

Inferential Analysis

1. NaN values are present only in the salary column. The students who are not placed, do not get salary, hence the NaN values in the salary column are replaced with 0.
 2. 67 students are not placed.
 3. From the dataset, it can be seen that the students who secured an average mark in mba_p are not placed.
 4. There is a direct relationship between mba_p and salary. As the mba_p increases, the salary also increases but only by 13.9%, hence there is a low degree of positive correlation between mba_p and salary.
 5. Marketing and Finance specialization is getting higher salary.
 6. 3 students are getting salary higher than 500000.
 7. Since p value < 5%, we reject null hypothesis. Hence there is significant difference between etest_p and mba_p at 5% significance level.
 8. At 5% significant level, the pvalue is 0.7%, hence alternate hypothesis is accepted.
- So, there is similarity between the degree_t (Sci&Tech) and specialization (Mkt&HR) with respect to salary.
9. Salary column is converted from normal distribution to standard normal distribution using z-score.
 10. The probability Density Function of the salary range from 700000 to 900000 is 0.59%.
 11. At 5% significant level, the pvalue < 5%, hence alternate hypothesis is accepted.
- So, there is similarity between the degree_t (Sci&Tech) with respect to etest_p and mba_p.
12. From the correlation graph, it is seen that ssc_p is highly correlated with salary.
 13. I have plotted a histogram for column "mba_p".

We get a bell-shaped curve so; it can be said that the mba_p is normally distributed. The mean is around 62 and the standard deviation is nearly 5 for mba_p.