GIT 1111:Basic Statistics

Topic: Measures of Central Tendencies as Data Analysis Tools

Gulu University

September 22, 2024

Objectives of the Lecture

After reading this unit, you will be able to:

- explain the concept of central tendency of data;
- describe the different measures of central tendency;
- discuss the properties, advantages and limitations of mean, median andmode; and
- compute measures of central tendency for ungrouped and grouped data.

Introduction

- Measures of central tendency provides a single value that indicates the general magnitude of the data and this single value provides information about the characteristics of the data by identifying the value at or near the central location of the data (Bordens and Abbott, 2011).
- King and Minium (2013) described measures of central tendency as a summary figure that helps in describing a central location for a certain group of scores.
- Tate (1955) defined measures of central tendency as “a sort of average or typical value of the items in the series and its function is to summarise the series in terms of this average value”

Main Functions of Measure of central Tendency

- They provide a summary figure that explain the central location of the whole data and gives an idea about the whole data.
- Large amount of data is reduced to a single figure. Mean, median and mode can be computed for a large data.
- When mean is computed for a certain sample, it will help gauge the population mean.
- The results from measures of central tendency help in making decisions not only with regard to research but have applications areas like policy making, marketing and sales and so on.
- Comparison can be carried out based on single figures computed with the help of measures of central tendency. For example, performance of students in mathematics test, the mean marks obtained by girls and the mean marks obtained by boys can be compared.

Characteristics of a good measure of central tendency

- The definition of the central tendency needs to be adequately specified/ rigid and should be clear.
- It should not be subject to varied interpretations and needs to be unaffected by any individual bias.
- The measure of central tendency should be easy to understand and easy to compute.
- For the value obtained from the computation of measures of central tendency to be representative of the data, the whole data needs to be computed.
- The data needs to be collected from a sample that truly represents the population. The sample thus needs to be randomly selected.
- The measure of central tendency should not be affected by outliers.
- The measure of central tendency should render itself to further mathematical computations

Different Measures of Central Tendency

We will discuss the three measures of central tendency namely:

- Mean or Arithmetic mean
- Median
- Mode

Mean or Arithmetic Mean

- Mean for sample is denoted by symbol $\overline{X}$ ('x-bar')' and mean for population is denoted by μ (mu).
- It is one of the most commonly used measures of central tendency and is often referred to as average.
- It can also be termed as one of the most sensitive measure of central tendency as all the scores in a data are taken in to consideration when it is computed (Bordens and Abbott, 2011).
- Mean is a total of all the scores in data divided by the total number of scores.

Different Measures of Central Tendency

Median

- Median is a point in any distribution below and above which lie half of the scores.
- Median is the middle score in an ordered distribution.
- The middle score in this distribution is then identified as median.

Mode

- Mode is the score in a distribution that occurs most frequently.
- For instance if there were other 10 students in this group of 100 students, who secured 47 marks, 47 is the value that is occurring as frequently as 35 and thus, will be termed as mode along with 35.
- Mode as such does not provide an adequate characterisation of the distribution because it just takes in to consideration the most frequent scores and other scores are not considered.
- Though if the scores in a distribution greatly vary then it is possible that there is no mode.

Measures of Central Tendency

How to choose a measure of central tendency?

- The choice of a measure of central tendency will depend on first of all, the scales of measurement that we discussed in the first unit.
  - For nominal scales one can compute mode but not mean or median. For example, in case of males and females, the males can be coded as 1 and females can be coded as 2 (or vice versa) in such a case, we can compute frequently occurring score, that will provide us information whether there are more males or more females.
  - However, it is not possible to compute mean or median with regard to ordinal scale, median or mode can be used. For example, if we rank the students based on their performance in mathematics test, it is possible to find median belowand above which lie half of the ranks.
  - Mode can also be computed if more than one student gets same rank. With regard to interval scale and ratio scale mean can be computed.

- While making a choice with regard to which measure of central tendency to use is, whether the data is normally distributed or not.
  - If the data is normally distributed we will compute mean and if it is not normally distributed, we will compute median or mode. This is because mean may not adequately represent the data when the data is not normally distributed.

Properties, Advantages and Limitations of Mean, Median and Mode

Properties of Mean

- Mean is sensitive to the actual position of each and every score in a distribution and if another score is included in the distribution, then the mean or average of that distribution will change.
  For example, mean of the scores 5, 4 ,6 ,3, 2 is 4 .
  But if we change the scores to 5, 4, 6, 3, 2, 8, the mean will be 4.67
- Mean denotes a balance point of any distribution and the total of positive deviations from the mean is equal to the negative deviations from the mean.
- Mean is especially effective when we want the measure of central tendency to reflect the sum of the scores.

Advantages of Mean

- The definition of mean is rigid which is a quality of a good measure of central tendency.
- It is not only easy to understand but also easy to calculate.
- All the scores in the distribution are considered when mean is computed.
- Further mathematical calculations can be carried out on the basis of mean.
- Fluctuations in sampling are least likely to affect mean.

Limitations of Mean

- Outliers or extreme values can have an impact on mean.
- When there are open ended classes, such as 10 and above or below 5, mean cannot be computed. In such cases median and mode can be computed. This is mainly because in such distributions mid point cannot be determined to carry out calculations.
- If a score in the data is missing or lost or not clear, then mean cannot be computed unless mean is computed for rest of the data by not considering the lost score and dropping it all together.
- It is not possible to determine mean through inspection. Further, it cannot be determined based on a graph.
- It is not suitable for data that is skewed or is very asymmetrical as then in such cases mean will not adequately represent the data

Properties, Advantages and Limitations of Median

Properties of Median

- When compared to mean, median is less sensitive to extreme scores or outliers.
- When a distribution is skewed or is asymmetrical median can be adequately used.
- When a distribution is open ended, that is, actual score at one end of the distribution is not known, then median can be computed.

Advantages of Median

- The definition of median is rigid which is a quality of a good measure of central tendency.
- It is easy to understand and calculate.
- It is not affected by outliers or extreme scores in data.
- Unless the median falls in an open ended class, it can be computed for grouped data with open ended classes.
- In certain cases it is possible to identify median through inspection as well as graphically

Limitations of Median

- Some statistical procedures using median are quite complex. Computation of median can be time consuming when large data is involved because the data needs to be arranged in an order before median is computed.
- Median cannot be computed exactly when an ungrouped data is even. In such cases, median is estimated as mean of the scores in the middle of the distribution.
- It is not based on each and every score in the distribution.
- It can be affected by sampling fluctuations and thus can be termed as less stable than mean.

Properties of Mode

- Mode can be used with variables that can be measured on nominal scale.
- Mode is easier to compute than mean and media. But it is not used often because of lack of stability from one sample to another and also because a single set of data may possibly have more than one mode. Also, when there is more than one mode, then the modes cannot be termed to adequately measure central location.
- Mode is not affected by outliers or extreme scores.

Advantages of Mode

- It is not only easy to comprehend and calculate but it can also be determined by mere inspection.
- It can be used with quantitative as well as qualitative data.
- It is not affected by outliers or extreme scores.
- Even if a distribution has one or more than one open ended classe(s), mode can easily be computed.

Limitations of Mode

- It is sometimes possible that the scores in the data vary from each other and in such cases the data may have no mode.
- Mode cannot be rigidly defined.
- In case of bimodal, trimodal or multimodal distribution, interpretation and comparison becomes difficult.
- Mode is not based on the whole distribution.
- It may not be possible to compute further mathematical procedures based on mode.
- Sampling fluctuations can have an impact on mode.

Computation of Measures of Central Tendency in ungrouped and grouped data

Definition

(Ungrouped data)

Any data that has not been categorised in any way is termed as an ungrouped data.

For example, we have an individual who is 25 years old, another who is 30 years old and yet another individual who is 50 years old. These are independent figures and not organised in any way, thus they are ungrouped data.

Definition

(Grouped data)

A data that is categories or organised is termed as grouped data. Mainly such data is organised in frequency distribution. For example, we can have age range 26- 30 years, 31- 35 years, 36- 40 years and so on. Grouped data are convenient especially when the data is large.

Computation of Mean for Ungrouped Data

The formula for computing mean for ungrouped data is

$$\overline{X} = \frac{\sum x}{N}$$

Where, $\overline{X}$ = Mean
$\sum x$ = Summation of scores in the distribution
N = Total number of scores.

Example

The scores obtained by 10 students on psychology test are as follows:
58 34 32 47 74 67 35 34 30 39

$$\overline{X} = \frac{\sum x}{N} = \frac{58 + 34 + 32 + 47 + 74 + 67 + 35 + 34 + 30 + 39}{10} = \frac{450}{10} = 45$$

Computation of Mean for Grouped Data

The formula for computing mean for grouped data is

$$\overline{X} = \frac{\sum fx}{N}$$

Where,
M= Mean
$\sum$= Summation
x= Midpoint of the distribution, $x = \frac{upper\ \ limit + lower\ \ limit}{2}$
f = The respective frequency
N = Total number of scores

Example

A class of 30 students were given a psychology test and the marks obtained by them were categorised in to six categories. The lowest marks obtained were 10 and highest marks obtained were 35. A class interval of 5 was employed.The data is given as follows:

Marks	Frequencies (*f*)	Midpoint (X)	*f*X
35- 39	5	37	185
30-34	7	32	224
25-29	5	27	135
20-24	6	22	132
15-19	4	17	68
10- 14	3	12	36
	N= 30		**ΣfX = 780**

$$\text{Mean}, \overline{X} = \frac{\sum fx}{N} = \frac{780}{30} = 26$$

Computation of Mean by Shortcut Method (with Assumed mean)

In certain cases data is very large and it is not possible to compute each product fx. In such situations, a short cut method with the help of assumed mean can be computed. A real mean can thus be computed with application of correction.
The formula is

$$\overline{X} = A + \frac{\sum fd}{N}$$

Where,
A = Assumed mean,
$\sum$ = Summation
$d = (x - A)$, x the midpoint of the scores in the interval
f = the respective frequency of the midpoint
$\sum f = N$ = The total number of frequencies or students.

Example

The table below contains information on the grades received by 150 students in an examination.

Class	$0-20$	$20-40$	$40-60$	$60-80$	$80-100$
Frequency	18	28	40	35	29

Solution

Class interval	Frequency, f	Mid point, x	$d = x - A$	fd
$0-20$	18	10	$10-50=-40$	-720
$20-40$	28	30	$30-50=-20$	-560
$40-60$	40	A=50	$50-50=0$	0
$60-80$	35	70	$70-50=20$	700
$80-100$	29	90	$90-50=40$	1160
	$N=150$			$\sum fd = 58$

$$\overline{X} = A + \frac{\sum fd}{N} = 50 + \left(\frac{580}{150}\right) = 50 + 3.8667 = 53.87$$

Computation of Median for Ungrouped Data

With regard to computation of median for ungrouped data, different procedures are followed for data that is odd and data that is even.
Odd Data: When the data is odd the median is computed in the following manner Data: 58 34 32 47 74 67 35 34 30 (N= 9)

Step 1 : Arranged the data in either ascending or descending order. Data in ascending order 30 32 34 34 35 47 58 67 74

Step 2 : The following formula is then used to compute Median:
$M_d = \frac{N+1}{2}^{th}$ score.
Thus $\frac{(9+1)}{2} = \frac{10}{2} = 5^{th}$ item In our data the 5^{th} item is 35, that is the median of this data.

Even data: When the data is even, the median is computed in thefollowing manner:
58 34 32 47 74 67 35 34 30 39 (N= 10)

Step 1 :First arranged the data in either ascending or descending order.The data in ascending order looks like this:
30 32 34 34 35 39 47 58 67 74

Step 2 :The following formula is used to compute median:

$$M_d = \frac{(\frac{N}{2})^{th} term + [\frac{N}{2} + 1]^{th} term}{2}$$

The $\left(\frac{N}{2}^{th}\right)$ term is the 5^{th} term, that is 35. The $\left(\frac{N}{2}^{th} + 1\right)$ term is the 6^{th} term, that is 39.
Thus

$$M_d = \frac{35 + 39}{2} = 37$$

The median thus obtained is 37

Computation of Median for Grouped Data

The formula used for computation of median for grouped data is as follows:

$$Md = L + \left[\frac{(\frac{N}{2}) - cf_b}{f_m}\right] \times i$$

Where,
L = The lower limit of the median class
N = Total of all the frequencies
cf_b = cummulative frequencies before the median class
f_m= frequency within the interval upon which the median falls
i = class interval of the median class.

Let us discuss the steps followed for computation of median with the help ofthe example given below:

Class Intervals	Frequency	Cummulative frequency
$35 - 39$	5	5
$30 - 34$	7	12
$25 - 29$	5	17
$20 - 24$	6	23
$15 - 19$	4	27
$10 - 14$	3	30
	N=30	

The steps in computing median for grouped data are as follows:

Step 1 : The first step is to compute $N/2$, that is $30/2$ so that we obtain one half of the scores in the data (15 in this case).

Step 2 :As the scores are even in number (N$=$ 30), the median should fall between 15^{th} and 16^{th} score. So the median will fall in the class interval 25-29.

L, the lower class boundary of the median class 25-29, will be 24.5.

Step 3 : Determine cf_b and i, from the table it is $cf_b = 12$ and $i = 5$.

Step 4 : Find f_m the frequency of the median falls. In the present example the median class interval is 25-29 and the frequency for this class interval is 5. So $f_m = 5$.

Step 5 : The values can now be put in the formula to obtain the median

$$M_d = L + \left[\frac{(\frac{N}{2}) - cf_b}{f_m}\right] \times i = 24.5 + \left(\frac{15 - 12}{5}\right) \times 5 = 27.5$$

Computation of Mode for Ungrouped Data

Let us now learn how to compute mode for an ungrouped data with the help of the following example:

58 34 32 47 74 67 35 34 30 39

The mode can be calculated in simple manner by just counting the scores that appears maximum number of times in the data.
In our example, the score occurring maximum number of times is 34, that occurs twice.
Thus the mode is 34

Note

- If two or more observations are repeated maximum number of times then each number is a mode.
- If there is no any observations repeated then there is no mode.

Computation of Mode for Grouped Data

The mode for grouped data can be computed using the formula: In the second method of computing mode for grouped data the following formula is used:

$$Mo = L + \left[\frac{d_1}{d_1 + d_2}\right] \times i$$

Where,
$L =$ Lower limit of the class interval in which the mode may lie, called as modal class
$i =$ Class interval of the modal class
$d_1 =$ difference between frequencies of modal class and class interval below it.
$d_2 =$ difference between frequencies of modal class and class interval above it,

Let us discuss the steps followed for computation of mode with the help of the example given below:

Class Intervals (Marks)	**Frequencies (*f*)**
35- 39	5
30-34	**7**
25-29	5
20-24	6
15-19	4
10- 14	3
	N= 30

Step 1 :The mode is most likely to fall in the the class intervals 30-34 as that has the highest frequencies (7). Thus this is our modal class and the lower class boundary of the same (L) will be 29.5.

Step 2 :The class interval (i) for this example is 5.

Step 3 :Compute d1 and d2

$$d_1 = 7 - 5 = 2, d_2 = 7 - 5 = 2$$

Step 4 Now lets compute the mode

$$\begin{aligned} Mo =& L + \left[\frac{d_1}{d_1 + d_2}\right] \times i \\ =& 29.5 + \left[\frac{2}{2+2}\right] \times 5 = 29.5 + 2.5 = 32 \end{aligned}$$

Exerceise

1. The heights of students in a certain class were recorded as shown in the table below

Height (cm)	Frequency (f)
149-152	5
153-156	17
157-160	20
161-164	25
165-168	15
169-172	6
173-176	2

Estimate the mean, mode and median height of the student and comment on your answer.

2. The marks below were obtained by students in a class in mathematics examination.

Marks	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69	70 – 79	80 – 89
No of students	3	8	18	24	20	12	5

Find the average mark, modal mark and median mark for the class and comment on your answers

3. The table below shows the marks obtained, out of 50, by students in a class test.

Mark (x)	No. of students(f)
1-10	9
11-20	10
21-30	12
31-40	8
31-50	7

Using an assumed mean, calculate the mean mark to the nearest whole number.

4. Given the following sets of numbers.

   (i) 1769, 1771, 1772, 1775, 1778, 1781, 1784
   (ii) 0.85, 0.88, 0.89, 0.93, 0.94, 0.96
   (iii) 23, 33, 34, 36, 38, 40, 41, 41, 44
   (iv) 1, 2, 1, 1, 3, 4, 100

   Compute the measures of central tendency and comment on which measure(s) is suitable for the data set

5. In an experiment, 50 people were asked to guess the weight of a mobile phone in grams. The guesses were as follows:

47	39	21	30	42	35	44	36	19	52
23	32	66	29	5	40	33	11	44	22
27	58	38	37	48	63	23	40	53	24
47	22	44	33	13	59	33	49	57	30
17	45	38	33	25	40	51	56	28	64

Compute the measure of central tendency and provide a descriptive statistics on the weight of the phone using the calculated values.

The End,
Thanks for attending, read the content in order to achieve the objectives of the topic