 Measures of Dispersion and Variability

The three measures of central tendency- mode, median, and mean-specify the center of a distribution, but provide no indication of the amount of spread or variability in a set of values Dispersion can be calculated in several different ways.

The level of measurement and nature of the frequency distribution determine which dispersion statistic is most appropriate.

The simplest measure of variability is the range, defined as the difference between the largest and smallest values in an interval-ratio set of data.

$$\text{Range} = X_{max} - X_{min}$$

 Because it is derived solely from the two most extreme or atypical values and ignores all other values, the range can be a misleading measure.

Example: find the range for the:
 (20,21,22,25,26,29,33,34,38)

R = 38 - 20 = 18

Example: find the range for the:

 (11,16,19,23,25,32,39,46,52)

R = 52 - 11 =41

In the two Examples above, dataset 1 has a range of 20 – 38 = 18 while dataset 2 has a range of 11 – 52 = 41. Dataset 2 has a broader range and, hence, more variability than dataset 1.

If data are grouped, the range is defined as the difference between the upper value in the highest numbered class interval and the lower value in the lowest numbered class interval.

Relative humidity %	Frequency
51 - 55	2
56 - 60	7
61 - 65	8
66 - 70	4

The range for the grouped Relative humidity % data, are

$$R = 70 - 51 = 19$$

While the range is easy to understand, it is based on only the two most extreme values in the dataset, which makes it very susceptible to outliers. If one of those numbers is unusually high or low, it affects the entire range even if it is atypical.

Quartile deviation

Quartiles are the values that divide a list of numbers into quarters:

Quartile deviation is based on the difference between the first quartile and the third quartile in the frequency distribution and the difference is also known as the interquartile range, the difference divided by two is known as quartile deviation or semi interquartile range.

$$Quartile\ deviation\ = \frac{(Q3 - Q1)}{2}$$

To calculate the Quartile deviation for a simple distribution or frequency distribution we have to put the data in order then cut the order data into four equal parts then takes half of the difference between the 3rd quartile and the 1st quartile.

The Quartile deviation is better than the range because it is not affected by the outliers, excluding the outliers from above and below.

Example:
Consider a data set of following numbers: 22, 12, 14, 7, 18, 16, 11, 15, 12.

You are required to calculate the Quartile Deviation.

Solution:

First, we need to arrange data in ascending order to find Q3 and Q1 and avoid any duplicates.

7, 11, 12, 13, 14, 15, 16, 18, 22

Calculation of Q1 can be done as follows,

$$Q1 = (N + 1) * \frac{1}{4}$$

$$Q1 = (9 + 1) * \frac{1}{4}$$

Q1=2.5 Term

Calculation of Q3 can be done as follows,

$$Q3 = (N + 1) * \frac{3}{4}$$

$$Q3 = (9 + 1) * \frac{3}{4}$$

Q3= 7.5 Term

Calculation of quartile deviation can be done as follows,
Q1 is an average of 2nd and 3rd term of order data which is:

$$Q1 = \frac{(11 + 12)}{2} = 11.5$$

Q3 is an average of 7th and 8th term of order data which is:

$$Q3 = \frac{(16 + 18)}{2} = 17$$

$$\boldsymbol{Quartile\ deviation} = \frac{(\boldsymbol{Q3 - Q1})}{\boldsymbol{2}} = \frac{(\boldsymbol{17 - 11.5})}{\boldsymbol{2}} = \boldsymbol{2.75}$$

Example 2:

calculate the Quartile Deviation for the following numbers:

155, 169, 188, 150, 177, 145, 140, 190, 175, 156.

Solution:

The number of observations here is 10 and our first step would be to arrange data n ascending order.

140, 145, 150, 155, 156, 169, 175, 177, 188, 190

Calculation of Q1 can be done as follows,

$Q1 = (N + 1) * \frac{1}{4} = (10+1) * \frac{1}{4} = 2.75^{th}$ Term

Calculation of Q3 can be done as follows,

$Q3 = (N + 1) * \frac{3}{4} = (10+1) * \frac{3}{4} = 8.25^{th}$ Term

Calculation of quartile deviation can be done as follows,

Q1 is an average of 2nd and 3rd term of order data which is:

$$Q1 = \frac{(145 + 150)}{2} = 148.75$$

Q3 is an average of 7th and 8th term of order data which is:

$$Q3 = \frac{(177 + 188)}{2} = 179.75$$

$$\boldsymbol{Quartile\ deviation} = \frac{(\boldsymbol{Q3 - Q1})}{2} = \frac{(\boldsymbol{179.75 - 148.75})}{2} = 15.5$$

Quartile Deviation for grouped data

Quartile Deviation is given by

$$\boldsymbol{Quartile\ deviation}\ (QD) = \frac{(\boldsymbol{Q3 - Q1})}{2}$$

where,

- Q1 is the first quartile
- Q3 is the third quartile

The formula for i^{th} quartile is

$$Q_i = L + \left(\frac{\frac{iN}{4} - F_<}{f} \right) \times h; \quad i = 1, 2, 3$$

where

- L: the lower limit of the ith quartile class.
- N=$\sum$f : total number of observations.
- f : frequency of the ith quartile class.
- F< : cumulative frequency of the class previous to i^{th} quartile class.
- h: the class width.

Example 1: Compute Inter quartile range for the following frequency distribution.

Solution

	Class Boundaries	Frequencies (fi)	Cumulative frequencies (cf)
1	9.5-12.5	3	3
2	12.5-15.5	12	15
3	15.5-18.5	15	30
4	18.5-21.5	24	54
5	21.5-24.5	2	56
Total		56	

Quartiles

The formula for ith quartile is

Qi=$(\frac{i(N)}{4})^{th}$ value, i=1,2,3

where N is the total number of observations.

First Quartile Q1

Q1 =$(\frac{1(N)}{4})^{th} = (\frac{1(56)}{4})^{th} = (14)^{th}$ value

The cumulative frequency just greater than or equal to 14 is 15. The corresponding class 12.5−15.5 is the 1st quartile class.
Thus

- L=12.5, the lower limit of the 1st quartile class
- N=56, total number of observations
- f=12, frequency of the 1st quartile class
- F<=3, cumulative frequency of the class previous to 1st quartile class
- h=3, the class width

The first quartile Q1 can be computed as follows:

$$Q_1 = l + \left(\frac{\frac{1(N)}{4} - F_<}{f} \right) \times h$$

$$= 12.5 + \left(\frac{\frac{1*56}{4} - 3}{12} \right) \times 3$$

$$= 12.5 + \left(\frac{14 - 3}{12} \right) \times 3$$

$$= 12.5 + (0.9167) \times 3$$

$$= 12.5 + 2.75$$

$$= 15.25$$

The third quartile Q3 can be computed as follows:

$$Q3 = \left(\frac{3(N)}{4}\right)^{th} = \left(\frac{3(56)}{4}\right)^{th} = (42)^{th} \text{ value}$$

The cumulative frequency just greater than or equal to 42 is 54. The corresponding class 18.5−21.5 is the 3rd quartile class.

Thus

- L = 18.5, the lower limit of the 3rd quartile class
- N = 56, total number of observations
- f = 24, frequency of the 3rd quartile class
- F< = 30, cumulative frequency of the class previous to 3rd quartile class
- h = 3, the class width

The third quartile Q3 can be computed as follows:

$$Q_3 = l + \left(\frac{\frac{3(N)}{4} - F_<}{f} \right) \times h$$

$$= 18.5 + \left(\frac{\frac{3*56}{4} - 30}{24} \right) \times 3$$

$$= 18.5 + \left(\frac{42 - 30}{24} \right) \times 3$$

$$= 18.5 + (0.5) \times 3$$

$$= 18.5 + 1.5$$

$$= 20$$

The quartile deviation (QD) is

$$\text{Quartile deviation} = \frac{(Q3 - Q1)}{2} = \frac{(20 - 15.25)}{2} = 2.37$$

standard deviation

Standard deviation is a measure of dispersement in statistics. "Dispersement" tells you how much your data is spread out. Specifically, it shows you how much your data is spread out around the mean or average.

It is the most robust and widely used measure of dispersion since, unlike the range and inter-quartile range, it takes into account every variable in the dataset.

For example, are all your scores close to the average? Or are lots of scores way above (or way below) the average score?

When the values in a dataset are pretty tightly bunched together the standard deviation is small. When the values are spread apart the standard deviation will be relatively large. The standard deviation is usually presented in conjunction with the mean and is measured in the same units.

$$S.D = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n}}$$

Where:
- xi = Value of each data point
- x̄ = Mean
- N = Number of data points

For example, suppose we have five climatic stations and have recorded rainfall in mm as follows (60,47,17,43,30). Calculate the standard deviation for them.

Solution:

$$\bar{X} = \frac{60 + 47 + 17 + 43 + 30}{5} = 39.4$$

so the mean (average) height is 39.4 mm.

Now we calculate each station's difference from the Mean:

$$= \frac{(60-39.4)^2 + (47-39.4)^2 + (17-39.4)^2 + (43-39.4)^2 + (30-39.4)^2}{5}$$

$$= \frac{424.36 + 57.76 + 501.76 + 12.96 + 88.36}{5}$$

$= 217.04$
Now the Standard Deviation is

$$S.D = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \bar{X})^2}{n}}$$

$$S.D = \sqrt{217.04}$$

$$S.D = 14.73$$

And the good thing about the Standard Deviation is that it is useful. Now we can show which heights are within one Standard Deviation (14.73 mm) of the Mean:

So, using the Standard Deviation we have a "standard" way of knowing what is normal rainfall, and what is extra-large rainfall or extra small rainfall.

Standard deviation for grouped data

$$S.D = \sqrt{\frac{\sum_{i=1}^{n} fi(Xi - \bar{X})^2}{\sum_{i=1}^{n} fi}}$$

$$\bar{x} = \frac{\sum fx}{n}$$

Find the standard deviation of the following data:

Classes	Frequency (f)
30-34	4
35-39	5
40-44	2
45-49	9
TOTAL	20

Solution:

Classes	Frequency (f)	Mid classes (x)	f.x	x- mean	(x-mean)²	f.(x-mean)²
30-34	4	32	128	32-41= -9	81	4*81= 324
35-39	5	37	185	37-41= -4	16	5*16=80
40-44	2	42	84	42-41= 1	1	2*1=2
45-49	9	47	423	47-41= 6	36	9*36= 324
TOTAL	20					730

Mean $\bar{x} = \frac{\sum f*x}{n} = \frac{820}{20} = 41$

$$S.D = \sqrt{\frac{\sum_{i=1}^{n} fi(Xi - \bar{X})^2}{\sum_{i=1}^{n} fi}} = \sqrt{\frac{730}{20}} = 6.04$$