Faculty of Veterianry Medicine

Damanhour University

**Project Title**

**Measures of dispersion for ungrouped data**

**Student Name:** فادي جبر علي صدقة

**Student ID:** 1443

**National ID:** 29807061801952

**Course Level:** الفرقة الاولي

**Department:** Statistics

## Abstract

Dispersion is a statistical term that describes the size of the distribution of values ex-pected for a particular variable. Dispersion can be measured by several different statis-tics, such as range, variance, and standard deviation. In finance and investing, disper-sion usually refers to the range of possible returns on an investment, but it can also be used to measure the risk inherent in a particular security or investment port-folio. It is often interpreted as a measure of the degree of uncertainty, and thus, risk, associated with a particular security or investment portfolio. In statistics, the measure of central tendency gives a single value that represents the whole value; however, the central ten-dency cannot describe the observation fully. The measure of dispersion helps us to study the variability of the items. In a statistical sense, dispersion has two meanings: first it measures the variation of the items among themselves, and second, it measures the variation around the average. If the difference between the value and average is high, then dispersion will be high. Otherwise it will be low. According to Dr. Bowley, “dispersion is the measure of the variation between items.” Researchers

use this tech-nique because it determines the reliability of the average. Dispersion also helps a re-searcher in comparing two or more series. It is also the facilitating technique to many other statistical techniques like correlation, regression, structural equation modeling, etc. In statistics, dispersion has two measure types. The first is the absolute measure, which measures the dispersion in the same statistical unit. The second type is the rela-tive measure of dispersion, which measures the ratio unit. In statistics, there are many techniques that are applied to measure dispersion.Range: Range is the simple measure of dispersion, which is defined as the difference between the largest value and the smallest value

## Introduction

Introduction

A measure of statistical dispersion is a nonnegative real number that is zero if all the data are the same and increases as the data become more diverse.

Most measures of dispersion have the same units as the quantity being measured. In other words, if the measurements are in metres or seconds, so is the measure of disper-sion. Examples of dispersion measures include:

- Standard deviation
- Interquartile range (IQR)
- Range
- Mean absolute difference (also known as Gini mean absolute difference)
- Median absolute deviation (MAD)
- Average absolute deviation (or simply called average deviation)
- Distance standard deviation

These are frequently used (together with scale factors) as estimators of scale parame-ters, in which capacity they are called estimates of scale. Robust measures of scale are those unaffected by a small number of outliers, and include the IQR and MAD.

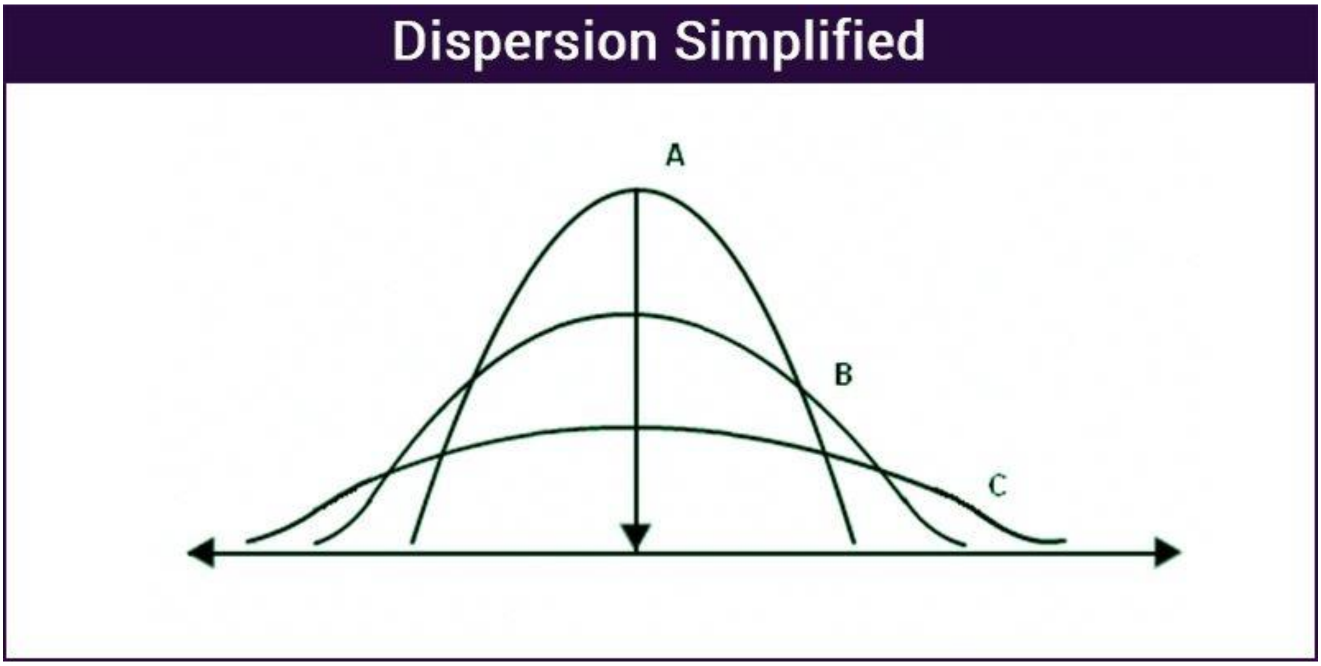
All the above measures of statistical dispersion have the useful property that they are location-invariant and linear in scale. This means that if a random variable X has a dispersion of SX then a linear transformation Y = aX + b for real a and b should have dispersion SY = |a| SX, where |a| is the absolute value of a, that is, ignores a preceding negative sign.

**Project Aim and Outline**

1. **Measures of Dispersion**
2. **Types of measure of dispersion**
   - **Absolute Measure of Dispersion**
   - **Relative Measure of Dispersion**
3. **Absolute Measure of Dispersion**
   - **Standard Deviation**
4. **Relative Measure of Dispersion**
   - **Coefficient of Variation:**
5. **Coefficient of Dispersion**
6. **Measures of Dispersion Formulas**

**Results**

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which a numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.




**Measures of Dispersion**

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogenous the data is. In simple terms, it shows how squeezed or scattered the variable is.

**Types of Measures of Dispersion**

There are two main types of dispersion methods in statistics which are:

- Absolute Measure of Dispersion
- Relative Measure of Dispersion

**Absolute Measure of Dispersion**

An absolute measure of dispersion contains the same unit as the original data set. Ab-solute dispersion method expresses the variations in terms of the average of deviations of observations like standard or means deviations. It includes range, standard devia-tion, quartile deviation, etc.

The types of absolute measures of dispersion are:

1. Range: It is simply the difference between the maximum value and the minimum value given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6
2. Variance: Deduct the mean from each data in the set then squaring each of

them and adding each square and finally dividing them by the total no of values in the data set is the variance. Variance

$$(\sigma 2) = \sum (X - \mu) \text{^}2/N$$

3. Standard Deviation: The square root of the variance is known as the standard deviation i.e.

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

4. Quartiles and Quartile Deviation: The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.

5. Mean and Mean Deviation: The average of numbers is known as the mean and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is known as the mean deviation.

**Standard Deviation:**

The standard deviation is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance. It is calculated as the square root of variance by determining the variation between each data point relative to the mean. If the data points are further from the mean, there is a higher deviation within the data set; thus, the more spread out the data, the higher the standard deviation.

Standard deviation is a statistical measurement in finance that, when applied to the an-nual rate of return of an investment, sheds light on the historical volatility of that in-vestment. The greater the standard deviation of securities, the greater the variance be-tween each price and the mean, which shows a larger price range. For example, a vola-tile stock has a high standard deviation, while the deviation of a stable blue-chip stock is usually rather low.

Using Standard Deviation

Standard deviation is an especially useful tool in investing and trading strategies as it helps measure market and security volatility—and predict performance trends. As it re-lates to investing, for example, one can expect an index fund to have a low standard deviation ver-sus its benchmark index, as the fund's goal is to replicate the index.

On the other hand, one can expect aggressive growth funds to have a high standard de-viation from relative stock indices, as their portfolio managers make aggressive bets to generate higher-than-average returns.

A lower standard deviation isn't necessarily preferable. It all depends on the invest-ments one is making, and one's willingness to assume the risk. When dealing with the amount of deviation in their portfolios, investors should consider their personal toler-ance for volatility and their overall investment objectives. More aggressive investors may be comfortable with an investment strategy that opts for vehicles with higher-than-average volatility, while more conservative investors may not.

Standard deviation is one of the key fundamental risk measures that analysts, portfolio managers, advisors use. Investment firms report the standard deviation of their mutual funds and other products. A large dispersion shows how much the return on the fund is deviating from the expected normal returns. Because it is easy to understand, this sta-tistic is regularly reported to the end clients and investors.

Standard Deviation vs. Variance

Variance is derived by taking the mean of the data points, subtracting the mean from each data point individually, squaring each of these results and then taking another mean of these squares. Standard deviation is the square root of the variance.

The variance helps determine the data's spread size when compared to the mean value. As the variance gets bigger, more variation in data values occurs, and there may be a larger gap between one data value and another. If the data values are all close together, the variance will be smaller. This is more difficult to grasp than are standard deviations, however, because variances represent a squared result that may not be meaningfully expressed on the same graph as the original dataset.

Standard deviations are usually easier to picture and apply. The standard deviation is expressed in the same unit of measurement as the data, which isn't necessarily the case with the variance. Using the standard deviation, statisticians may determine if the data has a normal curve or other mathematical relationship. If the data behaves in a normal curve, then 68% of the data points will fall within one standard deviation of the aver-age, or mean data point. Bigger variances cause more data points to fall outside the standard deviation. Smaller variances result in more data that is close to average.

## The Formula for Standard Deviation

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

**where:**

$x^i$=Value of the $i^{th}$ point in the data set
x=The mean value of the data set

n=The number of data points in the data set

**Relative Measure of Dispersion:**
The relative measures of depression are used to compare the distribution of two or more data sets. This measure compares values without units. Common relative dispersion methods include:

1. Coefficient of Range
2. Coefficient of Variation
3. Coefficient of Standard Deviation
4. Coefficient of Quartile Deviation
5. Coefficient of Mean Deviation

**Coefficient of Variation:**
The most important of all the relative measures of dispersion is the coefficient of variation. This word is variation not variance. There is no such thing as coefficient of variance. The coefficient of variation (C.V) is defined as:

$$(C.V) = \frac{S}{X} * 100$$

Thus C.V is the value of S when $\bar{X}$ is assumed equal to 100. It is a pure number and the unit of observation is not mentioned with its value. It is written in percentage form like 20% or 25%. When its value is 20%, it means that when the mean of the observations is assumed equal to 100, their standard deviation will be 20. The C.V is used to compare the dispersion in different sets of data particularly the data which differ in their means or differ in their units of measurement. The wages of workers may be in dollars and the consumption of meat in families may be in kilograms. The standard deviation of wages in dollars cannot be compared with the standard deviation of amount of meat in kilograms. Both the standard deviations need to be converted into a coefficient of variation for comparison. Suppose the value of C.V for wages is 10% and the values of C.V for kilograms of meat is 25%. This means that the wages of workers are consistent. Their wages are close to the overall average of their wages. But the families consume meat in quite different quantities. Some families consume very small quantities of meat and some others consume large quantities of meat. We say that there is greater variation in their consumption of meat. The observations about the quantity of meat are more dispersed or more variant.

**Coefficient of Dispersion**

The coefficients of dispersion are calculated along with the measure of dispersion when two series are compared which differ widely in their averages. The dispersion coeffi-cient is also used when two series with different measurement unit are compared. It is denoted as C.D.

The common coefficients of dispersion are:

| C.D. In Terms of | Coefficient of dispersion |
|---|---|
| Range | C.D. = (Xmax – Xmin) ⁄ (Xmax + Xmin) |
| Quartile Deviation | C.D. = (Q3 – Q1) ⁄ (Q3 + Q1) |
| Standard Deviation (S.D.) | C.D. = S.D. ⁄ Mean |
| Mean Deviation | C.D. = Mean devia-tion/Average |

**Measures of Dispersion Formulas**

The most important formulas for the different dispersion methods are:

| | |
|---|---|
| Arithmetic Mean Formula | Quartile Formula |
| Standard Deviation Formula | Variance Formu-la |
| Interquartile Range Formula | All Statistics Formulas |

Ungrouped data is the data you first gather from an experiment or study. The data is raw — that is, it's not sorted into categories, classified, or otherwise grouped. An ungrouped set of data is basically a list of numbers.

**Conclusions**

Without knowing something about how data is dispersed, measures of central tendency may be misleading. For example, a residential street with 20 homes on it having a mean value of $200,000 with little variation from the mean would be very different from a street with the same mean home value but with 3 homes having a value of $1 million and the other 17 clustered around $60,000. Measures of dispersion provide a more complete picture. Dispersion measures include the range, average deviation, variance, and standard deviation. Range The simplest measure of dispersion is the range. The range is calculated by simply taking the difference between the maximum and minimum values in the data set. However, the range only provides information about the maximum and minimum values and does not say anything about the values in between. Average Deviation Another method is to calculate the average difference between each data point and the mean value, and divide by the number of points to calcuate the average deviation (mean deviation). However, performing this calcuation will result in an average deviation of zero since the values above the mean will cancel the values below the mean. If this method is used, the absolute value of the difference is tak-en so that only positive values are obtained, and the result sometimes is called the mean absolute deviation. The average deviation is not very difficult to calculate, and it is intuitively appealing. However, the mathematics are very complex when using it in subsequent statistical analysis. Because of this complexity, the average deviation is not a very commonly used measure of dispersion. Variance and Standard Deviation A better way to measure dispersion is to square the differ-ences before averaging them. This measure of dispersion is known as the vari-ance, and the square root of the variance is known as the standard deviation. The standard deviation and variance are widely used measures of dispersion

## References


- Peavy, J. V., Dyal, W. W., & Eddins, D. L. (1981). Descriptive statistics: Measures of central tendency and dispersion. Atlanta, GA: U.S. Dept. of Health and Human Services/Public Health Service, Centers for Disease Control.
- Johnson, R., & Kuby, P. (2000). Elementary statistics. Pacific Grove: Duxbury.
- (1998, March). Retrieved May 25, 2020, from https://simon.cs.vt.edu/SoSci/converted/Dispersion_I/
- (2010, October 4). Retrieved May 12, 2020, from http://bayes.acs.unt.edu:8083/BayesContent/class /Jon/ISSS_SC/Module003/isss_m3_describingdata/node5.html
- Jin, G. (2019, September 18). Summary Measures for Quantitative Data. Retrieved May 29, 2020, from http://my.ilstu.edu/~gjin/hsc204-hed/Module-5-Summary-Measure-2/Module-5-Summary-Measure-25.html