\frac{\alpha}{2}}(r-1)(c-1)$; otherwise fail to reject H_0.
- Reject the null hypothesis at a specified level of significance if the computed value of chi-square exceeds the table value.

Example

A study is being conducted to determine whether there is a relationship between jogging and blood pressure. A random sample of 210 subjects is selected, and they are classified as shown in the contingency table. Using $\alpha = 0.05$, determine whether a relationship exists between jogging and blood pressure.

Observed Frequencies/Values

Jogging Status	Blood Pressure			
	Low	Moderate	High	Total
Jogger	34	57	21	112
Non-Jogger	15	63	20	98
Total	49	120	41	210

Step 1. Formulate the hypotheses

H_0 : There is no relationship between jogging and blood pressure.
H_a : There is a relationship between jogging and blood pressure.

Step 2: Obtained the expected frequencies for each cell

The expected frequency in each cell can be determined by getting the product of the row total and column total then divide the product by the grand total.

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

Observed Frequencies

	Low	Moderate	High	Total
Jogger	34	57	21	112
Non-Jogger	15	63	20	98
Total	49	120	41	210

Row Totals

Grand Total

Column Totals

Expected Frequencies

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

	Low	Moderate	High	Total
Jogger	$L_J = \frac{112 \times 49}{210} = 26.13$	$M_J = \frac{112 \times 120}{210} = 64$	$H_J = \frac{112 \times 41}{210} = 21.87$	112
Non-Jogger	$L_{NJ} = \frac{98 \times 49}{210} = 22.87$	$M_{NJ} = \frac{98 \times 120}{210} = 56$	$H_{NJ} = \frac{98 \times 41}{210} = 19.13$	98
Total	49	120	41	210

Step 3: Subtract the expected frequencies from the observed frequencies.

	O	E	$O - E$
L_J (Low/Jogger)	34	26.13	7.87
L_{NJ} (Low/Non-Jogger)	15	22.87	-7.87
M_J (Moderate/Jogger)	57	64	-7
M_{NJ} (Moderate/Non-Jogger)	63	56	7
H_J (High/Jogger)	21	21.87	-0.87
H_{NJ} (High/Non-Jogger)	20	19.13	0.87

Step 4: Square the difference.

	O	E	$O - E$	$(O - E)^2$
L_J (Low/Jogger)	34	26.13	7.87	61.94
L_{NJ} (Low/Non-Jogger)	15	22.87	-7.87	61.94
M_J (Moderate/Jogger)	57	64	-7	49
M_{NJ} (Moderate/Non-Jogger)	63	56	7	49
H_J (High/Jogger)	21	21.87	-0.87	0.76
H_{NJ} (High/Non-Jogger)	20	19.13	0.87	0.76

Step 5: Divide the squared difference by the expected frequencies.

	O	E	$O-E$	$(O-E)^2$	$\frac{(O-E)^2}{E}$
L_J (Low/Jogger)	34	26.13	7.87	61.94	2.37
L_{NJ} (Low/Non-Jogger)	15	22.87	-7.87	61.94	2.71
M_J (Moderate/Jogger)	57	64	-7	49	0.77
M_{NJ} (Moderate/Non-Jogger)	63	56	7	49	0.88
H_J (High/Jogger)	21	21.87	-0.87	0.76	0.03
H_{NJ} (High/Non-Jogger)	20	19.13	0.87	0.76	0.04

Step 6: Add the quotients to obtain the chi-square value.

	O	E	$O - E$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
L_J (Low/Jogger)	34	26.13	7.87	61.94	2.37
L_{NJ} (Low/Non-Jogger)	15	22.87	-7.87	61.94	2.71
M_J (Moderate/Jogger)	57	64	-7	49	0.77
M_{NJ} (Moderate/Non-Jogger)	63	56	7	49	0.88
H_J (High/Jogger)	21	21.87	-0.87	0.76	0.03
H_{NJ} (High/Non-Jogger)	20	19.13	0.87	0.76	0.04
					$\sum \frac{(O - E)^2}{E} = 6.80$

Step 7: Find the degrees of freedom

$\alpha = 0.05$

$df = (r-1)(c-1)$
$df = (2-1)(3-1)$
$df = (1)(2)$
$df = 2$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 5.99

Step 8: Compare the obtained chi-square value with the table value at 0.05 level of significance.

$x^2 = 6.80$; Table $x^2 = 5.99$ at 0.05 and df = 2

- The tabular chi-square value of 5.99 is less than the computed value of 6.80

Decision Rule: Reject the null hypothesis at a specified level of significance if the computed value of chi-square exceeds the table value.

Step 9: Make a Conclusion

- Since, the tabular chi-square value of 5.99 is less than the computed value of 6.80, then there is sufficient evidence to reject the null hypothesis and conclude that there is significant relationship between jogging and blood pressure.

03

TEST OF HOMOGENEITY

Test of Homogeneity

- It is used to test the homogeneity of the responses of the respondents with regard to certain issues and opinions; where responses are put in a contingency table.

Example:

- Impeachment trial of Pres. Estrada – the reactions of the Filipino people
- Opening of the second envelope favor; not favor or neutral

- To calculate the test statistic for a test for homogeneity, follow the same procedure as with the test of independence.

Hypotheses

- H_o: The distributions of the two populations are the same.
- H_a: The distributions of the two populations are not the same.

Example

President Arroyo made a nationwide announcement on television about her conversation with the COMELEC Commissioner and she asked for public apology. To determine the opinion of the public (Agree, Disagree, No Opinion) , a survey was conducted in 3 municipalities of La Union. The following table gives the opinion of 2000 parents from San Fernando, 1500 parents from Rosario and 1000 parents from San Juan.

At the 0.01 level of significance, test for homogeneity of opinion among the 3 municipalities concerning the public apology of President Arroyo.

Observed Frequencies/Values

Opinion	Municipalities			
	San Fernando	Rosario	San Juan	Total
Agree	650	660	360	**1670**
Disagree	420	300	260	**980**
No Opinion	930	540	380	**1850**
	2000	**1500**	**1000**	**4500**

Step 1. Formulate the hypotheses.

H_0: For each opinion, the proportions of municipalities are the same.
H_a : For at least one opinion the proportions of the municipalities are not the same.

Step 2: Determine the expected frequencies.

The expected frequency in each cell can be determined by getting the product of the row total and column total then divide the product by the grand total.

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

Opinion	Municipalities			
	San Fernando	Rosario	San Juan	Total
Agree	650	660	360	**1670**
Disagree	420	300	420	**1140**
No Opinion	930	540	220	**1690**
	2000	**1500**	**1000**	**4500**

Observed Frequencies

Expected Frequencies

$$EF = \frac{(row\ total)(column\ total)}{grand\ total}$$

Opinion	Municipalities			
	San Fernando	Rosario	San Juan	Total
Agree	**742.22**	**556.67**	**371.11**	**1670**
Disagree	**506.67**	**380**	**253.33**	**1140**
No Opinion	**751.11**	**563.33**	**375.56**	**1690**
	2000	**1500**	**1000**	**4500**

Step 3: Subtract the expected frequencies from the observed frequencies.

	O	E	$O - E$
San Fernando/Agree	650	742.22	-92.22
San Fernando/Disagree	420	506.67	-86.67
San Fernando/No Opinion	930	751.11	178.89
Rosario/Agree	660	556.67	103.33
Rosario/Disagree	300	380	-80
Rosario/No Opinion	540	563.33	-23.33
San Juan/Agree	360	371.11	-11.11
San Juan/Disagree	420	253.33	166.67
San Juan/ No Opinion	220	375.56	-155.56

Step 4: Square the difference.

	O	E	$O - E$	$(O - E)^2$
San Fernando/Agree	650	742.22	-92.22	8504.52
San Fernando/Disagree	420	506.67	-86.67	7511.69
San Fernando/No Opinion	930	751.11	178.89	32001.63
Rosario/Agree	660	556.67	103.33	10677.09
Rosario/Disagree	300	380	-80	6400
Rosario/No Opinion	540	563.33	-23.33	544.29
San Juan/Agree	360	371.11	-11.11	123.43
San Juan/Disagree	420	253.33	166.67	27778.89
San Juan/ No Opinion	220	375.56	-155.56	24198.91

Step 5: Divide the squared difference by the expected frequencies.

	O	E	$O - E$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
San Fernando/Agree	650	742.22	-92.22	8504.52	11.46
San Fernando/Disagree	420	506.67	-86.67	7511.69	14.83
San Fernando/No Opinion	930	751.11	178.89	32001.63	42.61
Rosario/Agree	660	556.67	103.33	10677.09	19.18
Rosario/Disagree	300	380	-80	6400	16.84
Rosario/No Opinion	540	563.33	-23.33	544.29	0.97
San Juan/Agree	360	371.11	-11.11	123.43	0.33
San Juan/Disagree	420	253.33	166.67	27778.89	109.65
San Juan/ No Opinion	220	375.56	-155.56	24198.91	64.43

Step 6: Add the quotients to obtain the chi-square value.

	O	E	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
San Fernando/Agree	**650**	**742.22**	**-92.22**	**8504.52**	**11.46**
San Fernando/Disagree	**420**	**506.67**	**-86.67**	**7511.69**	**14.83**
San Fernando/No Opinion	**930**	**751.11**	**178.89**	**32001.63**	**42.61**
Rosario/Agree	**660**	**556.67**	**103.33**	**10677.09**	**19.18**
Rosario/Disagree	**300**	**380**	**-80**	**6400**	**16.84**
Rosario/No Opinion	**540**	**563.33**	**-23.33**	**544.29**	**0.97**
San Juan/Agree	**360**	**371.11**	**-11.11**	**123.43**	**0.33**
San Juan/Disagree	**420**	**253.33**	**166.67**	**27778.89**	**109.65**
San Juan/ No Opinion	**220**	**375.56**	**-155.56**	**24198.91**	**64.43**
					$\sum \frac{(O-E)^2}{E} = 280.3$

Step 7: Find the degrees of freedom

$\alpha = 0.01$

$df = (r-1)(c-1)$
$df = (3-1)(3-1)$
$df = (2)(2)$
$df = 4$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 13.28

Step 8: Compare the obtained chi-square value with the table value at 0.01 level of significance.

$x^2 = 280.3$; Table $x^2 = 13.28$ at 0.01 and df = 4

- The tabular chi-square value of 13.28 is less than the computed value of 280.3

Decision Rule: Reject the null hypothesis at a specified level of significance if the computed value of chi-square exceeds the table value.

Step 9: Make a Conclusion

- Since, the tabular chi-square value of 13.28 is less than the computed value of 280.3, then there is sufficient evidence to reject the null hypothesis and conclude that at least the proportions of the opinions in each municipality are not the same. Meaning, people in different municipalities give different views with regards to the public apology of Pres. Arroyo.

Try This!

Considering a study in which the effectiveness of hypnosis as a means of improving the memory of the eyewitness to a crime is examined and the result is shown:

	Hypnotized	Control
Correct Identification	7	17
Incorrect Identification	33	23
Total	40	40

Step 1. Formulate the hypotheses.

Ho: Hypnosis does not affect the recognition memory of eyewitness to a crime.

Ha: Hypnosis affects the recognition memory of eyewitness to a crime.

Step 2: Obtained the expected frequencies for each cell

Observed Frequencies	Hypnotized	Control	Total
Correct Identification	7	17	24
Incorrect Identification	33	23	56
Total	40	40	80

Expected Frequencies	Hypnotized	Control	Total
Correct Identification	$H_c =$ [redacted]	$C_c =$ [redacted]	24
Incorrect Identification	$H_i =$ [redacted]	$C_i =$ [redacted]	56
Total	40	40	80

Calculate Test Statistic

	O	E	$O - E$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
H_C (Hypnotized/Correct)	7	12	-5	25	2.08
H_i (Hypnotized/Incorrect)	33	28	5	25	0.89
C_c (Control/Correct)	17	12	5	25	2.08
C_i (Control/Incorrect)	23	28	-5	25	0.89
					$x^2 = 5.94$

Perform Step 3 - 6

Step 7: Find the degrees of freedom

$\alpha = 0.05$

$df = (r-1)(c-1)$
$df = (2-1)(2-1)$
$df = (1)(1)$
$df = 1$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 3.84

Step 8: Compare the obtained chi-square value with the table value at 0.05 level of significance.

$x^2 = 5.94$; Table $x^2 = 3.84$ at 0.05 and df = 1

- The obtained chi-square value of 5.94 is greater than the tabular value of 3.84.

Decision Rule: Reject the null hypothesis at a specified level of significance if the computed value of chi-square exceeds the table value.

Step 9: Make a Conclusion

- Since, the obtained chi-square value of 5.94 is greater than the tabular value of 3.84, then we have sufficient evidence to reject the null hypothesis. The result suggests significant difference in the ability of hypnotized and control subjects in identifying a thief. The hypnotized subjects were less not more accurate in identifying the thief.

Try This!

Suppose that 250 randomly selected **male** college students and 300 randomly selected **female** college students were asked about their living arrangements: dormitory, apartment, with parents, other. Do male and female college students have the same distribution of living arrangements? Use a level of significance of 0.05. The results are shown in figure below.

	Low Income	Middle Income	High Income
Male	109	365	26
Female	192	249	9

Do male and female college students have the same distribution of living arrangements?

Step 1. Formulate the hypotheses.

Ho: The income distribution is the same for the males and females.
Ha: The income distribution is not the same for the males and females.

Step 2: Obtained the expected frequencies for each cell

Observed Frequencies	Low Income	Middle Income	High Income	Total
Male	109	365	26	500
Female	192	249	9	450
Total	301	614	35	950

Expected Frequencies	Low Income	Middle Income	High Income	Total
Male	158	323	18	500
Female	143	291	17	450
Total	301	614	35	950

Calculate Test Statistic

	O	E	$O - E$	$(O - E)^2$	$\frac{(O - E)^2}{E}$
Low Income/Male	109	158	-49	2401	15.20
Low Income/Female	192	143	49	2401	16.79
Middle Income/Male	365	323	42	1764	5.46
Middle Income/Female	249	291	-42	1764	6.06
High Income/Male	26	18	8	64	3.56
High Income/Female	9	17	-8	64	3.76
					$x^2 = 50.83$

Step 7: Find the degrees of freedom

$\alpha = 0.05$

$df = (r-1)(c-1)$
$df = (2-1)(3-1)$
$df = (1)(2)$
$df = 2$

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22

Tabular chi-square value is 5.99

Step 8: Compare the obtained chi-square value with the table value at 0.05 level of significance.

$x^2 = 50.83$; Table $x^2 = 5.99$ at 0.05 and df = 1

- ❑ The obtained chi-square value of 50.83 is greater than the tabular value of 5.99.

Decision Rule: Reject the null hypothesis at a specified level of significance if the computed value of chi-square exceeds the table value.

Step 9: Make a Conclusion

- Since, the obtained chi-square value of 50.83 is greater than the tabular value of 5.99, then we must reject the null hypothesis. Therefore, the income distribution among males and females is the same.

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

References:

- https://www.jmp.com/en_ch/statistics-knowledge-portal/chi-square-test/chi-square-goodness-of-fit-test.html
- https://www.investopedia.com/terms/g/goodness-of-fit.asp
- https://link.springer.com/referenceworkentry/10.1007%2F978-1-4020-5614-7_3475#:~:text=The%20chi%2Dsquare%20test%20of,the%20row%20and%20column%20labels.
- https://courses.lumenlearning.com/odessa-introstats1-1/chapter/goodness-of-fit-test/

THANK YOU FOR LISTENING !