Measures of Variation: Range, IQR

A measure of variation is a summary statistic that represents the amount of dispersion in a dataset. While a measure of central tendency describes the typical value, measures of variability define how far away the data points tend to fall from the center. We talk about variability in the context of a distribution of values. A low dispersion indicates that the data points tend to be clustered tightly around the center. High dispersion signifies that they tend to fall further away.

In statistics, variability, dispersion, and spread are synonyms that denote the width of the distribution. Just as there are multiple measures of central tendency, there are several measures of variability.

Range

Let's start with the range because it is the most straightforward measure of variability to calculate and the simplest to understand. The range of a dataset is the difference between the largest and smallest values in that dataset. For example, in the two datasets below, dataset 1 has a range of $20 - 38 = 18$ while dataset 2 has a range of $11 - 52 = 41$. Dataset 2 has a wider range and, hence, more variability than dataset 1.

Dataset 1	Dataset 2
20	11
21	16
22	19
25	23
26	25
29	32
33	39
34	46
38	52

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.

Additionally, the size of dataset affects the range. In general, you are less likely to observe extreme values. However, as you increase the sample size, you have more opportunities to obtain

these extreme values. Consequently, when you draw random samples from the same population, the range tends to increase as the sample size increases. Consequently, use the range to compare variability only when the sample sizes are similar.

Range is defined as the difference between the value of the smallest item and the value of the largest item included in the distribution. It is the simplest method of measuring dispersion. Symbolically,

• **Range= Largest value (L) – Smallest Value (S)**

The relative measure corresponding to range, called the coefficient of range, is obtained by applying the following formula:

• **Coefficient of Range= (L- S)/ (L S)**

Merits of Range

1. It is simple to understand and easy to calculate.

2. It is less time consuming.

Demerits of Range:

1. It is not based on each and every item of the distribution.

2. It is very much affected by the extreme values.

3. The value of Range is affected more by sampling fluctuations

4. Range cannot be computed in case of open-end distribution.

The Interquartile Range (IQR)

The interquartile range is the middle half of the data. To visualize it, think about the median value that splits the dataset in half. Similarly, you can divide the data into quarters. Statisticians refer to these quarters as quartiles and denote them from low to high as Q1, Q2, Q3, and Q4. The lowest quartile (Q1) contains the quarter of the dataset with the smallest values. The upper quartile (Q4) contains the quarter of the dataset with the highest values. The interquartile range is

the middle half of the data that is in between the upper and lower quartiles. In other words, the interquartile range includes the 50% of data points that fall in Q2 and

The IQR is the red area in the graph below.
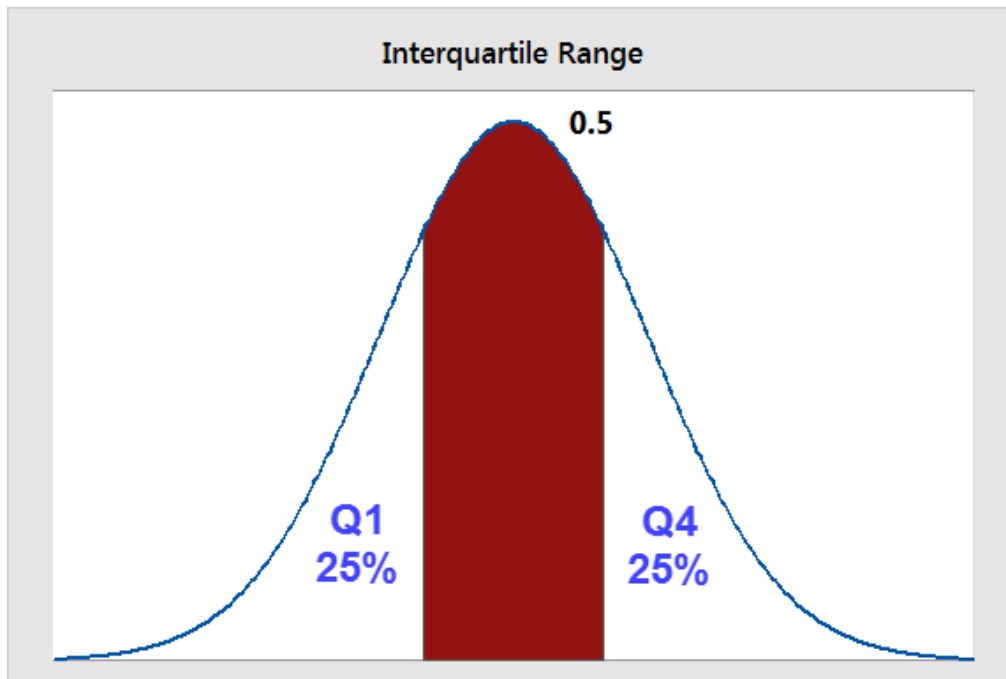
The interquartile range is a robust measure of variability in a similar manner that the median is a robust measure of central tendency. Neither measure is influenced dramatically by outliers because they don't depend on every value. Additionally, the interquartile range is excellent for skewed distributions, just like the median. As you'll learn, when you have a normal distribution, the standard deviation tells you the percentage of observations that fall specific distances from the mean. However, this doesn't work for skewed distributions, and the IQR is a great alternative.

I've divided the dataset below into quartiles. The interquartile range (IQR) extends from the low end of Q2 to the upper limit of Q3. For this dataset, the range is 21 – 39.
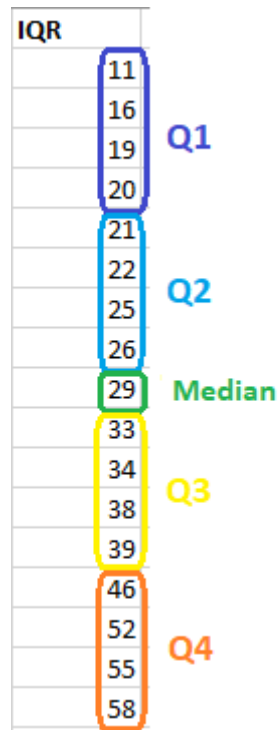
Quartile Deviation or Semi-Interquartile Range Merits:

1. It is easy to calculate and simple to follow.

2. It is not affected by the extreme values and is, therefore, useful in skewed distributions.

3. It is the only method of dispersion applicable in case of open-end classes.

Demerits:

1. Since Quartile Deviation is based on Quartiles, sometimes it is not rigidly defined.

2. It does not take into account all the observations in the series. Hence it is not representative.

3. It is not capable of further algebraic treatment.

4. It is not a stable measure of dispersion as it is affected very much by fluctuations of sampling.