UNIT 14
BIVARIATE MEASURES OF ASSOCIATION

This unit explores techniques used to examine the relationship between two variables (bivariate analysis) simultaneously. Understanding these relationships is critical for making predictive and causal claims in market research.

1. INTRODUCTION

Bivariate Measures of Association are statistical techniques used to summarize the strength and nature of the relationship between two variables. These relationships can be measured regardless of whether the variables are categorical (Nominal/Ordinal) or continuous (Interval/Ratio).

We want to see if one thing changes when another thing changes. For example: Does a customer's Gender relate to their Loyalty Card Usage? Or, does a high Satisfaction Score relate to high Visit Frequency?

2. CROSS TABULATIONS

Cross Tabulation (or contingency table) is a statistical technique used to simultaneously analyze two categorical variables (Nominal or Ordinal data). The results are presented in a matrix format, showing the frequency distribution of one variable within each category of the other. It's like sorting your data into a simple table to see if groups differ.

- **Coffee Shop Example:** We use cross-tabs to see if **Loyalty Card Ownership** (Variable 1: Yes/No) differs by **Gender** (Variable 2: Male/Female).

	Owns Loyalty Card (Yes)	**Does Not Own Card (No)**	**Total**
Male	50	70	120
Female	80	30	110
Total	130	100	230

- *Observation:* Females appear more likely to own a loyalty card than males. The Chi-Square () test would confirm if this observed difference is statistically significant.

3. SCATTER DIAGRAMS

A Scatter Diagram (or Scatter Plot) is a graphical representation of the relationship between two continuous variables (Interval or Ratio data). Each point on the graph represents a single observation, plotted on an X and Y axis.

It's the simplest way to visually check for correlation between two numerical things.

- **Coffee Shop Example:** We plot **Customer Age** (X-axis) against **Average Weekly Spend** (Y-axis).
    - If the dots rise from left to right, there's a **positive relationship** (older customers spend more).
    - If the dots are randomly spread, there's **no relationship**.

4. REGRESSION ANALYSIS

Regression Analysis is a powerful statistical technique used to analyze the linear relationship between a dependent variable (the variable to be predicted, Y) and one or more independent variables (the predictors, X). Bivariate Regression (Simple Linear Regression) innvolves only one independent variable (X) used to predict the dependent variable (Y).

Regression is used to predict the future or quantify the impact. It gives you a mathematical equation that allows you to answer the question: "If we change X by one unit, what specific change can we expect in Y?"

- **Coffee Shop Example:** We want to predict **Customer Visit Frequency (Y)** using the **Customer Satisfaction Score (X)**.
    - *Result:* If the resulting equation is Y=0.5+0.8X, the slope (b=0.8) means that for every 1-point increase in Satisfaction Score (X), we predict a 0.8 increase in Weekly Visit Frequency (Y).

5. CORRELATION ANALYSIS

Correlation Analysis measures the degree to which two continuous variables (Interval or Ratio data) vary together. The result is the Coefficient of Correlation (r), which ranges from −1.0 (perfect negative relationship) to +1.0 (perfect positive relationship).

Correlation tells you how strong and in what direction the two variables move, but it does *not* imply causation or give you a prediction formula.

- **Coffee Shop Example:** We calculate the correlation coefficient (r) between **Average Monthly Spend** and **Distance to the Store**.
    - If r = −0.75, it means there is a **strong negative correlation**: the further customers live from the store, the less they spend.
- **Key Difference from Regression:** Correlation just confirms a strong relationship exists (e.g., r=0.9). Regression determines the best way to *use* that relationship for prediction (e.g., Y=10+5X).

6. LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear Discriminant Analysis (LDA) is a multivariate statistical technique whose primary objective is to determine combinations of independent variables (predictors) that best discriminate (differentiate) between groups defined by a categorical dependent variable.

LDA helps classify customers into groups based on their measured characteristics. It creates a weighted equation (a "discriminant function") using multiple variables to determine if a customer belongs to **Group 1 (High Loyalty) or Group 2 (Low Loyalty).**

- **Coffee Shop Example:** We use customer characteristics (Satisfaction Score, Visit Frequency, Age, Distance) to determine which customers are most likely to fall into the **"Loyalty Tier" (Group 1)** versus the **"Regular Tier" (Group 2)**. The LDA tells us which factors are most important for making that classification.

7. AUTOMATIC INTERACTION DETECTOR (AID)

The Automatic Interaction Detector (AID) is a data analysis technique used to sequentially partition the sample into a series of subgroups that are maximally different with respect to the mean of a dependent variable. It is a precursor to modern decision tree methods.

AID automatically finds the most powerful segmentation rules for a single target goal. It creates an easy-to-understand tree structure by repeatedly asking, "Which factor creates the biggest difference in the average target outcome?"

- **Coffee Shop Example:** The goal (dependent variable) is **Average Monthly Spend**.
    - AID might find the most significant difference is based on **Age:** ≤30 **vs.** >30.
    - It then examines the "Age ≤30" group and finds the next best split is **Visits:** ≤4 **times/week vs.** >4 **times/week**.
    - This segmentation structure instantly tells the marketing team which combinations of factors define the highest-spending and lowest-spending customer groups.