Dr. Saeed Alghamdi
Room 542
Statistics Department
Faculty of Sciences
King Abdulaziz University
web sit
http://saalghamdy.kau.edu.sa

# Elementary Statistics

**A Step by Step Approach**
**Eighth Edition**

by
Allan G. Bluman

SLIDES PREPARED
BY

Dr. Saeed A Dobbah Alghamdi
Statistics Department
King Abdulaziz University

## CHAPTER 11

Other Chi-Square Tests



## Notes

- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………



## Objectives

- Test a distribution for goodness of fit using chi-square.
- Test two variables for independence using chi-square.
- Test proportions for homogeneity using chi-square.



## Notes

- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………
- ❑………………………………………………………………

# Introduction

- The chi-square distribution can be used for tests concerning frequency distributions, such as: " If a sample of buyers is given a choice of automobile colors, will each color be selected with the same frequency?"
- The chi-square distribution can also be used to test the independence of two variables. For example, "Are senators' opinions on gun control independent of party affiliations?"
- The chi-squared distribution can be used to test the homogeneity of proportions. For example: "Is the proportion of high school seniors who attend college immediately after graduating the same for the northern, southern, eastern, and western parts of the United States?"



# Notes

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………



# Test for Goodness-of-Fit

- The chi-square statistic can be used to see whether a frequency distribution fits a specific pattern.
- The formula for the chi-square goodness-of-fit test:

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

with d.f. = number of categories - 1

$O$ = observed frequency

$E$ = expected frequency

- This test is a right-tailed test, since when the ($O - E$) values are squared, the answer will be positive or zero.
- The assumptions includes that the data must be obtained from a random sample and the expected frequency for each category must be 5 or more.



# Notes

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

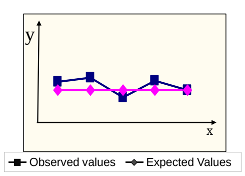
❑……………………………………………………………………

❑……………………………………………………………………

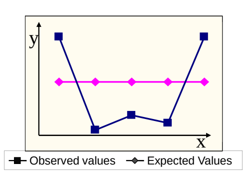


# Test for Goodness-of-Fit

- A Good Fit: When the observed values and expected values are close together, the chi-square test value will be small. Then the decision will be not to reject the null-hypothesis—hence, there is a "good fit."




- Not A Good Fit: When the observed values and the expected values are far apart, the chi-square test value will be large. Then, the null hypothesis will be rejected—hence, there is "not a good fit."






# Notes

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

❑……………………………………………………………………

## Chi-Square Goodness-of-Fit Procedure

- Step 1 State the hypotheses and identify the claim.
- Step 2 Find the critical value. The test is always right-tailed.
- Step 3 Compute the test value.
  Find the sum of the $\frac{(O-E)^2}{E}$ values.
- Step 4 Make the decision.
- Step 5 Summarize the results.



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………



## Goodness-of-Fit Results

- When there is a perfect agreement between the observed and the expected values, $\chi^2 = 0$, but $\chi^2$ can never be negative.
- The test is right-tailed because "$H_0$: Good fit" and "$H_1$: Not a good fit" means that $\chi^2$ will be small in the first case and $\chi^2$ will be large in the second case.



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………



A sample of 100 people provided the following data in preference of soda flavors:

| Cherry | Strawberry | Orange | Lime | Grape |
|---|---|---|---|---|
| 32 | 28 | 16 | 14 | 10 |

Is there enough evidence to reject the claim that there is no preference in the selection of fruit soda flavors with 5% significance level?

H0: Consumers show no preference for flavors

H1: Consumers show preference for flavors

| Cherry | Strawberry | Orange | Lime | Grape |
|---|---|---|---|---|
| 32 | 28 | 16 | 14 | 10 |
| 20 | 20 | 20 | 20 | 20 |

$$\chi^2 = \frac{1}{E}\sum O^2 - n$$

$$= \frac{1}{20}(32^2 + 28^2 + 16^2 + 14^2 + 10^2) - 100 = 18$$

$$\chi^2_4 = 9.488$$

Since 18>9.488, reject the null hypothesis. Thus, there is enough evidence to reject the claim that consumers show no preference for the flavors.



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………

1- Enter the observed frequencies in row 1

2- Enter the expected frequencies in row 2

3- From the toolbar, select Add-Ins, **MegaStat>Chi-Square/Crosstab>Goodness of Fit Test**.

4- In the dialog box, select the corresponding data. Then click [OK]

Goodness of Fit Test

| observed | expected | O - E | (O - E)² / E | % of chisq |
|---|---|---|---|---|
| 32 | 20.000 | 12.000 | 7.200 | 40.00 |
| 28 | 20.000 | 8.000 | 3.200 | 17.78 |
| 16 | 20.000 | -4.000 | 0.800 | 4.44 |
| 14 | 20.000 | -6.000 | 1.800 | 10.00 |
| 10 | 20.000 | -10.000 | 5.000 | 27.78 |
| 100 | 100.000 | 0.000 | 18.000 | 100.00 |

18.00 chi-square
4 df
.0012 p-value



## Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………



## Independence Test

- The chi-square *independence test* can be used to test the independence of two variables.
- $H_0$: There is no relationship between two variables.
- $H_1$: There is a relationship between two variables.
- If the null hypothesis is rejected, there is some relationship between the variables.
- In order to test the null hypothesis, one must compute the expected frequencies, assuming the null hypothesis is true.
- When data are arranged in table form for the independence test, the table is called a *contingency table*.



## Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………



## Independence Test

### Contingency Table

| | Column 1 | Column 2 | Column 3 |
|---|---|---|---|
| Row 1 | $C_{1,1}$ | $C_{1,2}$ | $C_{1,3}$ |
| Row 2 | $C_{2,1}$ | $C_{2,2}$ | $C_{2,3}$ |

- The degrees of freedom for any contingency table are d.f. = (rows – 1) (columns – 1) = $(R-1)(C-1)$.



## Notes

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

❑……………………………………………………………

## Independence Test

- The formula for the test value for the independence test is the same as the one for the goodness-of-fit test.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

with d.f. $= (R - 1)(C - 1)$
$O$ = observed frequency
$E$ = expected frequency



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………



## Homogeneity of Proportions Test

- *Homogeneity of proportions test* is used when samples are selected from several different populations and the researcher is interested in determining whether the proportions of elements that have a common characteristic are the same for each population.
- $H_0$: $p_1 = p_2 = p_3 = \ldots = p_n$.
- $H_1$: At least one proportion is different from the others.
- When the null hypothesis is rejected, it can be assumed that the proportions are not all equal.



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………



## Independence and Homogeneity

- The procedures for the chi-square independence and homogeneity tests are identical and summarized below.
- Step 1 State the hypotheses and identify the claim.
- Step 2 Find the critical value in the right tail.
- Step 3 Compute the test value. To compute the test value, first find the expected values. For each cell of the contingency table, use the formula

$$E = \frac{\text{(row sum)(column sum)}}{\text{grand total}}$$

to get the expected value. To find the test value, use the formula $\chi^2 = \sum \frac{(O - E)^2}{E}$

- Step 4 Make the decision and Summarize the results.



## Notes

- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………
- ……………………………………………………………

# Assumptions

- The assumptions for the chi-square independence and homogeneity tests:
  1. The data are obtained from a random sample.
  2. The expected value in each cell must be 5 or more.



# Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………



- A sociology wishes to see whether the number of years of college a person has completed is related to her or his place of residence. A sample of 88 people is selected and classified as shown.

| Location | No college | Four-year degree | Advanced degree | Total |
|---|---|---|---|---|
| Urban | **15 (11.53)** | **12 (13.92)** | **8 (9.55)** | **35** |
| Suburban | **8 (10.55)** | **15 (12.73)** | **9 (8.73)** | **32** |
| Rural | **6 (6.92)** | **8 (8.35)** | **7 (5.73)** | **21** |
| Total | **29** | **35** | **24** | **88** |

- At $\alpha = 0.05$, can the sociologist conclude that a person's location is dependent on the number of years of college?

$H_0$: A person's place of residence is independent of the number of years of college completed.

$H_1$: A person's place of residence is dependent on the number of years of college completed.

- The degrees of freedom are (3-1)(3-1)=4, hence the critical value is 9.488.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(15-11.53)^2}{11.53} + \frac{(13-13.92)^2}{13.92} + \frac{(8-9.55)^2}{9.55} + \frac{(8-10.55)^2}{10.55} + \frac{(15-12.73)^2}{12.73} + \frac{(9-8.73)^2}{8.73} + \frac{(6-6.92)^2}{6.92} + \frac{(8-8.35)^2}{8.35} + \frac{(7-5.73)^2}{5.73} = 3.01$$

- Since 3.01 < 9.488, the decision is not to reject the null hypothesis.



# Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………



1- Enter the row variable labels in column A, beginning at cell A2

2- Enter the column variable labels in cells B1, C1, D1, …

3- Enter the observed values in the appropriate cells

4- From the toolbar, select **Add-Ins, MegaStat>Chi-Square/Crosstab>Contingency Table.**

5- In the dialog box, select the corresponding range of data.

6- Check chi-square from the Output Options.

7- Click [OK].

Chi-square Contingency Table Test for Independence

| | No college | Four-year degree | Advanced degree | Total |
|---|---|---|---|---|
| Urban | **15** | **12** | **8** | 35 |
| Suburban | **8** | **15** | **9** | 32 |
| Rural | **6** | **8** | **7** | 21 |
| Total | 29 | 35 | 24 | 88 |

| | |
|---|---|
| 3.01 | chi-square |
| 4 | df |
| .5569 | p-value |



# Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………

A researcher selected a sample of 50 seniors from each of three area high schools and asked each senior, "Do you drive to school in a car owned by either you or your parents?" The data are shown in the table. At $\alpha = 0.05$, test the claim that the proportion of students who drive their own or their parents' car is the same at all three schools.

| | School 1 | School 2 | School 3 | Total |
|---|---|---|---|---|
| Yes | 18 | 22 | 16 | 56 |
| No | 32 | 28 | 34 | 94 |
| Total | 50 | 50 | 50 | 150 |

**Solution**

**Step 1** State the hypotheses.

$H_0$: $p_1 = p_2 = p_3$

$H_1$: At least one proportion is different from the others.

**Step 2** Find the critical value. The formula for the degrees of freedom is the same as before: (rows − 1)(columns − 1) = (2 − 1)(3 − 1) = 1(2) = 2. The critical value is 5.991.

**Step 3** Compute the test value. First, compute the expected values.

$$E_{1,1} = \frac{(56)(50)}{150} = 18.67 \qquad E_{2,1} = \frac{(94)(50)}{150} = 31.33$$

$$E_{1,2} = \frac{(56)(50)}{150} = 18.67 \qquad E_{2,2} = \frac{(94)(50)}{150} = 31.33$$

$$E_{1,3} = \frac{(56)(50)}{150} = 18.67 \qquad E_{2,3} = \frac{(94)(50)}{150} = 31.33$$



## Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………



The completed table is shown here.

| | School 1 | School 2 | School 3 | Total |
|---|---|---|---|---|
| Yes | 18 (18.67) | 22 (18.67) | 16 (18.67) | 56 |
| No | 32 (31.33) | 28 (31.33) | 34 (31.33) | 94 |
| Total | 50 | 50 | 50 | 150 |

The test value is

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(18-18.67)^2}{18.67} + \frac{(22-18.67)^2}{18.67} + \frac{(16-18.67)^2}{18.67} + \frac{(32-31.33)^2}{31.33} + \frac{(28-31.33)^2}{31.33} + \frac{(34-31.33)^2}{31.33}$$

$$= 1.596$$

**Step 4** Make the decision. The decision is not to reject the null hypothesis, since $1.596 < 5.991$.

**Step 5** Summarize the results. There is not enough evidence to reject the null hypothesis that the proportions of high school students who drive their own or their parents' car to school are equal for each school.



## Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………



1- Enter the row variable labels in column A, beginning at cell A2

2- Enter the column variable labels in cells B1, C1, D1, …

3- Enter the observed values in the appropriate cells

4- From the toolbar, select **Add-Ins, MegaStat>Chi-Square/Crosstab>Contingency Table.**

5- In the dialog box, select the corresponding range of data.

6- Check chi-square from the Output Options.

7- Click [OK].

Chi-square Contingency Table Test for Independence

| | School 1 | School 2 | School 3 | Total |
|---|---|---|---|---|
| Yes | **18** | **22** | **16** | 56 |
| No | **32** | **28** | **34** | 94 |
| Total | 50 | 50 | 50 | 150 |

| | |
|---|---|
| 1.60 | chi-square |
| 2 | df |
| .4503 | p-value |



## Notes

- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………
- ❑……………………………………………………………

## Summary

- There are three main uses of the chi-square distribution:
  1. The test of independence is used to determine whether two variables are related or are independent.
  2. It can be used as goodness-of-fit test, in order to determine whether the frequencies of a distribution are the same as the hypothesized frequencies.
  3. The homogeneity of proportions test is used to determine if several proportions are all equal when samples are selected from different populations.



## Notes

- ❑……………………………………………………………….
- ❑……………………………………………………………….
- ❑……………………………………………………………….
- ❑……………………………………………………………….
- ❑……………………………………………………………….
- ❑……………………………………………………………….