Chapter3

Measure of Central Tendency

Alemakef Wagnew

Institute of Public Health
Department of Epidemiology and Bio-statistics

March 23, 2019

Schedule

Measure of Central Tendency

measure of central Tendency

Definition:

- On the scale of values of a variable there is a certain stage at which the largest number of items tends to cluster/center around.
- Since this stage is usually in the center of distribution, the tendency of the statistical data to get concentrated at this stage/value is called" central tendency“. various measures determining the actual value at which the data tends to concentrate are called measures of central tendency.
- So, a measure of central location is the single value that best represents the whole series.

Types of Measure of Central Tendency

Types of Measure of Central Tendency

The most commonly used averages are:

- Mean (Arithmetic, Geometric and Harmonic)
- Mode
- Median

Average should posses the following:

- It should be rigidly defined.
- It should be based on all observation under investigation.
- It should be as little as affected by extreme observations.
- It should be capable of further algebraic treatment. It should be as little as affected by fluctuations of sampling.
- It should be ease to calculate and simple to understand. First point

Common Measure of Central Tendency

Common Measure of Central Tendency

A. The Arithmetic Mean

- The measure of central location you are probably most familiar with.
- It is the arithmetic average and is commonly called simply "mean" or "average."
- In formulas, the arithmetic mean is usually represented as for population mean, and, X (read as, x-bar) for sample mean

Ungrouped data: The formula for calculating the mean from individual data of a sample is:

$$Mean = \overline{x} = \sum \frac{Xi}{n} \quad (1)$$

- For population mean, substitute x-bar by $\mu andnbyN$.

The Arithmetic Mean

Example:

The following are the lengths (in cm) of a sample of six garment blanks chosen at random from a large batch of similar blanks: 54.5, 55.0, 55.7, 51.8, 54.2, 52.4
What is the mean length of the sample of garments?
Solution:

Solution:

$$\overline{x} = [\frac{1}{n}\sum_{i=1}^{n} Xi]$$

$$= \frac{1}{6}\sum_{i=1}^{6}(45.5 + 53.0 + 55.7 + 51.8 + 54.2 + 51.8)$$

$$\overline{X} = 53.6cm$$

The Arithmetic Mean for Grouped data

Grouped data:

- A mean can be determined for grouped data, or data that is placed in intervals.
- The sum of the products divided by the total number of values will be the value of the mean.
- To calculate the mean of grouped data, the first step is to determine the midpoint (also called a class mark) of each interval, or class

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

$$= \frac{1}{n}\sum_{i=1}^{n} (f_1 X_1 + f_2 X_2 + ... + f_n X_n)$$

Grouped mean Example

Example block

- The table below shows age group of 50 MOHA permanent workers.
- Calculate the arithmetic mean

Class limits	Frequency (f_i)	class mark (X_i)	
42-48	8	45	360
49-55	8	52	416
56-62	13	59	767
63-69	7	66	462
70-76	6	73	438
77-83	5	80	400
84-90	3	87	261

Figure: Caption

Solution

$$
\begin{aligned}
\overline{x} = \frac{1}{n}\sum_{i=1}^{n} f_i X_i \\
&= \frac{1}{50}\sum_{i=1}^{5} 0(360 + 416 + \dots\dots\dots + 261) \\
&= \frac{3104}{50} \\
&= 60.08
\end{aligned}
$$

Merits of Arithmetic Mean

- It is rigidly defined.
- It is based on all observation.
- It is suitable for further mathematical treatment.
- It is stable average, i.e. it is not affected by fluctuations of sampling to some extent.
- It is easy to calculate and simple to understand.

Demerits of Arithmetic Mean

- It is affected by extreme observations.
- It can not be used in the case of open end classes.
- It can not be determined by the method of inspection.
- It can not be used when dealing with qualitative characteristics, such as intelligence, honesty, beauty.
- Some times it leads to wrong conclusion if the details of the data from which it is obtained are not available.
- It gives high weight to high extreme values and less weight to low extreme values

The Mode

The Mode

- Mode is a value which occurs most frequently in a set of values.
- The mode may not exist and even if it does exist, it may not be unique.
- In case of discrete distribution the value having the maximum frequency is the model value.

Example:

1. Find the mode of 5, 3, 5, 8, 9 Mode =5
2. Find the mode of 8, 9, 9, 7, 8, 2, and 5. It is a bi modal Data: 8 and 9
3. Find the mode of 4, 12, 3, 6, and 7. No mode for this data.

Mode for Grouped data

In grouped data: we usually refer to the modal class, class with highest frequency. If a single value for the mode of grouped data must be specified, it is taken as

$$MODE = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * W$$

$$Where L = The lower class boundary of the modal class;$$

$$\Delta_1 = f_{mode} - f_1, \Delta_2 = f_{mod} - f_2$$

$$W = the size of the modal class$$

$$f1 = frequency of the class preceding the modal class.$$

$$f2 = frequency of the class succeeding the modal class$$

$$f_{mod} = frequency of the modal class.$$

Example of the mode

Calculate the modal age for the age distribution of 228 patients below.

classinterval	numberof women
15-19	6
20-24	19
25-29	50
30-34	57
35-39	48
40-44	27
45-49	21
total	228

solution Grouped Mode

$$MODE = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} * W$$

$$\Delta_1 = 57 - 50 = 7, \Delta_2 = 57 - 48 = 9$$

$$There\ for\ the\ Modal\ average$$

$$\hat{X} = 29.5 + \frac{7}{7+9} * 5\hat{X} = 31.7$$

Merits and Demerits of Mode

- **Merits:**
  - It is not affected by extreme observations.
  - Easy to calculate and simple to understand.
  - It can be calculated for distribution with open end class.
- **Demerits:**
  - It is not rigidly defined.
  - It is not based on all observations.
  - It is not suitable for further mathematical treatment.
  - It is not stable average, i.e. it is affected by fluctuations of sampling to some extent.
  - Often its value is not unique.

The Median

Median

The Median:

- The median is the middle value of a set of data that has been put into rank order.
- In a distribution, median is the value of the variable which divides it in to two equal halves.
- Useful measure of central tendency when data are skewed.

Ungrouped Data:

Let be $X_1, X_2, X_3, ...X_n$ *ordered observations*.

Then, the median value is:

$$The(\frac{n+1}{2})^{th} \text{ when } n \text{ is odd}$$

(2)

$$The\frac{(\frac{n}{2})th + (\frac{n}{2}+1)^{th}}{2} \text{ when even} \tag{3}$$

Example: Find the median of the following numbers.

1. 6, 5, 2, 8, 9, 4.
2. 2, 1, 8, 3, 5, 8.

The median

Grouped data:

Steps:

- Locate the class interval in which the median is located. We use the following procedure for this.
  - Find n/2 and see a class interval with a minimum cumulative frequency which contains n/2 this is the median class
- Find a unique median value, use the following interpolation format:

$$Median = L + \frac{W}{f_{med}}\left(\frac{n}{2} - CF\right) \tag{4}$$

- **Where:**
  - L= lower class boundary of the Median
  - W= the class width
  - CF= the cumulative frequency corresponding to the class preceding the median class and
  - $f_{med} = \textit{the frequency of median class}$
- **The Median class:** is the class with the smallest frequency (less than type) greater than or Equal to $n_{\overline{2}}$

Compute the median for the following distribution.	
Grade	Frequency
40-49	5
50-59	18
60-69	27
70-79	15
80-89	6

solution:

Compute the median for the following distribution.		
Class boundary	Frequency	less than CF
39.5-49.5	5	5
49.5-59.5	18	23
59.5-69.5	27	50
69.5-79.5	15	65
79.5-89.5	6	71

The Median

$$\text{Since } n = 71, \frac{71}{2} = 35.5, \text{ and} \tag{5}$$

$$\text{the smallest less than CF greater than or equal to } 35.5 \text{ is } 50; \tag{6}$$

$$\text{thus, the median class is the third class.} \tag{7}$$

$$\text{And for this class} \tag{8}$$

$$L = 59.5, w = 10, f_{med} = 27, CF = 23. \tag{9}$$

$$\text{Then applying the Formula we get :} \tag{10}$$

$$\hat{X} = 59.9 + \frac{10}{27}\left(35.5 - 23\right) \tag{11}$$

$$= 64.13 \tag{12}$$

Merits and Demerit's of Median

Merits:

- Median is a positional average and hence not influenced by extreme observations.
- Can be calculated in the case of open end intervals.
- Median can be located even if the data are incomplete.

Demerits:

- It is not a good representative of data if the number of items is small.
- It is not amenable to further algebraic treatment.
- It is susceptible to sampling fluctuations.

Exercises

Exercises for grouped median

Values	Frequency
140-150	17
151-160	29
161-170	42
171-180	72
181-190	84
191-200	107
201-210	49
211-220	34
221-230	31
231-240	16
241-250	12

Compute Mean, Mode and Median

Thank you