

Unit II

7. Converting to a Different Scale Type:

Converting data to a different scale type is a common data preprocessing technique in data analytics. It involves transforming variables from one scale type to another to meet specific requirements or assumptions for a particular analysis or modeling technique. Converting scale types can help improve the accuracy and effectiveness of the analysis and ensure that the data adheres to the assumptions of the chosen statistical methods.

1. Converting Categorical to Numerical (Ordinal to Interval/Ratio):
 - In some cases, categorical variables with an inherent order (ordinal scale) need to be transformed into numerical variables with equal intervals (interval or ratio scale). This transformation allows for the use of statistical techniques that require interval or ratio data, such as regression analysis.
2. Normalizing Numeric Data:
 - When dealing with data with significantly different ranges, normalization (scaling) can be applied to bring all variables to a similar scale. Common normalization techniques include Min-Max scaling and Z-score normalization. Normalization ensures that no variable dominates the analysis due to its larger magnitude.
3. Log Transformation:
 - Logarithmic transformation is used when the data distribution is highly skewed or the relationship between variables is multiplicative rather than additive. Log transformation can help stabilize variance and make the data more amenable to certain statistical analyses.
4. Binning and Discretization:
 - Continuous numerical variables can be transformed into categorical variables by binning or discretization. Binning involves dividing the range of values into predefined intervals (bins) and assigning each data point to the corresponding bin. This simplifies the data and may help identify patterns or trends.
5. Dummy Coding:
 - Categorical variables with multiple levels can be converted into binary (0/1) dummy variables for analysis. This process is known as dummy coding and is commonly used in regression and other modeling techniques.
6. Standardization:
 - Standardization (also known as z-score standardization) transforms numerical variables to have a mean of 0 and a standard deviation of 1. This technique is useful when variables have different units or scales and need to be compared on a common scale.

LEVELS OF MEASUREMENT

