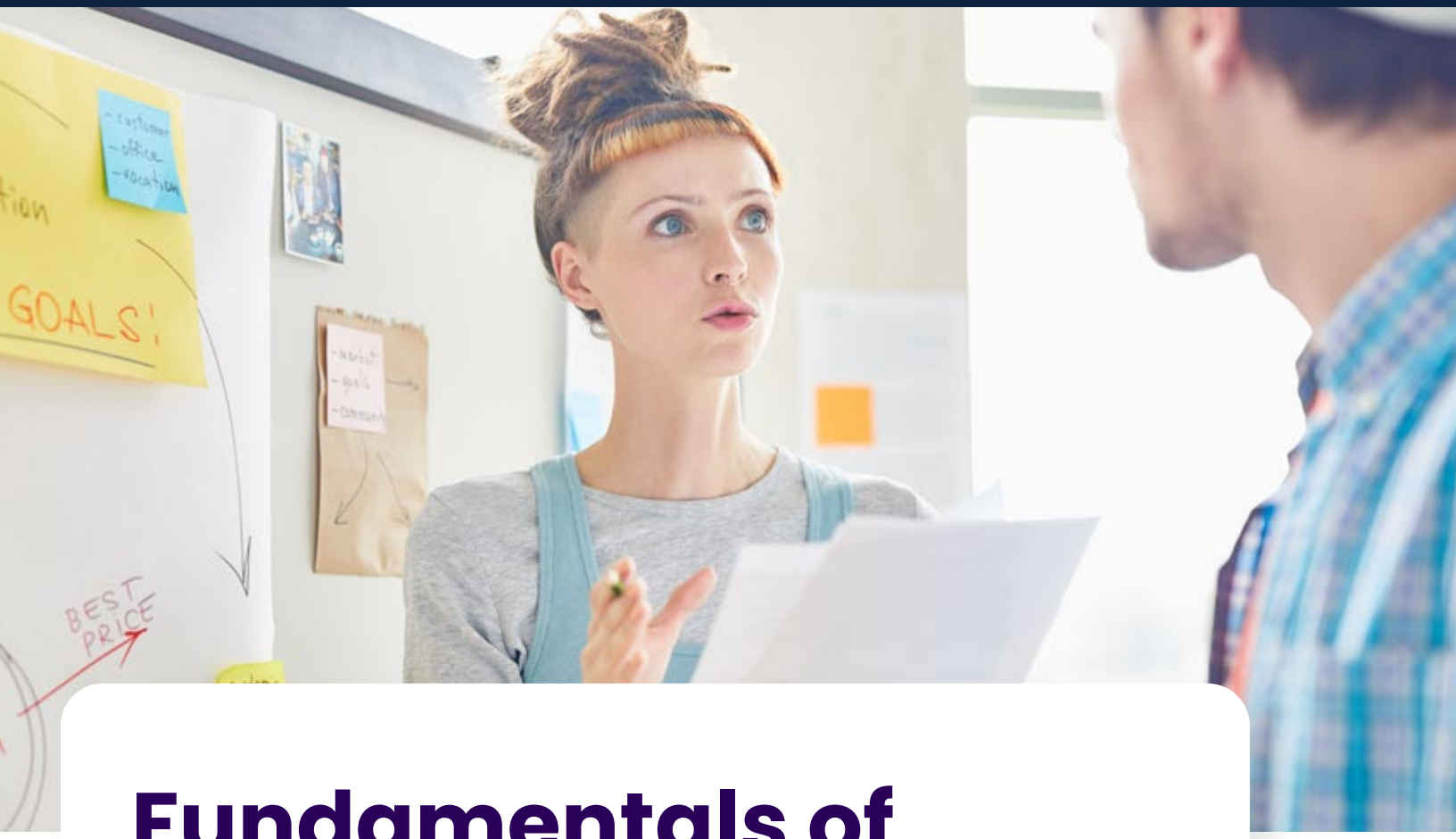

Fundamentals of Data Analytics

Module 1:

Introduction to Basic Numeric Descriptive Measures

ANALYTTICA TREASURE HUNT
LEARN . APPLY . SOLVE

Table of Contents

Collection of Data: Population and Sampling

Population and Sample

The universe of individuals or objects that are required to be analyzed is known as population. For example, if the goal is to analyze the nutritional status of children in an under developed country, then all children in that country comprise of the population.

In order to study the characteristics of any population of individuals or objects, information needs to be collected for every member (or item) in the population. This method of information collection for the entire population is called census. In the above example, one needs to reach out to every individual child in the country in question.

However, often it becomes difficult and sometimes even impossible to measure every unit in the population due to one or more constraint(s). In such cases, a subset of the population is considered for analysis, wherein, that subset is highly representative of the overall population. This subset is known as sample. In the above example, instead of collecting nutritional information for every child in the country, an alternative approach could be to randomly select 1% or 2% of children from the country and measure their nutritional status, assuming them to be representative of the entire population. This method of data collection is called sampling.

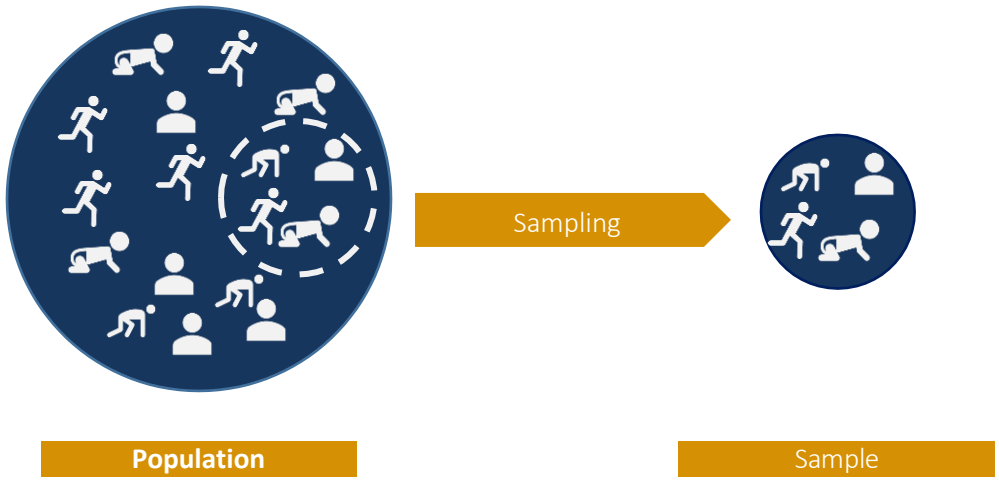

Figure 1: Population and Sample

Why Is Sampling Needed in Real Life?

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 for a small subset or sample of the population. At this stage, it is worthwhile to understand the reason behind the need for sampling.

For example, consider the population of a small country, say Country A. There are 726,680 children less than 5 years of age and 1,298,503 women 15-49 years of age. If one survey team could collect data on 13 women and 13 children per day, 6 teams would take 71 years to complete data collection! And this doesn't even include the travel time required by the survey team. 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

Before understanding the different types of sampling methods, it is also important to understand the difference between Parameter and Statistics. A formal definition is as follows:

A parameter is a value, usually unknown, and requires being estimated with minimum error using statistics from one or more samples. In simple terms, parameter is associated with Population (usually unknown) and Statistic is associated with sample (drawn from population, hence known).

Population is usually represented with Greek Letters (e.g. σ for Population Standard Deviation /μ for Population Mean). Statistic is represented by an English alphabet like S signifying standard deviation of sample.

There can be several types of sampling methods, which can be classified under Probability Sampling Methods and Non-Probability Sampling Methods. A diagrammatic representation has been shown below to help understand these better.

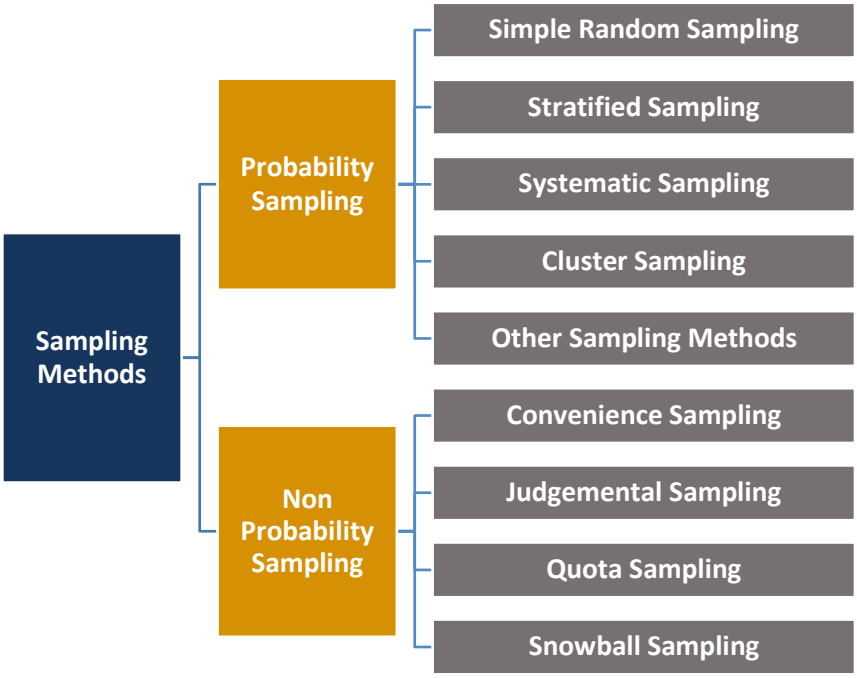

Figure 2: Types of Sampling

Each one of these have their own advantage and disadvantage. The method of sampling is decided considering various factors like cost, availability of data, applicability and resources required. But, for the purpose of this course, all sampling would be based on Simple Random Sampling, where each individual member or object in the population has an equal probability of being selected in the sample. This is the most commonly used Sampling technique.

Classification of Data

Data can be described as 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.

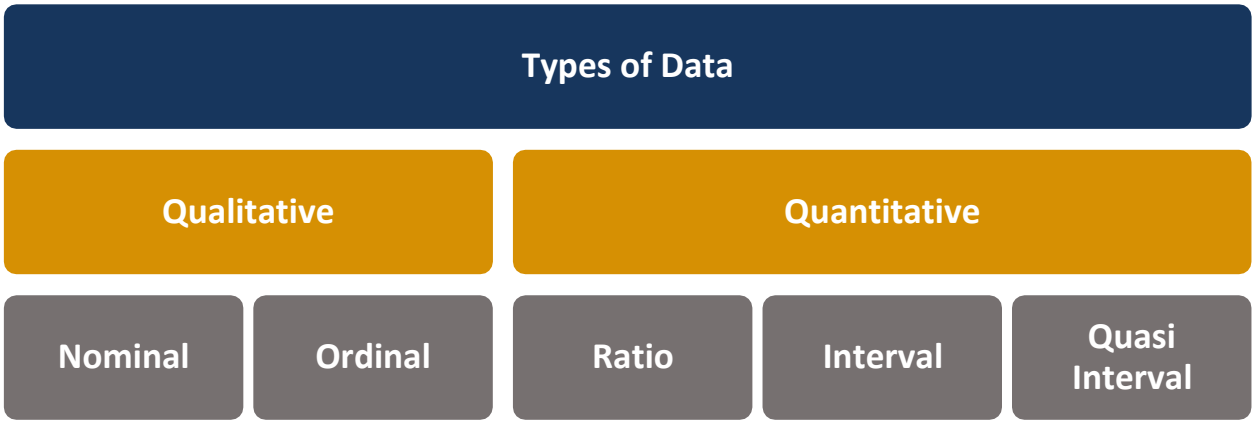

Figure 3: Types of Data

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. It is important to note the absence of order. 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 measures with order and establishes numerically equal distances on the scale. 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: An opinion poll with options from Strongly Disagree to Strongly Agree.

Ratio data measures have equal intervals and a 'true' zero point. It has all the properties of interval data with a clear definition of true 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.

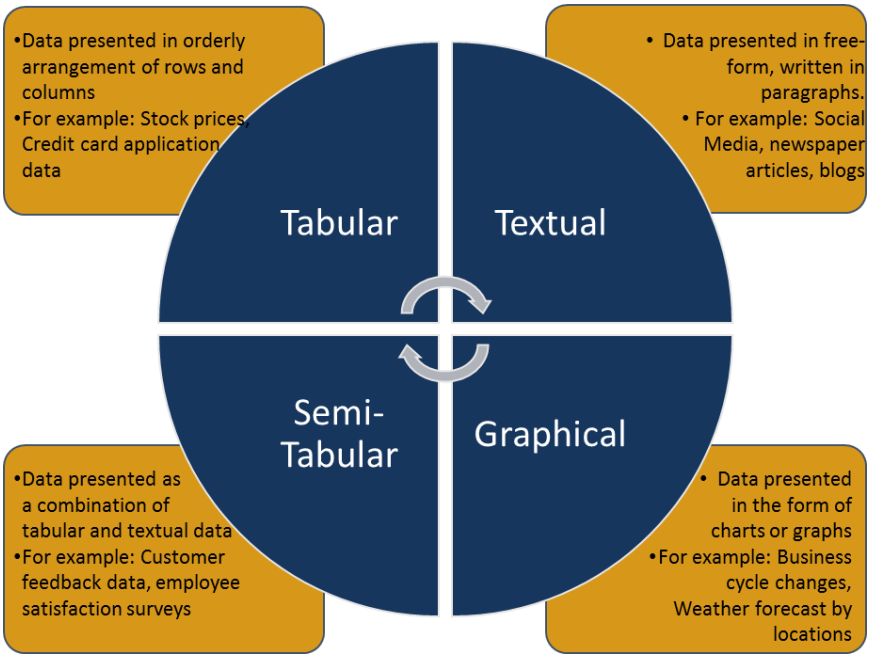

Figure 4: Representation of Data

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%

However, it is impossible to study the raw data for everyone in the survey. Hence the raw data needs to be summarized in some form that is readable and comprehensible.

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.

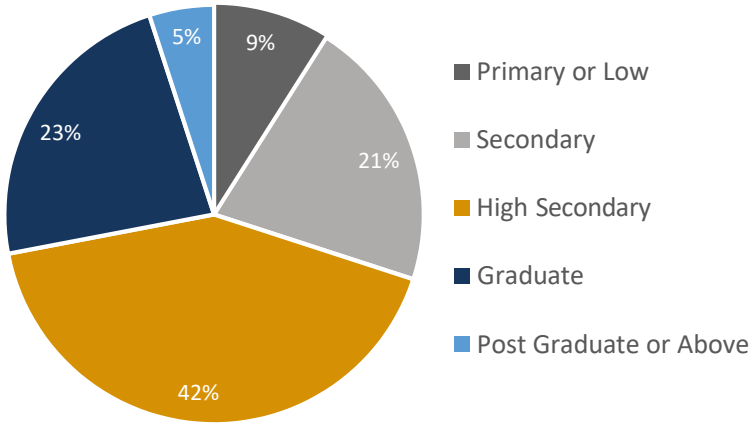

Figure 4: Graphical Representation of Data from Table 2

Table 2 can be represented as a Bar Chart, where height of each bar represents the total number of individuals within each score range.

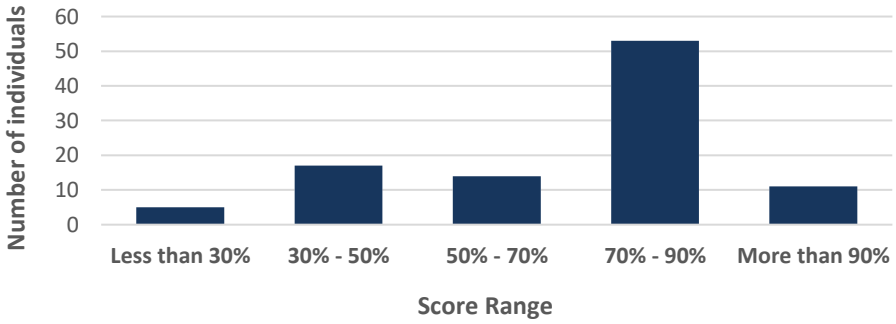

Figure 5: Graphical Representation of Data from Table 3

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

Whenever any metric of interest (e.g. share price of a stock, height of an individual, score of a student, delay in arrival of a flight, etc.) is measured, a fairly large number of observations in the sample tend to center around a single value. This value can be considered as a single representative value of the desired metric for the sample. As these values give the 'location' of a 'central' value of the sample, they are called "Measures of Location" or more commonly "Measures of Central Tendency" of the sample.

There are different statistics to measure the central tendency of any data. In this module, we will cover the most important measures of Central Tendency - Mean, Median, Mode, Percentile and Quartile.

Mean

Mean is the most common measure of central tendency. Mean 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 number of data points. For a data series, mean is simply all the numbers added and the sum being divided by the 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

Median is the middle number of any data series which has been sorted in ascending (i.e. lowest to highest) or descending (i.e. highest to lowest) order. If the series is odd, then the median is exactly the middle number. However, if the series is even, then it is the average of middle two numbers. Median is defined for ordinal data too, along with interval-scale or ratio-scale 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:

Mean is often not the actual value observed in a data series, however, it the most common value observed. Using mean for a data series for prediction purpose, it will produce least deviation/error from all other values in the data. It also includes all the data points in the series 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 we prefer Median over Mean or Mode?

If 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 mostly used for categorical (Nominal) data, where we need to find out which category is the most frequent. A general rule of thumb for finding the best measure of central tendency is listed below:

- For Symmetric interval or ratio data, mean is the central tendency measure to be used.
- For Skewed interval or ratio data, median is the central tendency measure to be used.
- For Ordinal data, median is the central tendency measure to be used.

Percentiles and Quartiles

Percentile is a way of providing estimation of proportions of the data that should fall above and below a given value. The p^{th} percentile is a value such that at most (p)% of the observations are less than this value and that at most (1 – p) % are greater, when the data is sorted. For example: A 99 percentile score in an exam means the examinee has scored higher than 99% of the other candidates who took this exam.

It is also important to remember the associated terms:

- 25^{th} percentile of the data is the First Quartile (Q1).
- 50^{th} percentile is the Median or Second Quartile (Q2).
- 75^{th} percentile is the Third Quartile (Q3).

Please refer to the video for further details.

Measures of Variability

To understand the data well, only learning about location is not enough. One must measure how the data is spread or scattered. A variable with less variability indicates more confidence in making any conclusion from data based on the location parameter. A high variability may indicate influence of some other unknown variable on the data. Hence, variability of a data indicates how much we still do not know about the data. The measures used to describe variability in a data are known as "Measures of Dispersion". In this section, we will cover the most fundamental Measures of Dispersion – Range, Interquartile Range (IQR), Variance, Standard Deviation and Coefficient of Variation.

Range

Range is the simplest measure of dispersion – it is the difference between the minimum value and the maximum value in the dataset. It offers a crude insight into the spread of the data, but is very susceptible to outliers. This measure does not make any assumption about the distribution of the data. This helps to capture the variation in the data. This is a helpful measure when you want to focus only on the extremes like weather reports – here understanding the range of temperature in a given day is sufficient.

For example: The temperature is measured every two hours during a given day.

Table 6:

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

It is evident that the minimum value for the temperature was 9^0C at 2:00 hours. The maximum is 22^0C at 14:00 hours. This could be an important measure if temperature was one of the deciding factors for an open-air event. However, this measure is distorted by presence of outlier or extreme

values in the data. Standard deviation 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)

Interquartile range or IQR is a related measure of range and it also measures the spread or variability of the data. It indicates how the data in a series is spread from the mean. Like range, IQR also does not make any assumption regarding distribution of data (i.e. it is non-parametric). IQR measures the spread of the middle half of the data. It is the difference between the third (75%) and first (25%) quartile of the data. So, by definition, it accounts for only 50% of the data.

As a rule of thumb, data points are more spread out as the IQR goes up. They are assumed to be uniformly spread around the mean if IQR is small. 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: 00011111223333333444444444445556

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 16th observation (i.e. 3). Now you need to identify the median of the first half of the data and the second half of the data. First quartile is the 8th student who owns 1 equipment; third quartile is the 24th 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 number of observations in the sample.

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

Simply explained, Variance measures the variability of the data around its mean. High variance means there is more variability or volatility whereas low variance indicates less variability. If variance of a series is zero that means all the numbers of that series are same.

Standard Deviation

Standard deviation is simply the square root of variance and it is the parametric equivalent of Interquartile range. This together with variance represents the variability in the data around mean. This is also often termed as 'volatility'.

Standard Deviation and Variance together are important measures of variability and are frequently used in statistics. In financial risk management, investors often worry about volatility of return i.e. how much the return spreads from the average. Standard deviation helps to provide measure of volatility of return and is considered to be a very important measure of risk.

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

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

The coefficient of variation (CV) is a measure of relative variation. It is the ratio of the standard deviation to the mean (average). It is a unit free measure and is measured in percentage. It is used to compare the variability in two or more data series with different units or very disparate averages.

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

At a first glance, these look incomparable, given how disparate their average revenue is. This is an appropriate application of "Coefficient of Variation". It is calculated as the ratio of standard deviation to mean of the data series being measured. It is represented as a percentage value.

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

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

In simple words, correlation is a measure that tells how closely two variables move in the same or opposite direction. A positive value indicates that they move in the same direction (i.e. if one increases other increases), where as a negative value indicates the opposite. It indicates the strength of association of two variables.

The most popular correlation measure for numerical data is Pearson's Correlation. This measures the degree of linear relationship between two numeric variables and lies between -1 to +1. It is represented by 'r'.

- r=1 means perfect positive correlation
- r=-1 means perfect negative correlation
- r=0 means no linear correlation (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:

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

Covariance can vary from negative infinity to positive 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.

- r_s=1 means perfect positive correlation
- r_s=-1 means perfect negative correlation
- r_s=0 means no monotonous 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.

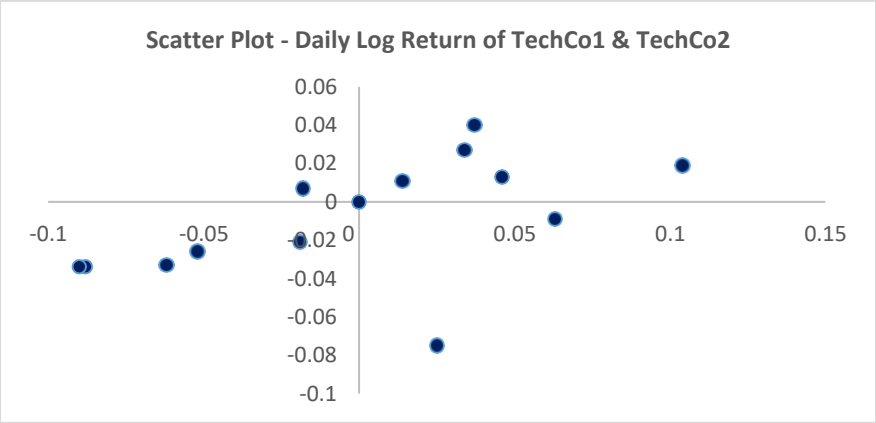

Figure 6: Scatter Plot

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.

Whereas, if there is a causal relationship between two random variables, 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