

Inferential Analysis_Question_Hope Artificial Intelligence

November 6, 2022

1 Justification or answer summary needed for every question

1.1 ** Download the dataset from here(Click Here)

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

I used `dropna(subset=['salary'])` to remove missing salary values because:

1. Replacing with 0 would **bias the results** (e.g., minimum, average salary).
2. The rest of the features are usable; so dropping only a few rows keeps the dataset meaningful and clean.

1.3 2)How many of them are not placed?

67 of them are not placed

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

The mean values of `degree_p`, `etest_p`, and `mba_p` are lower for 'Not Placed' candidates compared to 'Placed' candidates. This suggests that lower academic performance may be a reason for non-placement.

1.5 4)What kind of relation between salary and mba_p

The relation between salary and `mba_p` has positive correlation 0.1750129406952748.

1.6 5)Which specialization is getting minimum salary?

specialisation

Mkt&Fin 200000.0

Mkt&HR 200000.0

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

3 of them getting above 500000 salary

1.8 7)Test the Analysis of Variance between etest_p and mba_p at significance level 5%.(Make decision using Hypothesis Testing)

There is a significant difference between the average values of `etest_p` and `mba_p` ($p\text{-value}=4.672547689133573e-21$), based on the ANOVA test ($p\text{-value} < 0.05$).

1.9 8) Test the similarity between the degree_t(Sci&Tech) and specialisation(Mkt&HR) with respect to salary at significance level of 5%. (Make decision using Hypothesis Testing)

There is a statistically significant difference in average salary between Sci&Tech students and Mkt&HR students at the 5% significance level.

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

I converted the salary column to a standard normal distribution by calculating Z-scores. Then I plotted it to check. The graph shows that the data is now centered around 0 with standard deviation 1.

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

Based on the probability density function, the chance of a salary falling between ₹700,000 and ₹900,000 is very low (about 0.0000054), which means this range is uncommon in the dataset.

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)

There is a significant difference between the etest_p and mba_p scores for Sci&Tech students. We reject the null hypothesis at the 5% significance level using the t-test (p-value = 0.00005).

1.13 12) Which parameter is highly correlated with salary?

The parameter etest_p (employability test percentage) is most highly correlated with salary, with a correlation value of 0.1783.

1.14 13) Plot any useful graph and explain it.

The scatter plot shows no strong link between MBA percentage and salary. Students with similar MBA scores have very different salaries. This means MBA marks alone do not decide salary.