Chapter-statistics

Introduction to Statistics

Statistics is the science of collecting, organizing, analyzing, and interpreting data. In this chapter, we focus on measures of central tendency for grouped data.

1. Measures of Central Tendency

Measures of central tendency are values that represent the center or typical value of a dataset. For grouped data, we study three main measures: mean, median, and mode.

1.1 Mean for Grouped Data

The mean is the arithmetic average of a dataset. For grouped data, there are three methods to calculate the mean: Direct Method, Assumed Mean Method, and Step Deviation Method.

1.1.1 Direct Method

Formula:

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

Where:

- x_i = Class mark = $(\text{Lower Limit} + \text{Upper Limit})/2$
- f_i = Frequency of the class

Example 1: Calculate the mean using the direct method for the following data:

Class Interval	Frequency
0-10	5
10-20	8
20-30	12
30-40	7
40-50	3

Solution:

- **Given Data**:
  - Class Marks (x_i): [5, 15, 25, 35, 45]
  - Frequencies (f_i): [5, 8, 12, 7, 3]
- **Calculations**:
  - $\sum f_i x_i = (5 \times 5) + (15 \times 8) + (25 \times 12) + (35 \times 7) + (45 \times 3) = 825$
  - $\sum f_i = 5 + 8 + 12 + 7 + 3 = 35$
  - Mean ($\bar{x}$) = $\frac{\sum f_i x_i}{\sum f_i} = \frac{825}{35} = 23.57$

1.1.2 Assumed Mean Method

Formula:

$$\bar{x} = A + \frac{\sum f_i d_i}{\sum f_i}$$

Where:

- A = Assumed mean (any class mark chosen as a reference)
- $d_i = x_i - A$, the deviation of each class mark from the assumed mean

Example 2: Calculate the mean using the assumed mean method for the same data as above.

Solution:

1. Choose an assumed mean ($A = 25$)
2. Find deviations ($d_i = x_i - A$):
   $d_1 = -20, d_2 = -10, d_3 = 0, d_4 = +10, d_5 = +20$
3. Multiply frequencies by deviations ($f_i d_i$):
   $(-20 \times 5) + (-10 \times 8) + (0 \times 12) + (10 \times 7) + (20 \times 3) = -100 - 80 + 0 + 70 + 60 = -50$

---

4. Calculate mean:
   Mean $(\bar{x})$ = $25 + (-50/35)$ = $25 - 1.43 = 23.57$

1.1.3 Step Deviation Method

Formula:

$$\bar{x} = A + \frac{\sum f_i u_i}{\sum f_i} \times h$$

Where:

- A = Assumed mean
- $u_i = \frac{x_i - A}{h}$, where h is the class size
- h = Class interval size

Example 3: Calculate the mean using the step deviation method for the same data.

Solution:

1. Choose an assumed mean ($A = 25$) and class size ($h = 10$)
2. Find step deviations ($u_i = (x_i - A)/h$):
   $u_1 = -2, u_2 = -1, u_3 = 0, u_4 = +1, u_5 = +2$
3. Multiply frequencies by step deviations ($f_i u_i$):
   $(-2 \times 5) + (-1 \times 8) + (0 \times 12) + (+1 \times 7) + (+2 \times 3)$:
   Sum of $f_i u_i = -10 - 8 + 0 + 7 + 6 = -5$
4. Calculate mean:

Mean=25 + (-5 /35) X 10=23.57

1.2 Median for Grouped Data

The median is the middle value when the data is arranged in order.
Formula: Median = l + [(N/2 - cf) / f] × h

Where:

- l = lower limit of median class
- N = total frequency
- cf = cumulative frequency of class before median class
- f = frequency of median class
- h = class size

Steps to calculate:

1. Find N/2, where N is the total frequency
2. Locate the median class (where N/2 falls)
3. Apply the formula

Example 3: Find the median for the data in Example 1.
Solution:

1. N = 35, so N/2 = 17.5
2. Cumulative frequency: 5, 13, 25, 32, 35
3. Median class: 20-30 (as 17.5 falls here)
4. Median = 20 + [(17.5 - 13) / 12] × 10 = 23.75

Example 4: Calculate the median for this distribution:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	4	6	10	8	2

Solution:

1. N = 30, N/2 = 15
2. Cumulative frequency: 4, 10, 20, 28, 30
3. Median class: 20-30
4. Median = 20 + [(15 - 10) / 10] × 10 = 25

1.3 Mode for Grouped Data

The mode is the value that appears most frequently in a dataset.
Formula: Mode = l + [(f1 - f0) / (2f1 - f0 - f2)] × h

Where:

- l = lower limit of modal class
- f1 = frequency of modal class
- f0 = frequency of class before modal class
- f2 = frequency of class after modal class
- h = class size

Example 5: Find the mode for the data

Class Interval	Frequency
0-10	5
10-20	8
20-30	12
30-40	7
40-50	3

Solution:

1. Modal class: 20-30 (highest frequency)
2. l = 20, f1 = 12, f0 = 8, f2 = 7, h = 10
3. Mode = $20 + [(12 - 8) / (2\times12 - 8 - 7)] \times 10 = 23.33$

Example 6: Calculate the mode for this distribution:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	12	6	4

Solution:

1. Modal class: 20-30
2. l = 20, f1 = 12, f0 = 8, f2 = 6, h = 10
3. Mode = 20 + [(12 - 8) / (2×12 - 8 - 6)] × 10 = 24

2. Additional Concepts

2.1 Class Mark

Class Mark = (Lower Limit + Upper Limit) / 2

2.2 Class Size

Class Size = Upper Limit - Lower Limit

2.3 Cumulative Frequency

Cumulative Frequency = Sum of all frequencies up to and including the current class

2.4 Empirical Relationship

For a moderately skewed distribution:

$$\text{Mean} - \text{Mode} \approx 3(\text{Mean} - \text{Median})$$