Lecture on Descriptive Statistics in Research (1 Hour)

Introduction to Descriptive Statistics

Descriptive statistics help us summarize, organize, and interpret data to understand its basic features. In research, it is crucial to simplify large sets of data, making it easier to communicate findings. Descriptive statistics give us an idea about the data's central point, spread, and overall distribution, which are essential before diving into deeper analysis.

Example in Research:
Imagine you're conducting a study on the impact of sleep on productivity. You collect data on how many hours a group of employees sleep and their productivity levels. Descriptive statistics would help you summarize the hours of sleep and productivity scores to get an overall idea of the pattern.

KEY CONCEPTS IN DESCRIPTIVE STATISTICS

1. Distribution (Also Called Frequency Distribution)

A **frequency distribution** shows how often each value appears in the dataset. It helps in understanding how data is distributed across different categories or intervals. This is useful in research because it allows us to see how common certain values are.

Example in Research:
If you're analyzing the number of hours students study per week, a frequency distribution could show how many students study for 5, 10, 15, or 20 hours.

---

2. Measures of Central Tendency

Measures of central tendency are ways to summarize the "center" of your data. There are three main types:

- **Mean**: The average of all values.
  Example: In a survey about monthly expenses, if the mean expense is $3,000, it means that, on average, people spend around $3,000 per month.
- **Median**: The middle value when data is sorted.
  Example: If you're studying income levels, the median income is useful because it’s not affected by extremely high or low incomes (like the salary of a billionaire).
- **Mode**: The most frequent value.
  Example: If most participants in a study are aged 25, the mode of the ages is 25.

---

3. Variability (Also Called Dispersion)

Variability helps us understand how spread out the data is. It tells us whether the data points are close to the central value or scattered far apart.

- **Range**: The difference between the highest and lowest values.
  Example: If the range of study hours is 5 to 20 hours, you know that students' study time varies significantly.
- **Variance**: The average of the squared differences from the mean. This tells you how much the data points vary.
  Example: If you're studying the heights of basketball players, a large variance would mean that players' heights differ a lot.
- **Standard Deviation**: The square root of the variance. It gives a more understandable value for variability. A lower standard deviation means data points are close to the mean.
  Example: If the standard deviation of income in a community is small, it means most people earn similarly.

---

4. Univariate Descriptive Statistics & Bi variate

Univariate analysis focuses on one variable at a time. For example, you might study just one factor like income, age, or height. This helps in understanding patterns related to that specific variable.

Example in Research:
If you're studying the number of hours people exercise per week, univariate analysis could help you determine the average exercise hours and identify common patterns in the dataset.

---

5. Graphical Representations of Data

Visual representations of data help make descriptive statistics easier to understand. Here are common graphical techniques:

- **Histograms**: Show the frequency of data points in continuous intervals.
  Example: If you collect data on how many hours people work each day, a histogram could show how many people work between 1-2 hours, 2-3 hours, etc.
- **Bar Charts**: Compare different categories.
  Example: You might use a bar chart to compare the number of students in different majors.

- **Pie Charts**: Show proportions within a whole.
  Example: A pie chart would show what percentage of people prefer different types of smartphones.
- **Scatter Plots**: Show the relationship between two variables.
  Example: If you're analyzing the relationship between study hours and test scores, a scatter plot would help visualize whether more study hours lead to higher scores.
- **Box Plots**: Show the distribution of data based on quartiles.
  Example: In research on employee salaries, a box plot can show the range of salaries, the median salary, and any outliers.

Difference Between Histplot and Distplot

- **Histplot**: This plot shows the distribution of a dataset by breaking the data into bins and showing the frequency of data points in each bin. It's great for visualizing the shape of the data distribution.
- **Distplot**: This combines the histogram with a density plot, showing not only the frequency of values but also a smooth curve that represents the data distribution. It's useful when you want to visualize both the frequency and the probability distribution.

Example:
If you're analyzing exam scores, a **histplot** would show how many students scored within certain score ranges (bins), while a **distplot** would show this and also give an idea of how scores are distributed smoothly across the range.

Research Question:

"How many hours per week do university students spend on social media, studying, and exercising?"

Data Table (Question Table)

Student	Social Media Hours (hrs)	Study Hours (hrs)	Exercise Hours (hrs)
1	12	25	5
2	15	20	3
3	10	30	6
4	18	15	2

5	14	22	4
6	16	18	3
7	9	28	5
8	13	21	4
9	11	26	6
10	17	17	2

Research Question:

"How much time do individuals spend on different activities in a day (in hours)?"

Data Table (Question Table)

Person	Work Hours (hrs)	Leisure Hours (hrs)	Sleep Hours (hrs)
1	8	4	7
2	9	3	8
3	7	5	6
4	10	2	6
5	6	6	8
6	9	4	7
7	8	5	6
8	7	4	7
9	8	3	8
10	9	4	6

Task:

Calculate the following for **Work Hours**, **Leisure Hours**, and **Sleep Hours**:

1. **Mean**
2. **Median**
3. **Mode** (if applicable)
4. **Range**
5. **Variance**
6. **Standard Deviation**

Solve this table live using the formulas for each statistical measure.