**Chapter Three**

**Measures of Central Tendency**

**3. Introduction**

* When we want to make comparison between groups of numbers it is good to have a single value that is considered to be a good representative of each group. This single value is called the **average** of the group. Averages are also called measures of central tendency.
* A single value that describes the characteristics of the entire mass of data is called measures of central tendency or average.
* An average which is representative is called typical average and an average which is not representative and has only a theoretical value is called a descriptive average. A typical 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.

**3.1. Objectives of Measure of Central Tendency**

Objectives of measuring central tendency are:

* To get a single value that represent(describe) characteristics of the entire data
* To summarizing/reducing the volume of the data
* To facilitating comparison within one group or between groups of data
* To enable further statistical analysis

3.2. The Summation Notation

Let a data set consists of a number of observations, represents by  where *n* (the last subscript) denotes the number of observations in the data and  is the *ith* observation. Then the sum

* The symbol is a mathematical shorthand for

For instance a data set consisting of six measurements 21, 13, 54, 46, 32 and 37 is represented by and where = 21, = 13, = 54, = 46, = 32 and = 37.

Their sum becomes 21+13+59+46+32+37=208.

Similarly = 

**Some Properties of the Summation Notation**

1. =  where  is a constant number.
2.  where is a constant number
3.  where and are constant numbers
4.  0
5. 3+
6. 

Example:

Let 

Find 

Solution: 



**3.3. Types of Measures of Central Tendency**

Several types of averages or measures of central tendency can be defined, the most commons are

* + the mean
  + the mode
  + the median

1. **The Mean**

* There are four types of means: Arithmetic mean, Weighted arithmetic mean, Harmonic mean and Geometric mean.
* **Arithmetic mean** is defined as the *sum of the measurements of the items divided by the total number of items*.
* The arithmetic mean of a sample is the sum of all the observation divided by the number of observations in the sample. i.e.

* Suppose that are n observed values in a sample of size n taken from a population of size N. Then the ***arithmetic mean of the sample***, denoted by , is given by

* If we take an entire population, the ***population mean*** denoted by **µ** is given by

**Arithmetic Mean for Ungrouped Frequency Distribution**

When the data are arranged or given on the form of ungrouped frequency distribution, then the formula for the mean is

***Example 1:*** The net weights of five perfume bottles selected at random from the production line

What is the arithmetic mean weight of the sample observation?

***Solution:***

= = .

*Example 2:* You measure the body lengths (in inches) of 10 full-term infants at birth and record the following:

17.5, 19.5, 17.5, 19, 20, 21, 18, 19.5, 18, 10.75

Compute the mean length of the infants for these data.

Solution:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Xi | 10.75 | 17.5 | 18 | 19 | 19.5 | 20 | 21 | Total |
| Fi | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 10 |
| Xifi | 10.75 | 35 | 36 | 19 | 39 | 20 | 21 | 179.5 |



*Example 3:* Monthly incomes of fourth year regular students are given in the following frequency distribution.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Monthly income (birr) | 54.5 | 64.5 | 74.5 | 84.5 | 94.5 | 104.5 | 114.5 |
| Number of students | 6 | 9 | 15 | 25 | 13 | 7 | 5 |

Compute the mean for these data.(Exercise)

**Arithmetic Mean for Grouped Frequency Distribution**

If data are given in the form of continuous frequency distribution, the sample mean can be computed as



Where  is he class mark of the class; *i* = 1, 2, …, *k*

 = the frequency of the class and *k* = the number of classes

Note that= the total number of observations.

*Example1:* The following table gives the daily wages of laborers. Calculate the average daily wages paid to a laborer.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Wages in birr | 11-13 | 13-15 | 15-17 | 17-19 | 19-21 | 21-23 | | 23-25 |
| Number of laborers | 3 | 4 | 5 | 6 | 6 | 4 | | 3 |
| Class mark(mi) | 12 | 14 | 16 | 18 | 20 | | 22 | 24 |
| Fimi | 36 | 56 | 80 | 108 | 120 | | 88 | 72 |
| Total |  |  |  |  |  | |  | 560 |



***Example 2:*** The net income of a sample of large importers of Urea was organized into the

following table. What is the arithmetic mean of net income?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Net income (CI) | 2-4 | 5-7 | 8-10 | 11-13 | 14-16 |
| Number of importers ()) | 1 | 4 | 10 | 3 | 2 |

***Solution****:* is the sum of the frequencies or total number of observations.

To calculate consider the following table.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Net income (CI) | 2-4 | 5-7 | 8-10 | 11-13 | 14-16 | Total |
| Number of importers () | 1 | 4 | 10 | 3 | 2 | 20 |
| Class marks () | 3 | 6 | 9 | 12 | 15 |  |
|  | 3 | 24 | 90 | 36 | 30 | 183 |

.

***Example 3:*** From the following data, calculate the missing frequency? The mean number of tablets to cure ever was 29.18.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Number of tablets |  |  |  |  |  |  |  |
| Number of persons cured | 6 | 13 | 19 |  | 18 | 12 | 9 |

***Solution****:* is the sum of the frequencies or total number of observations.

To calculate consider the following table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CI |  |  |  |  |  |  |  | Total |
|  | 6 | 13 | 19 |  | 18 | 12 | 9 |  |
|  | 20 | 23 | 26 | 29 | 32 | 35 | 38 |  |
|  | 120 | 299 | 494 |  | 576 | 420 | 342 |  |

Exercises:

1. Marks of 75 students are summarized in the following frequency distribution:

|  |  |
| --- | --- |
| Marks | No. of students |
| 40-44 | 7 |
| 45-49 | 10 |
| 50-54 | 22 |
| 55-59 | f4 |
| 60-64 | f5 |
| 65-69 | 6 |
| 70-74 | 3 |

If 20% of the students have marks between 55 and 59

1. Find the missing frequencies f4 and f5.
2. Find the mean.

**Properties of the Arithmetic Mean**

* The sum of the deviations of the items from their arithmetic mean is zero. This means, the algebraic sum of the deviations of a set of numbers from their mean is zero.

That is

E.g., the mean of then

* The sum of the squares of the deviations of a set of observations from any number, say A, is minimum when A= . That is,
* When a set of observations is divided into *k* groups and  is the mean of observations of group 1, is the mean of observations of group2, …,  is the mean of observations of group *k* , then the combined mean ,denoted by , of all observations taken together is given by

***Example 1:***  Compute the ***combined*** mean for the following two sets.

***Solution:***

* If a wrong figure has been used in calculating the mean, we can correct if we know the correct figure that should have been used. Let
* denote the wrong figure used in calculating the mean
* be the correct figure that should have been used
* be the wrong mean calculated using , then the correct mean, , is given by



*Example 2:* An average weight of 10 students was calculated to be 65 kg, but latter, it was discovered that one measurement was misread as 40 kg instead of 80 kg. Calculate the corrected average weight.

Solution:



* If the mean of is, then
  1. the mean of  will be
  2. The mean of will be.

*Example 1:* Last year there were three sections taking Stat 273 course in Alemaya University. At the end of the semester, the three sections got average marks of 80, 83 and 76. There were 28, 32 and 35 students in each section respectively. Find the mean mark for the entire students.

Solution:

79.54

*Exercise:* The average score on the mid-term examination of 25 students was 75.8 out of 100. After the mid-term exam, however, a student whose score was 41 out of 100 dropped the course. What is the average/mean score among the 24 students?

**Weighted Mean**

In finding arithmetic mean, all items were assumed to be of equally importance (each value in the data set has equal weight). When the observations have different weight, we use weighted *average*. Weights are assigned to each item in proportion to its relative importance.

If represent values of the items and are the corresponding weights, then the weighted mean, is given by

*Example1:* A student’s final mark in Mathematics, Physics, Chemistry and Biology are respectively 82, 80, 90 and 70.If the respective credits received for these courses are 3, 5, 3 and 1, determine the approximate average mark the student has got for one course.

Solution: We use a weighted arithmetic mean, weight associated with each course being taken as the number of credits received for the corresponding course.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | |  | 82 | 80 | 90 |  | 70 | |  | 3 | 5 | 3 |  | 1 | |

Therefore 

Average mark of the student for one course is approximately 82.

***Example2:*** If a student scores **"A"** in a **3 EtCTS** course, **"B"** in a **6 EtCTS** course, **"B"** in another **5 EtCTS** course and **"D"** in a **2 EtCTS** course. Compute his /her GPA for the semester.

***Solution****:* Here the numerical values of the letter grades are the values (i.e. ) and the corresponding EtCTS of the course are their respective weights.

i.e.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade values () | 4 | 3 | 3 | 1 |
| Weight () | 3 | 6 | 5 | 2 |

=

**Merits of Arithmetic Mean**

* Arithmetic mean has a rigidly defined mathematical formula so that its value is always definite.
* It is calculated based on all observations.
* Arithmetic mean is simple to calculate and easy to understand.
* It doesn’t need arrangement of data in increasing or decreasing order.
* Arithmetic mean is also capable of further algebraic treatment.
* It affords a good standard of comparison.

**Drawbacks of Arithmetic Mean**

* It is highly affected by extreme (abnormal) values in the series.
* It can be a number which does not exist in the series.
* It sometime gives such results which appear almost absurd. For example it is likely that we can get an average of ‘3.6 children’ per family.
* It can’t be calculated for open-ended classes.

**Geometric Mean:** It used when observed values are measured as ratios, percentages, proportions, indices or growth rates.

,

If the observed have frequencies 

**Example**: compute the geometric mean of the following values: 2, 8, 6, 4, 10, 6, 8, 4

Solution:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Values | 2 | 4 | 6 | 8 | 10 | Total |
| frequencies | 1 | 2 | 2 | 2 | 1 | 8 |



* In general, the sample geometric mean is calculated by

**Harmonic Mean**: is a suitable measure of central tendency when the data pertains to speed, rate and time. 

If the data arranged in the form of frequency distribution



**Example:** A motorist travels 480km in 3 days. She travels for 10 hours at rate of 48km/hr on 1st day, for 12 hours at rate of 40km/hr on the 2nd day and for 15 hours at rate of 32km/hr on the 3rd day. What is her average speed?



**Relations among different means**

1. 
2. For two observations 
3. if all observation have equal magnitude
4. **The Median**

The median of a set of items (numbers) arranged in order of magnitude (i.e. in an array form) is the middle value or the arithmetic mean of the two middle values. We shall denote the median of by. For ungrouped data the median is obtained by



***Example 1:*** Find the median for the following data:

***Solution***: First arrange the given data in increasing order. That is

***Example 2:*** Find the median for the following data:

***Solution:*** First arrange the given data in increasing order. That is

* For grouped data the median, obtained by interpolation method, is given by

Where lower class boundary of the median class

Sum of frequencies of all class lower than the median class (in other words it is the cumulative frequency immediately preceding the median class)

Frequency of the median class and  is class width

The *median class* is the class with the smallest cumulative frequency greater than or equal to.

*Examples 2:* The following table gives the distribution of the weekly wages of employees of a small firm.

|  |  |
| --- | --- |
| Wages in birr | No. of employees |
| 126 and below | 3 |
| 127 – 135 | 5 |
| 136 – 144 | 9 |
| 145 – 153 | 12 |
| 154 – 162 | 5 |
| 163 – 171 | 4 |
| 172 and above | 2 |

* 1. Find the median weekly wage.
  2. Why is the median a more suitable measure of central tendency than the mean in this case?

Solution:

* First find the less than cumulative frequency.
* Identify the median class.
* Find median using formula.

|  |  |  |
| --- | --- | --- |
| Wages in birr | No. of employees | Cumu.Freq(less than type |
| 126 and below | 3 | `3 |
| 127 – 135 | 5 | 8 |
| 136 – 144 | 9 | 17 |
| 145 – 153 | 12 | 29 |
| 154 – 162 | 5 | 34 |
| 163 – 171 | 4 | 38 |
| 172 and above | 2 | 40 |







**Merits of median**

* It is not influenced by extreme values.
* Arithmetic mean is rigidly defined a mathematical formula so that its value is always definite.
* Median can be calculated even in case of open-ended intervals.
* It can be computed for ratio, interval, and ordinal level of data.

**Demerits of median**

* It is not capable of further algebraic treatment.
* It is not a good representative of the data if the number of items (data) is small.
* The arrangement of items in order of magnitude is sometimes very tedious process if the number of items is very large.

1. **The Mode**

The mode or the modal value is the most frequently occurring score/observation in a series and denoted by. Note that the mode may not exist in the series or, even if it does exist, it may not be unique.

For grouped data, the mode is found by the following formula:



Wherelower class boundary of the modal class

The difference between the frequency of the modal class and frequency of the class

immediately preceding the modal class

The difference between the frequency of the modal class and frequency of the class

Immediately follows the modal class

is the class width

The *modal class* is the class with the highest frequency in the distribution.

*Examples 1:* The marks obtained by ten students in a semester exam in statistics are: 70, 65, 68, 70, 75, 73, 80, 70, 83 and 86. Find the mode of the students’ marks.

Mode=70

* 1. Find the mode of 5, 3, 5, 8, 9

Mode =5

* 1. Find the mode of 8, 9, 9, 7, 8, 2, and 5.

It is a bimodal Data: 8 and 9

* 1. Find the mode of 4, 12, 3, 6, and 7.

No mode for this data

*Example 2:* Find the mode for the frequency distribution of the birth weight (in kilogram) of 30 children given below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Weight | 1.9-2.3 | 2.3-2.7 | 2.7-3.1 | 3.1-3.5 | 3.5-3.9 | 3.9-4.3 |
| No. of children | 5 | 5 | 9 | 4 | 4 | 3 |

**Solution:** 2.7-2.3 is the modal class since it has the highest frequency

 



***Example 3:*** The ages of newly hired, unskilled employees are grouped into the following distribution. Then compute the modal age?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Ages |  |  |  |  |  |
| Number | 4 | 8 | 11 | 20 | 7 |

***Solution****:* First we determine the modal class. The modal class is , since it has the highest frequency.

***Interpretation:*** The age of most of these newly hired employees is 27.7 (27 years and 7 months).

**Merits of mode**

* Mode is not affected by extreme values.
* Mode can be calculated even in the case of open-end intervals. And it is not necessary to know all observations.
* It can be computed for all level of data i.e. ratio, interval, ordinal or nominal.

**Demerits of mode**

* Mode may not exist in the series and if it exists it may not be a unique value.
* It does not fulfill most of the requirements of a good measure of central tendency

**3.4. Measure of non-central tendency or Position (Quantiles)**

Quantiles are values which divides the data set arranged in order of magnitude in to certain equal parts. They are averages of position (non-central tendency). Some of these are quartiles, deciles and percentiles.

1. **Quartiles:** are values which divide the data set in to four equal parts, denoted byand . The first quartile is also called the lower quartile and the third quartile is the upper quartile. The second quartile is the median.

* For Ungrouped data:

Let be the quartile value for. Then



* For grouped data

We can apply the following formula:



Wherethequartile we are going to calculate

Lower class boundary of thequartile class

Sum of frequencies of all classes lower than the  quartile class

Frequency of thequartile class and Class width

**The quartile class** is the class with the smallest cumulative frequency greater than or equal to**.**

1. **Deciles:** are values dividing the data in to ten equal parts, denoted by. The fifth decile is the median.
   * For Ungrouped data

Let be thepercentile value for. Then



* For grouped data

We can apply the following formula:



Define the symbols similar way as we did in the case of quartiles.

**The  decile class** is the class with the smallest cumulative frequency greater than or equal to**.**

**Percentiles:** are values which divide the data in to one hundred equal parts, denoted by. The fiftieth percentile is the median.

* + - * + For ungrouped data

Letbe the percentile value for. Then



* For grouped data

We can use the following formula:



Define the symbols similar way as we did in the case of quartiles.

**The percentile class** is the class with the smallest cumulative frequency greater than or equal to.

**Interpretations**

1. is the value below which of the observations in the series are found (where). For instance means the value below which 75 percent of observations in the given series are found.
2. Is the value below whichof the observations in the series are found (where). For instance is the value below which 40 percent of the values are found in the series.
3. is the value below whichof the total observations are found (where). For example 73 percent of the observations in a given series are below.

**Example**: Considering the following distribution

Calculate:

1. All quartiles.
2. The 7th decile.
3. The 90th percentile.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | Values | |  | Frequency | | 140- 149 | |  | 17 | | 150- 159 | |  | 29 | | 160- 169 | |  | 42 | | 170- 179 | |  | 72 | | 180- 1189 | |  | 84 | | 190- 199 | |  | 107 | | 200- 209 | |  | 49 | | 210- 219 | |  | 34 | | 220- 229 | |  | 31 | | 230- 239 |  | | 16 | | 240- 249 |  | | 12 | |

Solutions:

* First find the less than cumulative frequency.
* Use the formula to calculate the required quantile.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | | Cum.Freq(less than type) | |
| Values |  | Frequency |  |  |
|  |  |  |  |  |
| 140- 149 |  | 17 |  | 17 |
| 150- 159 |  | 29 |  | 46 |
| 160- 169 |  | 42 |  | 88 |
| 170- 179 |  | 72 |  | 160 |
| 180- 1189 |  | 84 |  | 244 |
| 190- 199 |  | 107 |  | 351 |
| 200- 209 |  | 49 |  | 400 |
| 210- 219 |  | 34 |  | 434 |
| 220- 229 |  | 31 |  | 465 |
| 230- 239 |  | 16 |  | 481 |
| 240- 249 |  | 12 |  | 493 |

1. Quartiles:
   1. Q1

- determine the class containing the first quartile.



 

* 1. Q2 - determine the class containing the second quartile.





* 1. Q3

- determine the class containing the third quartile.



|  |
| --- |
|  |





1. D7

- determine the class containing the 7th decile.







1. P90

- determine the class containing the 90th percentile.







Exercise: The following table presents the male population of a certain region in Ethiopia.

Find a) all quartiles

b) The anddecile and

c) and percentiles

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age groups (in years) | 0 – 5 | 5 – 10 | 10 – 15 | 15 – 20 | 20 – 25 | 25 – 30 | 30 – 35 | 35 - 40 |
| Male population | 2580 | 3737 | 4620 | 5200 | 7250 | 620 | 297 | 355 |