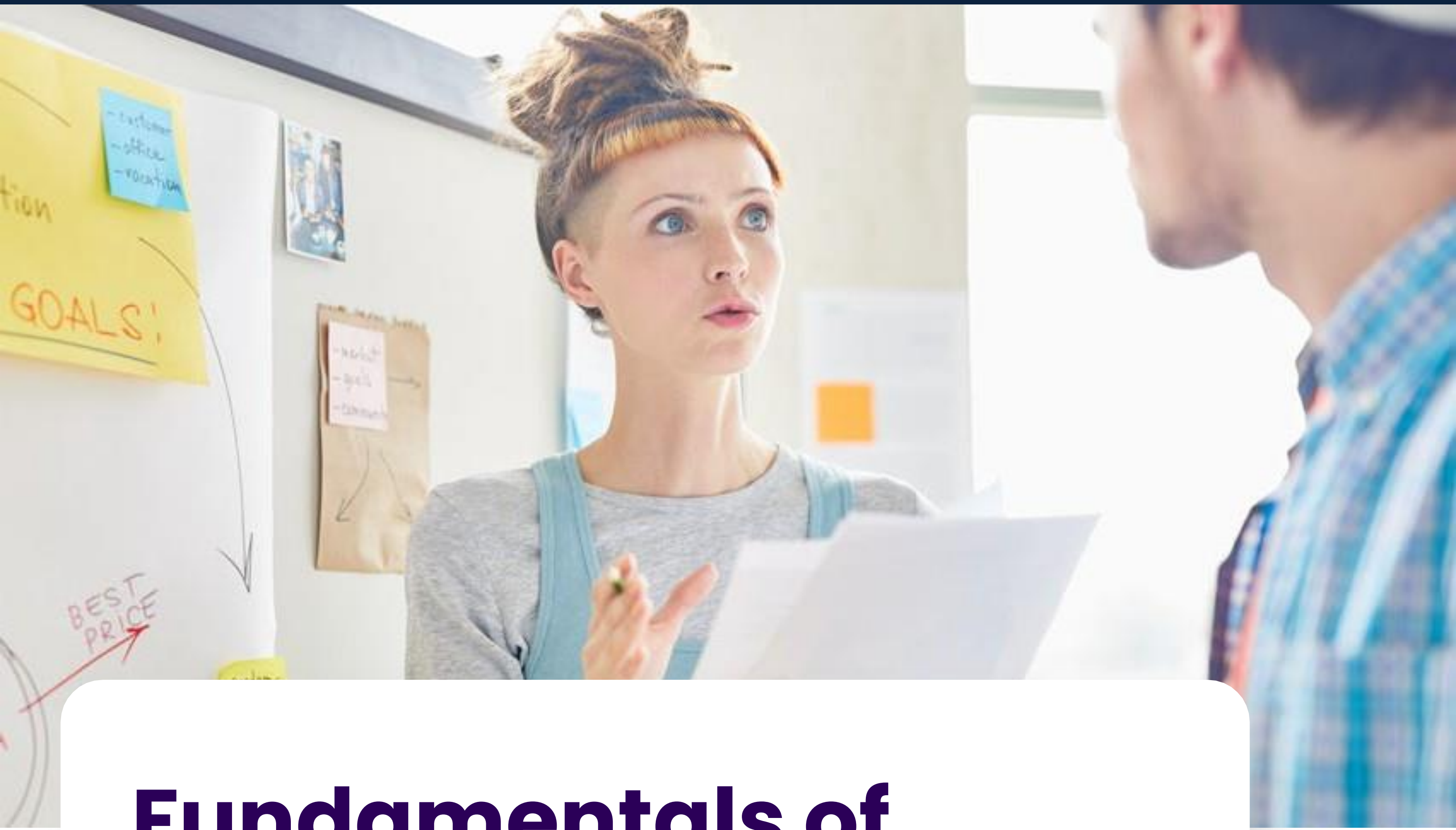
Analyttica TreasureHunt® LEAPS

www.analyttica.com



Fundamentals of Data Analytics

Module 1:

Introduction to Basic Numeric Descriptive Measures

ANALYTTICA TREASURE HUNT
LEARN . APPLY . SOLVE

Table of Contents

Data Collection

Introduction to Population and Sample

The group of people or things that need to be examined is referred to as a "population." For instance, all children in the population would be included if the objective was to examine the nutritional status of children in a developing nation.

Every member (or item) of the population must have information collected in order to examine the characteristics of the population as a whole. Census is the name for this process of gathering data on the entire population. In the aforementioned illustration, it is necessary to interact with each and every child in the relevant nation.

However, due to a number of limitations, it is frequently challenging, if not impossible, to measure every unit in the population (s). In these situations, a population subset that is highly representative of the entire population is taken into consideration for study. Sample refers to this subgroup. In the aforementioned illustration, an alternative method to gathering nutritional data on every child in the nation would be to randomly select 1 or 2 percent of children from the nation and assess their nutritional condition, presuming that they are typical of the entire population. Sampling is the term for this data collection technique.

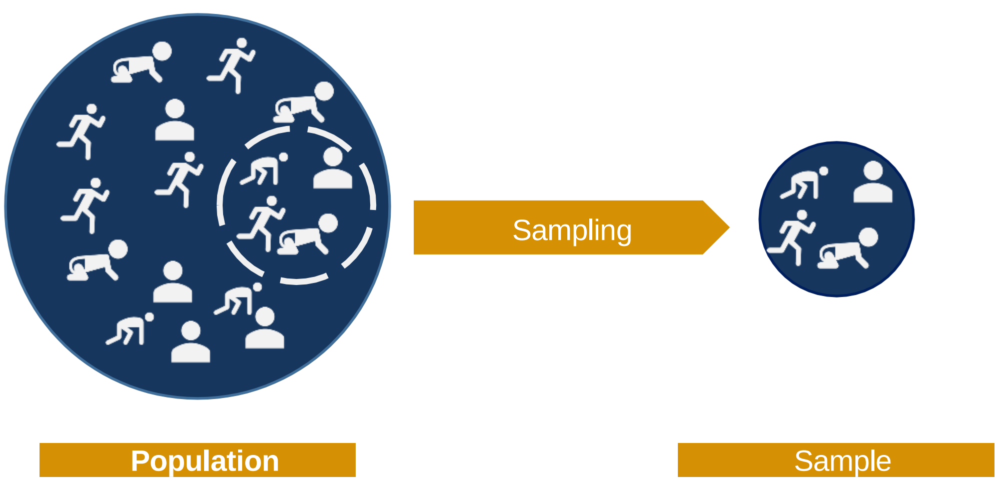


Figure 1: Population and Sample

Needed of Sampling?

In most real-world situations, gathering data from the entire population is not possible. Sampling is applicable in such situations. Even if there is a small population, gathering data about every unit could be a difficult and lengthy process. When sampling is scientifically done, inference about the population can be drawn by collecting data from a sample of the population. At this stage, it is worthwhile to understand the reason behind the need for sampling.

Consider the population of, say, Country A, a small nation. There are 1,298,503 women between the ages of 15 and 49 and 726,680 children under the age of five. Six teams would need 71 years to finish data collection if each team could survey 13 women and 13 children each day. Additionally, the survey team's travel time is not even included in. Do you think it would be wise and cost effective to collect the data for the entire population given all the resources that would get consumed during this survey? In addition, do ponder over the fact that by the time the survey would be completed, the children would have grown up and many women would have died, thereby eliminating the relevance of the very survey.

These are the kind of situations where Sampling can enable an efficient and effective data collection.

Listed below are few advantages of sampling:

- Surveying or measuring everyone is not cost effective.
- Using sampling, one can produce information faster and hence sampling enables timely decision making.
- In a sample survey, as smaller set of individuals or objects are being managed, more accurate and in-depth information can be collected than a census.
- A smaller set of individuals often results in lesser data collection errors.
- In destructive tests like experiments with chemical reactions, or machine longevity tests, census is not an acceptable option.

A statistically optimal sample size can be estimated subject to an acceptable cost.

Different Types of Sampling

Prior to comprehending the various sampling techniques, it's critical to comprehend the distinction between statistics and parameters. Here is a formal definition:

A parameter is a value that must be accurately calculated using statistics from one or more samples, and is typically unknown. Simply said, a parameter is linked to the population, which is typically unknowable, while a statistic is linked to the sample (drawn from population, hence known).

Greek letters are typically used to denote population, such as for population standard deviation and for population mean. The English alphabet's S, which stands for the sample standard deviation, is used to indicate statistics.

Numerous sampling techniques can be categorised as either probability sampling techniques or non-probability sampling techniques. To help with understanding, a diagrammatic illustration of these has been provided below.

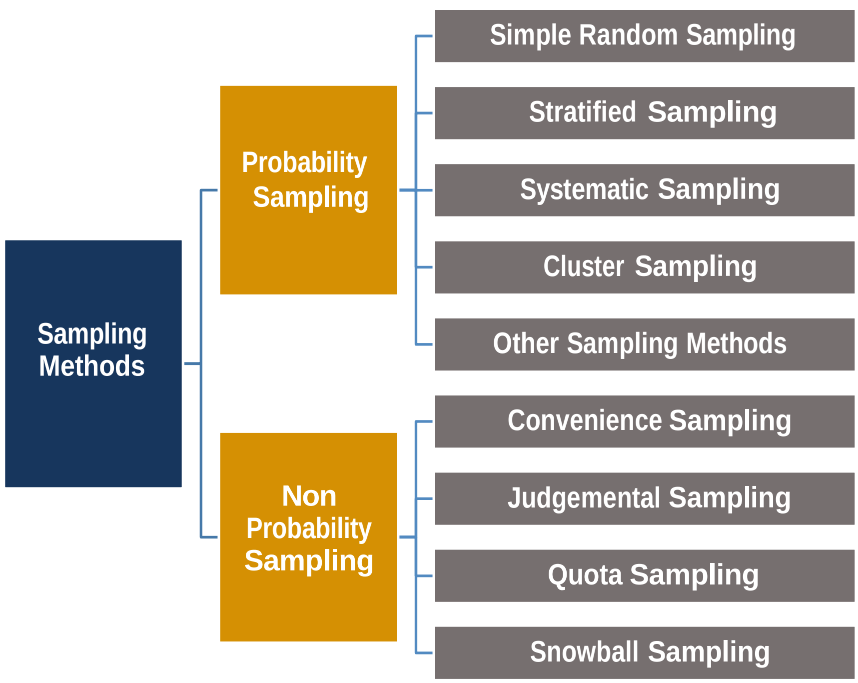


Figure 2: Types of Sampling

These all have advantages and drawbacks of their own. The sample strategy is chosen after taking into account a number of variables, including cost, data accessibility, applicability, and the required resources. However, for the purposes of this course, every sampling procedure would be based on Simple random sampling. In this type of sampling, every person has an equal chance of being selected from the population, hence it is the most commonly used sampling method

Data Classification

Data is nothing but a collection of values either qualitative or quantitative. Data can be numbers or measurement or scale or rank or it can even be purely qualitative.

Data points can be classified into five broad categories - Nominal, Ordinal, Interval, Ratio and Quasi-Interval.

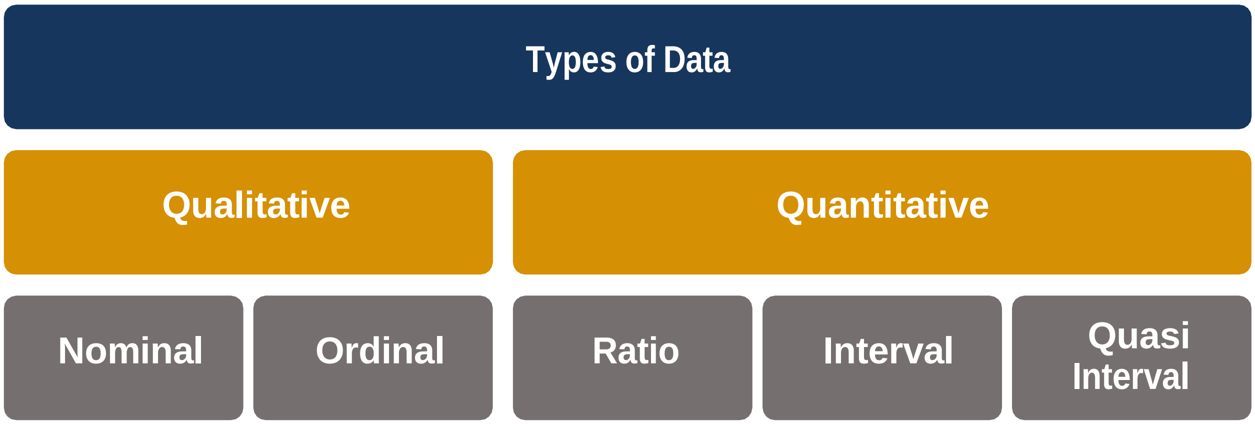


Figure 3: Types of Data

Nominal data represents qualitative information without order. It indicates two or more different classifications which doesn’t follow any particular order. For example: Type of school could be either vocational, private, state etc. One school being Private and the other one being State does not tell if one of them is better than the other. They just capture the type of school without any rank.

Ordinal data represents qualitative information without order, indicates that the measurement classifications are different and can be ranked. For example: The grading system of A, B, C, D. When Nominal data has some measure of rank built within, it is referred to as Ordinal data. A letter grade of A in exam is ranked higher than a grade of B.

Both the above types of qualitative data are also known as categorical data in data analysis, Nominal data is also called categorical data while ordinal data is also known as ordered categorical data.

Interval data establishes numerically equivalent distances on the scale and measures with order. For example: In the performance in a GMAT exam, difference between 800 and 700 marks is equal to difference between 600 and 500. In other words, if the difference between two values is meaningful, then the data is classified as an interval data. Quasi-Interval data is a special case of Interval Data, that falls between ordinal and interval. For example: A survey with Strongly Disagree to Strongly Agree.

The intervals and "genuine" zero point for ratio data measures are equivalent. It possesses all the characteristics of interval data and clearly defines the genuine zero point. For example: Weight, height, price are all ratio variables. A package of 100 grams is twice in weight of a package of 50 grams. Similarly, a height of 5 ft. is 5 times the height of 1 ft. However, temperature in either F or C are not ratio variables as 0 in those scales do not imply that there is no heat. An alternative is the Kelvin scale which is ratio data.

Representation of Data

The main purpose of data collection and data classification is to analyze data and draw inferences from them. Representation of data in the most relevant manner is a key step towards analysis. Data can be represented in various ways, as shown by the diagram below.

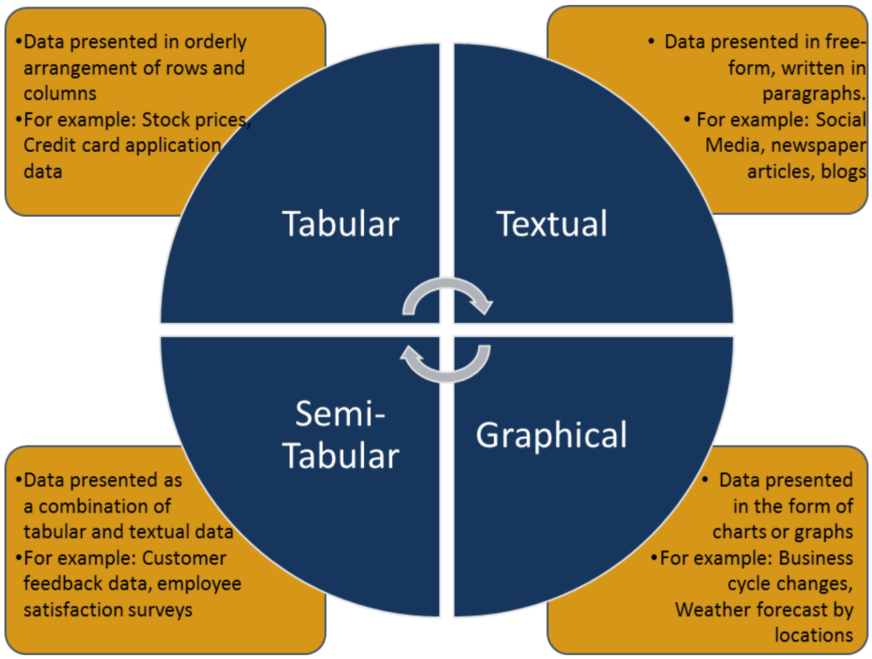


Figure 4: Representation of Data

The two most commonly used methods for data representation are tabular and graphical. Below is an example that will demonstrate both these methods.

A survey has been conducted to collect the educational qualification of 100 individuals. The raw data for one individual straight from the survey could look like the following:

Table 1:

Name	John
School Type	Public
Education Level	Graduate
Score (in %)	54%

It is, however, not practical to examine the raw data for every respondent to the poll. Therefore, a legible and understandable summary of the raw data is required.

Tabular Representation of Data

The survey data for Education level can be summarized as shown below – this provides a view of all the 100 individuals in the data.

Table 2:

Education Level	Number of Observations
Primary or Low	9
Secondary	21
High Secondary	42
Graduate	23
Post Graduate or Above	5
Total	**100**

Similarly, the score data can be summarized as below - this is called a grouped frequency table.

Table 3:

Score Range (In %)	Number of Observations
Less than 30%	5
30% - 50%	17
50% - 70%	14
70% - 90%	53
More than 90%	11
Total	**100**

Graphical Representation

The same two tables above can also be represented in different graphical forms.

Table 1 can be represented as a Pie-chart, where each share of pie represents a specific education level while the size of pie represents the relative presence of the education level in the sample.

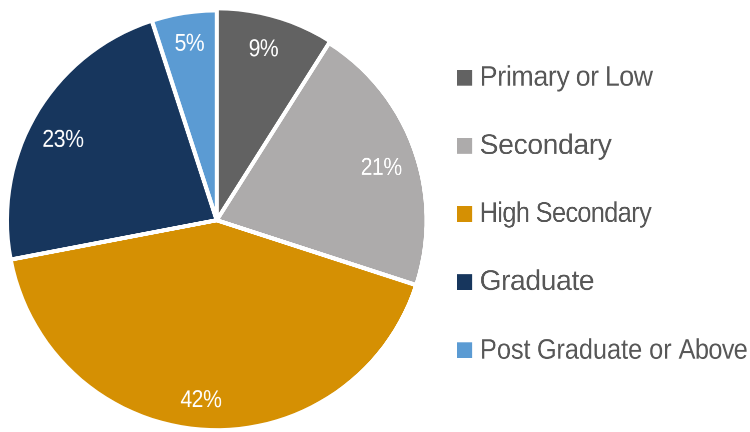


Figure 4: Graphical Representation of Data from Table 2

The height of each bar in Table 2's bar chart can be interpreted as the total number of respondents who fell within each score range.

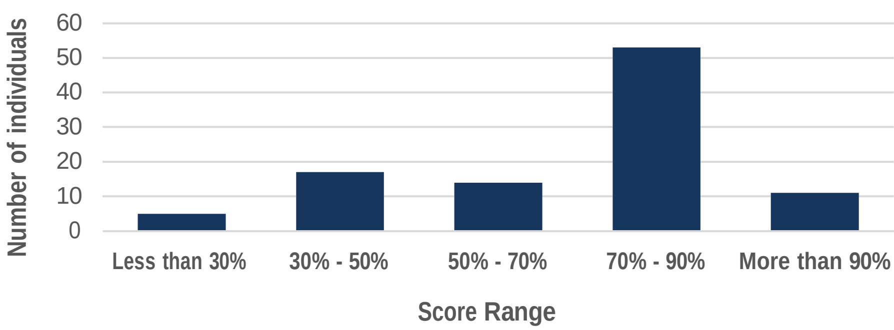


Figure 5: Graphical Representation of Data from Table 3

The type of graph to be used depends on the data and the specific scenarios. Nominal data is better represented in Pie-Charts. Ordinal data or discrete data in often represented in form of bar charts. For continuous variables, histogram, frequency polygon, or frequency charts are commonly used. For summarizing data over various time windows (also called time-series data), line plot is the best plot.

Please refer to the video for further details.

Measures of Location

Every time an interesting metric is observed (such as the share price of a stock, an individual's height, a student's grade, a flight's delay, etc.), a sizable portion of the observations in the sample tend to cluster around a single value. One typical value of the desired metric for the sample can be thought of as this value. These values are known as "Measures of Location" or, more generally, "Measures of Central Tendency" of the sample since they reveal the "location" of a "central" value of the sample.

The central tendency of any data can be measured using a variety of statistics. The most significant measures are Mean, Median, Mode, Percentile, and Quartile

Mean

The commonly used measure of central tendency is Mean. It can be defined for all ratio-scale and interval-scale data. Mathematically, it is defined as below:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Where, $\bar{X}$ would be mean, X_i are the data points, n is total number of data points. it is simply all the numbers added and the sum being divided by the total number of data points. Simply put, Mean is the average value of any data series. Note that the total deviation around mean is always zero. In other words, $\sum_{i=1}^{n} (X_i - \bar{X}) = 0$

Median

The middle number of any series of data, when sorted in either ascending or descending order is known as Median. If the data series is odd, then the median is the exact middle number. The average of the middle two numbers is used if the sequence is even. Along with interval-scale or ratio-scale data, the median is also defined for ordinal data.

Mode

Mode is the most frequent number in an array i.e. the number with highest occurrences or frequency. Depending on the frequency of the data, a series might or might not have a mode. For example, in a series like 10,10,10,10, 20, 30, mode is obviously 10 since it's the most frequent (it has appeared 4 times out of 6 numbers). However, a series like: 10, 20, 30, 40; It has no mode since all the numbers have appeared with frequency 1. Mode is defined for all kinds of data, viz. nominal, ordinal, interval-scale, and ratio-scale.

Why are three Measures of Central Tendency needed?

A natural question to arise would be, why are three different measures for central tendency needed when all three serve the same purpose of providing a representative number for the sample. The answer is that while each of these measures do serve the same purpose, they differ in their application. They have their own advantages and disadvantages. Weighing in their respective merits (and demerits), you can select when to use which measure.

For example, mode is the simplest measure of all three, but does not make much sense or can give misleading value for a continuous variable like "Departure Delay of a Flight", or "Time to React in a Chemical Reaction", or "Monthly Income of an Individual", etc.

Let's discuss some important features of Mean:

Although the mean is frequently not the true value seen in a data series, it is almost always seen. The least amount of departure will be produced when the mean for a data series is used for prediction. Additionally, it contains all of the series' data points needed for calculation.

When not to use Mean?

One major disadvantage of Mean is that it is particularly susceptible to extreme values or outliers in data.

Example: Consider the series of students with exam scores:

Table 4:

Student	Exam Score	
George	50	
Sheila	56	
Meera	60	
Amit	62	
Patrick	**150**	→ **Outlier**

Mean of the series = 75.6, which is significantly higher barring Patrick.

Mean of first 4 students score = 57, which is revealing the true picture in terms of the average score.

From the above scenario, you can see how an outlier or one extreme value (in this case, the 5th student, Patrick with an exam score of 150) can distort this measure. So, you need to be careful when summarizing data using mean as outliers or extreme values present in the data can distort this measure. Outliers need to be treated or data needs to be normalized in presence of extreme values.

What should we be careful about when using median?

Median is based on only the relative positioning of the observations, not their actual value. Hence, often it may not reflect the slow shift in the sample or population. For example: Below are the scores for 9 selected students in first term and second term.

Table 5:

Term 1	35	37	42	45	50	55	56	59	64
Term 2	41	46	48	48	50	65	73	84	93

Note that in both the cases the Median is 50. But, 8 out of 9 students got more marks in Term 2 than Term 1. This is not reflected in the Median measure.

When to use Median over Mean or Mode?

When the data is normally distributed or symmetric, then mean, median and mode are the same. However, for non-symmetric/skewed data, median is preferred, as that remains unchanged and is not affected by skewness in the data. This will be discussed in detail in the second module. Also, Mode is used majorly for categorical (Nominal) data, where it is important to find the most frequent category.

- Symmetric interval or ratio data - Mean is the central tendency measure to be used.
- Skewed interval or ratio data - Median is the central tendency measure to be used.
- Ordinal data - Median is the central tendency measure to be used.

Percentiles and Quartiles

Using percentiles, you may estimate the percentages of the data that should fall above and below a certain value. When the data is sorted, the pth percentile states that at most (p)% of the observations are less than this number and at most (1-p)% are more. For instance: A 99 percentile result on an exam indicates that the test-taker outperformed 99 percent of the other participants.

It is also important to remember the associated terms:

- The 1st quarter = the 25th percentile (Q1).
- The median or second quarter = 50% percentile (Q2).
- The 3rd Quartile = 75th percentile (Q3).

Please refer to the video for further details.

Measures of Variability

Only understanding location is insufficient for a thorough understanding of the data. The distribution or scattering of the data must be measured. An indicator of greater confidence in drawing any conclusions from data based on the location parameter is a variable with lower variability. A high level of variability may point to the impact of another unidentified variable on the data. Therefore, a data's variability reflects how much we still don't know about it. "Measures of Dispersion" are the metrics used to describe data variability.

Range

It generally measures the distance between the highest and lowest value of the given data series and is the most basic way to assess dispersion. It provides a rough understanding of how the data is distributed but is particularly vulnerable to outliers. The distribution of the data is not assumed in any way by this measurement. This aids in capturing the data's variance. This is a useful strategy when you want to concentrate just on the extremes, such as in weather reports, where knowing the temperature range for a particular day is sufficient.

As an illustration, the temperature is checked every two hours throughout the day.

Table 6:

Hour	Temperature	
0:00	10ºC	
2:00	9ºC	
4:00	11ºC	
6:00	12ºC	
8:00	14ºC	
10:00	16ºC	
12:00	20ºC	
14:00	22ºC	**Maximum Value**
16:00	19ºC	
18:00	16ºC	
20:00	15ºC	
22:00	12ºC	

It is certain that around 2:00 p.m., the temperature was 90 °C at its lowest point. At 14:00, the highest temperature is 220C. If the temperature was a deciding factor for an outside event, then this might be a crucial measurement. However, this measure is distorted by presence of outlier or extreme values in the dataset. It is a better measure of spread due to less susceptibility to outliers; however, calculation of standard deviation is more complex than calculation of range.

Interquartile Range (IQR)

A similar measure of range known as the interquartile range, or IQR, similarly gauges the dispersion or variability of the data. It shows how far away from the mean the data in a series are. Like range, IQR does not assume anything about how the data are distributed (i.e. it is non-parametric). IQR calculates the range of the data's middle half. It is the difference between the 75% percentile and 25% percentile. Therefore, by definition, it only accounts for half of the data.

As the IQR increases, data points tend to be more dispersed. If IQR is modest, it is presumed that they are evenly distributed around the mean. For normally distributed data, this is closely related to standard deviation and is about 35% greater than standard deviation. You will learn more about Normal distribution in the second module. IQR can also help to determine the outlier in the data series. Data values that deviate from twice the IQR are often defined as outlier, where as those deviating more than 3.5 times of IQR are far outliers. Note, data needs to be sorted in ascending order to calculate IQR.

For example: The following table shows the number of digital equipment (phone, tablet etc.) owned by students in a freshman class.

Table 7:

Number of Equipment	Frequency	Cumulative Frequency
0	3	3
1	5	8
2	2	10
3	7	17
4	10	27
5	3	30
6	1	31
>6	0	31
Total	**31**	**31**

You can also think of this data to be represented as: 0001111122333333344444444445556

To calculate IQR, you first need to order the dataset (as in the representation above), then find the median. In this case there are 31 observations, hence the mean is the 16^{th} observation (i.e. 3). Now you need to identify the median value of the first and second half of the data. First quartile is the 8^{th} student who owns 1 equipment; third quartile is the 24^{th} student who owns 4 equipment. So IQR in this example is 4-1= 3.

Population & Sample Variance

The previous two measures discussed - Range and Inter Quartile Range, do not use all the data points directly. To get the 'variability' of the data from the 'central' value, we introduce the concept of deviation. Deviation (representing variability) of a data point is usually measured from mean, the most popular measure of Central Tendency.

But, we have seen before that the total deviation around mean is always zero. Hence, we measure the "average squared deviation" from mean, μ(Mu). We call this measure as "Variance". It is represented by σ^2(sigma square).

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

In the above formula, the assumption is that the actual population mean, i.e. μ, is known. Hence it is called "Population Variance". If the value of μ is not known, then it is estimated using $\bar{X}$, the sample mean. Then the formula for variance changes slightly and we call that statistic as "Sample Variance". Please note that the denominator in a Sample Variance formula is 1 less than the total observations in the sample.

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2}{n-1}$$

Variance, put simply, measures how variable the data are around the mean. High variance denotes greater volatility or variability, whereas low variance denotes less volatility. If a series' variance is 0, all of the numbers in that series are same.

Standard Deviation

The parametric equivalent of interquartile range, standard deviation can be calculated by taking the variance's square root. The variability in the data around the mean is represented by this along with variance. This is also frequently known as "volatility."

Together, the standard deviation and variance are significant indices of variability that are commonly applied in statistics. Investors frequently fret about the return's volatility, or how far it deviates from the average, while managing financial risk. The standard deviation is a key indicator of risk since it helps to quantify the return's volatility.

$$\text{Standard Deviation} = \sqrt{Variance}$$

For example: Consider the following example of two technology company stocks, TechCo1 & TechCo2.

Table 8:

	TechCo1: Adjusted Closing price	TechCo1: Daily Log Return	TechCo2: Adjusted Closing price	TechCo2: Daily Log Return
	1455.22	0.000	44.57	0.000
	1399.42	-0.039	43.06	-0.034
	1402.11	0.002	43.52	0.011
	1403.45	0.001	42.06	-0.034
	1441.47	0.027	42.61	0.013
	1457.6	0.011	42.92	0.007
	1438.56	-0.013	41.82	-0.026
	1432.25	-0.004	40.46	-0.033
	1449.68	0.012	41.22	0.019
	1465.15	0.011	42.92	0.040
	1455.14	-0.007	44.09	0.027
Standard Deviation (in%)		**1.696**		**2.664**

SPX is an Index that measures the composite stock value for 500 large companies in the US and considered as the market benchmark. TechCo1 is the ticker for a large technology firm. From the above example, it is clear that volatility or standard deviation of SPX is 1.7% compared to TechCo1 (2.7%). This indicates that SPX is less volatile compared to TechCo1 for investors. Note that in this computation we have considered the variable “Daily Log Return” for both the stocks to calculate standard deviation. This is a commonly used metric for assessing volatility of a given stock.

Coefficient of Variation

It is a measurement of relative variance. It is the standard deviation to mean ratio (average). It is a percentage-based, unit-free measurement. When two or more data series have extremely different averages or distinct units, it is used to compare the variability in those series.

For example: Consider the following example of two companies A and B, with their Revenue details for the last 10 years:

Table 9:

Revenue (In $)	Company A	Company B
Mean	$105. 5	$2.3
Standard Deviation	$30.2	$1.0

These appear to be incomparable at first glance given how different their average revenue is. It is the ratio of the data series' mean to standard deviation.

Coefficient of Variation = $(\sigma/\bar{x})$ * 100

For the above example, the CV would be as below:

Table 10:

Revenue (In $)	Company A	Company B
Mean	$105. 5	$2.3
Standard Deviation	$30.2	$1.0
Coefficient of Variation	28.59%	44.21%

Lower Coefficient of Variation indicates lower dispersion around the mean, indicating higher stability. In this example, Company A has higher stability than Company B over the period considered.

Please refer to the video for further details.

At this stage, you should work on Simulation 1 to apply your learnings from the previous sections.

Basic Bivariate Analysis

Throughout the previous sections, you learnt about understanding and interpreting one variable at a time. This type of analysis is called Univariate Analysis. However, in many real-world situations, descriptive analytics is done through simultaneous study of two or more variables. The analysis using two variables is referred to as Bivariate Analysis. When more than two variables are analyzed together, that is referred to as Multivariate Analysis. More advanced techniques are used to conduct multivariate analyses. In this section, you will learn about Bivariate analysis.

Pearson's Correlation and Covariance

Correlation can be defined as an indicator of how closely two variables move in the same or opposite directions. If both values are positive, they are moving in the same direction (i.e., if one increases, so does the other), whereas if both values are negative, they are moving in the opposite direction. It shows how strongly two variables are associated with one another.

The most used correlation statistic for numerical data is Pearson's Correlation. This gauges the strength of the linear relationship between two numerical variables and has a range of -1 to +1. It is denoted by the letter "r".

- Perfect positive correlation, r=1.
- Perfect negative correlation, r=-1.
- There is no linear association when r=0 (note, it does not mean no correlation)

Pearson's correlation is calculated as below:

$$r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

The numerator of the above formula represents another measure called 'Covariance' It is calculated by:

$$Cov(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n(n-1)}$$

The range of covariance is from -infinity to infinity. Pearson's correlation is a scaled measure and does not have any unit. Hence, often it is preferred over covariance to understand association between variables.

Rank Correlation

Spearman Correlation is a rank correlation that works on ordered data. Rather than looking at the absolute value of an observation, it looks at the order of the observation in the entire data. Unlike Pearson's correlation coefficient, Spearman Rank Correlation measure the degree of monotonic (move in the same direction but not at a constant rate) relationship between two variables. It is calculated using the following formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Where d_i = difference in paired ranks and n = number of observations

Like Pearson's correlation, Spearman Correlation (also shown as $r_{s)}$ also lies between -1 to +1.

- The perfect positive correlation is one, or r_s=1.
- Perfect negative correlation is denoted by r_s =-1.
- r_s =0 denotes a lack of monotonic correlation (note, it does not mean no correlation)

For example: Consider the relationship between the stock movement of three stocks: TechCo1, SPX and TechCo2.

Table 11:

TechCo1: Adjusted Closing Price	TechCo1: Daily Log Return	SPX: Adjusted Closing Price	SPX: Daily Log Return	TechCo2: Adjusted Closing Price	TechCo2: Daily Log Return
27.87	0	1455.22	0	44.57	0.000
25.52	-0.088	1399.42	-0.039	43.06	-0.034
25.89	0.014	1402.11	0.002	43.52	0.011
23.65	-0.090	1403.45	0.001	42.06	-0.034
24.77	0.046	1441.47	0.027	42.61	0.013
24.33	-0.018	1457.6	0.011	42.92	0.007
23.09	-0.052	1438.56	-0.013	41.82	-0.026
21.7	-0.062	1432.25	-0.004	40.46	-0.033
24.08	0.104	1449.68	0.012	41.22	0.019
25	0.037	1465.15	0.011	42.92	0.040
25.87	0.034	1455.14	-0.007	44.09	0.027
26.53	0.025	1455.9	0.001	40.91	-0.075
28.25	0.063	1445.57	-0.007	40.53	-0.009
27.71	-0.019	1441.36	-0.003	39.67	-0.021

- Correlation between TechCo1 & SPX daily log return = 0.567 (Daily Log Return of TechCo1 and SPX)
- Correlation between SPX & TechCo2 = 0.446 (Daily Log Return of SPX and TechCo2)
- Correlation between TechCo1 & TechCo2 = 0.535 (Daily Log Return of TechCo1 and TechCo2)

Please note that all the above correlation coefficients are through the Pearson correlation method. We see that all the correlations are positive, meaning that in any pair, if one stock goes up the other one goes up as well. Order of magnitude suggests that there is healthy interrelation in all the pairs. However, TechCo1 is more correlated to market return (SPX) than TechCo2 is.

Correlation is often visually represented through a scatter plot.

Consider the above example for better understanding this. Plotted in Scatter-plot below, consider the two stocks TechCo1 and TechCo2.

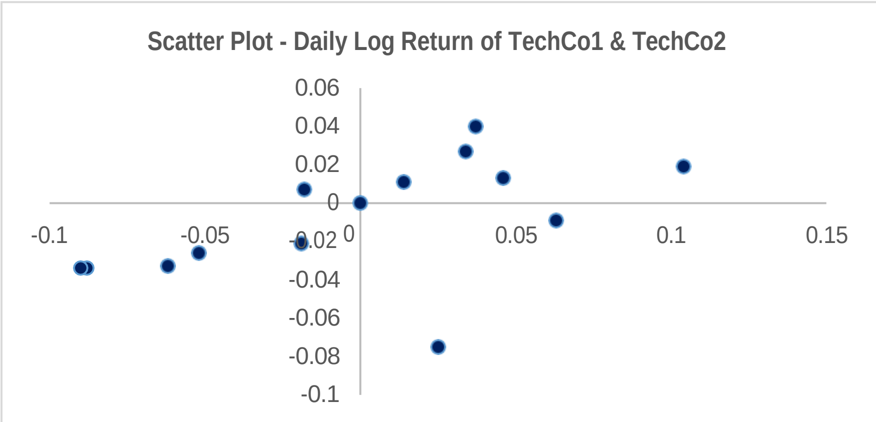


Figure 6: Scatter Plot

Correlation does not Signify Causation.

In two correlated variables, change in magnitude of one variable does not indicate that it will cause change in the other variable. Correlation does not signify causation. Rather, correlation merely suggests of related movement in the same direction of these two variables.

Conversely, if two random variables are causally related, it will automatically imply that if one changes that will cause change in the other one. It is very important to remember the difference between association and causation. It is also interesting to note that causation is an asymmetric relationship whereas correlation is symmetric relationship. For example, how much you earn might be highly correlated with your level of education. However, we cannot say that one causes another. It will require further investigation.

Please refer to the video for further details on Bivariate Analysis.

At this stage, you should work on Simulation 2 to apply your learnings from the previous sections.

Write to us at

support@analyttica.com

USA Address

Analyttica Datalab Inc.
1007 N. Orange St, Floor-4, Wilmington, Delaware - 19801
Tel: +1 917 300 3289/3325

India Address

Analyttica Datalab Pvt. Ltd.
702, Brigade IRV Centre,2nd Main Rd, Nallurhalli,
Whitefield, Bengaluru - 560066.
Tel : +91 80 4650 7300