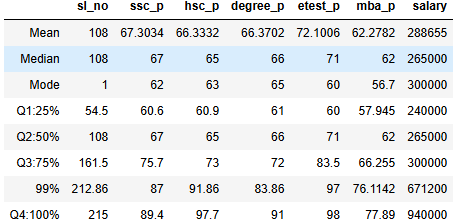
Report on Central Tendency and Percentiles



# 1. Introduction

This report summarizes the mean, median, mode, and percentiles (Q1, Q2, Q3, Q4) of a placement dataset. The dataset includes academic percentages (ssc\_p, hsc\_p, degree\_p, etest\_p, mba\_p) and salary.

# 2. Measures of Central Tendency

• Mean (Average): The mean gives the average marks/salary. For example, the average salary is 288,655, and the average E-Test percentage is 72.10.

• Median (Middle Value): The median shows the central value when data is arranged in order. For example, the median salary is 265,000, meaning half of the students earned below this amount and half earned above.

• Mode (Most Frequent Value): The mode represents the most common value. For example, the mode of MBA percentage is 56.7, which means this score appeared most frequently in the dataset.

# 3. Percentile Distribution

• Q1 (25% Percentile): 25% of the data lies below this value. For example, the 25th percentile of salary is 240,000.

• Q2 (50% Percentile or Median): 50% of the data lies below this value. For example, the 50th percentile of degree\_p is 66, meaning half the students scored below 66%.

• Q3 (75% Percentile): 75% of the data lies below this value. For example, the 75th percentile of salary is 300,000, meaning 75% of students earn below this and only 25% earn more.

• Q4 (100% or Maximum): Represents the maximum value. For example, the maximum salary is 940,000.

# 4. Key Observations

• Students scored an average of 66–67% in SSC, HSC, and Degree exams.

• The highest average score is in E-Test (72.1%).

• Salaries are highly skewed: although the mean salary is 288,655, the maximum salary is 940,000, indicating some students earn much higher than the average.

• The median salary (265,000) is lower than the mean, showing that a few very high salaries increased the average.

# 5. Conclusion

This statistical analysis highlights the use of mean, median, mode, and percentiles in understanding data distribution. It shows central tendencies of student performance and salary, as well as variations and skewness in outcomes.

* **IQR (Interquartile Range)** = Q3 − Q1
* It shows the **spread of the middle 50% of the data**.
* Example: If Q1 = 20 and Q3 = 40 → IQR = 20.

**🧐 Why multiply by 1.5?**

The **1.5 × IQR rule** is a **convention** (a standard guideline) to decide which points are *too far* from the middle of the data.

* **Lower Bound (Lesser Outlier):**  
  Q1 − 1.5 × IQR
* **Upper Bound (Greater Outlier):**  
  Q3 + 1.5 × IQR

Anything outside these bounds is considered an **outlier**.

**🔎 But why 1.5, not 2 or 3?**

* 1.5 was chosen by **John Tukey**, a famous statistician, when he invented the **boxplot**.
* It’s a **balance point**:
  + If we choose **smaller than 1.5** → too many normal points will look like outliers.
  + If we choose **much bigger than 1.5** (like 3) → we’ll miss real outliers.
* 1.5 is not a magic number, but it works well for many types of data as a **rule of thumb**.

**🎒 Example for School Student**

Imagine most of your classmates score between **60 and 80 marks** (Q1 = 60, Q3 = 80 → IQR = 20).

* Lower Bound = 60 − (1.5 × 20) = 30
* Upper Bound = 80 + (1.5 × 20) = 110

👉 So if someone scores **below 30 or above 110**, we call it an **outlier** (very unusual score).

✅ In short:  
We multiply by **1.5 × IQR** because it gives a **reasonable cutoff** to catch unusual values without being too strict or too loose.