

15CSE481 Machine Learning and Data Mining Lab

Lab 4- Aug 12- 2020

Practice Chi-square statistical test

- [Go through the manual:](#)
- Practice the coding
 - Using implementation from scratch
 - Using Scipy package
 - Compare the results

Evaluation Problem

- Download dataset: [titanic.csv](#)
- Preprocess the data

Write programs for the following

1. Count the total number of passengers on board the Titanic (2 Mark)
2. Count the number of passengers who survived the sinking of the Titanic. (2 Mark)
3. Measure the percentage of passengers who survived the sinking of the Titanic. (2 Mark)
4. Count the number of passengers based on gender (2 Mark)
5. Run a Pearson's Chi-squared test, **without using Scipy function**, to test the following hypothesis: (4 mark)
Hypothesis: The proportion of females onboard who survived the sinking of the Titanic was higher than the proportion of males onboard who survived the sinking of the Titanic.
6. What is the inference based on the test? (1 mark)
7. Repeat questions 5 and 6 using Scipy function. (2 marks)

After solving the questions, submit both ipynb file and PDF.

Deadline: 12/08/2020, 5:00 PM

