

# ASSIGNMENT INFERENTIAL ANALYSIS

## 1.1 The dataset from here

### **Key Variables:**

- Gender, Board, Specialisation, Work Experience, Placement Status: **Categorical**
- SSC, HSC, Degree, MBA, Etest Scores, Salary: **Numeric**
- Target Variable: status (**Placed / Not Placed**)

### **Missing Data:**

- salary has **67 missing values**, likely for students who were **not placed**.

## 1.2 1) Replace the NaN values with correct value. And justify why you have chosen the same.

Replace NaN with 0 in salary

- Students who are **not placed** don't have a salary.
- Replacing with 0 is logical (not statistical imputation like mean/median).

## 1.3 2) How many of them are not placed?

There are 67 students in the dataset who are not placed. All these 67 students have "Not Placed" in the status column.

In real placement scenarios:

- Placed students → offered salary
- Not placed students → no salary (hence, NaN or 0)

## 1.4 3) Find the reason for non placement from the dataset?

**Lower Academic Performance**

On average, students who were *not placed* scored lower in:

- SSC% (Secondary School %)
- HSC% (Higher Secondary %)
- Degree %

#### **1.5 4)What kind of relation between salary and mba\_p**

Correlation between salary and mba\_p = 0.139823 (Positive correlation)

#### **1.6 5)Which specialization is getting minimum salary?**

On calculating the mean salary for each specialization:

- Mkt&HR shows a lower average salary than Mkt&Fin (Marketing and Finance).

Specialization-wise average salaries:

Mkt&HR    270377.358491

Mkt&Fin    298852.631579

Lowest average salary specialization: 270377.358490566

#### **1.7 6)How many of them getting above 500000 salary?**

Number of students with salary above 500000: 3

#### **1.8 7)Test the Analysis of Variance between etest\_p and mba\_p at significance level 5%.(Make decision using Hypothesis Testing)**

**Null Hypothesis (H<sub>0</sub>):** There is no significant difference between the of etest\_p and mba\_p.

**Alternative Hypothesis ( $H_1$ ):** There is a significant difference between the of etest\_p and mba\_p.

Here, F-Statistic: 98.64487057324706

P-Value: 4.672547689133573e-21

here  $p\_value > 0.05$ , so fails to reject  $H_0$  and therefore "There is no significant difference between the of etest\_p and mba\_p"

### **1.9 8) Test the similarity between the degree\_t (Sci&Tech) and specialisa tion (Mkt&HR) with respect to salary at significance level of 5%. (Make decision using Hypothesis Testing)**

**Null Hypothesis ( $H_0$ ):** There is no significant difference in salary between Students with degree\_t = Sci&Tech and Students with specialisation = Mkt&HR.

**Alternative Hypothesis ( $H_1$ ):** There is a significant difference in salary between the two groups.

Here, T-Statistic: 2.692041243555374

P-Value: 0.007897969943471179

here  $p\_value < 0.05$ , so reject  $H_0$  and therefore there is a significant difference in salary between the two groups.

### **1.10 9) Convert the normal distribution to standard normal distribution for salary column**

To convert a normal distribution to a standard normal distribution for the salary column, essentially performing Z-score normalization.

### **Standard Normal Distribution Formula:**

For each salary value x:

$$z = \frac{x - \mu}{\sigma}$$

Where:

- $\mu$  = Mean of salary
- $\sigma$  = Standard deviation of salary
- $z$  = Standardized (Z-score) value

### **1.11 10) What is the probability Density Function of the salary range from 700000 to 900000?**

Mean=198702.326, Standard Deviation=154780.927

The area between range (700000,900000):  
0.0005973310593974868

### **1.12 11) Test the similarity between the degree\_t(Sci&Tech) with respect to etest\_p and mba\_p at significance level of 5%. (Make decision using Hypothesis Testing)**

**Null Hypothesis ( $H_0$ ):** There is no significant difference between etest\_p and mba\_p for the students with degree\_t(Sci&Tech)

**Alternative Hypothesis ( $H_1$ ):** There is a significant difference between etest\_p and mba\_p for the students with degree\_t(Sci&Tech)

Here, T-Statistic: 5.0049844583693615

P-Value: 5.517920600505392e-06

here  $p\_value > 0.05$ , so fails to reject  $H_0$  and There is no significant difference between etest\_p and mba\_p for the

students with degree\_t(Sci&Tech) are statistically similar at the 5% significance level.

### 1.13 12) Which parameter is highly correlated with salary?

The parameter etest\_p (employability test score) is most highly correlated with salary.

### 1.14 13) plot any useful graph and explain it.

#### Explanation of the Graph

- The boxplot shows how salary is distributed for the two MBA specializations:
  - **Mkt&Fin (Marketing & Finance)**
  - **Mkt&HR (Marketing & Human Resource)**
- From the results:
  - Mkt&Fin students have a **higher median salary** and larger upper range.
  - Mkt&HR salaries are generally **lower** and less spread out.