

# Chi-Square Test

Prepared by Dr. Hesham Mousli

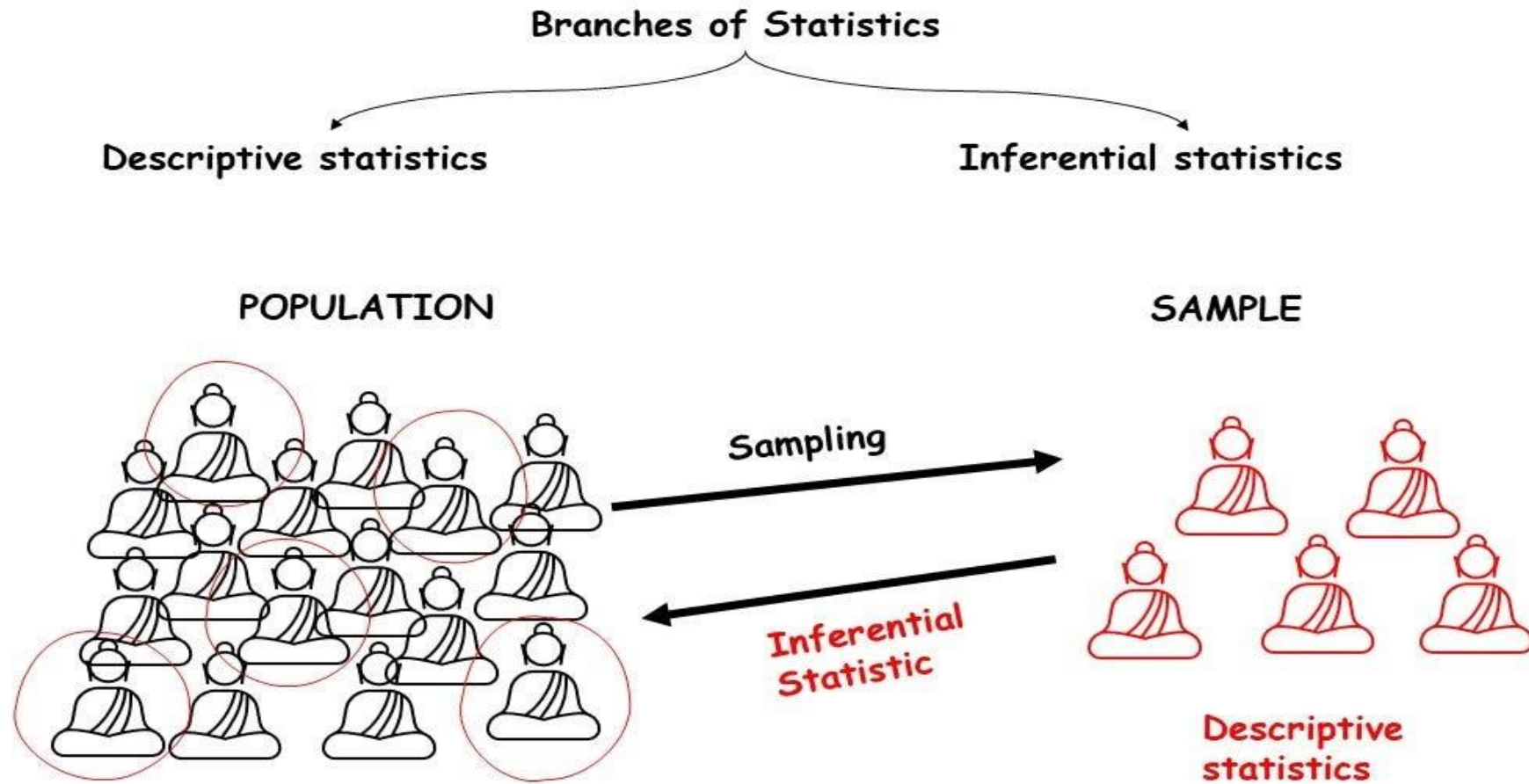
MHA 613  
Statistics in Healthcare

**Lecture 6**



## INTENDED OBJECTIVES

1. Understand the purpose of the Chi-square test and its applications in healthcare management.
2. Learn the step-by-step process of conducting a Chi-square test.
3. Identify and apply the assumptions required for a valid Chi-square test.
4. Develop skills to accurately interpret and communicate the results of a Chi-square test.



## Research Question

- In a rural healthcare setting, health officials are investigating whether there is an association between vaccination status and the incidence of a particular infectious disease. Data is collected from individuals, noting **whether they have been vaccinated** and **whether they contracted the disease** within the past year.

## Research Question

Is there a significant association between **vaccination status** and **disease status**?



## Research Question

**Does Vaccination status have  
an impact on Contracting the  
disease?**



## Hypotheses Statements

- A. Null Hypothesis:** There is no significant association between vaccination status and the incidence of the disease.
- B. Alternative Hypothesis:** There is a significant association between vaccination status and the incidence of the disease.

# Comparing two or more proportions (Categorical data) (Chi-Square test) ( $\chi^2$ )

Chapter

06

- A Pearson's **Chi-Square test** is a statistical hypothesis test for comparing the proportions of two categorical variables.
- It allows you to test whether the **two categorical variables** compared are related to each other (**dependent**) or they are unrelated (**independent**).

## Chi-Square test

- The chi-square test is a statistical method used to determine if there is a significant **association** between **two categorical variables**.
- It assesses whether the **observed frequencies** in a **contingency table** match the **expected frequencies** if there is no relationship between the variables.

## Observed

|   | A      | B      |
|---|--------|--------|
| C | N = 50 | N = 80 |
| D | N = 50 |        |
|   |        | N = 20 |

?

## Expected

|   | A      | B      |
|---|--------|--------|
| C | N = 65 | N = 65 |
| D | N = 35 | N = 35 |

## Chi-Square test steps summary

1. The data are obtained, initially, as **frequencies**, i.e. the numbers with and without the characteristic in each group.
2. A table in which the entries are frequencies is called a **contingency table**; when this table has two rows and two columns it is called a  **$2 \times 2$  table**.

## Chi-Square test steps summary

3. The **contingency table** shows the **observed frequencies** in the four cells corresponding to **each row/column combination**, the four **marginal totals** (the frequency in a specific row or column, and the **overall total (Grand total) =  $n$** .
4. We can calculate the frequency that we would **expect** in each of the four cells of the table if  $H_0$  was true (**the expected frequencies**).

Chapter  
**06**

## Chi-Square test steps summary

| Contingency Table     |     |      |     |             |
|-----------------------|-----|------|-----|-------------|
|                       | Boy | Girl | Sum | Row total   |
| like Snickers         | 43  | 30   | 73  |             |
| doesn't like Snickers | 8   | 19   | 27  |             |
| Sum                   | 51  | 49   | 100 | Grand total |
| Column total          |     |      |     |             |

## Chi-Square test steps summary

5. Then we go on to calculate the **Chi-Square test statistic** as well as the **resulting test P-value**.
6. Compare the **resulting P-value** versus **the value of 0.05** to decide which hypothesis to **accept** and which one to **reject**.

|              | Vaccinated | Not vaccinated | Total |
|--------------|------------|----------------|-------|
| Diseased     | 36         | 12             | 48    |
| Not diseased | 4          | 8              | 12    |
| Total        | 40         | 20             | 60    |



**Observed values**

**Vs.**

|              | Vaccinated                     | Not vaccinated                 |
|--------------|--------------------------------|--------------------------------|
| Diseased     | $\frac{48 \times 40}{60} = 32$ | $\frac{48 \times 20}{60} = 16$ |
| Not diseased | $\frac{12 \times 40}{60} = 8$  | $\frac{12 \times 20}{60} = 4$  |



**Expected values**

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

$p < 0.05 \rightarrow$  statistically  
significant

Association between the two categorical  
variables

$p > 0.05 \rightarrow$  no statistically  
significant

Association between the two categorical  
variables

## Chi-Square test Assumptions

1. The **rows (and columns)** of the table are **mutually exclusive**, implying that **each individual can belong to only one row and only one column**.
2. The usual, **although being conservative**, approach **requires** that the **expected frequency** in each **of the four cells** is **at least five**.

Estrogen supplementation to delay or prevent the onset of Alzheimer's disease in postmenopausal women.

|                   |     | Alzheimer's onset during 5-year period |     |       |
|-------------------|-----|----------------------------------------|-----|-------|
|                   |     | No                                     | Yes |       |
| received estrogen | Yes | 147                                    | 9   | 156   |
|                   | No  | 810                                    | 158 | 968   |
|                   |     | 957                                    | 167 | 1,124 |

**The null hypothesis ( $H_0$ ):** Estrogen supplementation in postmenopausal women is unrelated to Alzheimer's onset.

**The alternate hypothesis ( $H_A$ ):** Estrogen supplementation in postmenopausal women delays/prevents Alzheimer's onset.

$p < 0.05 \rightarrow$  statistically  
significant

Association between the two categorical  
variables

$p > 0.05 \rightarrow$  no statistically  
significant

Association between the two categorical  
variables

**How can I apply  
Chi-Square ( $\chi^2$ )  
Practically?**



## Case Scenario (1)

- A hospital wants to determine if there is a difference in patient satisfaction between **telemedicine** and **in-person** consultations. The responses were collected in the form of **satisfied/not satisfied** responses.

## Case Scenario (1)

- **Null Hypothesis (H0):** There is no association between the type of consultation (**telemedicine vs. in-person**) and patient satisfaction levels.
- **Alternative Hypothesis (H1):** There is an association between the type of consultation and patient satisfaction levels.

## Chi-Square test steps summary

|               | Telemedicine<br>consultation | In-person<br>consultation |
|---------------|------------------------------|---------------------------|
| Satisfied     | 54                           | 20                        |
| Not satisfied | 10                           | 66                        |

|    | Satisfaction status | Consultation type |
|----|---------------------|-------------------|
| 1  | Satisfied           | In person         |
| 2  | Satisfied           | In person         |
| 3  | Satisfied           | In person         |
| 4  | Satisfied           | In person         |
| 5  | Satisfied           | In person         |
| 6  | Not satisfied       | In person         |
| 7  | Not satisfied       | In person         |
| 8  | Not satisfied       | In person         |
| 9  | Not satisfied       | In person         |
| 10 | Not satisfied       | In person         |
| 11 | Not satisfied       | In person         |
| 12 | Not satisfied       | In person         |
| 13 | Not satisfied       | In person         |
| 14 | Not satisfied       | In person         |
| 15 | Satisfied           | In person         |
| 16 | Satisfied           | In person         |
| 17 | Satisfied           | In person         |

JASP\*

Menu: Edit Data | Descriptives | T-Tests | ANOVA | Mixed Models | Regression | Frequencies | **Contingency Tables**

|    | Satisfaction status | Consultation type |
|----|---------------------|-------------------|
| 1  | Satisfied           | In person         |
| 2  | Satisfied           | In person         |
| 3  | Satisfied           | In person         |
| 4  | Satisfied           | In person         |
| 5  | Satisfied           | In person         |
| 6  | Not satisfied       | In person         |
| 7  | Not satisfied       | In person         |
| 8  | Not satisfied       | In person         |
| 9  | Not satisfied       | In person         |
| 10 | Not satisfied       | In person         |
| 11 | Not satisfied       | In person         |

**Classical**

- Binomial Test
- Multinomial Test
- Contingency Tables**
- Log-Linear Regression

**Bayesian**

- Binomial Test
- A/B Test
- Multinomial Test
- Informed Multinomial Test
- Informed Multi-Binomial Test
- Contingency Tables
- Log-Linear Regression

The screenshot shows a statistical software interface with a central panel and a right-hand sidebar. The central panel has four sections: Rows, Columns, Counts, and Layers, each with a play button icon. The Rows section contains 'Satisfaction status' with a blue arrow pointing to it. The Columns section contains 'Consultation type' with a blue arrow pointing to it. The Counts section is empty. The Layers section is empty. Below these sections is a 'Statistics' dropdown menu. A blue arrow points to the 'Statistics' section. The 'Statistics' section is expanded, showing several options: 

- ☒  $\chi^2$
- ☐  $\chi^2$  continuity correction
- ☐ Likelihood ratio
- ☐ Vovk-Sellke maximum p-ratio
- ☐ Odds ratio (2x2 only)
- ☒ Log Odds Ratio
- Confidence interval  %
- Alt. Hypothesis (Fisher's exact test)
- ☐ Group one  $\neq$  Group two

## Results ▼

### Contingency Tables ▼

#### Contingency Tables ▼

| Satisfaction status | Consultation type |               | Total |
|---------------------|-------------------|---------------|-------|
|                     | Satisfied         | Not satisfied |       |
| Telemedicine        | 54                | 10            | 64    |
| In person           | 20                | 66            | 86    |
| Total               | 74                | 76            | 150   |

#### Chi-Squared Tests

|                | Value  | df | p      |
|----------------|--------|----|--------|
| X <sup>2</sup> | 54.838 | 1  | < .001 |
| N              | 150    |    |        |

## Case Scenario (1)

- Since  $p = 0.01 < 0.05$ , we reject the null hypothesis ( $H_0$ ).
- There is sufficient evidence to conclude that there is a statistically significant association between the type of consultation (**telemedicine vs. in-person**) and patient satisfaction levels.

## Case Scenario (2)

- A healthcare facility assesses whether compliance with infection control protocols differs among staff in different departments (**e.g., Surgery, ICU, General Wards**).

## Case Scenario (2)

- **Null Hypothesis (H0):** There is no significant association between compliance with infection control protocols and the department.
- **Alternative Hypothesis (H1):** There is a significant association between compliance with infection control protocols and the department.

## Chi-Square test steps summary

|                      | Compliant | Not compliant |
|----------------------|-----------|---------------|
| <b>Surgery</b>       | 11        | 15            |
| <b>ICU</b>           | 25        | 23            |
| <b>General Wards</b> | 27        | 24            |



|  Department |  Compliance status |
|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |
| Surgery                                                                                      | Compliant                                                                                             |



Edit Data

Descriptives

T-Tests

ANOVA

Mixed Models

Regression

Frequencies

Factor

| Department | Compliance status |  |
|------------|-------------------|--|
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |
| Surgery    | Compliant         |  |

**Classical**

Binomial Test

Multinomial Test

Contingency Tables

Log-Linear Regression

**Bayesian**

Binomial Test

A/B Test

Multinomial Test

Informed Multinomial Test

Informed Multi-Binomial Test

Contingency Tables

Log-Linear Regression

## Contingency Tables

Rows

Department

Columns

Compliance status

Counts

Layers

Statistics

☒  $\chi^2$

☐ Odds ratio (2x2 only)

## Contingency Tables

*Contingency Tables* ←

| Department    | Compliance status |               | Total |
|---------------|-------------------|---------------|-------|
|               | Compliant         | Not compliant |       |
| Surgery       | 11                | 15            | 26    |
| ICU           | 25                | 23            | 48    |
| General Wards | 27                | 24            | 51    |
| Total         | 63                | 62            | 125   |

*Chi-Squared Tests* ←

|                | Value | df | p     |
|----------------|-------|----|-------|
| X <sup>2</sup> | 0.867 | 2  | 0.648 |
| N              | 125   |    |       |

## Case Scenario (2)

- Since  $p = 0.648 > 0.05$ , we reject the alternative hypothesis ( $H_1$ ).
- There is insufficient evidence to conclude that compliance with infection control protocols differs among staff in different departments. Compliance appears to be independent of the department.

## TO WRAP UP

1. The Chi-square test is a statistical method used to determine if there is a significant association between categorical variables.
2. Data for a Chi-square test is organized into a contingency table, which displays observed frequencies in each cell for the combinations of categories.
3. This resulting P-value is then compared to a significance level, commonly 0.05, to determine whether to reject or retain the null hypothesis.
4. The Chi-square test requires **mutually exclusive categories** & that the **expected frequencies in each cell should generally be at least five**.



# Thank You