DATA ANALYSIS

Univariate and Bivariate

Dataset:Placement

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

Ans:

In the Placement dataset **salary** column has some nan values. So we can replace the nan with "zero".Because the salary column is depends on status column. Status column has 2 values. one is "placed "another one is "not placed".If status column fill with "placed" then salary column contain some values otherwise salary column has nan values.

Why I choosed "Zero"? if the person had placed then we will give salary otherwise not.so I will put "zero"instead of nan values.

1.3)How many of them are **Not Placed** from the dataset?

Ans:

67 persons are Not Placed.

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

Ans:

Failing to prepare for interviews.

1.5)What kind of relation between salary and mba_p?

Ans:

Correlation between salary and mba_p is: **0.139823**

This **is positive correlation** that means independent variable is directly proportional to dependent variable.

If the person pass in mba_p then salary will increase only 13 percentage.

1.6)Which specialization is getting minimum salary?

Ans:

Minimum salary is=200000

mkt & hr, mkt & fin specialization is getting minimum salary.

1.7)How many of getting above 500000 salary?

Ans:

Three Employees are getting above 500000 salary

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

Ans:

Using one way classification the answer is

F_onewayResult(statistic=98.64487057324706, pvalue=4.6725476891335 73e-21)

p-value <0.05 reject null hypothesis.In this case accept alternative hypothesis.There is significance difference between e-test and mba_p.

1.9)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)

Ans:

Using T-test(Unpaired Test-Different Group but Same Condition) the output is

TtestResult(statistic=2.692041243555374, pvalue=0.007897969943471179, df=152.0)

Pvalue is less than 0.05. So we reject null hypothesis. There is no significant difference respect to salary**.**

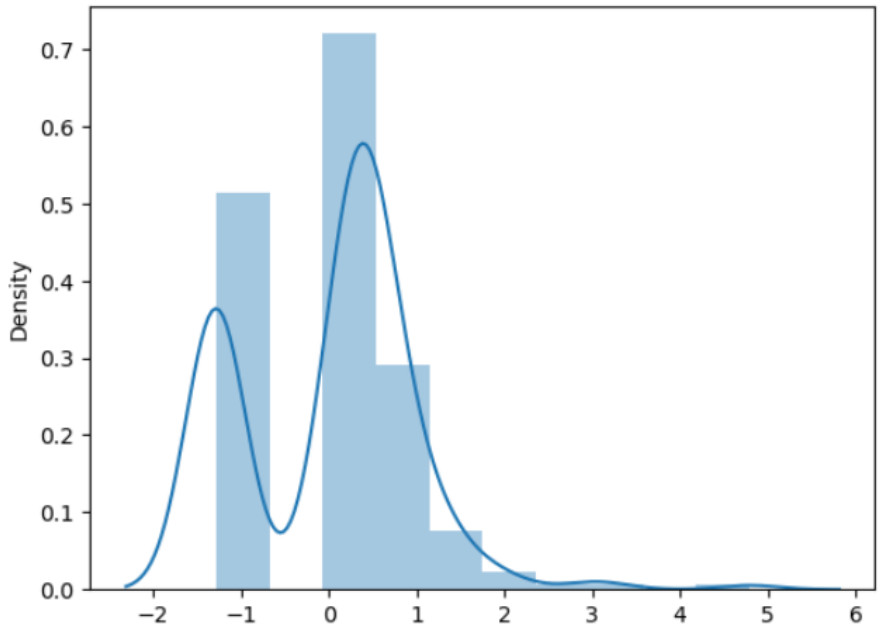
1.10) Convert the normal distribution to standard normal distribution for salary column

Ans:

Using Z-score we can convert the normal distribution to standard normal distribution

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

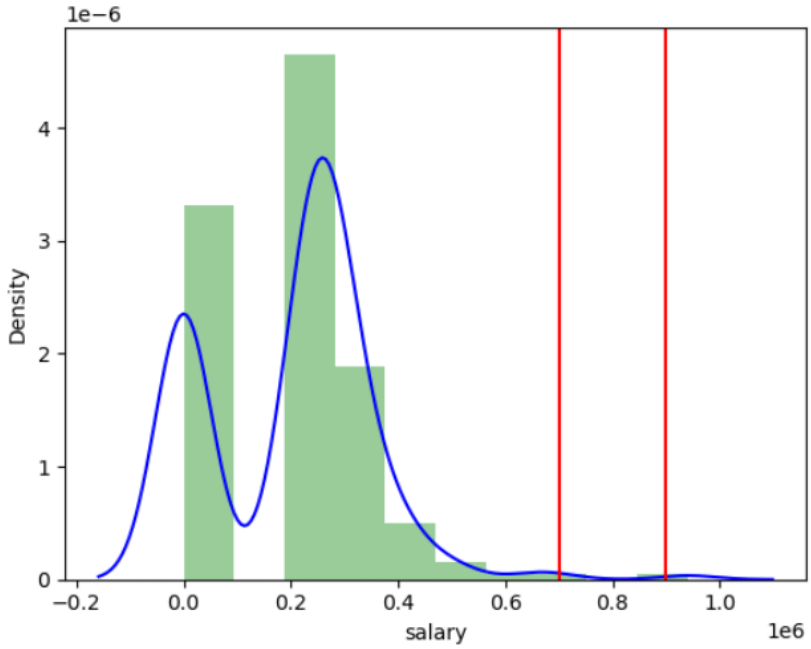
μ = Mean
σ = Standard Deviation



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

Ans:

The area between range(700000,900000):0.00059733105939749 01



70 to 90 percentage of probability to increased the salary from 700000 to 900000.

1.12)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)

Ans:

Using T-test(Paired Test-Same Group but Different Condition) the output is

TtestResult(statistic=5.0049844583693615, pvalue=5.517920600505392e-06, df=58)

Pvalue is greater than 0.05. So we accept alternative hypothesis. There is significant difference respect to to etest_p and mba_p.

1.13)Which parameter is highly correlated with salary?

Ans:

Using variance_inflation_factor we can find answer.

VIFvalue comes between 1 and 5.So ssc_p, hsc_p,degree_petest_p and mba_p are Moderately Correlated**.**

1.14) plot any useful graph and explain it.

Ans:

Seaborn's pairplot function is a powerful tool for visualizing pair wise relationships in a dataset. Using pairplot we can find multicolinearity between the columns**.** Multicolinearity means any Linear Relationship between columns.