STATISTICS FOR DATA SCIENCE

HYPOTHESIS and INFERENCE

Dr. Deepa Nair
Department of Science and Humanities

STATISTICS FOR DATA SCIENCE

UNIT-4 HYPOTHESIS and INFERENCE

Session-9

Chi-squared Test

Dr. Deepa Nair
Department of Science and Humanities

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- There was a tender for collecting tool for a newly opened bridge. To- the minimum rates mentioned in the tender ,following is the data given to them.Before bidding for the contract a firm wanted to check how correct or accurate the data is .

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
No	170	20	90	130	200	170	220

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- So they mentioned the traffic for a week and collected the related data.

Day	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
No	190	50	100	130	200	150	200

- You all understand that the vehicle movement on any road is variable and depending on the day time, month season and many other factors.

- So the decisions that we need to make whether the data given by the authorities are reliable or not very easy to it first by looking the data.so we use a statistical tool called λ^2 –test.

STATISTICS FOR DATA SCIENCE

Chi-squared Test

For example:

- A gambler wants to test a die to see whether it deviates from fairness.

- Let p_i be the probability that the number i comes up. The null hypothesis will state that the die is fair.

- The null hypothesis is $H_0 : p_1 = p_{01}, p_2 = p_{02}, \ldots p_6 = p_{06} = 1/6.$

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- A generalization of the Bernoulli trial is the multinomial trial
- Which is an experiment that can result in any one of k outcomes, where $k \geq 2$.
- The probabilities of the k outcomes are denoted $p_1, \ldots, p_k$.
- In this section, we generalize the tests for a Bernoulli probability to multinomial trials.
- The null hypothesis has the form

$$H_0 : p_1 = p_{01}, p_2 = p_{02}, \ldots p_k = p_{0k}.$$

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- The gambler rolls the die 600 times and The results obtained are called the observed values.
- To test the null hypothesis, we construct a second column, labeled “Expected.” This column contains the expected values.
- The expected value for a given outcome is the mean number of trials that would result in that outcome if H_0 were true.

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- The idea behind the hypothesis test is that if H_0 is true, then the observed and expected values are likely to be close to each other.

- Therefore we will construct a test statistic that measures the closeness of the observed to the expected values.

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- The statistic is called the chi-square statistic. To define it, let k be the number of outcomes ($k = 6$ in the die example),
- Let O_i and E_i be the observed and expected numbers of trials, respectively, that result in outcome i .
- The chi-square statistic is

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

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

- When the expected values are all sufficiently large, a good approximation is available.
- It is called the chi-square distribution with $k - 1$ degrees of freedom, denoted $\chi 2_{k-1}$
- A table for the chi-square distribution is available

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

Example:

The gambler rolls the die 600 times and The results obtained are as shown in the table:

Source: Text Book Chapter6

STATISTICS FOR DATA SCIENCE

Chi-squared Test

Example:

Consider the following table

Catogory	Observed	Expected
1	115	100
2	97	100
3	91	100
4	101	100
5	110	100
6	86	100
Tot	600	600

$$\chi^2 = \frac{(115-100)^2}{100} + \cdots \frac{(86-100)^2}{100}$$

$$= 2.25 + \cdots + 1.96$$

$$= 6.12$$

The upper 10% point is 9.236. We conclude that $P > 0.10$. There is no evidence to suggest that the die is not fair.

Dr. Deepa Nair

Department of Science and Humanities

deepanair@pes.edu