**ASSIGNMENT 4**

**1.Write the formula of the line of best fit and state its scientific name**

The regression line is called the r-squared and the formula is:

Y=mx + b

**2.Write the two formulas of median with regards to grouped and ungrouped data**

For grouped data:

If the total number of observations (n) is odd, then the median is (n+1)/2 th observation.

If the total number of observations (n) is even, then the median will be average of n/2th and the (n/2)+1 th observation.

For ungrouped data:

Median=l+(n2−cff)×h

Where:

l is the lower limit of the median class

n is the number of observations

f is the frequency of median class

h is the class size

cf is the cumulative frequency of class preceding the median class.

**3. Define homoscedasticity and why it’s important in analysis**

Homoscedasticity is when there is equal variance or variance of error term is constant while heteroscedasticity is when there is unequal variance

The important of homoscedasticity is it shows that a model is accurate hence fulfilling the requirements of OLS. Also it shows that the model is well defined

**4. What is the difference between a boxplot and countplot**

A boxplot is a graphical representation that displays the distribution of a continuous variable. It shows the median, quartiles, and outliers, providing a quick overview of the data's central tendency and variability.

A countplot is a categorical plot that displays the frequency or count of observations for each category. It's used to visualize the distribution of a categorical variable, highlighting the proportion of each category.

**6. What distribution usually has same mean, median and mode**

A normal distribution has the same mean, median and mode. It is perfectly symmetrical with most data points clustering around the central point.