**Moments and Measures of Skewness and Kurtosis**

We have learned so far the measure of central tendency and dispersion. The most appropriate measure of central tendency and dispersion are arithmetic mean () and the variance (, where

or, , where

or,

The most general measure of this type is

,

which is called, moment of around the origin

If the given values are classified into a frequency table, the formula takes the form

,

where is the class-mark of class and is its frequency

**Moments**

Given observations , and an arbitrary constant ,

is called the 1st moment about

is called the 2nd moment about

is called the 3rd moment about

and so on. Sometimes, these are also represented by ,, , etc.

*Moments about zero* (i.e. when ) and *moments about mean* (i.e. when ) are particularly important.

**Moments about zero (or Raw moments)**

1st moment about zero,

2nd moment about zero,

3rd moment about zero,

4th moment about zero,

.

.

rth moment about zero,

Note that the 1st moment about zero is the mean , i.e. .

**Moments about mean (or Central moments)**

1st moment about mean,

2nd moment about mean,

3rd moment about mean,

4rth moment about mean,

.

.

rth moment about mean,

Sometimes, the central moments are represented by , , , etc.

Note that the central moment is always zero and the central moment is the variance , i.e.

**and**

It may be noted that **,** and the mean of a variable is its first moment about 0, while the variance is the second central moment.

The 3rd central moment is used to measure skewness and the 4th central moment is used to measure kurtosis. Higher order moments , etc are seldom used.

**Central moments expressed in terms of moments about an arbitrary origin**

Summing both sides for all from 1 to , and dividing by , we get

(1)

[Since, ]

*It may be noted that the relation holds for moments obtained from grouped data as well.*

**Important relation between central and non-central moments**

In particular, putting 1, 2, 3 and 4 in equation (1) we get,

In most practical cases, it will be sufficient to calculate , , and . These computations are greatly facilitated by first computing moments about a suitably chosen origin and then using these relations.

**Problem 1:** The yields in gm. of 12 tomato plants are given as follows: 1216, 1374, 1167, 1232, 1407, 1453, 1202, 1372, 1278, 1141, 1221 and 1329. Compute , , and for the data on yields of the tomato plants.

***Solution:*** Here the values are quite large. So the direct computation of the first four moments will be extremely laborious, as this will require obtaining squares, cubes and forth powers of the given quantities and finding their totals. The computational labour can be reduced to a great extent by taking deviations of the given values about a suitable origin and then computing moments about that origin.

In the present case, we may take the origin at 1300 gm. The different steps in the calculation of moments about 1300 gm. are indicated in the following table.

**Table 1:** Calculation of moments for the data on yield of tomato plants

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Yield (gm.) () |  |  |  |  |
| 1216 | -84 | 7056 | -592704 | 49787136 |
| 1374 | 74 | 5476 | 405224 | 29986576 |
| 1167 | -133 | 17689 | 2352637 | 312900721 |
| 1232 | -68 | 4624 | 314432 | 21381376 |
| 1407 | 107 | 11449 | 1225043 | 131079601 |
| 1453 | 153 | 23409 | 3581577 | 547981281 |
| 1202 | -98 | 9606 | -941192 | 92236816 |
| 1372 | 72 | 5184 | 373284 | 26873856 |
| 1278 | -22 | 484 | -10648 | 234256 |
| 1141 | -159 | 25281 | -4019679 | 639128961 |
| 1221 | -79 | 6241 | -493039 | 38950081 |
| 1329 | 29 | 841 | 24387 | 707281 |
| Total | -208 | 117338 | -31114850 | 1891247942 |

From this table, we get

gm

(gm)2

(gm)3

(gm)4

Hence the mean, the SD and the third and fourth central moments will be as follows:

gm

(gm)2

Therefore, gm

(gm)3

(gm)4

**In case of grouped data**, simplifications can be made in the calculations of moments if the deviations about the chosen origin are divided by a suitable number. The class-width may be used for this purpose when the class-intervals are equally wide.

**Problem 2:** Frequency distribution of height for 177 Indian adult males is given:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 144.55-149.55 | 149.55-154.55 | 154.55-159.55 | 159.55-164.55 | 164.55-169.55 | 169.55-174.55 | 174.55-179.55 | 179.55-184.55 |
| 1 | 3 | 24 | 58 | 60 | 27 | 2 | 2 |

Compute , , and for the data on height of Indian adult males.

***Solution:***

We may take the origin at 162.05 gm and divide values by the class interval, i.e. 5. Then, the calculation of moments can be carried out.

**Table 2:** Calculation of moments for the data on heights

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class-mark  ( | Frequ-ency  ( |  |  |  |  |  |  |
| 147.05 | 1 | -3 | -3 | 9 | -27 | 81 | 16 |
| 152.05 | 3 | -2 | -6 | 12 | -24 | 48 | 3 |
| 157.05 | 24 | -1 | -24 | 24 | -24 | 24 | 0 |
| 162.05 | 58 | 0 | 0 | 0 | 0 | 0 | 58 |
| 167.05 | 60 | 1 | 60 | 60 | 60 | 60 | 960 |
| 172.05 | 27 | 2 | 54 | 108 | 216 | 432 | 2187 |
| 177.05 | 2 | 3 | 6 | 18 | 54 | 162 | 512 |
| 182.05 | 2 | 4 | 8 | 32 | 128 | 512 | 1250 |
| Total | 177 | - | 95 | 263 | 383 | 1319 | 4986 |

**Charlier’s check**

To guard against computational mistakes Charlier gave a simple and effective check on the calculation for moments from a grouped frequency distribution. For a frequency distribution these checks are based on the following relations:

and so on.

*Application of Charlier’s check on the calculation in Table 2.*

Here,

while

The values being equal, the computations may be supposed to be free from errors.

Now we can move forward for calculations of the required moments.

We have, therefore,

cm

(cm)2

Therefore, cm

(cm)3

(cm)4

**Sheppard corrections for moments**

* The raw and central moments for data grouped into class-intervals are obtained by means of the following formulae:

, and

* Here we are acting ***as if the observations falling in a class were all equal to the class-mark, although the observations may be really unequal.***
* The assumption naturally introduces some errors, which are called the **errors due to grouping**.
* **Sheppard** has developed a method for correcting the errors due to grouping. It is applicable only if the class intervals are equal.

**Sheppard’s corrections for raw moments**

* =
* , ( is the width of each class-interval)

**Sheppard’s corrections for central moments**

**The conditions for validity of Sheppard’s corrections**

The Sheppard’s corrections will be valid only if certain conditions are fulfilled. These are:

* The observations should relate to a continuous variable.
* The frequency curve of the distribution should be continuous and should have moderate asymmetry.
* The total frequency should be sufficiently large.
* The width of the class-interval should not be too small compared to the range of variation of the data. In other words, the number of classes should not be too large.

*As a general rule, the Sheppard corrections should not be applied unless the total frequency is higher than 1000 and the number of classes is smaller than 20.*

*These corrections are not applicable to J or U shaped distributions or to very skew distribution.*

**Problem 3:** Show that if and are the first two moments about an arbitrary origin , then mean and standard deviation may be obtained as and .

***Proof:***

**Problem 4**: The first four moments of a distribution about the value 3 are 2, 10, 40 and 218. Find the moments about the origin and the mean.

***Solution:*** Given, , ,

and

We have to find (a) , , , ; and (b) , , and .



or, , or ,

or

or

or

or

or

or

or

or

or

or

or

or

or

1. Using relation between central and non-central moments, we get

(in all cases)

**Exercise 1:** The arithmetic mean of a certain distribution is 5. The second and the third moments about the mean are 20 and 140 respectively. Find the third moment of the distribution about 10. [Ans. -285]

**Exercise 2:** If and are respectively the second moment about an arbitrary origin and that about , then show that , where .

**Exercise 3:** Apply Sheppard’s corrections to the central moments calculated in problem 2.

**Moments about an arbitrary origin expressed in terms of central moments**

, where

Thus,

Summing both sides for all from 1 to , and dividing by , we get

(2)

The last term but one is, of course, zero since

In particular,