Statistics Notes

B.Tech. CSE

Anant Srivatsava

Instructor:
Mrs. Nancy

Contents

1 Measures of Central Tendency

1. Mean
2. Median
3. Mode

1.1 Mean

It is the ratio of sum of all the observations to the total number of observations. let $x_1, x_2, \ldots, x_n$ be all the observations. then:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

1.1.1 Properties of Mean

- The sum of deviation of observations from mean is always zero
- the sum of square of deviations of observations is minimum as compared to any other measure.
- suppose there are two sequences:

	Series 1	Series 2
Number of observations	n_1	n_2
mean of the observations	$\overline{x}_1$	$\overline{x}_2$

then

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2}$$

Problem 1

If there are 5 and 8 number of observations of 2 series with mean 15 and 18, find the combined mean

Solution:
We can get the solution by taking the weighted mean of the two sequences.
so the required mean is :

$$\frac{5 \times 15 + 8 \times 18}{5 + 8}$$
$$= \frac{75 + 144}{13}$$
$$= \frac{219}{13}$$
$$= 16.846154$$

Problem 2

Class	frequency
0-10	3
10-20	5
20-30	7
30-40	4
40-50	1

Solution:
change of origin:

Class	frequency	X	d=X-A	f·d
0-10	3	5	-20	
10-20	5	15	-10	
20-30	7	25	0	
30-40	4	35	10	
40-50	1	45	20	

$$\overline{x} = A + \frac{\sum fd}{n}$$

change of scale

Class	frequency	X	d=X/n	f·d
0-10	3	1	-20	
10-20	5	3	-10	
20-30	7	5	0	
30-40	4	7	10	
40-50	1	9	20	

$$\overline{x} = A + \frac{\sum fd}{n}$$

1.2 Median

Steps to find Median in case of Discrete and continuous data:

1. Arrangement of data
2. if n is odd then the median is the $\frac{n+1}{2}$th term
3. if n is even then the median is the mean of the $\frac{n}{2}$th term and $\frac{n}{2} + 1$th term

Problem 3

find the median for the data :

1. 9,9,10,10,12,13,15
2. 9,9,10,10,12,13,14,15

Solution:

1. 9,9,10,10,12,13,15 has 7 elements. Therefore our median will be the 4th term in the arranged order
$\therefore Median = 10$

2. 9,9,10,10,12,13,14,15 has 8 elements. Therefore our median will be the mean of the 4th and 5th terms.
$\therefore Median = \frac{10+12}{2} = 11$

Problem 4

Finding the median of discrete data.

X	f	cf(cumulative frequency)
1	5	5
2	8	13
3	9	22
4	**12**	**34**
5	6	40
6	7	47
7	4	51
Total	51	

find the value of x which has cumulative frequency just greater than $\frac{n}{2}$

In case of continuous data:

$$Median = l + \frac{\left(\frac{n}{2} - cf\right) h}{f}$$

where cf is the cumulative frequency and f is the frequency of the chosen class, h is the class size

1.3 Mode

The observation which occurs the most is called the mode of the data.
In more general terms, the most probable observation in a dataset is the mode of the data.

Problem 5

Find mode for the following data: 10,11,15,18,18,18,15,10,18,20

Problem 6

Find the mean, median and mode for the following data

CI	f
0-10	3
10-20	5
20-30	7
30-40	2
40-50	1
Total	51

How to find the mode for continuous data

1. Find the modal class which is having the maximum frequency.
2. based on that input the values into the following formulae:

$$mode = l + h\left(\frac{f_1 - f_2}{2f_1 - f_0 - f_2}\right)$$

1.4 The interconnection between the measures of central tendency

$$Mode = 3Median - 2Mean$$

1.5 Geometric and Harmonic mean

Def^n :
Geometric mean is defined as the nth root of the product of n observations
Mathematically:

$$GM = \sqrt[n]{\prod_{i=0}^{n} x_i}$$

Problem 7

Find the Geometric Mean for the values 2,4,8

Def^n :
Harmonic mean is defined as the reciprocal of arithemetic mean of the reciprocal of all the observations

$$HM = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Theorem:
The following inequality is always true:

$$AM \geq GM \geq HM$$