Use of Scatter diagram in the study of correlation

A scatter diagram (also known as a scatter plot) is a type of graph that shows the relationship between two continuous variables. Each data point is plotted as a pair of (x, y) coordinates on a 2-dimensional graph in the scatter diagram, which is frequently used to analyse if there is a link between the variables. If there is a positive correlation, the points on the scatter diagram will form an upward sloping line from left to right, and if there is a negative correlation, the points will form a downward sloping line from left to right. If there is no association, the points will create a random pattern with no discernible direction.

A scatter plot can provide insights into the nature of the relationship between two variables. Some additional aspects of scatter plots are:

1. Outliers: Points that fall outside the main pattern of the scatter plot are called outliers. They may indicate measurement errors or unusual observations.
2. Strength of correlation: The strength of the correlation between two variables can be measured using statistical measures such as Pearson's correlation coefficient. The coefficient ranges from -1 to 1, with -1 indicating a strong negative correlation, 0 indicating no correlation, and 1 indicating a strong positive correlation.
3. Non-linear relationships: A scatter plot can also show non-linear relationships between variables, where the relationship is not well described by a straight line. In such cases, a non-linear regression model may be more appropriate to describe the relationship.
4. Multivariate relationships: A scatter plot can also be used to investigate relationships between more than two variables. This is called a multivariate scatter plot, and it can be helpful in understanding complex relationships between variables.

In summary, a scatter plot is a valuable tool for visualizing the relationship between two continuous variables and can provide insights into the strength, direction, and nature of the relationship.

Differences between correlation and regression.

Correlation and regression are related statistical concepts, but they are not the same thing.

1. Correlation measures the relationship between two variables, and it can be positive, negative, or zero.
2. Regression, on the other hand, involves fitting a line or a model to the relationship between two or more variables, with the goal of predicting one variable based on the values of the other(s).
3. Correlation only tells us the strength and direction of a relationship between variables, but not the cause-and-effect relationship.
4. Regression, on the other hand, can help establish the cause-and-effect relationship by controlling for other variables.

Sure, here are some additional points to consider:

5. Correlation can be calculated for both continuous and categorical variables, but regression is typically only used for continuous variables.
6. Correlation is determined using a correlation coefficient, which ranges from -1 to 1 and represents the strength of the relationship between the two variables.

7. Regression, on the other hand, is determined by fitting a line to the data and finding the line of best fit that minimizes the residuals (the differences between the observed values and the values predicted by the model).
8. Correlation does not imply causation, meaning that just because two variables are correlated, it does not necessarily mean that one variable causes the other.
9. Regression can help establish causality by controlling for other variables that may confound the relationship between two variables, making it possible to determine the true effect of one variable on another.
10. In summary, correlation is used to describe the relationship between two variables, while regression is used to model and predict the relationship between two or more variables.