SLIIT UNI
THE KNOWLEDGE UNIVERSITY

Sri Lanka Institute of Information Technology

Faculty of Computing

IT2120 - Probability and Statistics

Ms. K. G. M. Lakmali

Year 02 and Semester 01

Lecture 10

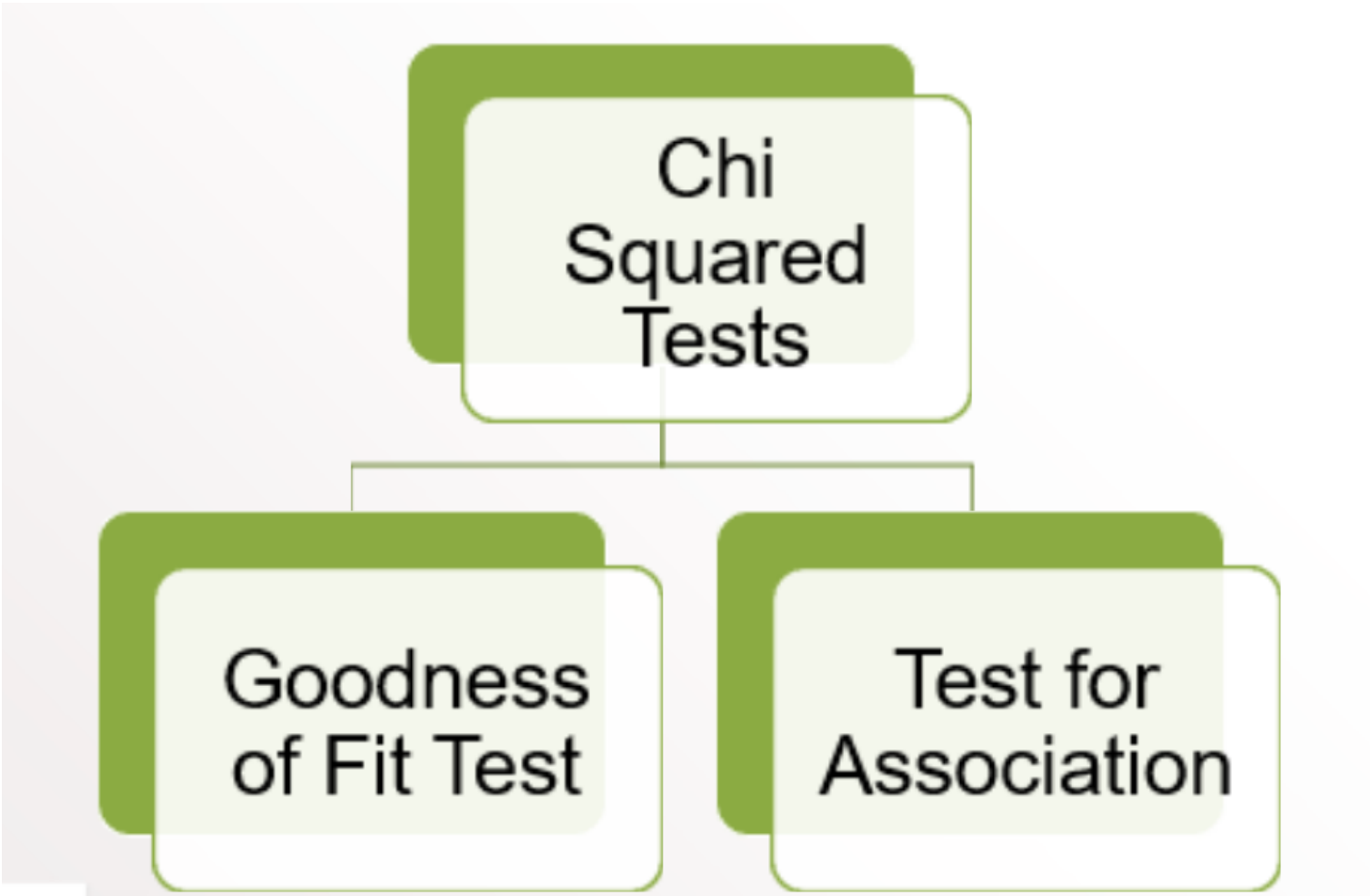
CHI SQUARED TESTS

Chi-squared tests

Chi-squared tests are used for,

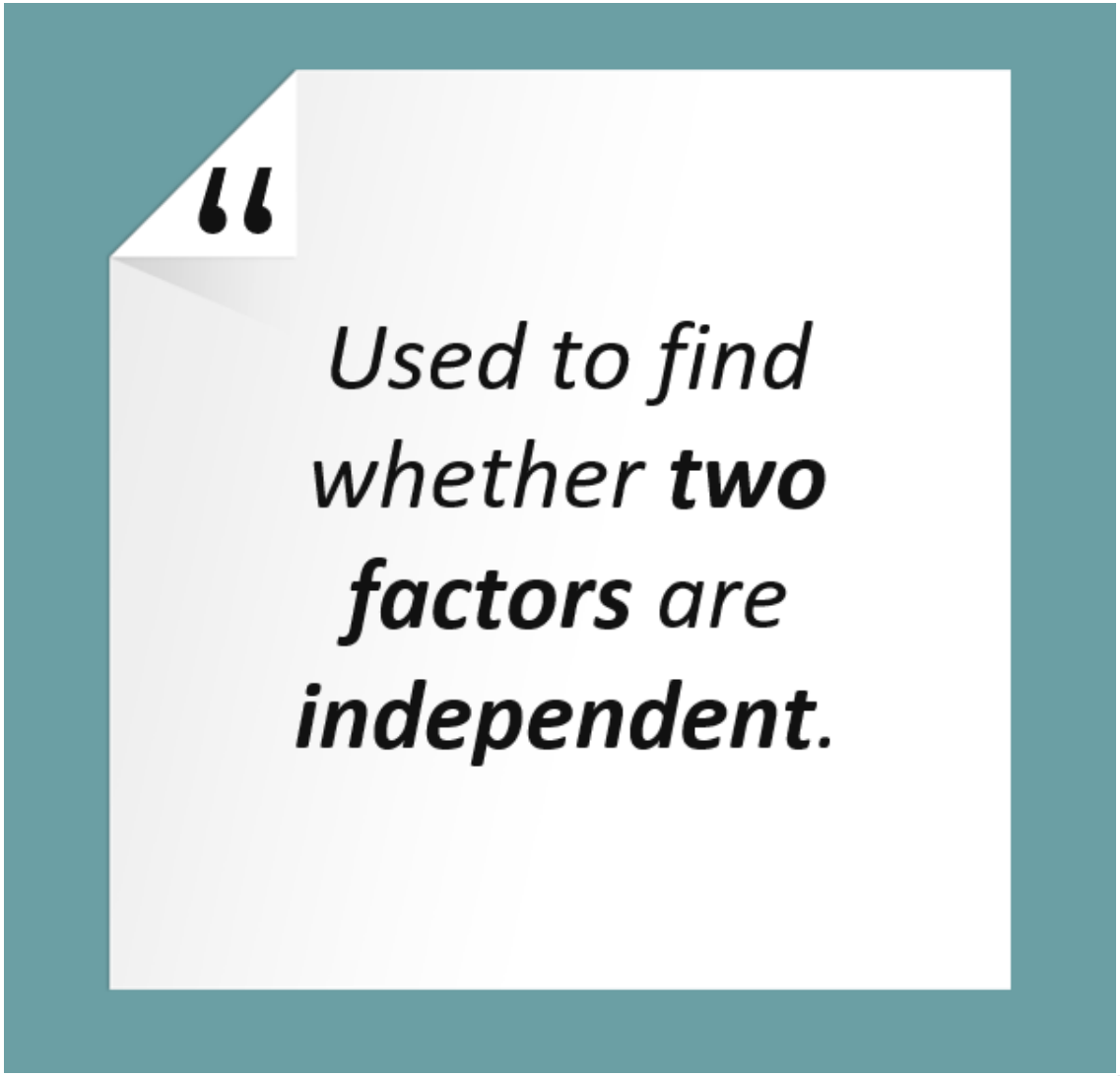
- Discrete data
- Categorical data

Chi Squared Tests



Test for Association

Test for Association



- The hypothesis for the test is,

$$\boldsymbol{H_0}: \textit{The factors are independent.}$$
$$\boldsymbol{H_1}: \textit{The factors are not independent.}$$

- Test Statistic,
  Under $\mathbf{H_0}$,

$$X^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim X^2_{d.f.}$$

- O_{ij} - Observed frequency for cell ij
- E_{ij} - Expected frequency for cell ij
- $df = (No\ of\ \ rows - 1)(No\ of\ \ columns - 1)$

- Reject $\mathbf{H_0}$, if $X^2_{cal} > X^2_{df,\alpha\%}$ (critical value)
- Test:
  - Find the **expected frequencies** for each cell.
  - Calculate **test statistic value**.
- Conclusion:
  Compare calculated **test statistic value** with **critical value** and give the conclusion.

Important

Rule 01

- **All expected counts** should be **greater than 5**.

Rule 02

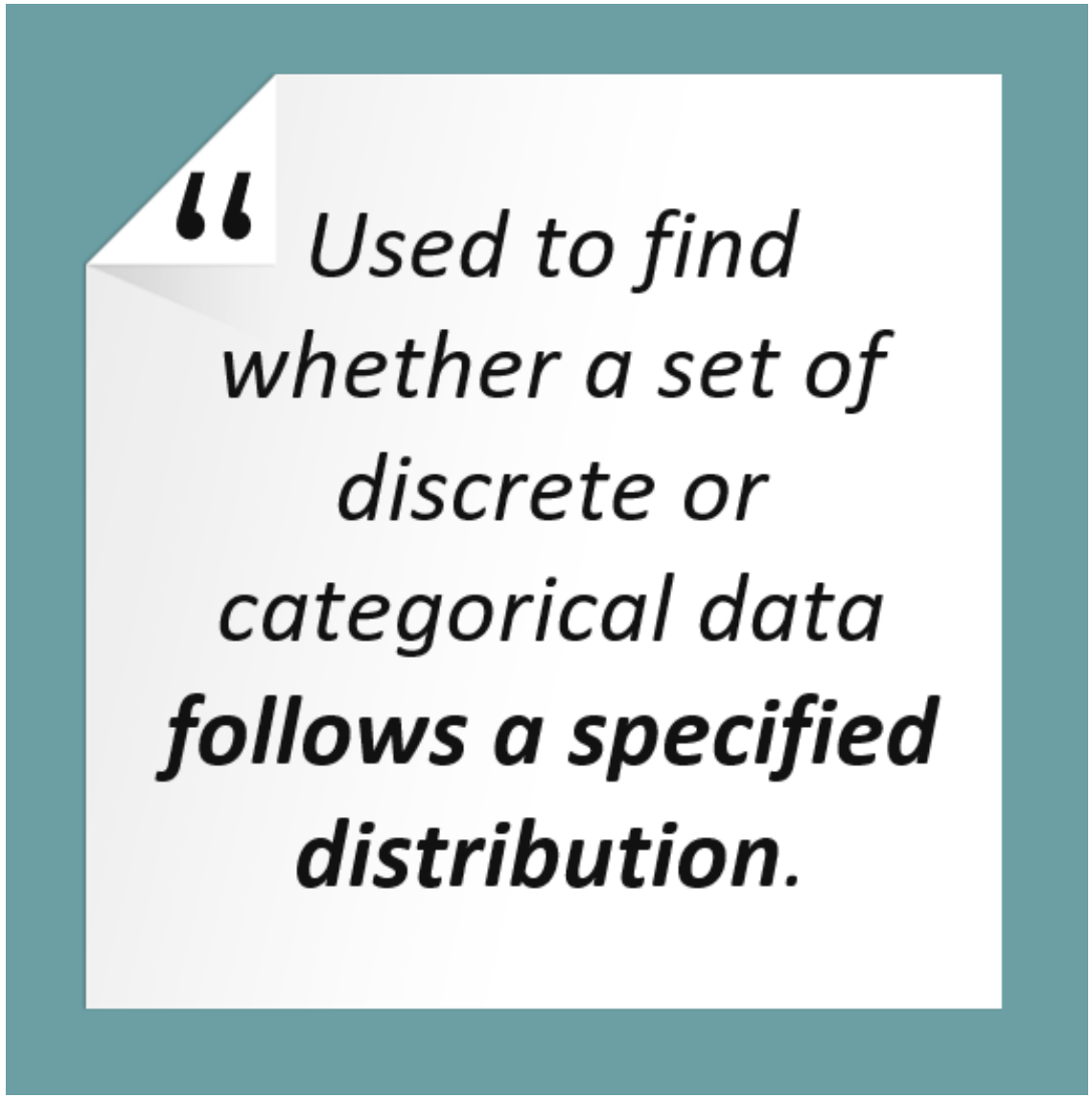
- **All expected counts** should be **greater than 1 & at least 80% of the cells** should have **expected count** which is **greater than or equal to 5.**
- If not, categories can be joined.

Example

The following table gives a classification according to religious affiliation and marital status for 500 randomly selected individuals. For $\alpha = 1\%$, test the null hypothesis that marital status and religious affiliation are independent.

		Religious Affiliation				
		A	B	C	D	None
Marital Status	Single	39	19	12	28	18
	Married	172	61	44	70	37

Goodness of Fit Test


“Used to find whether a set of discrete or categorical data follows a specified distribution.

- The hypothesis for the test is,

H_0: *The data are consistent with the specified distribution.*

H_1: *At least one category deviates from the specified distribution.*

- Test Statistic,
  Under $\mathbf{H_0}$,

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim X^2_{d.f.}$$

- O_i - Observed frequency for cell i
- E_i - Expected frequency for cell i

- **d.f.** = No of classes – No of parameters estimated - 1
- **Reject H_0**, if $X^2_{cal} > X^2_{df,\alpha\%}$ (critical value)
- Test:
  - Find the **expected frequencies** for each category.
  - Calculate **test statistic value**.
- Conclusion:
  Compare calculated **test statistic value** with **critical value** and give the conclusion.

Example

1) A die is rolled 60 times and the face values are recorded. The results are as follows.

Up Face	1	2	3	4	5	6
Frequency	8	11	5	12	15	9

Is the die balanced? Test using $\alpha = 0.05$.

Example

2) The number of accidents in a month observed over a period of 10 years is given below.

No of accidents	0	1	2	3	4	5	6	≥7
Frequency	41	40	22	10	6	0	1	0

Is the data following a Poisson distribution? Test using $\alpha = 0.05$.

Example

3) The grades of students in a class of 51 are given in the following table. Test the hypothesis that the grades are normally distributed with a mean of 75 and a standard deviation of 8. Use $\alpha = 0.05$.

Range	0-59.5	59.5-69.5	69.5-79.5	79.5-89.5	89.5-100
No of students	8	11	5	12	15

Thanks!

Any questions?