Unit 16

Measures of Central Tendency

Contents

Learning Objectives

It is expected that after reading Unit 16 you would be able to

- Understand the procedure of arriving at measures of central tendency of the data collected
- Work out the ways of finding out mean, mode and median measures of central tendency
- Decide which of the three measures is more appropriate in the case of your data.

16.1 Introduction

After dealing with the skills of sampling techniques for studying large complex social groups, we would now discuss the matter of measuring central tendenc and its application.

Unit 16 deals with the basic measures of central tendency and their application for those of you who may lack a strong background in mathematics. In doing so, complex mathematical derivations of formulae have been omitted. Besides a minimal number of essential 'shorthand' mathematical symbols, and familiar examples drawn from social science data are presented in a non-mathematical form.

16.2 Mean

Mean® is the most common and widely used measure of central tendency. Each observation in a population may be referred to as X_i (read "X sub i") value. Thus, one observation might be denoted as X_1, another as X_2, a third as X_3, and so on. The subscript i might be any integer value up through *N*, the total number of X_i values in the population. The mean of the population is denoted by the Greek letter μ (lower case mu).

Calculating the mean from ungrouped data

Mean *(M)* is the most familiar and useful measure used to describe the central tendency average of a distribution of scores for any group of individuals, objects or events. It is computed by dividing the sum of the scores by the total number of scores.

$M = \Sigma X_i / N$ 1

Where, M is the mean (sample), X_i are the scores, *N* is the total number of scores and Σ is 'the sum of'. See Box 16.1 and Box 16.2 for examples 1 and 2.

Box 16.1

Example 1: The Number of Cattle Owned by Members of a Community is Recorded Below.

12, 11, 13, 20, 16, 18, 19, 17, 22 and 23

$\Sigma X_i = 12 + 11 + 13 + 20 + 16 + 18 + 19 + 17 + 22 + 23 = 170$

N = 10

$M = \Sigma X_i / N$; $M = 170 / 10 = 17$

The mean is the balance point in a distribution such that if you subtract each value in the distribution from the mean and add all these deviation scores, the result will be zero.

Calculating mean from grouped data

Calculation of mean from grouped data is slightly different from calculation from ungrouped data.

$M = \Sigma F_i * X_i / \Sigma F_i$ 2

where, M is the mean, X_i are the midpoint of class intervals, F_i are the number of cases in various intervals, ΣF_i is the total number of scores or sum of frequencies of various intervals.

Box 16.2

Example 2: Following is the frequency (8, 9, 12, 9, 7, and 5) of households in a community owning numbers of chickens, arranged in six groups (1-3, 4-6, 7-9, 10-12, 13-16 and 16-18).

Number of Chickens	Mid-Point of the Interval (X_i)	Frequency: Number of Households (F_i)	$F_i * X_i$
1 -3	2	8	16
4 - 6	5	9	45
7 - 9	8	12	96
10 - 12	11	9	99
13 - 16	14	7	98
16 - 18	17	5	22
		50	376

$\Sigma F_i * X_i = 376$ $\Sigma F_i = 50$

$M = \Sigma F_i * X_i / \Sigma F_i = 376 / 50 = 7.52$

A short method of calculating mean from grouped data

There is a shorter way of calculating mean from grouped data, which saves time and labour in computation, particularly when one has to deal with a large number of cases. It involves the assumption of mean and making a guess at identifying the interval in which the mean probably falls (generally among the central groups of intervals). A different guess of the interval alters calculations, but not the mean.

Mean (M) = $AM + ((\Sigma F_i * D_i \ / \ \Sigma F_i))*i$ 3

Also,

$D_i = (AM - X_i) \ / \ i$ 4

Where, M is the mean, AM = Assumed mean, X_i are the midpoint of class intervals, F_i are the number of cases in various intervals, Σ is the symbol of sum total, D_i are the deviations of the midpoints of the various classes from the midpoints of the class having the assumed mean divided by the size of the class interval (equation 4) and i is the size of the class intervals. See Box 16.3 for example 3.

Box 16.3 Example 3: Marital Distance (the distance between the villages of the spouse)

The marital distance was investigated in a community. Following was the frequency (88, 93, 72, 97, 79, and 54) when the data were arranged in six groups according to marital distance (25 - 30, 30 - 35, 35 - 40, 40 - 45, 45 - 50, 50 -55). Let us find the mean marital distance.

Marital Distance (km)	Frequency (F_i):	Mid-Point of the Interval (X_i)	$D_i = (AM - X_i)/ i$	$F_i * D_i$
25 -30	88	27.5	+3	264
30 -35	93	32.5	+2	186
35 -40	72	37.5	+1	72
40 -45	97	AM=42.5	0	0
45 -50	79	47.5	-1	-79
50 -55	54	52.5	-2	-108
	483			335

AM =42.5 $\Sigma F_i = 483$ $\Sigma F_i * D_i = 335$ i = 5 (30-25)

Mean (M) = $AM + ((\Sigma F_i * D_i \ / \ \Sigma F_i))*i$

= 42.5 + (335/ 483) *5 = 42.5 + 3.468 = 45.968

After the three examples for calculating mean for ungrouped and grouped data, we would now discuss the technique of finding the Median.

16.3 Median

Median® is the score that divides the distribution into halves; half of the scores are above the median and the other half are below it when the data are arranged in a numerical order. Median is also referred to as the score at the 50th percentile in the distribution.

Calculating median from ungrouped data

- Arrange the series in numerical order (ascending or descending).
- Find the median location of N numbers by the formula $(N + 1) / 2$. When N is an odd number, for example 7 then the value of the 4th item ((7+1)/2 = 4) is the median. For example in the following ordered distribution the value of 4th item, i.e. 9 is the median.

  2, 5, 8, 9, 12, 16, 16
- Whereas, when N is an even number, say 12 then the median is half-way between the 6th and 7th items ((12+1)/2 = 6.5).

See Box 16.4 for examples 4 and 5.

> **Box 16.4 Example 4: Finding the Median**
> When N is an odd number: Find the median in the distribution of numbers: 1, 13, 8, 3, 4, 11, and 7.
> The median location is $(N + 1) / 2$ or $(7 + 1) / 2 = 4$.
> The ordered distribution is: 1, 3, 4, 7, 8, 11 and 13.
> The value of 4th item in the distribution is 7 and thus median is 7.
> **Example 5: *When N is an even number:***
> Find the median in the distribution of numbers: 1, 8, 3, 13, 11, and 7.
> The median location is $(6 + 1) / 2 = 3.5$.
> The ordered distribution is 1, 3, 7, 8, 11 and 13.
> The halfway value between the 3rd and 4th item is 7.5 ((7+8) / 2), and thus median is 7.5.

Calculating median from grouped data

Finding the median score in the frequency distribution below involves five steps.

Step 1: Divide the total number (N or ΣF_i) by two.

Step 2: Start at the low end of the frequency distribution and sum the scores in each interval until the interval containing the median is reached (C. F.).

Step 3: Subtract the sum obtained in step two above from the number necessary (calculated at step 1) to reach the median (N/2 - C. F.).

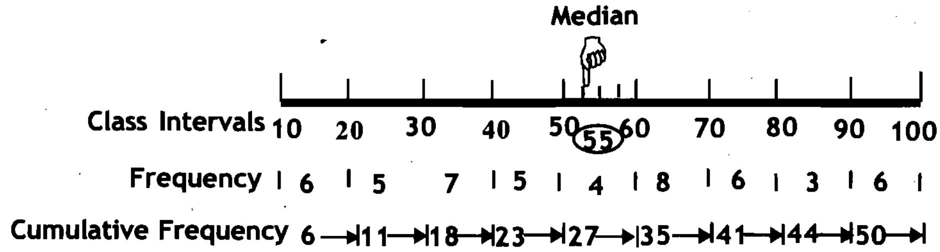
Step 4: Now calculate the proportion of the median interval that must be added to its lower limit in order to reach the median score. This is done by dividing the number obtained in step 3 above by the number of scores (f) in the median interval and then multiplying by the size of the class interval (i), i.e. $[(N/2 - C.F.) / f] * i$.

Step 5: Finally, add the number obtained in step 4 above to the exact lower limit of the median interval.

Median = L + [(N /2 - C.F.) / f] *i

Where, *L* = the exact lower limit of the median interval, *N* = the total number of scores; C.F. = the sum of the scores in the intervals below the median interval, *f* = the number of scores in the median interval; i = the size of the class interval.

Graphical representation of calculating the median from grouped data



See Box 16.5 for example 6 for finding the media for grouped data.

Box 16.5 Example 6: Find the Median of the following distribution

Class Interval	18-21	21-24	24-27	27-30	30-33	33-36 36-39 39-42 42-45 45-48 48-51
Frequency						

Class Interval	Frequency	Cumulative Frequency
18-21	1	1
21-24	2	3
24-27	3	6
27-30	6	12
30-33	7	19
33-36	8	27
36-39	8	35
39-42	6	41
42-45	4	45
45-48	3	48
48-51	2	50
Total	ΣF_i =50	

Median = L + [(N /2 - C.F.) / f] *i

N or ?F_1 / 2 =50/2 =25
Lower limit of the median class (L) = 33
Cumulative frequency of the class preceding the median class (C.F.) = 19

Cumulative frequency of the class preceding the median class (C.F.) = 19
Frequency of the median class (f) = 8
Size of the class interval = 3
Median = 33 + [(25-19) / 8)] *3 = 33 + 2.25 = 35.25
Let us now complete Reflection and Action 16.1 for checking if the calculation methods have now become clearer and easier to perform.

After the Reflection and Action 16.1, you would learn about calculating mode from ungrouped and grouped data.

Reflection and Action 16.1

Following the examples given in the text for calculating the mean and median for ungrouped and grouped data and the short method of calculating mean of grouped data, provide your own examples of each of the five calculations in the manner similar to examples in the text. This exercise would provide you an opportunity of practicing such calculations. These calculation exercises would come in handy while you would carry out your own mini research project.

16.4 Mode

Mode® of a distribution is simply defined as the most frequent or common score in the distribution. Mode is the point (or value) of X that corresponds to the highest point on the distribution. If the highest frequency is shared by more than one value, the distribution is said to be **multimodal**. It is not uncommon to see distributions that are bimodal reflecting peaks in scoring at two different points in the distribution.

Calculating mode from ungrouped data

The most frequent data in the series is the mode. It can be determined by viewing the series (if the series is small) or looking at the frequency distribution (if the series is large). See Box 16.6 for example 7.

Box 16.6 Example 7: Find the Mode of the following Distribution.

Serial number of family	1	2	3	4	5	6	7	8	9	10
Number of Children	1	2	3	4	3	3	2	1	2	3

In the above example 3 occurs the maximum number of times (4 times), and hence 3 is the mode of the distribution.

Calculating mode from grouped data

Mode of the grouped data can be calculated using the following steps:
Step 1: Identify the modal class (class with maximum frequency) by inspection or analysis.

Step 2: Apply the following formula

Mode = $L + [(f_m - f_1) / (f_m - f_1) + (f_m - f_2)] * i$

Or

Mode = $L + [(f_m - f_1) / (2f_m - f_1 - f_2)] * i$

Where, L = the exact lower limit of the modal interval, f_m = frequency of the modal class, f_1 = frequency of the class preceding modal class, f_2 = frequency of the class succeeding modal class, i = the size of the class interval.

You can find the graphical representation of mode in grouped data in Figure 16.1.

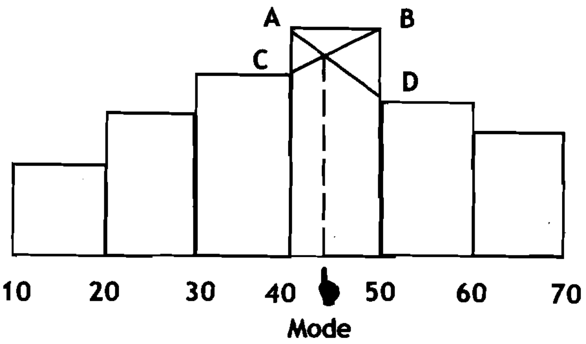


Figure 16.1 Graphical Representation of Mode in Grouped Data

The sample mode is the best estimate of population mode. When one samples a symmetrical unimodal population, mode is an unbiased and consistent estimate of mean and median, but it is relatively inefficient and should not be so used. As a measure of central tendency, mode is affected by skewness less than is mean or median, but it is affected by sampling more than these other two measures. Mode, but neither median nor mean, may be used for data on nominal, as well as the ordinal, interval, and ratio scales_of measurement. Mode is not used often in social or biological researches, although it is often interesting to report the number of modes detected in a population, if there are more than one. See Box 16.7 for example 8.

Box 16.7 Example 8: Find the Modal Income on the Basis of the Following Data.

Income (in Thousands)	5 - 10	10 - 16	16 - 20	20 - 25	25 - 30	30 -35
No. of Households	8	16	29	22	14	12

Income (in Thousands)	No. of Households
5 - 10	8
10 - 15	15 f_1
Modal Class 15 - 20	29 f_m
20 - 25	22 f_2
25 - 30	14
30 - 35	12

Mode lies in the (16 - 20) having the maximum frequency (29)
Lower limit of the modal class = 16
Frequency of the modal class (f_m) = 29
Frequency of the class preceding modal class (f_1) = 16
Frequency of the class succeeding modal class (f_2) = 22
Size of the class interval = 5
Mode = $L + [(f_m - f_1) / (2f_m - f_1 - f_2)] * i$
Mode = 16 + [(29 - 16) / (2*29 - 16 - 22)] *5 = 16 + (14 / 21) * 5 = 16 + 3.33 = 18.33
The modal income is 18.33 thousands.

After learnign about mean, median and mode, we will discuss in Section 16.5 the relationship among the three measures of central tendency. But before going on to Section 16.5, let us complete Reflection and Action 16.2.

Reflection and Action 16.2

Make a graphical representation of mode in grouped data of your choice along the lines of Figure 16.1. You may then use similar type of graphic representation of grouped data in your own mini research project.

16.5 Relationship between mean, mode and median

Mean, mode and median (the three measures of central tendency) are related to each other and can be calculated using the following equation.

Mode = 3 * Median - 2 * Mean

The values of mean, mode and median are the same when the frequency is normally distributed, but their values differ when the frequency is positively or negatively skewed.

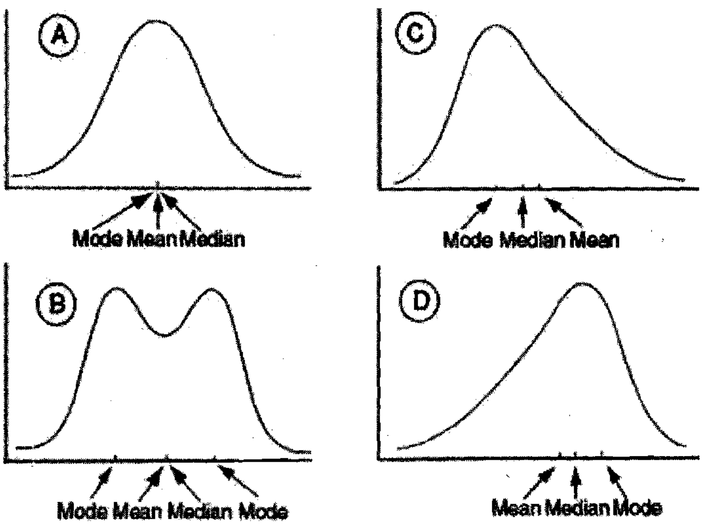


Fig. 16.2: Relationship of Mean, Mcde and Median in various Types of Frequency Distributions

Fig. 16.2 shows the relationship of mean, mode and median in various types of frequency distributions: (A) Normal distribution (B) Bimodal distribution (C) Positively skewed distribution (D) Negatively skewed distribution. Values of the variables are along x axis and the frequencies are along y axis.

After learning about the relationship among the three measure of central tendency, let us find out how to decide which of the three to choose for one's research.

16.6 Choosing a measure of central tendency

Sometimes the researcher has to decide which of the three measures of central tendency to use. The following advice may be of help.

Mean is doubtless the most commonly used measure of central tendency. It is the only one of the three measures which uses all the information available in a set of data, that is to say, it reflects the value of each score in a distribution. It has the decided advantage of being capable of combining with the means of other groups measured on the same variable. For example, from the average unemployment levels in various states of India one can compute the overall mean unemployment rate of India. Since neither the median nor the mode is based on arithmetic, this useful application is not possible. The precisely defined mathematical value of the mean allows the other advanced statistical techniques to be based on it too.

There are occasions, however, when taking into account the value of every score in a distribution can give a distorted picture of the data. For example, marriage distance (the distance between the places of residence of the two partners) in five cases is 40, 60, 60, 80 and 810. Without the very atypical score of 810, the mean score of the group is 60 and the median, likewise, is 60. The effect of introducing the score of 810 is to pull the mean in the direction of that extreme value. The mean now becomes 210, a value that is unrepresentative of the series. The median remains 60, providing a more realistic description of the distribution than the mean.

With these observations in mind:

Use the mean

i) When the scores in a distribution are more or less symmetrically grouped about a central point.

ii) When the research problem requires a measure of central tendency that will also form the basis of other statistics (such as measures of variability or measures of association).

iii) When the research problem requires the combination of mean with the means of other groups measured on the same variable.

iv) To measure the central tendency in a sample of observations when one needs to estimate the value of a corresponding mean of the population from which the sample is taken.

v) When the interval level or ratio level data providing that the distribution of scores approximates a normal curve.

Use the median

i) When the research problem calls for knowledge of the exact midpoint of a distribution.

ii) When extreme scores are there in the series, as they distort the mean, but not the median. Particularly, when dealing with 'oddly-shaped' distributions, for example, those in which a high proportion of extremely high scores occur as well as a low proportion of extremely low ones.

Use the mode

i) When all that is required is a quick and appropriate way of determining central tendency.

ii) When in referring to what is 'average', the word is used in the sense of the 'typical' or the 'most usual'. For example, in talking about the average take-home pay of the coffee plantation worker, it is the modal wage that is being alluded to rather than an exact arithmetic average.

Reflection and Action 16.3

Provide examples of data that require mean, median and mode type of calculations for reflecting the central tendency of the data.

16.7 Conclusion

Succinctly, mode would be the appropriate statistic to use as a measure of the 'most fashionable' or 'most popular' when data are collected using a nominal scale. *Median* would generally be associated with the ordinal level data. *Mean* will be used with interval level or ratio level data providing that the distribution of scores approximates a normal curve.

You can take mean to be a mathematical measure and median mode to be the positional measures. You can always cluster your observations around a central value. A central value manifests both the distribution and the comparison of various distributions. It is always useful for a researcher to provide measures that indicate the average feature of a frequency distribution. Unit 16 has discussed the three measure of central tendency and provided skills of basic statistical tools for application in your research.

It would have become apparent to you that the three measures of the central tendency, namely, i) average of all the values in the distribution or mean, ii) mid-point of the distribution or median and iii) highest density in the distribution or mode, are not to applied in a mechanical way. In the light of the objective of your study you would need to determine when you are to use which measure. You have learnt in Unit 16 that a graphic representation of distributions shows either a symmetrical or a skewed pattern. In symmetrical type, you will find that the three values coincide. This provides you the option of using the mean. In the case of bi-modal or multi-modal representation, you would do better to use the mode. In skewed distribution, if the tail is on the right side, it indicates the positive skewing of distribution. If the tail is on the left side, it shows the negative skewing of distribution. For both the negative and the positive types of skewing of distribution, you would do better by using the median measure of central tendency. You may want to work out with the help of Unit 16 the type of measure of central tendency you will use in your mini research project.

Further Reading

Black, Thomas R. 1999. *Doing Quantitative Research in the Social Science*. An Integrated Approach to Research Design, Measurement and Statistics

Nachmias, David and Chava Nachmias 1981. *Research Methods in Social Sciences*. St. Martin Press: New York.