Small Project

Student ID: 23031641

Dataset Description:

The objective of this study is to perform a comparative assessment of student examination scores for the years 2020 and 2024. For this purpose, we have been provided with two datasets. The first dataset, *2020input1.csv*, is a grouped data of 300 students' results in the module's 2020 exam. The second dataset filename *2024input1.csv*, is ungrouped CSV file with 296 individual student grades obtained in 2024 exam.

Statistical Formulae:

To compute the mean for the 2020 and 2024 exams based on grouped and ungrouped data, respectively, we may utilize the subsequent formulas:

$$\mu_{2020} = Mean\ (2020\ Exam) = \frac{\sum f\,.\,m}{\sum f}$$

where m is the midpoint of the interval and f is the frequency of the interval.

$$\mu_{2024} = Mean\ (2024\ Exam) = \frac{\sum_{k=1}^{N} x_k}{N}$$

where $x_k\ (k = 1, 2, \cdots, N)$ are the grades of students and N is the number of individuals counted in the data set.

The standard deviation (SD) can be calculated for 2020 and 2024 exam data with the help of the following formulas respectively:

$$\sigma_{2020} = SD\ (2020\ Exam) = \sqrt{\frac{\sum f\left(m - \mu_{2020}\right)^2}{\sum f}} \text{ and } \sigma_{2024} = SD\ (2024\ Exam) = \sqrt{\frac{\sum_{k=1}^{N}\left(x_k - \mu_{2024}\right)^2}{N}}.$$

Using the above formulas and running the Python code, we get the following values with three significant figure.

Table 1: 2020 and 2024 Exams (Mean and Standard Deviation)

Exam Year	Mean	SD	Coefficient of Variation (C.V)
2020 Exam	$\mu_{2020} = 62.4$	$\sigma_{2020} = 15.5$	$C.V_{2020} = \frac{\sigma_{2020}}{\mu_{2020}} \times 100\% \approx 24.8\%$
2024 Exam	$\mu_{2024} = 56.7$	$\sigma_{2024} = 12.8$	$C.V_{2024} = \frac{\sigma_{2024}}{\mu_{2024}} \times 100\% \approx 22.6\%$

Assuming that these data sets are obtained from the same standard class, if the average grade for the 2020 tests (62.4) is greater than the average grade for the 2024 exams (56.7), it implies that students obtained better results in the 2020 examinations. On the other hand, the lower value of $\sigma_{2024} = 12.8$ and $C.V_{2024} \approx 22.6\%$ for 2024 exam indicates consistent performance as compared to $\sigma_{2020} = 15.5$ and $C.V_{2020} \approx 24.8\%$ for the 2020 exam. Thus, in terms of comparative variability, the second dataset might be deemed superior.

Discussion of V Value:

We have to calculate the V which is the proportion of students with grade of 70 or higher in 2024 exam as:

$$V_{2024} = \frac{Number\ of\ students\ with\ grades\ greater\ than\ or\ equal\ to\ 70\ in\ 2024\ exam}{Total\ number\ of\ students\ in\ 2024\ Exam} = \frac{40}{296} = 0.135$$

The data from the 2024 exam reveals that just a small proportion 13.5% of the students obtained scores of 70 or better. The variation in exam scores can be attributed to factors such as the

complexity of the exams, the student's academic achievement, and other variables like the grading standards etc.

Plot (Histogram):

The following histogram plot has been obtained from running Python code, and we have tested the code using Spyder.

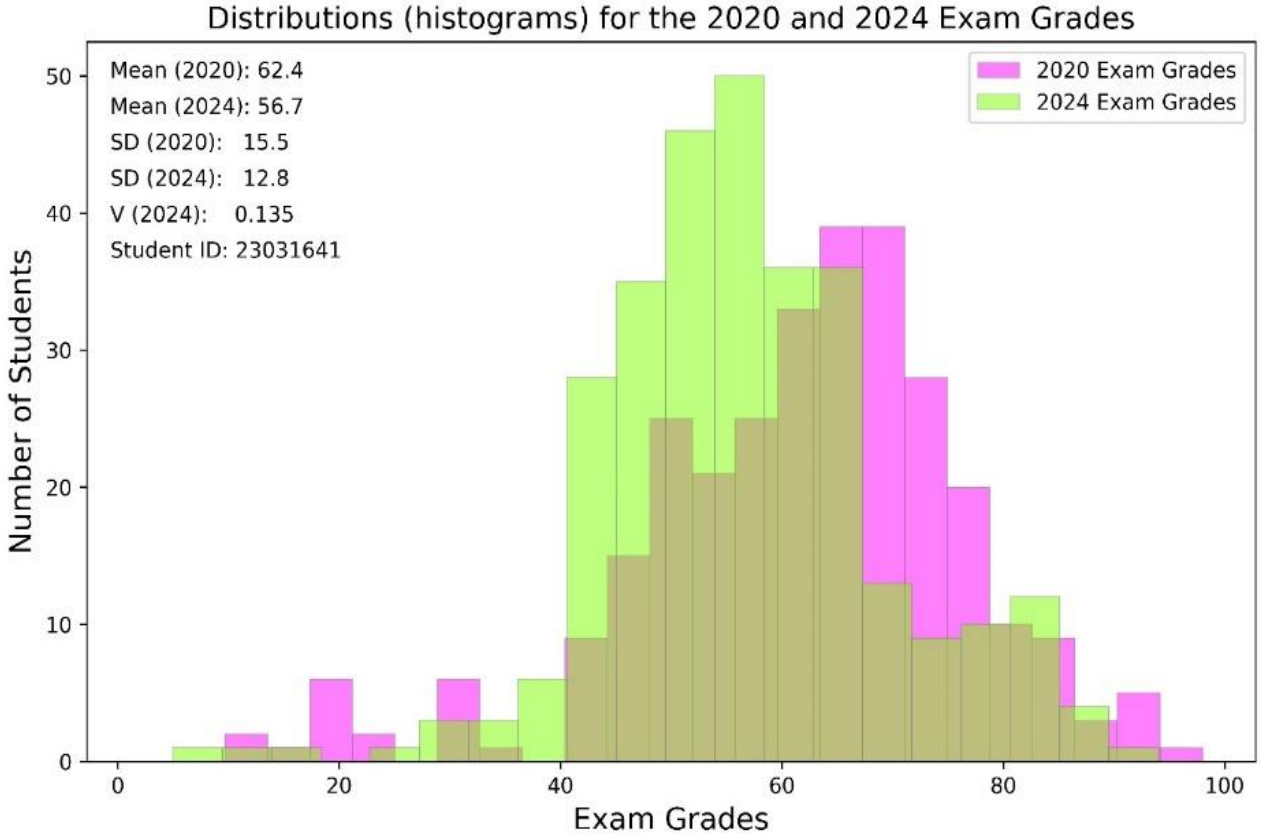

Figure 1: Histogram Plot of 2020 and 2024 Exams Grades