FUNDAMENTALS OF DATA SCIENCE

Coding assignment - Semester B 2023/24 (30 points, 30%)

Student Name: Ruthvik Adimulapu

Student ID Number: 23032641

Description Of Datasets:

The datasets provided consists of exam grade distributions for the years 2020 and 2024.The 2020 dataset includes tabulated distributions of student grades in a module's exam, where each row represents an interval(bin) of grades along with the corresponding number of students who received grades within that interval. On the other hand, the 2024 dataset comprises individual student grades, approximately 300 entries in total. The first dataset contains of 3 columns in which the third columns are counts (i.e. student numbers). And, the second dataset is a one column CSV file with individual student grades.

Formula Used for Calculations:

For the calculation of mean values and standard deviations, the following formulas were employed:

For the 2020 distribution:

Weighted Mean:

$$\bar{x} = \frac{\sum (x_i \times w_i)}{\sum w_i}$$

Weighted Standard Deviation:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2 \times w_i}{\sum w_i}}$$

After creating the distribution for the 2020 exam grades the mean value and standard deviation of the distributions are:

The Mean value of the students in 2020 is 60.39

The Standard deviation of 2020 is 15.54

For the 2024 distribution:

- Mean:

$$\bar{x} = \frac{\sum x_i}{n}$$

- Standard Deviation SD:

$$SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

- The Mean value 2024
  After creating the distribution for the 2024 grades of the students, the mean value and standard deviation of the distributions are 56.73 and standard deviation of 2024 is 12.84

Discussion of V value:

The value ‘V’ signifies the proportion of students who achieved a grade of 70 or higher in the 2024 exam “0.14”.

V=It is obtained by dividing the

Number of students with grade >=70 / Total number of students in 2024 exam

Discussion of Differences between 2024 and 2020 Distributions:

Upon comparing the mean and standard deviation values calculated for the 2024 and 2020 distributions ,significant differences emerge. The variation suggest discrepancies in student performances or exam difficulty between the two years.

The Plot Produced by the Code:

The generated plot depicts the distribution of the exam grades for the years 2020 ad 2024.It includes histograms with clearly labeled axes ,titles ,and and legeds. Additionally, markers denoting the mean values, standard deviations, and V Value enhance the interpretability of the graph, facilitating a visual comparision of the two distributions

Conclusion:

The analysis conducted provides a comprehensive understanding of the exam grade distributions for the years 2020 and 2024.By calculating key statistics and metrics ad visualizing the distributions through histograms, the report sheds light on the difference between the two years and offers valuable insights into student performance and exam trends.