Chapter 6
Chi-Square Tests

Department of Statistics and Operations Research



21- October-2026

Bayan Almukhlif

Contents

1 Introduction

2 Goodness-of-Fit Test χ^2

3 Goodness-of-Fit Examples χ^2

4 Test of Independence χ^2

5 Test of Homogeneity χ^2

6 Summary of the Tests

1

Introduction

What a chi-square test is · the two types of tests

Chi-Square Tests

A **Chi-square** test is a hypothesis testing method. It compares the observed values (O) with the expected values (E) to check a claim about an experiment.

There are three types of Chi-square tests:

1. **Goodness of fit test with a probability distribution.**
   Does the data follow a given probability distribution?

2. **Independence between two variables.**
   Are two variables A and B related?

3. **Homogeneity between two variables.**
   Are several populations alike?

2

Goodness-of-Fit Test

Does a population have a specified distribution?

Goodness-of-Fit Test

We want to know if a population has a specified theoretical distribution. The test checks how close the **observed** frequencies in a sample are to the **expected** frequencies from the hypothesized distribution.

We test the Hypothesis

$\mathbf{H_0}$ **:** The data follows the specified probability distribution.

$\mathbf{H_1}$ **:** The data does not follow the specified probability distribution.

Goodness-of-Fit Test: Steps (1)

1 Hypotheses

$\mathbf{H_0}$: The data follows the specified probability distribution.

$\mathbf{H_1}$: The data does not follow the specified probability distribution.

2 Observed frequencies

The data range is divided into $\boldsymbol{k}$ intervals (or cells or classes) and the frequency count $\mathbf{O_i}$ for each interval (cell, or class) is tabulated.

3 Expected frequencies

The expected frequency $\mathbf{E_i}$, based on the theoretical **pdf** $\boldsymbol{f}$ with **cumulative distribution** $\boldsymbol{F}$, is calculated for each interval (cell, or class) **i** as:

$$\boldsymbol{E_i = np_i}$$

where $\boldsymbol{n}$ is the number of data points, and

$$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = n$$

For a **class interval** with end points $\mathbf{a_{i-1}}$ and $\mathbf{a_i}$, use the cumulative distribution function **F**:

$$p_i = F(a_i) - F(a_{i-1})$$

For a **single value** $\boldsymbol{x_i}$, use the probability function $\boldsymbol{f}$:

$$p_i = f(a_i)$$

Rule

Each expected frequency must be ($\mathbf{E_i \geq 5}$).
If $\mathbf{E_i < 5}$, combine the class with a neighbor class. Then **k** becomes **smaller.**

Goodness-of-Fit Test: Steps (2)

4 Test statistic

Under the fact that $\mathbf{E_i \geq 5}$, i = 1, …, k, the test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

The test statistic (χ_0^2) is approximated to the chi-square distribution with ν degrees of freedom:

$$\nu = k - m - 1$$

k is the number of classes (after combining), ***m*** is the number of unknown parameters of the theoretical probability distribution estimated from sample data.

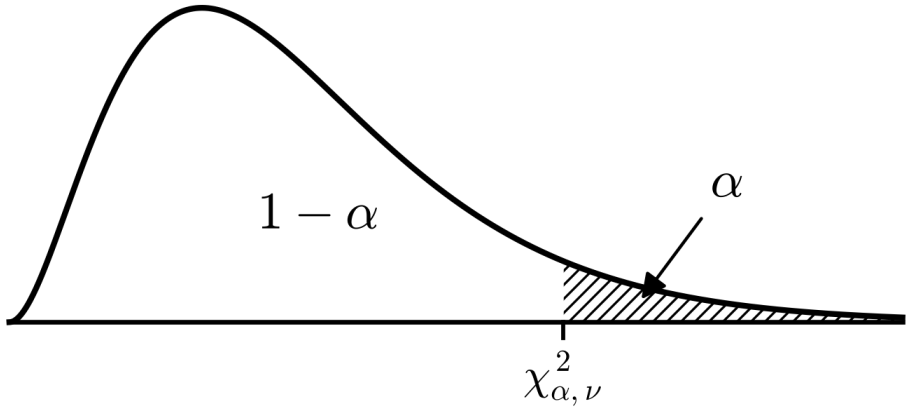
- If $E_i < 5$, E_i is combined to a neighbor level.

5 Decision

Reject H_0 if the test statistic is greater than the critical value:

$$\boldsymbol{\chi_0^2 > \chi^2_{\alpha,\, k-m-1}}$$

$\chi^2_{\alpha,\nu}$ is the value of Chi-Square distribution for significance level α with df = ν in the Chi-Square distribution table.



Shaded area = rejection region (right tail)

3

Goodness-of-Fit: Worked Examples

Die · machines (binomial) · battery lives and women weights (normal)

Example 1

We consider the tossing of a die. We hypothesize that the die is honest(balanced), which is equivalent to testing the hypothesis that the distribution of outcomes is the **discrete uniform distribution**

$$f(x) = \frac{1}{6}, \qquad x = 1,2, \dots, 6$$

Suppose that the die is tossed 120 times and each outcome is recorded. Theoretically, if the die is balanced, we would **expect** each face to occur 20 times ($E_i = 120 * \frac{1}{6} = 20$). The results are given by

Face	1	2	3	4	5	6	Total
Observed (O_i)	20	22	17	18	19	24	120
Expected (E_i)	20	20	20	20	20	20	120

By comparing the observed frequencies with the corresponding expected frequencies, the hypotheses:

H_0 : The die is fair ($f(x) = 1/6, x = 1, 2, \dots, 6$).

H_1 : The die is not fair ($f(x)$ is not the discrete uniform distribution).

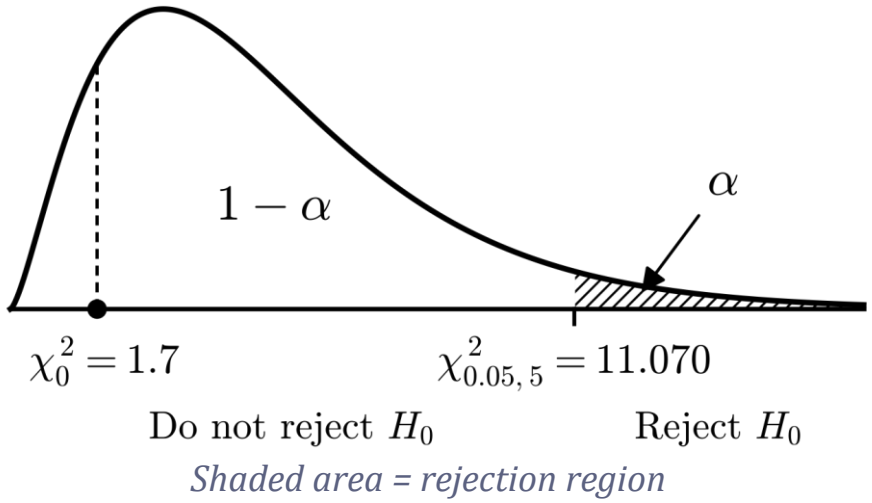
Example 1: Solution

We find that the χ_0^2 test statistic value is

$$\chi_0^2 = \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(24-20)^2}{20} = 1.7$$

Here **k = 6** and **m = 0**, so **ν =** k − m − 1 = 6 − 0 − 1 = **5.** Using the chi-squared table (Table A.5) with α = 0.05, we find

$$\boldsymbol{\chi^2_{0.05,5} = 11.070}$$



Shaded area = rejection region

Decision

Since 1.7 is less than the critical value ($\chi^2_{0.05,5}$), we fail to reject H_0. We conclude that the die is fair.

Example 2

In a study for the quality control for a certain machine to be used for paving the public roads, a researcher selects a sample of 4 machines within a period of 200 days for the production line and records the number of machines needed maintenance in each day. He obtained the following records:

No. of machines needed maintenance (x_i)	0	1	2	3	4	Total
No. of days (O_i)	101	79	19	1	0	200

Does this data suggest that the number of machines that need maintenance follows **Bin(4, p)** at level 0.01?

Hint

The parameter p is not given, so we must estimate it from the data.

Example 2: Solution

The estimation of the parameter p

$$\mu = np \;\Rightarrow\; \hat{\mu} = 4\hat{p}$$

$$\hat{\mu} = \frac{\sum O_i x_i}{n} = \frac{0(101) + 1(79) + 2(19) + 3(1) + 4(0)}{200} = 0.6$$

$$\hat{p} = \frac{\hat{\mu}}{4} = \frac{0.6}{4} = 0.15$$ **, So m=1**

Hypotheses

H_0 : No. of machines that need maintenance follows Bin(4, p).

H_1 : No. of machines that need maintenance does not follow Bin(4, p).

Expected frequencies under H_0

$$E_i = np_i = 200\binom{4}{i}(0.15)^i(0.85)^{4-i}, \qquad i = 0,1,2,3,4$$

Example 2: Solution

No. of machines needed maintenance (x_i)	0	1	2	3	4	Total
No. of days (O_i)	101	79	19	1	0	200
Probability (p_i)	0.5220	0.3685	0.0975	0.0115	0.0005	1.0000
Expected values ($E_i = 200\ p_i$)	104.4	73.7	19.5	2.3	0.1	200.0

For example, $p_0 = (0.85)^4 = 0.5220$ and $E_0 = 200 \times 0.5220 = 104.4$.

Combine classes

We note that the $E_3 = 2.3$ and $E_4 = 0.1$, which correspond to i = 3, 4, are less than 5. So we combine these two categories with the category before them (we add both the observed and the expected values).

The new category is "**x ≥ 2**", with **O = 19 + 1 + 0 = 20** and **E = 19.5 + 2.3 + 0.1 = 21.9.**

Example 2: Solution

No. of machines (x_i)	0	1	≥ 2	Total
No. of days (O_i)	101	79	20	200
Expected values (E_i)	104.4	73.7	21.9	200.0

The number of observed values became k = 3, and the test statistic is

$$\chi_0^2 = \frac{(101-104.4)^2}{104.4} + \frac{(79-73.7)^2}{73.7} + \frac{(20-21.9)^2}{21.9} = 0.657$$

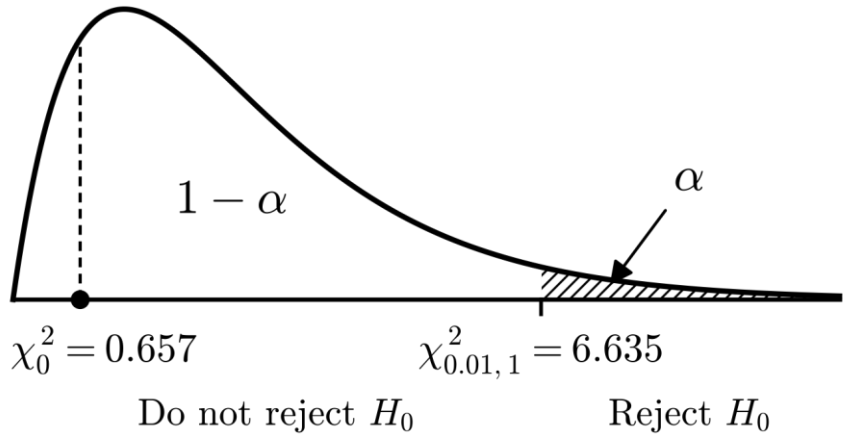
We estimated one parameter (**m = 1**), so

$$\nu = k - m - 1 = 3 - 1 - 1 = 1$$

$$\text{R. R.} = \left(\chi^2_{0.01,1}, \infty\right) = (6.635, \infty)$$

Decision

Since 0.657 < 6.635, we fail to reject H_0 at the level 0.01. Therefore, the No. of machines that need maintenance follows Bin(4, p)



Shaded area = rejection region

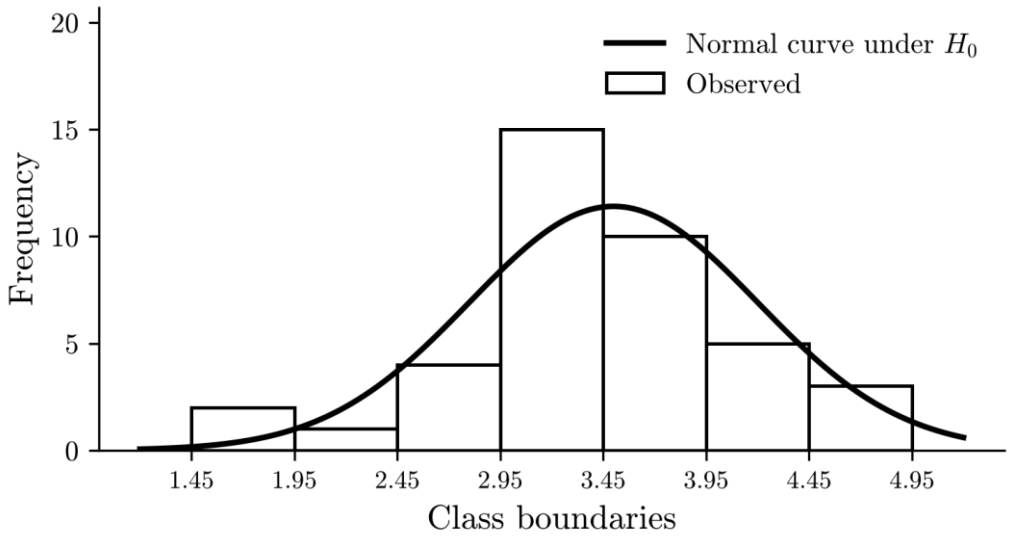
Remark

If p is not estimated but is given in advance, then m = 0 and ν = k − 1 = 2.

Example 3

A third example is to test the hypothesis that the frequency distribution of battery lives given in this Table

Class	Class boundaries	Frequency (O_i)
1	1.45 – 1.95	2
2	1.95 – 2.45	1
3	2.45 – 2.95	4
4	2.95 – 3.45	15
5	3.45 – 3.95	10
6	3.95 – 4.45	5
7	4.45 – 4.95	3
Total		**40**



The frequency may be approximated by a normal distribution with mean **μ = 3.5** and standard deviation **σ = 0.7**. Test the hypothesis

H_0 : The data is normal $N(3.5, 0.7^2)$.
H_1 : The data is not normal.

at level of significance α = 0.05.

Example 3: Solution

Under H_0, the z-values of the class boundaries are $z = (x - 3.5)/0.7$. The first and last classes are treated as open, so they take all the area to the left or to the right. For the upper boundary of first class is $z_2 = (1.95 - 3.5)/0.7 = -2.21$.

From the standard normal table (Table A.3), the area to the left of z_2 gives

$$p_1 = P(Z < -2.21) = F(-2.21) = 0.0136 \quad ; \quad E_1 = np_1 = 40(0.0136) \approx 0.5$$

The first and last classes take all the area to the left and to the right. Similarly, we get:

Class	Boundaries	O_i	z-interval	From Table A.3	p_i	$E_i = 40\ p_i$
1	1.45 – **1.95**	2	z < -2.21	F(-2.21) = 0.0136	0.0136	0.5
2	1.95 – 2.45	1	-2.21 < z < -1.50	F(-1.5)-F(-2.21)=0.0668 – 0.0136	0.0532	2.1
3	2.45 – 2.95	4	-1.50 < z < -0.79	F(-0.79)-F(-1.50)=0.2148 – 0.0668	0.1480	5.9
4	2.95 – 3.45	15	-0.79 < z < -0.07	F(-0.07)-F(-0.79)=0.4721 – 0.2148	0.2573	10.3
5	3.45 – 3.95	10	-0.07 < z < 0.64	F(0.64)-F(-0.07)=0.7389 – 0.4721	0.2668	10.7
6	3.95 – 4.45	5	0.64 < z < 1.36	F(1.36)-F(0.64)=0.9131 – 0.7389	0.1742	7.0
7	**4.45** – 4.95	3	z > 1.36	1 – F(1.36) = 1 – 0.9131	0.0869	3.5
Total		**40**			**1.0000**	**40.0**

The classes 1, 2 and 7 have $E_i < 5$ (shaded). We combine them in the next step.

Example 3: Solution

We note that $E_1 = 0.5$, $E_2 = 2.1$ and $E_7 = 3.5$ are less than 5. So we combine classes 1, 2 and 3 (E = 0.5 + 2.1 + 5.9 = 8.5), and we add class 7 to class 6 (E = 7.0 + 3.5 = 10.5). Now **k = 4.**

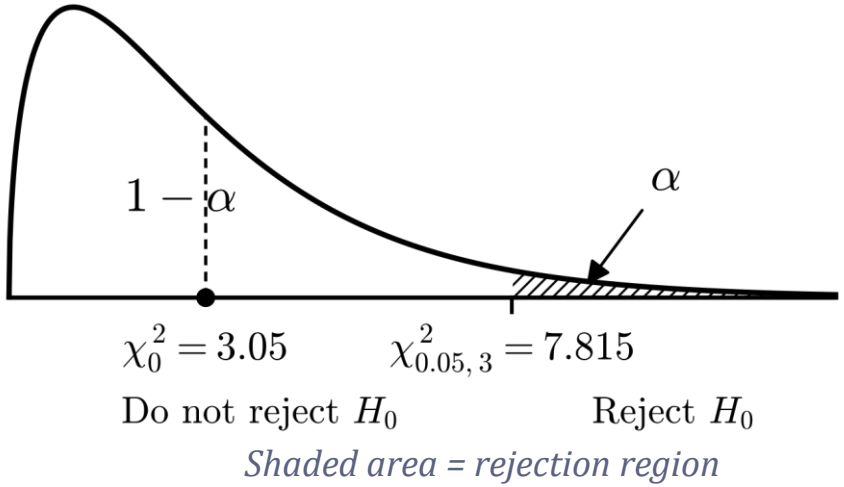
Class boundaries	1.45–2.95	2.95–3.45	3.45–3.95	3.95–4.95	Total
Observed (O_i)	**7**	15	10	**8**	40
Expected (E_i)	**8.5**	10.3	10.7	**10.5**	40.0

$$\chi_0^2 = \frac{(7-8.5)^2}{8.5} + \frac{(15-10.3)^2}{10.3} + \frac{(10-10.7)^2}{10.7} + \frac{(8-10.5)^2}{10.5} = \mathbf{3.05}$$

μ and **σ** are given and **not estimated** (m = 0),

so ν = k − m − 1 = 4 − 0 − 1 = 3.

$$\text{R. R.} = \left(\chi^2_{0.05,3}, \infty\right) = (7.815, \infty)$$



Shaded area = rejection region

Decision

Since the computed $\chi_0^2 = 3.05 < 7.815$, we have no reason to reject H_0. The normal distribution with μ = 3.5 and σ = 0.7 provides a good fit for the distribution of battery lives.

Example 4

The frequency distribution table of 120 women weights is:

Class interval	42.5–47.5	47.5–52.5	52.5–57.5	57.5–62.5	62.5–67.5	67.5–72.5	72.5–77.5	Total
Frequencies (O_i)	5	16	28	32	23	11	5	120
Class Midpoint **(x_i)**	45	50	55	60	65	70	75	

Does this data suggest that the women weight is normally distributed at level 0.05?

Note

$$\text{Class Midpoint} = \frac{lower\ bound + upper\ bound}{2}$$

$$\text{First class midpoint} = \frac{42.5+47.5}{2} = 45$$

Example 4: Solution

The mean μ and the variance σ^2 of the weight population are estimated by

$$\bar{x} = \frac{\sum O_i x_i}{n} = \frac{7125}{120} = 59.375$$

and

$$S^2 = \frac{\sum O_i x_i^2 - n\bar{x}^2}{n-1} = \frac{429225 - 120(59.375)^2}{119} = 51.9$$

$\widehat{\mu} = 59.4$ *and* $\widehat{\sigma^2} = 51.9$ **, so m=2 .**

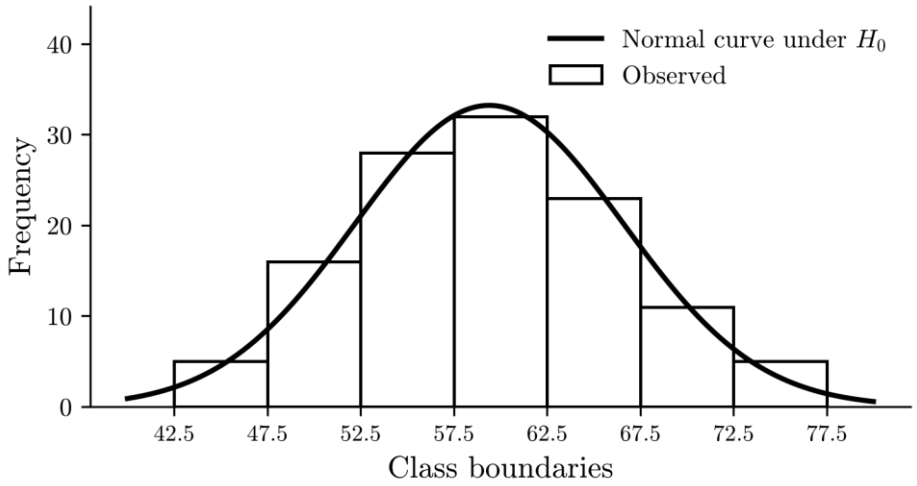
The hypothesis:

H_0 : Women weight follows Normal distribution.

H_1 : Women weight does not follow Normal distribution.

Under that H_0 is correct, the expected number in each class interval (a_i, b_i) is:

$$E_i = np_i = 120\, P(a_i \le X \le b_i) = 120\, [F(b_i) - F(a_i)]$$



Observed frequencies and the fitted normal curve

Example 4: Solution

Here $\hat{\sigma} = \sqrt{51.9} \approx 7.2$ and $\mathbf{z = (x - 59.4)/7.2}$. The first and last classes take all the area to the left and to the right. The sum of the $\mathbf{E_i}$ must be $\mathbf{n = 120}$. Using the standard normal table (Table A.3):

Class interval	O_i	z-interval	From Table A.3	p_i	$E_i = 120\ p_i$
42.5– **47.5**	5	$z < -1.65$	$F(-1.65) = 0.0495$	0.0495	5.94
47.5–52.5	16	$-1.65 < z < -0.96$	F(-0.96)-F(-1.65)=0.1685 – 0.0495	0.1190	14.28
52.5–57.5	28	$-0.96 < z < -0.26$	F(-0.26)-F(-0.96)=0.3974 – 0.1685	0.2289	27.47
57.5–62.5	32	$-0.26 < z < 0.43$	F(0.43)-F(-0.26)=0.6664 – 0.3974	0.2690	32.28
62.5–67.5	23	$0.43 < z < 1.12$	F(1.12)-F(0.43)=0.8686 – 0.6664	0.2022	24.26
67.5–72.5	11	$1.12 < z < 1.82$	F(1.82)-F(1.12)=0.9656 – 0.8686	0.0970	11.64
72.5 –77.5	5	$z > 1.82$	$1 - F(1.82) = 1 - 0.9656$	0.0344	4.13
Total	**120**			**1.0000**	**120.00**

Only the last class has $E_i < 5$ (shaded, $E_7 = 4.13$). We add it to class 6 in the next step.

Example 4: Solution

Class interval	42.5–47.5	47.5–52.5	52.5–57.5	57.5–62.5	62.5–67.5	67.5–77.5	Total
Observed (O_i)	5	16	28	32	23	**16**	120
Expected (E_i)	5.94	14.28	27.47	32.28	24.26	**15.77**	120

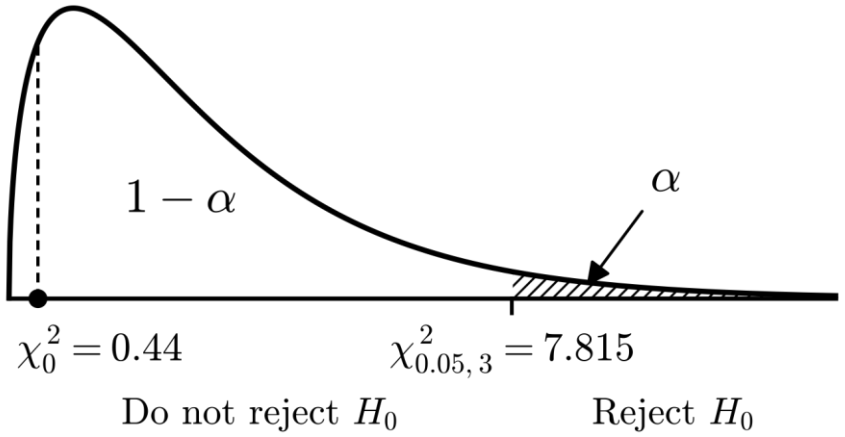
The number of observed values is **k = 6**. The test statistic is

$$\chi_0^2 = \frac{(5-5.94)^2}{5.94} + \frac{(16-14.28)^2}{14.28} + \cdots + \frac{(16-15.77)^2}{15.77} = \mathbf{0.44}$$

The mean and the variance of the normal distribution are not given, so we estimate both from the data (m = 2).

$\nu = k - m - 1 = 6 - 2 - 1 = 3$.

$$\boldsymbol{\chi^2_{0.05,3} = 7.815}$$



Shaded area = rejection region

Decision

Since 0.44 < 7.815, we fail to reject H_0 at the level 0.05, and the data follow N(59.4, 51.9).

4

Test of Independence

Two classification variables measured on one random sample

Test of Independence: The Data

We select one random sample and classify each unit by two variables: A with r levels (rows) and B with c levels (columns). The table of counts is called the **r × c association table**.

A \ B	1	2	...	c	Total
1	O_{11}	O_{12}	...	O_{1c}	$O_{1.}$
2	O_{21}	O_{22}	...	O_{2c}	$O_{2.}$
⋮	⋮	⋮		⋮	⋮
r	O_{r1}	O_{r2}	...	O_{rc}	$O_{r.}$
Total	$\boldsymbol{O_{.1}}$	$\boldsymbol{O_{.2}}$	...	$\boldsymbol{O_{.c}}$	$\boldsymbol{O_{..} = n}$

Examples

- Weight and academic performance
- Ability in Mathematics and in Statistics
- Gender and academic performance
- Smoking habit of a student and of the parents

Row, column and grand totals

$$O_{i.} = \sum_{j=1}^{c} O_{ij}$$

$$O_{.j} = \sum_{i=1}^{r} O_{ij}$$

$$O_{..} = \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij} = n$$

Test of Independence: Test Statistic

Sampling is done **before** classification, so the row totals and the column totals are determined by chance. We test

H_0 : A and B are independent.

H_1 : A and B are not independent.

Expected frequency under H_0

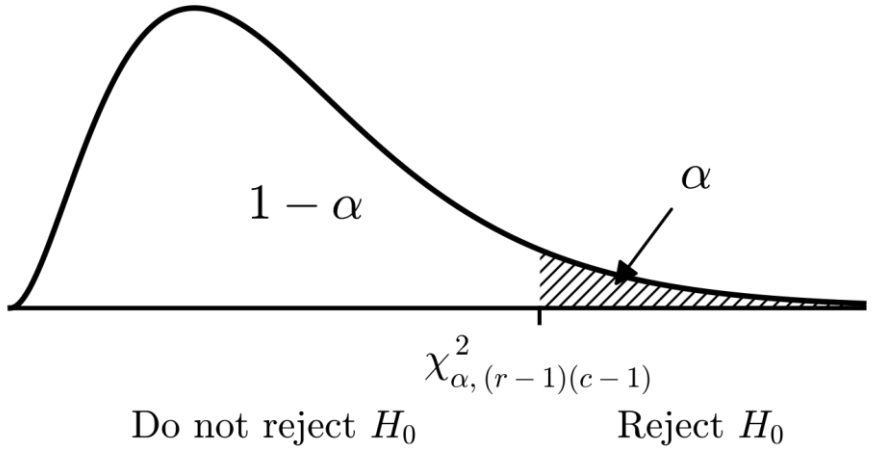
$$E_{ij} = n \times \frac{O_{i\cdot}}{O_{\cdot\cdot}} \times \frac{O_{\cdot j}}{O_{\cdot\cdot}} = \frac{O_{i\cdot} \times O_{\cdot j}}{O_{\cdot\cdot}}$$

(row total × column total) / grand total

Test statistic

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_\nu^2$$

$$\nu = (r-1)(c-1)$$



Shaded area = rejection region

Decision rule

Reject H_0 if

$$\chi_0^2 > \chi_{\alpha,(r-1)(c-1)}^2$$

Every expected frequency E_{ij} should be at least 5 (Eij ≥ 5).

Example 5

In a planned study to investigate the relationship between the **weight** (A) and the **academic level** (B) for the students, in a sample of 120 students we got the following table:

	Excellent	Very Good	Pass	Fail	Total
Skinny	14	11	10	5	**40**
Average	10	16	16	14	**56**
Fat	3	4	7	10	**24**
Total	**27**	**31**	**33**	**29**	**120**

Does this data suggest that there is a relationship between weight (phenomena A) and the academic level (phenomena B)? Use $\alpha = 0.05$.

Example 5: Solution

The null hypothesis is

H_0 : The two random variables (phenomena A and B) are independent.

H_1 : (A) and (B) are not independent.

The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi_\nu^2$$

$$\nu = (r-1)(c-1) = 2\times 3 = 6$$

$$\text{R. R.} = \left(\chi^2_{0.05,6}, \infty\right) = (12.592, \infty)$$

Expected frequencies

Now, we add the expected values

$$E_{ij} = O_{i.} \times O_{.j} / O_{..}$$

E_{ij} = (row total × column total) / n

in the following table for our calculation. For example:

$$E_{11} = \frac{40\times 27}{120} = 9$$

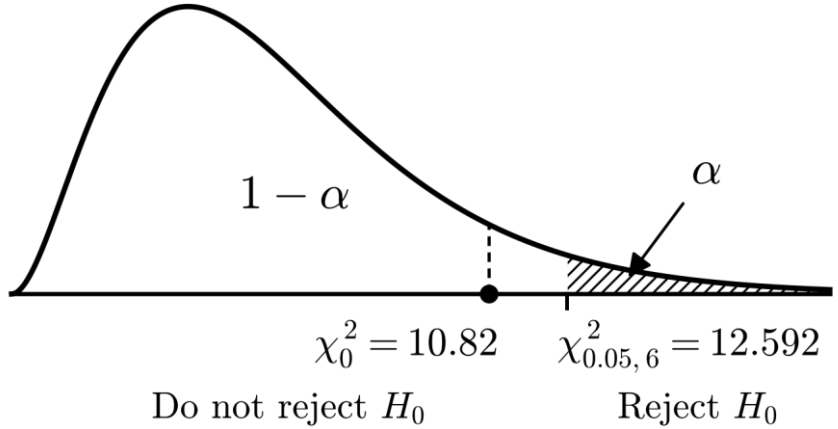
$$E_{12} = \frac{40\times 31}{120} = 10.33$$

All the expected values are shown in the next slide.

Example 5: Solution

Observed values and (expected values):

	Excellent	Very Good	Pass	Fail	Total
Skinny	14 (**9.00**)	11 (**10.33**)	10 (**11.00**)	5 (**9.67**)	**40**
Average	10 (**12.60**)	16 (**14.47**)	16 (**15.40**)	14 (**13.53**)	**56**
Fat	3 (**5.40**)	4 (**6.20**)	7 (**6.60**)	10 (**5.80**)	**24**
Total	**27**	**31**	**33**	**29**	120



Shaded area = rejection region

$$\chi_0^2 = \frac{(14-9)^2}{9} + \frac{(11-10.33)^2}{10.33} + \cdots + \frac{(10-5.8)^2}{5.8} = 10.82$$

Decision

As $\chi_0^2 = 10.82$ does not lie in the R.R. ($10.82 < 12.592$), we cannot reject H_0 given $\alpha = 0.05$. The independence of the phenomena A and B cannot be rejected.

(There is not enough evidence that A and B are related)

5

Test of Homogeneity

Several populations · one classification

Test of Homogeneity

There are **r** populations, and we select an **independent sample from each one**. The sample sizes (row totals) are known in advance. Then each unit is classified into **c** levels of the phenomenon **B**. We test

H_0 : The proportions of the categories of B are the same across all populations. (homogeneity)

H_1 : At least one population has different category proportions.

We use the same expected frequency, test statistic and degrees of freedom as in the independence test. The difference is in how the data are collected:

	Test of independence	Test of homogeneity
Sampling	One random sample from one population	An independent sample from each of r populations
Row totals	Random (determined by chance)	Fixed in advance
Question	Are A and B independent?	Are the r populations alike in B?

Example 6

A researcher was inspecting devices in three different institutes due to a certain ministry in order to check the similarity of them. He selected three samples to study their quality (Low, Average and High), and the following table shows the results obtained:

Institute (A)	Low	Average	High	Total
I	12	23	89	**124**
II	8	12	62	**82**
III	25	30	119	**174**
Total	**45**	**65**	**270**	**380**

Does this data give an evidence to say the qualities of the devices in the three institutes are the same? Use $\alpha = 0.01$.

Example 6: Solution

Note that this is an application for testing whether the two phenomena (A and B) are homogeneous. We are testing the null hypothesis:

H_0 : The proportions of the quality categories are the same across the three institutes " populations".

H_1 : At least one institute "population" has different quality category proportions.

The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_\nu^2$$

$$\nu = (r-1)(c-1) = 2\times 2 = 4$$

$$\text{R. R.} = \left(\chi^2_{0.01,4}, \infty\right) = (13.277, \infty)$$

Expected frequencies

Now, we add the expected values

$$E_{ij} = O_{i.} \times O_{.j} / O_{..}$$

E_{ij} = (row total × column total) / n.

in the following table for our calculation. For example:

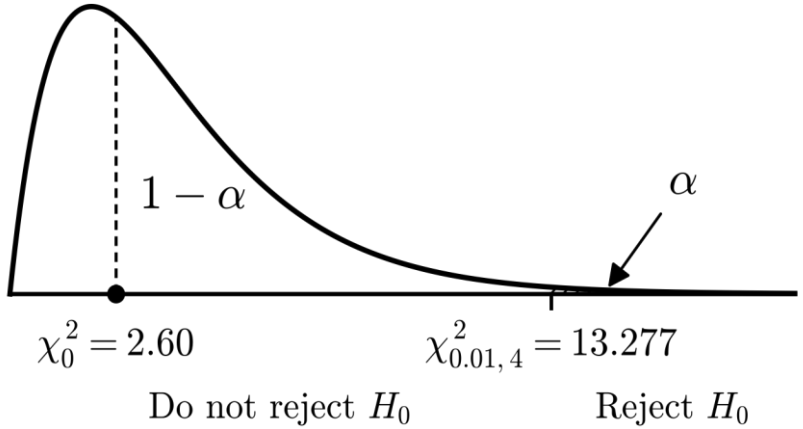
$$E_{11} = \frac{124\times 45}{380} = 14.68$$

$$E_{12} = \frac{124\times 65}{380} = 21.21$$

Example 6: Solution

Observed values and (**expected values**):

Institute	Low	Average	High	Total
I	12 (**14.68**)	23 (**21.21**)	89 (**88.11**)	**124**
II	8 (**9.71**)	12 (**14.03**)	62 (**58.26**)	**82**
III	25 (**20.61**)	30 (**29.76**)	119 (**123.63**)	**174**
Total	**45**	**65**	**270**	**380**



Shaded area = rejection region

$$\chi_0^2 = \frac{(12-14.68)^2}{14.68} + \frac{(23-21.21)^2}{21.21} + \cdots + \frac{(119-123.63)^2}{123.63} = 2.60$$

Decision

$\chi_0^2 = 2.60$ does not lie in the R.R. ($2.60 < 13.277$), so we cannot reject H_0. The quality of the devices is the same in the three institutes at $\alpha = 0.01$.

6

Summary of the Tests

Goodness of fit · independence · homogeneity

Summary of the Chi-Square Tests

	Goodness-of-fit	Independence	Homogeneity
Question	Does the data follow a given distribution?	Are A and B independent?	Are r populations alike in B?
Data	One sample, k classes	One sample, r × c table	r independent samples, r × c table
Expected frequency	$E_i = np_i$	E_{ij} = (row total × column total) / n	E_{ij} = (row total × column total) / n
Degrees of freedom	$\nu = k - m - 1$	$\nu = (r-1)(c-1)$	$\nu = (r-1)(c-1)$
Test statistic	$\Sigma\,(O_i - E_i)^2 / E_i$	$\Sigma\,\Sigma\,(O_{ij} - E_{ij})^2 / E_{ij}$	$\Sigma\,\Sigma\,(O_{ij} - E_{ij})^2 / E_{ij}$
Reject H_0 if	$\chi_0^2 > \chi^2_{\alpha,\nu}$	$\chi_0^2 > \chi^2_{\alpha,\nu}$	$\chi_0^2 > \chi^2_{\alpha,\nu}$

Always check

Every expected frequency must be at least 5. Combine neighbor classes when it is not.