UNIT-III MEASURES OF CENTRAL TENDENCY: Objective of averaging, characteristics of good average, types of average, arithmetic mean of grouped and ungrouped data, correcting incorrect values, weighted arithmetic mean, median - median of grouped and ungrouped data merit and limitation of median, computation of quartile, decile and percentile Mode - calculation of mode of grouped and ungrouped data, merits and limitation of mode, relationship between mean, median and mode. Geometric mean and Harmonic mean.

Content

Objective of averaging, characteristics of good average, types of average, arithmetic mean of grouped and ungrouped data, correcting incorrect values, weighted arithmetic mean

1. Objective of Averaging

The objective of averaging is to provide a single value that represents the central tendency of a data set. It summarizes a dataset of numbers into one representative number, making it easier to understand the data as a whole. Averages are widely used in fields like statistics, economics, and social sciences to describe the "typical" or "central" value in a collection of data. It facilitate us to compare dataset.

2. Characteristics of a Good Average

A good average should have the following characteristics:

- **Simplicity**: It should be easy to calculate and understand.
- **Representativeness**: It should reflect the central value of the data set, representing the majority of the values.
- **Stability**: A small change in the data should not drastically affect the average.
- **Sufficiency**: The average should represent the entire data set well, not just one subset of it.
- **Consistency**: It should work well with various types of data and give reliable results across different data sets.
- It should enable us make comparison between data.

3. Types of Averages

There are several types of averages used in statistics, each serving a different purpose:

- **Arithmetic Mean**: The sum of all values divided by the number of values.
- **Median**: The middle value when the data set is ordered from smallest to largest (or vice versa).
- **Mode**: The most frequent value in the data set.
- **Geometric Mean**: The nth root of the product of all the values in the data set.
- **Harmonic Mean**: The reciprocal of the arithmetic mean of the reciprocals of the values.
- **Weighted Mean**: Each value has a weight, and the mean is calculated by multiplying each value by its weight and dividing by the total weight.

4. Arithmetic Mean of Grouped and Ungrouped Data

- **Ungrouped Data**: This is when you have individual data points. The arithmetic mean (AM) is calculated by adding up all the values and dividing by the number of values:

Arithmetic Mean (AM) $=\sum x / n$

here $\sum x$ is the sum of all values andnn is the number of values.

- **Grouped Data**: When the data is grouped into classes or intervals, the AM is calculated using the midpoints of the classes. There are two methods available for calculations.
- **Direct Method**

  The formula is:

  Arithmetic Mean (AM) $=\sum f \cdot x / N$

  where f is the frequency of each class, and x is the midpoint of each class. This gives an approximation of the mean based on the class intervals.

- **Indirect Method** let A be assumed mean

  The formula is:
  Arithmetic Mean (AM) $=A+ \sum f \cdot d / N$
  Where $d=X-A$

5. Correcting Incorrect Values

Incorrect values (often called outliers or errors) can distort the average. To correct them:

- **Identify the error**: Check the data for any values that seem unrealistic or extreme.
- **Correct the error**: If the value is due to a typographical error or a miscalculation, correct it manually.
- **Exclude outliers**: If a value is legitimately incorrect (i.e., an outlier), it might be best to exclude it from the analysis or adjust the data accordingly.
- **Recalculate the average** after correcting the values to ensure the mean reflects the data correctly.

6. Weighted Arithmetic Mean

The **Weighted Arithmetic Mean** is an average where each value has a different level of importance or frequency. In this case, the mean is calculated by multiplying each value by its corresponding weight, asumming those products, and dividing by the sum of the weights:

Weighted Mean$=\sum wi \cdot xi/N$

where wi is the weight of each value xi, and the summation is over all values in the dataset. This is particularly useful when some values in the data set are more significant or frequent than others.

Median of grouped and ungrouped data, Merit and limitation of median

The **median** is the middle value in an ordered list of numbers. It divides the data set into two equal halves: one half of the data is less than the median, and the other half is greater.

Median of Ungrouped Data

Steps to find the median for ungrouped data:

1. Arrange the data in ascending (or descending) order.
2. If the number of observations (n) is odd, the median is the middle value.

If n is odd: Median= Size of (N+1)/2 th item

3. If n is even, the median is the average of the two middle values.

If n is even: Median= {size of N/2th + size of (N/2+1)th}/2

Median of Grouped Data

For **grouped data**, the median is calculated using the **median class** (the class interval that contains the median). Grouped data is usually presented in frequency distributions, so the following steps are used to find the median:

Steps to find the median for grouped data:

1. **Calculate the cumulative frequency** (CF) for each class.
2. **Identify the median class**, which is the class where the cumulative frequency is greater than or equal to (n/2) but less than (n/2) + 1.
3. Use the following formula to calculate the median:

$$\text{Median} = L + \left(\frac{(\frac{N}{2} - PCF)}{f}\right) \times h$$

Where:

- L = Lower boundary of the median class
- N = Total frequency
- CF = Cumulative frequency of the class before the median class
- f = Frequency of the median class
- h = Class width

Merit of Median

1. **Resistant to Outliers:** The median is not affected by extreme values (outliers), unlike the mean. This makes it a better measure of central tendency when there are outliers in the data.
2. **Simple to Calculate:** The median can be easily calculated for both ungrouped and grouped data, and it does not require complex formulas or calculations like the mean (which needs all data points).
3. **Represents Central Value:** The median divides the data into two equal halves, providing a good representation of the "central" value in skewed distributions.

Limitation of Median

1. **Not Suitable for Further Calculations:**
   Unlike the mean, the median does not allow for further statistical calculations like variance or standard deviation because it doesn't take into account the values of all data points.
2. **Can Be Less Precise for Grouped Data:** In the case of grouped data, the median is an estimate, as it uses class intervals, and you lose exact values.
3. **May Not Represent the True Central Tendency in Symmetric Distributions:** In distributions that are symmetric, the mean is a more accurate measure of central tendency than the median, since the two will be similar.

Mode

Mode: The most frequently occurring value in a dataset.

Example (ungrouped)

- For the data set 2, 3,4,4,5,5,5,6 the mode is 5 (since it appears three times).

Formula (for grouped data)

$$\text{Mode} = L + \frac{f_1 - f_0}{2\, f_1 - f_0 - f_2}\ X\ h$$

Computation of Quartiles, Deciles, and Percentiles

These measures divide a dataset into intervals or groups based on its distribution. Following provides steps how each of them is computed.

1. Quartiles

Quartiles divide the dataset into four equal parts, with each part containing 25% of the data. There are three quartiles:

- **Q1 (First Quartile)**: 25% of the data lies below this value.
- **Q2 (Second Quartile or Median)**: 50% of the data lies below this value. This is the median.
- **Q3 (Third Quartile)**: 75% of the data lies below this value.

Quartiles (ungrouped data):

1. **Arrange the data** in ascending order.
2. **Find Q2** (the median of the dataset):
   - If the number of data points (n) is odd, Q2 is the middle value.
   - If n is even, Q2 is the average of the two middle values.
3. **Find Q1** (the median of the lower half of the data): This is the median of the data points before Q2.
4. **Find Q3** (the median of the upper half of the data): This is the median of the data points after Q2.

Quartiles (grouped data):

$$Q_i = L + \left(\frac{(i.\frac{N}{4} - PCF)}{f}\right) \times h,$$ where i= 1, 2, 3

Example Breakdown:

Let's compute Q1, Q2, and Q3 for a small dataset:

Data: 3, 7, 8, 5, 12, 15, 6

1. **Sort the data** in ascending order: 3, 5, 6, 7, 8, 12, 15
2. **Q2 (Median)** is the middle value: 7
3. **Q1** is the median of the lower half: (3, 5, 6) → Q1 = 5
4. **Q3** is the median of the upper half: (8, 12, 15) → Q3 = 12

For example:

- Given data: 2, 4, 5, 7, 9, 10, 12, 14, 15
- **Q2 (Median)**: 9 (the middle value)
- **Q1**: 4.5 (the median of the lower half: 2, 4, 5, 7, 9)
- **Q3**: 12 (the median of the upper half: 9,10, 12, 14, 15)

2. Deciles (ungrouped data)

Deciles divide the data into 10 equal parts, with each part containing 10% of the data.
The 9 deciles (D1, D2, … D9) are the values that divide the data into these 10 parts.

Steps to compute deciles:

1. **Arrange the data** in ascending order.
2. **Find the position of each decile** using the formula: Dk=k(n+1)10D_k = \frac{k(n+1)}{10}Dk=10k(n+1) Where:
   - k = the decile number (1 for D1, 2 for D2, … 9 for D9)
   - n = the total number of observations in the data set
3. Find the value at the corresponding position in the ordered dataset.

For example:

- Given data: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- To compute **D1** (the first decile), we use the formula:
  $D1=\frac{1(10+1)}{10}=1.1$
  The value at position 1.1 is between 1 and 2, so D1=1.1.
- Similarly, **D2** and other deciles can be computed using the same formula and approach.

. **Deciles** (Grouped data)

$$Dk= L+ \left(\frac{(k.\frac{N}{10}-PCF)}{f}\right) \times h \qquad \text{where i= 1, 2, 3.....9}$$

3. Percentiles (ungrouped data)

Percentiles divide the data into 100 equal parts. Each percentile corresponds to a specific percentage of the data. The **kth percentile** is the value below which k % of the data fall.

Steps to compute percentiles:

1. **Arrange the data** in ascending order.
2. **Find the position of the kth percentile** using the formula: Pk=k(n+1)100P_k = \frac{k(n+1)}{100}Pk=100k(n+1) Where:
   - k = the percentile number (1 for the 1st percentile, 50 for the 50th percentile, etc.)
   - n = the total number of observations in the dataset
3. Find the value at the corresponding position.

For example:

- Given data: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- To compute the **25th percentile (P25)**:
  $P25=\frac{25(10+1)}{100}=2.75$

The 25th percentile will be between the values 2 and 3, so we calculate it accordingly (interpolation can be used if needed).

Percentiles (Grouped data)

Pk= L+ $\left(\frac{(k.\frac{N}{100} - PCF)}{f}\right) \times h$ where i= 1, 2, 3.....99

Geometric mean and Harmonic mean

1. **Geometric Mean**: Used when data points are multiplicative in nature, the geometric mean is the nth root of the product of all data points.

   Ungrouped Data

   Geometric Mean= $(\prod Xi)^{1/n}$

   Or Geometric Mean= Antilog $\left(\frac{\sum \log X}{N}\right)$

Grouped Data

Geometric Mean= Antilog $\left(\frac{\sum f\,.\,\log X}{N}\right)$

2. H**armonic Mean**: The reciprocal of the arithmetic mean of the reciprocals of the data points, typically used for rates or ratios.

   Ungrouped Data

   Harmonic Mean= $\frac{N}{\sum \frac{1}{Xi}}$

   Grouped Data

   Harmonic Mean = $\frac{N}{\sum f.\ \frac{1}{Xi}}$

Relationship B/w above measures

3. mode = 3 median -2 mean
4. AM x HM = GM^2
5. AM≥ GM ≥ HM

*Note- Queries, doubts, Numerical problems will be discussed only in class.