**Chapter Four**

**4. Measures of Dispersion (Variation) and Shape**

**4.1. Introduction**

Variation (dispersion) is the scatter or spread of observations /values/ in a distribution. The average or central value is of little use unless the degree of variation, which occurs about it, is given. If the scatter about the measure of central tendency is very large, the average is not a typical value. Therefore it is necessary to develop a quantitative measure of the dispersion (or variation) of the values about the average.

* Measures of variation are statistical measures, which provide ways of measuring the extent to which the data are dispersed or spread out.

**4.2. Objectives of Measuring Variation**

Measures of variation are needed for the following basic objectives.

* To judge the reliability of a measure of central tendency
* To compare two or more sets of data with regard to their variability
* To control variability itself
* To make further statistical analysis

**Properties of a good measure of dispersion**

A good measure of dispersion should:

* be rigidly defined by a mathematical formula,
* be simple to understand and easy to calculate,
* be unique,
* calculated based on all observations in the series,
* not be affected by some extreme values existing in the series,
* have sampling stability property, and
* be capable of further algebraic treatment as well as further statistical analysis.

**4.3. Types of Measures of Dispersion**

**4.3.1. Absolute Measures of Dispersion**

The measures of dispersion which are expressed in terms of the original unit of a series are termed as absolute measures. Such measures are not suitable for comparing the variability of two distributions which are expressed in different units of measurement and different average size.

4.3.2. Relative Measures of Variation

Relative measures of dispersions are a ratio or percentage of a measure of absolute dispersion to an appropriate measure of central tendency and are thus pure numbers independent of the units of measurement.

* For comparing the variability of two distributions (even if they are measured in the same unit), we compute the relative measure of dispersion instead of absolute measures of dispersion.

**I. The Range and Relative Range**

*Range (R)* is defined as the difference between the largest and the smallest observation in a given set of data. That is, where *xmax* and *xmin* are the largest and the smallest observations in the series respectively.

In case grouped data, range is found by taking the difference between the class mark of the last class and that of the first class. That is, where  and  are the class marks of the last class and that of the first class respectively.

**A *relative range (RR****)*, also known as *coefficient of range*, is given by



**Properties of Range and Relative Range**

* Range and relative range are easy to calculate and simple to understand.
* Both cannot be computed for grouped data with open ended classes.
* They do not tell us anything about the distribution of values in the series.

Example 1: Find the range and relative range for the monthly salary of ten workers in a certain paint factory given below.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 462 | 480 | 534 | 624 | 498 | 552 | 606 | 588 | 516 | 570 |

Solution:



Example 2: Find the values of the range and relative range for the following frequency distribution: which shows the distribution of the maximum loads supported by a certain number of cables.

|  |  |
| --- | --- |
| Maximum load  (in kilo-Newton) | Number of cables |
| 93 – 97 | 2 |
| 98 – 102 | 5 |
| 103 – 107 | 12 |
| 108 – 112 | 17 |
| 113 – 117 | 14 |
| 118 – 122 | 6 |
| 123 – 127 | 3 |
| 128 – 132 | 1 |

Solution:



**II. The Mean Deviation and Coefficient of Mean Deviation**

The *mean deviation (MD)* measures the average deviation of a set of observations about their central value, generally the mean or the median, ignoring the plus/minus sign of the deviations.

The mean deviation of a sample of n observations  is given as

 Where A is a central measure (the mean or the median)

In case of grouped data, the formula for MD becomes

 Where is the class mark of theclass, is the frequency of the class and.

* The mean deviation about the arithmetic mean is, therefore, given by

for ungrouped data

 for grouped frequency distribution; where is the class mark of theclass, is the frequency of the class and 

* The mean deviation about the median is also given by

for ungrouped data

 for grouped frequency distribution; whereis the class mark of theclass, is the frequency of the class and .

**Examples**:

1. The following are the number of visit made by ten mothers to the local doctor’s surgery. 8, 6, 5, 5, 7, 4, 5, 9, 7, 4

Find mean deviation about mean, median and mode.

**Solutions**:

First calculate the three averages



Then take the deviations of each observation from these averages.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Xi | 4 | 4 | 5 | 5 | 5 | 6 | 7 | 7 | 8 | 9 | total |
|  | 2 | 2 | 1 | 1 | 1 | 0 | 1 | 1 | 2 | 3 | 14 |
|  | 1.5 | 1.5 | 0.5 | 0.5 | 0.5 | 0.5 | 1.5 | 1.5 | 2.5 | 3.5 | 14 |
|  | 1 | 1 | 0 | 0 | 0 | 1 | 2 | 2 | 3 | 4 | 14 |





1. Find mean deviation about mean, median and mode for the following distributions.(exercise)

|  |  |  |
| --- | --- | --- |
| Class |  | Frequency |
|  |  |  |
| 40-44 |  | 7 |
| 45-49 |  | 10 |
| 50-54 |  | 22 |
| 55-59 |  | 15 |
| 60-64 |  | 12 |
| 65-69 |  | 6 |
| 70-74 |  | 3 |

**Remark**: Mean deviation is always minimum about the median.

**The *coefficient of mean deviation (CMD)*** is the ratio of the mean deviation of the observations to their appropriate measure of central tendency: the arithmetic mean or the median.

In general, where A is a measure of central tendency: the arithmetic mean or the median.

That is, CMD about the arithmetic mean is given bywhere MD is the mean deviation calculated about the arithmetic mean. On the other hand CMD about the median is given by in which case MD is calculated about the median of the observations.

**Properties of Mean Deviation and coefficient of mean deviation**

* It is easy to understand and compute
* It is based on all observations
* It is not affected very much by the values of extreme value(s).
* It is not capable of further mathematical treatments and it is not a very accurate measure of dispersion.

Example: calculate the C.M.D about the mean, median and mode for the data in example 1 above.

**Solutions:**









**III. The Variance, the Standard Deviation and Coefficient of Variation**

**The Variance**

*Variance* is the arithmetic mean of the square of the deviation of observations from their arithmetic mean.

* Population Variance ()

**For ungrouped data**

Whereis the population arithmetic mean and *N* is the total number of observations in the population.

**For grouped data**

Whereis the population arithmetic mean, is the class mark of theclass, is the frequency of theclass and.

* Sample Variance ()

**For ungrouped data**

Whereis the sample arithmetic mean and *n* is the total number of observations in the sample.

**For grouped data**

Whereis the sample arithmetic mean, is the class mark of theclass, is the frequency of theclass and.

**The Standard Deviation**

*Standard deviation* is the positive square root of the variance.

* Population Standard Deviation ()

whereis the population variance.

* Sample Standard Deviation ()

where is the sample standard deviation.

**Example 1:** compute the variance for the following data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **value** | **3** | **6** | **9** | **12** | **15** | **Total** |
| **frequency** | **1** | **4** | **10** | **3** | **2** | **20** |
|  | **3** | **24** | **90** | **36** | **30** | **183** |
|  | **-6.15** | **-3.15** | **-0.15** | **2.85** | **5.85** |  |
|  | **37.8225** | **9.9225** | **0.0225** | **8.1225** | **34.2225** |  |
|  | **37.8225** | **39.69** | **0.225** | **24.3675** | **68.445** | **170.55** |



And 

**Coefficient of Variation**

The standard deviation is an absolute measure of dispersion. The corresponding relative measure is known as the *coefficient of variation (CV)*.

**Coefficient of variation** is used in such problems where we want to compare the variability of two or more different series. Coefficient of variation is the ratio of the standard deviation to the arithmetic mean, usually expressed in percent.

. Where *S* is the standard deviation of the observations.

A distribution having less coefficient of variation is said to be less variable or more consistent or more uniform or more homogeneous.

Example: Last semester, the students of Biology and Chemistry Departments took *Stat 273* course. At the end of the semester, the following information was recorded.

|  |  |  |
| --- | --- | --- |
| **Department** | **Biology** | **Chemistry** |
| Mean score | 79 | 64 |
| Standard deviation | 23 | 11 |

Compare the relative dispersions of the two departments’ scores using the appropriate way.

Solution:

|  |  |
| --- | --- |
| Biology Department | Chemistry Department |
|  |  |

**Interpretation**: Since the CV of Biology Department students is greater than that of Chemistry Department students, we can say that there is more dispersion relative to the mean in the distribution of Biology students’ scores compared with that of Chemistry students.

**Example: 1.** The following table illustrates the frequency distribution of masses of 100 male students in Gander University.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mass (kg) | 60-62 | 63-65 | 66-68 | 69-71 | 72-74 |
| No. of students | 5 | 18 | 42 | 27 | 8 |

Find: a) the variance b) the standard deviation c) the coefficient of variation

d) Calculate mean deviation?

**Solution:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mass (kg) | 60-62 | 63-65 | 66-68 | 69-71 | 72-74 | total |
| No. of students(fi) | 5 | 18 | 42 | 27 | 8 | 100 |
| class mark(mi) | 61 | 64 | 67 | 70 | 73 |  |
| *fi mi* | 305 | 1152 | 2814 | 1890 | 584 | 6745 |
| *fi mi2* | 18605 | 73728 | 188538 | 132300 | 42632 | 455803 |
|  | 6.45 | 3.45 | 0.45 | 2.55 | 5.55 |  |
| *fi* | 32.25 | 62.1 | 18.9 | 68.85 | 44.4 | 226.5 |

,  , 

and 

1. 
2. 2.93
3. 
4. 

2. A meteorologist interested in the consistency of temperatures in three cities during a given week collected the following data. The temperatures for the five days of the week in the three cities were

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| City 1 | 25 | 24 | 23 | 26 | 17 |
| City2 | 22 | 21 | 24 | 22 | 20 |
| City3 | 32 | 27 | 35 | 24 | 28 |

Which city have the most consistent temperature, based on these data? (Exercise)