

Chapter 6: Data Distribution

Data distribution refers to how data values are spread or distributed across a range. It is an essential concept in statistics that helps visualize, describe, and summarize the characteristics of a dataset. Understanding the distribution allows us to identify patterns, trends, and outliers.

Frequency Distribution

A **frequency distribution** is a summary of data that shows the number of occurrences (frequency) of each unique value or range of values in a dataset. It helps in organizing and understanding the dataset by providing a visual or tabular representation of the distribution of data points.

Components of Frequency Distribution:

1. **Class Intervals:**
Ranges of values into which data is grouped (e.g., 0-10, 11-20).
2. **Frequency (f):**
Number of data points that fall within each class interval.
3. **Cumulative Frequency:**
Running total of frequencies up to each class interval.
4. **Relative Frequency:**
Proportion or percentage of the total frequency for each class interval.

Formula:

$$\text{Relative Frequency} = (\text{Frequency of Class} / \text{Total Frequency}) \times 100$$

Types of Frequency Distribution:

1. **Univariate Frequency Distribution:**
Summarizes the frequency of one variable.
 2. **Bivariate Frequency Distribution:**
Displays frequencies of two variables simultaneously.
Example: Height vs. Weight distribution.
-

Constructing a Frequency Distribution Table:

Example Dataset:

4,8,9,10,12,12,13,14,16,18,18,20

Steps:

1. **Determine the Range:**
Maximum value =20= 20=20, Minimum value =4= 4=4

$$\text{Range} = 20 - 4 = 16$$

2. **Choose Class Intervals:**
Divide the range into equal intervals (e.g., 4 units).
3. **Create the Table:**

Class Interval	Frequency (f)	Cumulative Frequency	Relative Frequency (%)
4-8	2	2	16.67
9-12	3	5	25.00
13-16	3	8	25.00
17-20	4	12	33.33

1. Univariate Frequency Distribution

A **univariate frequency distribution** summarizes the frequencies of a single variable in a dataset. It helps in understanding how data points are distributed across the values or intervals of a single characteristic, such as age, income, or test scores.

Characteristics of Univariate Frequency Distribution:

1. **Single Variable:** Focuses on one variable at a time.
2. **Frequency:** Counts how often each value or range of values occurs.
3. **Representation:** Can be presented in a table or visualized using charts like histograms, bar graphs, or pie charts.

Types of Univariate Frequency Distribution:

1. **Ungrouped Frequency Distribution:**
Lists individual values and their frequencies. Suitable for small datasets.

Example:

Data: 4,4,5,5,5,6,6,7

Value	Frequency (f)
4	2
5	3
6	2
7	1

2. **Grouped Frequency Distribution:**

Groups data into intervals, especially useful for large datasets with many unique values.

Example:

Data: 10,12,15,18,19,22,25,27,30,35

Intervals: 10–14,15–19,20–24,25–29,30–34

Class Interval	Frequency (f)
10–14	2
15–19	3
20–24	1
25–29	2
30–34	1

Visualizing Univariate Frequency Distributions:

1. **Histogram:**

- Represents grouped frequency distribution.
- X-axis: Class intervals, Y-axis: Frequencies.

2. **Bar Graph:**

- Represents ungrouped data.
- Bars are used to show the frequency of individual values.

3. **Pie Chart:**

- Represents relative frequency distribution.
- Each slice corresponds to a category or value.

Bivariate Frequency Distribution

A **bivariate frequency distribution** summarizes the relationship between two variables by showing how often each combination of their values occurs. This type of distribution is useful for identifying patterns, correlations, or dependencies between two variables.

Components of a Bivariate Frequency Distribution:

1. **Two Variables:**
 - One variable is represented along the rows (usually X).
 - The other variable is represented along the columns (usually Y).
2. **Frequency Table:**
 - Each cell in the table contains the frequency of occurrences for the corresponding combination of X and Y.

Example Dataset:

Data: Students' scores in Math (XXX) and Science (YYY):

(50,60),(50,70),(60,70),(60,80),(70,80),(50,70),(60,60),(70,80)

Frequency Table:

Science (Y)	60	70	80
Math (X)			
50	1	2	0
60	1	1	1
70	0	0	2

In this table:

- X=50,Y=60 occurs 1 time.
- X=50,Y=70 occurs 2 times

Visualizing Bivariate Frequency Distribution:

1. **Heatmap:**
 - Displays frequencies as color intensities in a matrix form.
 - Useful for identifying clusters or patterns.
2. **Scatter Plot:**
 - Plots individual data points on a 2D plane.
 - Used for continuous data to visualize relationships or trends.
3. **Stacked/Grouped Bar Chart:**
 - Represents frequencies for categorical variables.