ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

Differentiating Strong Data and Evidence from Weak Data and Evidence: Another Heuristic for use in General and Social Sciences Research

Sujay Rao Mandavilli[1]

[1]Instittute for the Study of the Globalization of Science

Publication Date: 2025/07/03

Abstract: This paper is as usual tied to many of our earlier papers in the business. The core objective of this paper is to show why techniques to isolate and differentiate strong evidence from weak evidence need to be formulated, orchestrated and developed and communicated to the wider public and masses, and why this is in the best interest of science and society i.e. why it has the potential to catapult science and scientific temper to an altogether higher league and trajectory in many different parts of the world. We begin this paper by reviewing what data is, and examine the different types of data in common and widespread use. The different types of evidence are also critically and thoroughly examined, and the differences between data, information and evidence suitably brought out. The importance of data collection, data evaluation and examination data modeling, data correlation, and data synthesis are also brought out, and the importance of strong methods in this regard duly stressed and emphasized. A host and plethora of related mathematical and statistical techniques are also probed and investigated as they add meat and substance to the paper. The importance of research design is also emphasized, and different types of research design are probed and investigated. All the concepts in this paper are linked in a continuous chain, and the essential requirements of rock solid and high-quality research laid bare. While also providing suitable examples to bolster our case, we stress and emphasize the need to rank and rate different aspects of scientific activity on the basis of their inherent strengths. As such, we expect this paper to be a crucial cog in the wheel of our globalization of science movement.

How to Cite: Sujay Rao Mandavilli (2025) Differentiating Strong Data and Evidence from Weak Data and Evidence: Another Heuristic for use in General and Social Sciences Research. *International Journal of Innovative Science and Research Technology*, 10(6), 2368-2383. https://doi.org/10.38124/ijisrt/25jun1831

I. INTRODUCTION

"There is nothing more deceptive than an obvious fact." — Arthur Conan Doyle

"Extraordinary claims require extraordinary evidence." — Carl Sagan

This paper is as usual tied to many of our earlier papers in the business. The core objective of this paper is to show why techniques to isolate and differentiate strong evidence from weak evidence need to be formulated, orchestrated and developed and communicated to the wider public and masses, and why this is in the best interest of science and society i.e. why it has the potential to catapult science and scientific temper to an altogether higher league and trajectory in many different parts of the world. We begin this paper by reviewing what data is, and examine the different types of data in common and widespread use. The different types of evidence are also critically and thoroughly examined, and the differences between data, information and evidence suitably brought out. The importance of data collection, data evaluation and examination data modeling, data correlation, and data synthesis are also brought out, and the importance of strong methods in this regard duly stressed and emphasized. A host and plethora of related mathematical and statistical techniques are also probed and investigated as they add meat and substance to the paper. The importance of research design is also emphasized, and different types of research design are probed and investigated. All the concepts in this paper are linked in a continuous chain, and the essential requirements of rock solid and high-quality research laid bare. While also providing suitable examples to bolster our case, we stress and emphasize the need to rank and rate different aspects of scientific activity on the basis of their inherent strengths. As such, we expect this paper to be a crucial cog in the wheel of our globalization of science movement.

➢ *What is Data?*

This paper is a crucial cog in our attempts to build a rational, sane, and more scientifically-inclined and scientifically-oriented society. This is because it allows readers to figure out for themselves what strong evidence is

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

and what weak evidence is, and in the process help them to build much saner and more rational worlds. Many aspects of this paper must even be taught from a young age to school and college going students. Let us begin this paper by discussing and debating what exactly data is. To put it in simple words, data is nothing but a collection of raw, random, and unorganized facts, transactions, characters, symbols, figures or analytical descriptions of objects or things that record events and activities, and can be further meaningfully be processed and analyzed, and often thematically and systematically so, in order to extract meaningful and usable information for further downstream processing and complex analysis. Data may also be used as input variables for complex analytical processes. Such data may be qualitative, quantitative, discrete, continuous, ordinal, cardinal, statistical or non-statistical. It usually expresses and denotes concrete concepts, but in some cases, and rather unusually, represents vague and abstract concepts as well. A datum is an individual value in a collection of data, though this term is not often or commonly used. Data in its most common and typical connotation must already have been collected, recorded, stored in a meaningful and a readily accessible format such as tables or arrays, but is not yet processed in a systematic and a meaningful fashion so as to be more useful.

Data may exist in various forms, and is represented by different data types such as numbers, text, images, videos, and sound images, and can be used as a basis for further reasoning, discussion, investigation, or computation. Data is widely used in many walks of daily life. Data is commonly used in scientific research, economics, and many other forms of human activities. Examples of data from the field of economics include price indices (such as the consumer price index), inflation rates, unemployment rates, GDP values, GDP growth rates, literacy rates, and census data. In this context, data represent the raw facts and figures from which useful information can be extracted. Data has become so indispensible in today's world that it has been described as "the new oil of the digital economy". Data science has also become a much sought after profession in today's world, and data scientists are in hot and constant demand all over the globe. [1] [2] [3] [4] [5]

Let us now briefly review some definitions of data below. According to the Australian bureau of statistics, "Data are measurements or observations that are collected as a source of information. There are a variety of different types of data, and different ways to represent data." According to an official website of the United States government, "Data is defined as a value or set of values representing a specific concept or concepts. Data become 'information' when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary depending on its context. 'Data' includes, but is not limited to, geospatial data, unstructured data, structured data, etc. Information, means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms." In the famous and preeminent Oxford Dictionary, "data" is defined as "a collection of facts or organized information. It can encompass numbers, words, images, or premises used to draw conclusions, and it's often gathered through observation, experience, or experiments. In the field of economics, data refers to factual information, often numerical, that describes an economy, past or present, and is used for analysis and decision-making." According to the Oxford learners dictionary, "Data in its simplest sense, refers to facts or information, especially when examined and used to find out things or to make decisions". The above definitions bring out the meaning of the term data rather comprehensively, but this is just a small and an indicative list. More definitions have been attempted from time to time, though these are rather redundant and superfluous, and will not be presented in this paper.

➢ *What is Information?*

Information is systematically processed data, or data that has been processed on some firm and logical basis with a view to use it for some downstream purpose or application. Information is most often obtained after subjecting data to a series of structured and non-ambiguous processing operations which convert related groups of data (what we may call raw facts or data) into a meaningful and coherent form. Processing could be in the form of simple operations such as addition, subtraction, multiplication, division, or more complex mathematical formulae or functions, and other operations such as sorting, inversion, complementarity, rearrangement etc. This makes information useful and meaningful in the context of the real-world. In other words, information may be defined as the desired form to which data is finally transformed after undergoing a series of sequential processing steps. There are several important distinctions between data and information. Information is an abstract but well-oiled concept that refers to something which has the power to inform people, and plays a crucial and a demonstrably critical role in the decision-making process. Data and information, though related, represent different stages in the process of understanding and utilizing facts. Data is raw, unprocessed

[1] P. Checkland and S. Holwell (1998). *Information, Systems, and Information Systems: Making Sense of the Field*. Chichester, West Sussex: John Wiley & Sons. pp. 86–89

2 Vines, Timothy H.; Albert, Arianne Y. K.; Andrew, Rose L.; Débarre, Florence; Bock, Dan G.; Franklin, Michelle T.; Gilbert, Kimberly J.; Moore, Jean-Sébastien; Renaut, Sébastien; Rennison, Diana J. (2014-01-06). "The availability of research data declines rapidly with article age". Current Biology. 24 (1): 94–97

[3] American Psychological Association (2020). "6.11". *Publication Manual of the American Psychological Association: the official guide to APA style*. American Psychological Association

[4] Tuomi, Ilkka (2000). "Data is more than knowledge". *Journal of Management Information Systems*. **6** (3): 103–117

[5] P. Beynon-Davies (2002). *Information Systems: An introduction to informatics in organisations*. Basingstoke, UK

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

facts and figures, while information is data that has been processed, organized, and given context to make it meaningful and useful for further derivation and analysis. We may also state that data is the input into a process, while information is the output. Information is derived from data through processing, analysis, and interpretation. Data has relatively less practical use, while information is much more useful practically. [6 7 8 9]

Information is usually detailed enough to allow for smooth and effective decision making. Information must contain an appropriate level of details for the user and further downstream processing of data, and must be presented in an easily understandable and concise and structured format avoiding duplication and redundancies. When used by high-end users, the information must be very broad in scope while at the low-end user level, the information must be of a much more detailed nature and exhaustive in scope; Information must relate to the current situation or requirements and have acceptable level of integrity, and internal and external consistency. It must be precise, accurate, and free from error and bias; It must be produced through an optimum level of effort, and must be capable of being processed effectively as well. Both data and information constitute evidence from the point of view of this paper because they may be either strong or weak as the case may be.

II. DATA COLLECTION TECHNIQUES

Data collection is the process of gathering and structuring information in a preliminary but systematic way, allowing researchers to answer basic questions to begin with, test hypotheses in a preliminary fashion, and predict or list out desirable outcomes. Data collection is a crucial and critical step in both research and decision-making processes across diverse fields, and one that is essential for ensuring the availability of accurate and appropriate data for analysis and interpretation. Variables, or at least the more important and crucial ones, are listed out and defined very early in the research process. Data **is** often collected using techniques such as measurement, observation, query, or analysis, and is typically represented (often quantified) as numbers, ordinal or cardinal values or characters that may be further used for downstream processing. There are many different types of data collection situations. For example, field data is collected spontaneously in an uncontrolled, in-situ environment, though it must be fairly reliable for further downstream use. Experimental data is data that is generated in the course of a controlled scientific experiment. Data is then further analyzed using techniques such as calculation, logic, reasoning, debate, discussion, presentation, visualization, and other suitable and apposite forms of analysis. Prior to analysis, raw data or unprocessed data must be cleaned by removing irrelevant data, non-validatable data or information such as outliers or other errors.

To reiterate and put it in a nutshell, a systematic approach, i.e. a planned and organized approach must be adopted. Variables of Interest must be defined very early and research questions must be addressed through the data gathered. Accuracy, precision and consistency of data methods must be maintained, and integrity must be ensured at all times. Trends and patterns must also be identified, and outcomes must be predicted and evaluated very early in the research process as far as practically possible. Data collection methods can be broadly categorized into primary and secondary data collection methods, with each method being further divided into quantitative and qualitative approaches. Common methods include surveys, interviews, (structured, semi-structured and semi-structured interviews) questionnaires, observations, focus group discussions, ethnography, fieldwork, participant observation, and qualitative experiments. Document reviews may also be adopted as a technique, as also may be narrative analysis. This technique involves analyzing existing documents, such as reports, reviews, articles, or public documentation and records from gazettes, bulletins etc, in order to gather relevant and discernible information. Pre-existing datasets from various sources may also be used in order to answer research questions. We may also describe quantitative versus qualitative data collection strategies at this juncture. Quantitative data collection primarily focuses on numerical data, by using surveys, experiments, and observations to measure and analyze statistical relationships, while qualitative data collection focuses on descriptive data, using interviews, focus groups, and observations to explore in-depth understanding of experiences and perspectives. Mixed method data collection techniques typically use a judicious and harmonious blend of both.

The key considerations when choosing a data collection method are research objectives, time and cost constraints and considerations, resource availability, geographical location or convenience, etc. Reliability of data collection techniques is also an important factor for consideration. Data reliability refers to the consistency and accuracy of data, ensuring it can be trusted for decision-making. A reliable data collection method consistently produces the same results under the same conditions. Factors such as consistent procedures and rhythm in data collection including sample size selection, accurate and error-free measurement, proper calibration of instruments, and freedom from bias and subjectivity greatly contribute to data reliability. Standardized procedures and processes must be also used in data collection, and data validation and cleaning must be performed. All assumptions and exceptions must formally be recorded and documented,

---

[6] Nielsen, Sandro (2008). "The Effect of Lexicographical Information Costs on Dictionary Making and Use". *Lexikos.* **18**: 170–189

[7] *Stewart, Thomas (2001). Wealth of Knowledge. New York, NY: Doubleday.*

[8] Liu, Alan (2004). *The Laws of Cool: Knowledge Work and the Culture of Information.* University of Chicago Press

[9] *Floridi, Luciano (2005). "Semantic Conceptions of Information". In Zalta, Edward N. (ed.). The Stanford Encyclopedia of Philosophy (Winter 2005 ed.). Metaphysics Research Lab, Stanford University.*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

and there must be both internal and external consistency, also on a temporal basis. This is a core and crucial component of this paper, and must as such be treated with a great level of seriousness. There are four other tests that need to be borne in mind and these include, A) Test-retest reliability, i.e. assessing the consistency of results when the same test, approach or method is administered to the same set or group of participants at different points in time. B) Inter-rater reliability, i.e. evaluating the consistency of results when different individuals are using the same method to collect data. C) Parallel-forms reliability, i.e., comparing the results of two different versions of a test designed to measure the same construct. D) Internal consistency reliability i.e., assessing the consistency of results across different items within a measurement instrument (e.g., a survey).

It is only by focusing on these aspects, that researchers will be able to enhance and augment the reliability of their data, leading to more trustworthy, credible and reliable findings across time and space. Of course, data must also be cleaned by getting it rid of unwanted elements, and data must be brought into a proper format to make it readily usable, and accessible. Data cleaning must also be done properly, and data entry errors, interviewer errors, and respondent errors must be corrected. For this to be accomplished, best practices in data entry must also be followed and proper validations and checks and balances must be in place to prevent input of erroneous data, and to maintain referential integrity and data consistency. Outliers also known as wild codes and out of range codes, must be re-verified through box plot diagrams or box plot whisker diagrams and other techniques, and this is also known as an extreme case check. Irrelevant data is sometimes removed through data sanitization methods, though these represent irreversible methods. Data must also be properly authorized though input, and access rights and segregation of duties must also be properly enforced. Data entry formats and templates must be properly and carefully designed. [10] [11] [12]

III. TYPES OF DATA

Types of data include quantitative data which is numerical data that can be measured or counted, often used for statistical analysis, and further downstream mathematical computation. Quantifiable data is numerical, measurable, and objective always. It is also known as cardinal data. Qualitative data refers to non-numerical information that captures the qualities, characteristics, and attributes of something. It is descriptive in nature, focusing on understanding concepts, experiences, and opinions rather than providing numerical measurements. Qualitative data is non-numerical, descriptive, subjective, and non-quantifiable directly at least. It is often contextual, and helps us understand the how's, and why's of a complex issue. Qualitative research can also be used in social science studies such as behavioural analysis. Examples of qualitative data include photographs, audios, videos, interview transcripts, narratives, etc. How reliable **is** qualitative data? The credibility of qualitative data also known as dependability refers to the consistency of qualitative research, and can be ensured through multiple counter-balanced and integrated perspectives throughout the data collection process to ensure that data is pertinent, relevant, adequate, and appropriate. This may be done through data, investigator, or methodological triangulation; participant validation or carefully instituted checks and balances; not to miss out on the rigorous techniques used to obtain, gather, vet or validate data. Often secondary, mitigating, and compensating checks and balances may also be instituted. This will ensure that both the data used in research, and the research process and output itself is seamless, efficient, consistent and dependable. Quantification techniques such as Likert's scale and Thurstone's scale are often used, but the reliability of such techniques may be disputed. From our perspective, qualitative techniques, are indispensible in social sciences research, and mathematical and statistical techniques may be used only where needed.

Statistical data refers to numerical facts and figures collected, analyzed, and interpreted to understand trends, patterns, and relationships within a dataset. It is used in various fields for informed decision-making, policy development, and research. Key aspects of statistical data include systematic collection, organization, analysis, interpretation and interpretation. Non-statistical data usually refers to information that is not based on or derived from statistical methods. This type of data often involves qualitative information, subjective assessments, or data that does not easily lend itself to numerical analysis or probability-based conclusions. It focuses heavily on descriptive or textual information rather than numerical assessments, summaries or patterns. Some amount of subjectivity may be involved in the analysis and interpretation of such data. Probabilistic data throws up results in terms of an event occurring on a scale of between 0 and 1 with values tending towards one being the most likely values. The results thrown up in such as case are often approximate and non-precise.

Categorical data is a form of qualitative data, which represents characteristics or qualities that can be sorted into distinct groups or categories. Nominal data, also known as nominal scale data, is a type of categorical data where variables are divided into groups or categories that cannot be ordered or ranked. Nominal data is categorical data with no inherent order or ranking. Examples of such data include colors (such as red, blue, yellow, orange, or green) or types of fruit (apple, banana, mango, grape, watermelon, papaya, or orange). These categories are simply labels, and there is no inherent numerical value or quantitative relationship

[10] Most, Marlene M.; Craddick, Shirley; Crawford, Staci; Redican, Susan; Rhodes, Donna; Rukenbrod, Fran; Laws, Reesa (October 2003). "Dietary quality assurance processes of the DASH-Sodium controlled diet study". *Journal of the American Dietetic Association*. **103** (10): 1339–1346

[11] Data Collection and Analysis By Dr. Roger Sapsford, Victor Jupp

[12] Ziafati Bafarasat, A. (2021) Collecting and validating data: A simple guide for researchers. Advance. Preprint

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

between them. In case of nominal data, there is no implicit order or hierarchy. Some characteristics of nominal data are that is categorical, order-free, descriptive, non-analytical, qualitative, and non-numerical. It cannot also easily lend itself to mathematical or statistical analysis.

Ordinal data is categorical data with a meaningful order or ranking i.e. one that makes sense from a practical point of view. Examples or ordinal data include product satisfaction levels (very satisfied, satisfied, neutral, somewhat dissatisfied, or highly dissatisfied) or educational levels (high school dropout, bachelor's degree, master's degree or doctorate). Ordinal data is a type of data that can be categorized and ranked, but the intervals between the categories are not necessarily equal or easily identifiable and quantifiable. In sum, ordinal data represents ordered categories where the order is of great importance, but the precise differences between the categories might not be meaningful or calculable thereby not lending themselves to a mathematical or a statistical analysis.

Discrete Data refers to numerical data that can only take specific, separate values (usually though not always represented by whole numbers, and decimals are sometimes used). Examples include the number of students in a school or the number of trees in a garden. Discrete data is a count that involves integers — therefore, only a limited number of values is permissible or possible. Discrete data includes discrete variables that are finite, numeric, and countable. Continuous data on the other hand refers to any numerical data that can take any value (including non-integer values) within a given range. The number of such values is usually infinite. Examples of continuous include temperature, height, or weight measured upto a certain number of decimals. Alternatively, such data may also be represented in terms of fractions, though this is increasingly less common in today’s world.

Continuous data represents measurements that can take any value within a given range, including fractions and decimals. It contrasts with discrete data, which can only be counted in whole numbers. Examples of continuous data include height, weight, temperature, and time, as these can be measured with varying degrees of precision and can fall anywhere within a range. There are some other Important data types as well. For example, we may have date time data which represents dates and times in an acceptable or a widely used format such as a ddmmyy format. Boolean data represents true or false values. These are logical values that are widely used in computer programming and decision making. This term was named after the famous mathematician George Boole. More complex data types such as data arrays, data dictionaries, data libraries, and data frames that can hold multiple pieces of data. Interval data is a type of quantitative data where values are measured along a scale with equal intervals between them, allowing for meaningful comparisons of differences. This type of data possesses order and direction allowing for the execution of different types of analysis. Ratio data is a type of numerical data that possesses all the main characteristics and features of interval data (for example, it is ordered, and has equal intervals) but with the added feature of a true zero point. In other words it is data classified relative to a fixed zero point on a linear scale. Mathematical operations can then be performed easily and with highly predictable and meaningful results. Examples of ratio measurements include those pertaining age, distance, height, weight, and volume. [13] [14] [15] [16]

We must also distinguish at the very outset between primary data and secondary data. Primary data refers to original information that is collected directly from a source for the purposes of a specific research study or analysis. Primary data therefore is raw data that has not been already pre-processed or interpreted by other researchers. Primary data stands in stark contrast with secondary data, which is information that has already been collected and is available for public use. Examples of primary data may include period and historical data – for example, we knew that the Delhi Bombay motor reliability trials were held in 1905 (December 26th 1904 to January 2nd 1905) because it is attested to by large primary and period data – it was organized under the aegis of the then Viceroy General of India, Lord Curzon to expose India to the automobile, and test its suitability to Indian conditions. The concept of idiographic data which is very important for social sciences research, largely focuses on the unique characteristics and experiences of an individual or specific case, rather than seeking to establish generalizable laws or patterns. This approach and technique which leads to rich and in-depth understanding of issues, must be contrasted with nomothetic and nomological approach to research design. Examples of this kind of research technique include cases studies and narratives; also content and discourse analysis. However, this approach requires personalized interventions, has limited generalizability, is resource intensive, and may lead to highly subjective results. A dataset is a collection of data in a structured form, usually in the form of a table. Other concepts such as data segments are highly technical, and are not useful or applicable for the purposes of our study.

Data may also take on the shape of a normal distribution, binomial distribution, or a Poisson distribution. A normal distribution, which is also sometimes known as a Gaussian distribution, is a type of probability distribution that is symmetrical and bell-shaped with data points clustering around its central value. This type of distribution is often used in Six sigma techniques, and is associated with parametric tests. A binomial distribution is a common discrete probability distribution that models the probability

[13] Rudas, Tamas (2010). "Probability Theory: An Outline". In Lovric, Miodrag (ed.). *International Encyclopedia of Statistical Science*. Springer. pp. 1123–1126

[14] Williams, David (2001). "Preface". *Weighing the Odds: A Course in Probability and Statistics*. Cambridge University Press. pp. xi–xvii

[15] Hays, William Lee, (1973) *Statistics for the Social Sciences*, Holt, Rinehart and Winston, p. xii

[16] *Dodge, Yadolah (2003). The Oxford Dictionary of Statistical Terms. Oxford University Press*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

of achieving a specific number of successes in a specified number of independent trials, where each trial has two possible outcomes (namely success or failure). In this type of distribution, there are two parameters: 'n', which is the number of trials, and 'p', which the probability of success in a single trial. Q, representing the possibility of failure is given by 1-p. The Poisson distribution is a popular and commonly used discrete probability distribution that expresses the probability of a given number of events occurring within a fixed or a stipulated interval of time. Datasets must also be assessed for complexity from time to time and factors contributing to complexity include volume, variety, data velocity, and veracity among other factors.

IV. DATA MODELING

Data modeling which is a commonly and widely used concept in data science refers the process of creating a visual representation of data along with its multifaceted internal and external relationships, which helps it serve as a template and a blueprint for how data is organized, structured, stored, archived, and accessed within a system. Data modeling typically involves defining entities, attributes, and their interdependent relationships in order to ensure consistency, transparency, clarity, completeness and efficiency in data management. Data modeling is critical in many elements of data structuring and data governance. Key aspects of data modeling include visual representation through the use of symbols, notations, and diagrams, definition of entities and attributes, definition of relationships between entities. Data modeling can be accomplished at multiple levels such as conceptual data modeling, logical data modeling, (with data types defined and relationships established) physical data modeling, (with table structures and indexing), etc. Benefits of data modeling include improved data quality, enhanced data management, better and more informed and reliable decision making, etc. While many of these concepts originated in the field of computer science, they are applicable in general research as well. Data modeling and data analysis must be performed reliably and accurately, and there must be level of consistency involved.

➢ *Other Related Mathematical Concepts*

Let us now review some mathematical concepts extremely briefly, as a detailed analysis will be somewhat out of place here. In mathematics, a constant is a value or number that remains fixed and does not change within a specific context or a specific problem. In other words, it remains constant and non-changing throughout the problem lifecycle. It can be a numerical value like 15, -33, or a symbol representing a fixed, and an unchanging quantity with which it is equated. Constants are quite fundamental and central in many fields of mathematics, because they provide stable reference points within equations and expressions. A variable on the other hand, is a symbol, usually a letter of the English i.e. Latin alphabet that represents a mathematical object or entity whose value can change or vary within a specific context. In other words, it can take on multiple values and values can readily and easily be assigned to it. In the equation $y = 7x + 7$, x and y are variables. Given that there are two variables in this case, this problem can only be solved using simultaneous equations. In other words, another equation must be present. Variables are usually quantitative. However, in some cases, they can be qualitative too.

In mathematics, the independent variable is the input value that is wholly independent and not dependant on any other variable. It is changed or controlled in an equation or experiment, while the dependent variable is the output value that depends on the changes in the independent variable, and whose value changes in accordance to changes values in the independent variable. This may be understood in cause and effect terms as well. An intervening variable, also known as a mediating variable, is a variable that explains the relationship between an independent and dependent variable and acts as an intervening mechanism through which the independent variable influences the dependent variable. A random variable is a variable whose numerical value is determined by the outcome of a random phenomenon, and can consequently take on a random value. This is another extremely important and crucial concept in the field of mathematics. A function is another very important concept in mathematics. A function is a rule that assigns an output value to each input value. Functions are used to model relationships between quantities in mathematics and as such this is an extremely useful and crucial concept with many downstream uses and implications. The set of all input values of a function is called a domain, and the set of all output values is called a range. We also have concepts such as linear functions, quadratic functions, trigonometric functions, polynomial functions besides one-to-one (Injective) function where each output corresponds to a unique input, onto (Surjective) function where every element in the range has a corresponding element in the domain, many-to-one function where multiple inputs can produce the same output, into function where not all elements in the codomain have a corresponding element in the domain, and inverse function which is a function that reverses the action of another function. Independent and dependant variables must be initialized correctly, and the interrelationship between them clearly established.

We also have integers which are whole numbers, including positive numbers, negative numbers, and zero. They do not include fractions or decimals unlike non-integers which include fractions and decimals. Real numbers are all the numbers that can be found on the number line, encompassing both rational and irrational numbers. They include integers, fractions, and decimals, representing a continuous set of values. All numbers except for complex numbers, is a real number. It may be stated here that complex numbers are a way to represent numbers that combine real and imaginary components. Rational numbers are those numbers that can be expressed as a fraction, while irrational numbers are those which cannot. A fraction represents a part of a whole, or a quantity that is not a whole number, and is written as two numbers separated by a line, with the top number (known as the numerator) indicating the number of parts, and the bottom number

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

(known as the denominator) indicating the total number of equal parts the whole is divided into. Wherever the denominator is equal to zero, the value of the fraction is equal to infinity. Types of fractions include proper Fraction i.e. numerator is less than the denominator, improper fraction i.e. numerator is greater than or equal to the denominator, mixed fraction i.e., a combination of an integer and a fraction, equivalent Fractions i.e. fractions that represent the same value, even if they have different numerators and denominators, like fractions i.e. fractions with the same denominator, and unlike fractions i.e. fractions with different denominators. While all these represent valid and bona fide values, data must be inputted, manipulated and used properly. Wherever required, approximation techniques may be used based on the level of precision required, and this is often achieved through truncation. The latter is particularly useful if a high degree of precision is unwarranted, computationally expensive or impractical.

Decimals are a practical and a highly intuitive way to represent numbers that include both whole number and fractional parts, using a decimal point to separate them. The decimal system is based on powers of ten, hence the name "decimal" which stems or originates from the old Latin term "decimus" meaning tenth. There are essentially four important types of decimals: namely terminating, non-terminating, recurring, and non-recurring decimals. As the name suggests or implies, terminating decimals have a finite and countable number of digits occurring after the decimal point, while non-terminating decimals have an infinite and uncountable number of digits occurring after the decimal points. Recurring decimals have a repeating pattern of digits (series of five, seven, or ten, for example) after the decimal point, while on the other hand, non-recurring decimals do not. It is preferable to use decimals over fractions in the interests of precision except in some cases. The decimal system is also compatible, additionally with the Metric system or the SI system of units. We also then have negative numbers whose values fall below zero, and indeterminate numbers such as zero divided by zero, or infinity divided by infinity. In case of the latter, values cannot be known with certainty. [17] [18] [19] [20]

[17] *Ginammi, Michele (February 2016). "Avoiding reification: Heuristic effectiveness of mathematics and the prediction of the Ω⁻ particle". Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics. **53**: 20–27*

[18] *Wilson, Edwin B.; Lewis, Gilbert N. (November 1912). "The Space-Time Manifold of Relativity. The Non-Euclidean Geometry of Mechanics and Electromagnetics". Proceedings of the American Academy of Arts and Sciences. **48** (11): 389–507*

[19] Rossi, Richard J. (2006). *Theorems, Corollaries, Lemmas, and Methods of Proof*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. John Wiley & Sons. pp. 1–14, 47–48

[20] Mueller, I. (1969). "Euclid's Elements and the Axiomatic Method". *The British Journal for the Philosophy of Science*. **20** (4): 289–309

V. TEMPORAL DATA

Temporal data refers to information that period to a specific point or period in time with start and end dates captured. Temporal data captures how data changes over a specific period in time, making it extremely important for understanding trends, patterns, changes in data, research on underlying reasons for change, and the historical context of changes. Examples of temporal data include time series data, event logs with date, time, author and modifier stamps, and data with associated validity periods, or expiry dates. Examples of such data include stock process, weather trends, etc, and there are many more common examples that we can cite here. Temporal data is important for time series analysis, trend analysis, predictive modeling, performance evaluation, anomaly detection, historical analysis, exponential smoothing, and data extrapolation. Time series analysis is an important and widely used statistical method that is used to analyze data points collected over a period of time, typically at consistent, regular or pre-specified intervals, to identify patterns, trends, and make predictions about future values. This is an important technique that is used to understand how a variable changes over time, why it changes, and to forecast and predict its behavior into the future. Trend analysis is the process of examining historical data and changes to historical data over time in order to identify patterns and predict future trends as well. It is widely used in various fields, including finance, marketing, and social sciences, to understand how data or events change with the passage of time and to make informed decisions or predictions.

A cyclical trend refers to a pattern of recurring fluctuations in data, also known as cyclical fluctuations, usually attributed to economic cycles or other repeating events, with a period of more than a year and not relatively fixed or uniform like seasonal patterns which are also tied to (and are generally subservient to) calendar years. These fluctuations are often characterized by non-uniform periods of expansion and contraction of unequal lengths or durations, also known as peaks and troughs, and are typically attributed to different types of economic, political, social, or cultural factors. Another important concept that we must mention here is that of exponential smoothing. Exponential smoothing is a time series forecasting method that uses weighted averages of past data points, while at the same time giving greater weightage to more recent observations. The core objective of this technique is to smoothen out fluctuations in data in order to make it easier to analyze. Time series analysis – and all its components including exponential smoothing- must be carried out properly and a great deal of judgment must be used in the process. This is necessary to prevent grave errors of omission and commission. This is a principle that we will stand by steadfastly in this paper and beyond. No blind and half-based manipulation and jugglery of data must be resorted to at any cost, and under any circumstances. The researcher must have a crucial and critical feel for the entire

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

research process from head to toe, and must rule over it with an iron fist or an iron grip. [21] [22] [23] [24]

Spatial data, which is also sometimes known as geospatial or geographic data, represents information about the location, size, direction, and shape of a geographic entity, regardless of whether they are natural and man-made. Examples of such data include latitude, longitude, and altitude. This data is sometimes used in different crucial and critical applications, from location mapping and urban planning to navigation and mapmaking. It is also used in Global Positioning Systems (GPS) which are satellite-based navigation systems that provides their users with reliable and accurate information on various parameters such as location, time, and velocity information anywhere on or near the Earth's surface. Vector data is also widely used in geographical information systems to represent various geographic features using points, lines, and polygons, among other shapes. In addition, three-dimensional datasets refer to data points that are conceived and represented in a three-dimensional space, with each data point being identified by three coordinates namely x, y, and z. Data integrity, data consistency and data reliability must be ensured regardless of data type, and this is but a foregone conclusion. This is a crucial spoke in the wheel for accurate, reliable and precise research. We do not wish to enter into discussions regarding other types of data such as real data and virtual data given the fact that they are used only in specialized contexts, and are by and large irrelevant for the purposes of this paper. Nonetheless and nevertheless, data and information are very crucial and critical from our point of view, given that they constitute a form of evidence that is used in further downstream research and analysis. Data may also be classified based on its centrality and criticality to research; for example, black box data is crucial and central to an aircraft accident investigation along with other qualitative data such as videos and testimonials, while peripheral data may not. [25] [26]

- *What is Evidence?*

Evidence refers to something – usually a body of reliable and rock solid data - that furnishes proof, and offers incontrovertible evidence for an act or an occurrence. This is how it differs from data which is often raw and purposeless. It may quite often refer to a testimony, and something that is submitted to a tribunal or a court of justice in order to ascertain the truth of a matter within a legal framework, and on legal grounds. Evidence may also be provided by an individual who bears or carries witness and provides a detailed account and narrative of a crime and leads to guilty people being punished or prosecuted. Evidence, including scientific evidence or one that is used in scientific evidence varies greatly in quality, and latent or inherent strength. High quality or strong evidence is that for which the scientists' belief in the truth of the claim is relatively high or large, weak evidence is that for which the belief is relatively small or weak. Key characteristics of strong evidence include relevance and support to the question or topic in hand, (or the argument or discussion currently being debated) and the innate ability to override negative or contradictory evidence that may manifest itself or be presented from time to time. The source of the evidence must also be authoritative and trustworthy, and its provenance must be clearly established. Reliable evidence often stems from an expert or an authority in the field, academic or non-academic, with or without institutional affiliation, whether legally admissible or non-admissible, and the evidence is often pre-published in a peer-reviewed journal. It must also be relevant and crucial for the purposes of the study.

The evidence must also be verifiable and replicable; it must be accurate and precise, it must be internally and externally valid, coherent and consistence, and must satisfy the principle of consistency over space and time. It must also satisfy the axiom of sufficiency and completeness. It must be sufficient and complete with respect to the issue or topic at hand. Therefore, both the quality and quantum of evidence are extremely important. The evidence must also be current and must not be outdated or obsolete. The data collection method must be transparent, methodological and systematic. There must be a total absence of bias, including researcher bias, data bias, or sampling bias, and researchers must possess objective in mindset. Evidence must also provide direct support to the claim, and must stem or emanate from credible sources. Evidence must be corroborated through multiple sources, (including wherever possible, eyewitness accounts) and must be actionable as well. In must also not be transactional, anecdotal or episodical. Evidence may also be oral or written, but as far as possible, must be documented properly, thoroughly, comprehensively and exhaustively.

On the other hand, unreliable evidence includes hearsay evidence, stand alone evidence, evidence from unreliable sources, unnecessary or irrelevant evidence, and evidence that is not corroborated through multiple sources. By thoroughly understanding and assimilating the innate and underlying characteristics of strong evidence, it becomes easier to understand what weak and non-admissible evidence is as well. For example, it is sometimes said that Sivananda was born on the 8th of August 1896, although this has not been independently verified. He was allegedly

[21] De Gooijer, Jan G.; Hyndman, Rob J. (2006). "25 Years of Time Series Forecasting". *International Journal of Forecasting*. Twenty Five Years of Forecasting. **22** (3): 443–473

[22] Weigend A. S., Gershenfeld N. A. (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992)

[23] Woodward, W. A., Gray, H. L. & Elliott, A. C. (2012), *Applied Time Series Analysis*, CRC Press.

[24] Shumway R. H., Stoffer D. S. (2017), *Time Series Analysis and its Applications: With R Examples (ed. 4)*, Springer, ISBN 978-3-319-52451-1

[25] M. R. Spiegel; S. Lipschutz; D. Spellman (2009). *Vector Analysis*. Schaum's Outlines (2nd ed.). US: McGraw Hill

[26] Rolfsen, Dale (1976). *Knots and Links*. Berkeley, California: Publish or Perish

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

born in the Sylhet District of undivided Bengal Presidency of British India. It this claim were true, he would have died at the age of 128 years, making this a potential world record for longevity. Evidence is widely used in legal contexts. It is used to determine whether a person is guilty or not. It is also widely used in forensic audits, and in document analysis – i.e. to establish the provenance of ancient texts and date them reliably. It is also widely used in scientific research. [27] [28] [29] Scientific research is a systematic and empirical process of investigation and critical examination that is conducted in order to establish new facts or to arrive at a new set of conclusions, in a process that involves making observation, conducting experimentation, and performing analysis. The objective of scientific research is to expand understanding of natural and social worlds, validate existing theories, or develop new applications through the use of a scientific methodology that often involves hypothesis preparation and testing.

Use of evidence in forensics is a very important application of evidence analysis. Forensic studies refer to scientific and multidisciplinary approaches that are gainfully employed to collect, analyze, and interpret evidence for the purposes of legal and forensic investigations. Forensics encompasses a wide range of mutually interdependent fields of practice under the broad umbrella of forensics, including forensic investigation, psychology, chemistry, criminal law, biology, sociology, and many more allied and interrelated fields of study. Forensic science always involves a systematic approach to crime scene investigation and evidence analysis with distinct and well-defined steps such as securing the scene of the crime, documenting the evidence carefully and systematically, collecting and preserving evidence, analyzing the evidence in a laboratory using scientific methods, and presenting findings in court. We also have other concepts such as information provided by the entity, and information used by the entity, and as such, these must be applied very carefully. [30] [31] [32]

[27] American College of Forensic Examiners Institute. (2016). *The Certified Criminal Investigator Body of Knowledge*. Boca Raton, Florida: CRC Press. pp. 112–113

28 Dogan, Aysel (2005). "Confirmation of Scientific Hypotheses as Relations". Journal for General Philosophy of Science. 36 (2): 243–259

29 Reiss, Julian; Sprenger, Jan (2020). "Scientific Objectivity". The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University.

[30] Reid, Donald L. (2003). "Dr. Henry Faulds – Beith Commemorative Society". *Journal of Forensic Identification*. **53** (2)

[31] Tewari, RK; Ravikumar, KV (2000). "History and development of forensic science in India". *J Postgrad Med*. **46** (46): 303–308

[32] Keith Inman, Norah Rudin, *Principles and Practice of Criminalistics: The Profession of Forensic Science* (p. 32), CRC Press, 2000

VI. RESEARCH DESIGN

Research design refers to the general and overarching plan or framework that guides a researcher and other people involved or associated with the study on how a research study will be conducted, and how its various components will be structured. The research design outlines the overall approach to research, including but not limited to research questions, data collection methods and techniques, and data analysis methodologies and techniques that must be employed. A research design ensures that the research will be conducted as efficiently and effectively as possible, leading to overall valid and reliable answers to the research problem or question at hand. We may have both qualitative and quantitative research designs. Examples of quantitative research designs are descriptive, experimental and correlational research design, while example of qualitative research design are exploratory research design, grounded theory, phenomenology, narrative analysis and case study methods. We also have other important concepts such as cross-sectional research design and longitudinal research design. A cross-sectional research design is a type of study that involves analyzing data from a population at a single point in time by taking a snapshot of a population at a given point in time in order to understand the prevalence of certain characteristics or associations between sets of defined variables. A longitudinal research design on the other hand, involves studying the same group of participants over an extended period, and carefully observing and systematically recording changes and trends over time. The research design and data analysis methods must also be apt in relation to the research problem in question, and the data and evidence must be handled and analyzed appropriately and reliably for proper consistency of results. Data triangulation, method triangulation and investigator triangulation must also be employed as this will lead to more reliable research results. This involves changing the data, method and investigator as required in order to ensure reliability and repeatability of results. [33] [34] [35]

VII. TYPES OF EVIDENCE

Let us now analyze the different types of evidence in brief and see how they impact our study for all practical purposes. The strength of limitations of each kind of evidence must be established on a case to case basis, and preferably formally through an explicit analysis. These categorizations and classifications only represent the more common types of evidence in use. In legal contexts especially though also in other contexts, the term "stand-alone evidence" generally refers to direct evidence that can prove a fact independently and on its own without needing

[33] Creswell, J.W. (2012). Educational research: Planning, conducting, and evaluating quantitative and qualitative research. Upper Saddle River, NJ: Prentice Hall

[34] Bell, J. (1999). Doing your research project. Buckingham: OUP.

[35] Robson, C. (1993). Real-world research: A resource for social scientists and practitioner-researchers. Malden: Blackwell Publishing

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

to be combined with other evidence or inferences. Circumstantial evidence is indirect evidence that suggests a fact or event happened, but is not able to prove or establish it directly. It relies on inferences and logical reasoning to connect the dots, namely evidence to the conclusion. This type of evidence must be used alongside other evidence to build a case, especially when direct evidence is not available. For example, we may have an alibi which clearly demonstrates that the alleged criminal was in another location when the crime was purported to have taken place. The exact opposite of circumstantial evidence is direct evidence which links the criminal directly to the rime. Evidence may also be oral or written, presented formally or informally, and may be tangible or intangible. Tangible evidence is also known as physical evidence or real evidence in common parlance. Hearsay evidence is often culled from secondary sources, and may be unreliable. Evidence may also be based on testimonies, and this is known as testimonial evidence. Evidence may also be strong or weak and the strongest form of evidence may come either from first-hand evidence or forensic evidence. The latter is infrequently used in scientific research.

Corroborative evidence in both scientific research and criminology refers to evidence that supports or confirms pre-existing or pre-presented evidence, although it makes it more credible, trustworthy or probable. This kind of evidence cannot stand independently on its own legs or feet, but instead strengthens other evidence by providing additional support for a claim, assertion or a statement. Empirical evidence is information acquired through rigorous and credible observation and experimentation alone, relying on direct, verifiable, and often quantifiable and measurable experience and evidence gathered through the senses alone rather than on abstract theory. It's the foundation of scientific inquiry and is crucial for validating or refuting claims and hypotheses. Examples of empirical evidence may include observations, and first-hand scientific investigations, though this is only a short list. We had also written extensively about the need to use contradictory data as far as possible. We had also introduced the concepts of self-canceling contradictory data or evidence and non-self-canceling contradictory data or evidence. For example, Steve Farmer's research on the Indus script was dubious and one-sided, and Asko Parpola's was utterly outdated. Let us cite another interesting example. If the Harappans had moved eastward in 1900 BC, it does not mean that Aryan culture could not have moved westward again in 1600 BC. Both these represent distinct pieces of evidence, and one does not automatically cancel out another. Both must be considered for building historical models, and no valid or bonafide data can be brushed under the carpet at any rate. We had also written previously about the concepts of cross-cultural research design, and dialectical approaches to introduce multivocality, but these concepts are somewhat at variance with the objectives of this paper.

Prima facie evidence also known as presumptive evidence, uses evidence gathered from a crime scene to make a plausible and a realistic assumption. This may however be subsequently refuted through the gathering or collection of contradictory or opposing data by rival teams. Evidences may often be demonstrable through aids such as charts or graphs, and this is known as demonstrative evidence. This concept is related to documentary evidence which is properly and systematically recorded. Analogical evidence often draws analogies from similar types of evidence. Admissible evidence is a type of evidence that judges allow lawyers to present in court based on relevance, authenticity and value. Admissible evidence is always factual, pertains to case in question and possesses a significant intrinsic value as opposed to inadmissible evidence that lawyers cannot present to a jury. Examples of such evidence could include hearsay evidence, partial evidence or improper and biased or prejudiced evidence. Statistical evidence refers to numerical and statistical evidence, and this type of evidence is more common in scientific research. Evidence, in order for it to be highly reliable must also be tied to primary sources, and must as far as possible, be formally and comprehensively documented and recorded. Secondary and tertiary sources of data are generally unreliable, and are at best avoided. Data must be in its original form throughout the research process. It must also additionally be untampered, unaltered or unfalsified.

Scientific research must also be carried out from primary sources of data as far as possible. This will make it much more reliable, dependable and of course consistent, but cost and time considerations must be factored in as well. Primary research involves collecting new, original and previously unused data in order to address a specific research question, or solve a research problem. It involves first-hand interaction with the subject of research, unlike secondary research which relies on pre-existing data. This type of research allows for better control over the data collection process besides allowing for new vistas and horizons for unique insights. Secondary research involves reusing pre-existing data that was collected by another researcher for another purpose. Secondary research often provides a cost-effective and time-effective approach to gathering insights by analyzing information from various sources like books, literature, research articles, research reports, newspaper articles. monographs, and databases. In research, tertiary sources are works that compile, index, or organize information from primary and secondary sources, and more common examples of it include , dictionaries, treatises, dissertations, encyclopedias, textbooks, research reports with a lot of secondary data, and bibliographies. [36] [37] [38]

[36] Porter, Roy; Lorraine Daston; Katharine Park. *The Cambridge History of Science: Volume 3, Early Modern Science*. p. 805

[37] McCrie, Robert D. "General Managerial Fundamentals and Competencies". *Security Operations Management*. 1st ed. Amsterdam: Butterworth-Heinemann/Elsevier, 2007. 93

[38] *Reid, Donald L. (2003). "Dr. Henry Faulds – Beith Commemorative Society". Journal of Forensic Identification.* ***53*** *(2).* See also this on-line article on Henry Faulds: *Tredoux, Gavan (December 2003)*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

➢ *Interpretation and Analysis of Data*

Data interpretation refers to a systematic and structured process of reviewing, analyzing, and identifying patterns in datasets in order to draw definite and meaningful conclusions and derive crucial and critical insights. It involves understanding metadata or data about data, including purpose and scope, cleaning and transforming the data into a useful format, and then applying analytical methods to discern meaningful patterns and trends. This process therefore transforms raw data into understandable information that can greatly aid in decision-making. Data analysis on the other hand, refers to the process of inspecting, analyzing, cleansing, transforming, and appropriately modeling data in order to discover useful and not so readily apparent information, draw meaningful insights and conclusions, and support and aid in informed decision-making. Data analysis involves a plethora of techniques such as statistical analysis, to extract insights from raw data. This process can greatly aid in discerning patterns, trends, and complex relationships within data, for further meaningful action. Techniques such as data mining, dimensional analysis and slice and dice analysis can be used here. Usage of statistical techniques such as computation of mean, median, mode, standard deviation, and variance can greatly help, and appropriate sampling strategies must be applied wherever required and necessary. Factor analysis can be used to identify critical variables, (This greatly aids in simplification) and alternatively techniques such as correlation analysis, regression analysis, cluster analysis, cohort analysis and data segmentation may also be used. Data synthesis, the process of combining and aggregating data from various diverse sources in order to create new information in a new and better (more usable) format , by employing a range of techniques such as meta analysis, data aggregation, data summarization, data transformation, etc. Data interpretation and analysis must be performed intelligently, and without losing sight of the overall, larger objective. This includes, as always data reliability and reliable and consistent communication to other users, researchers and stakeholders. Public must also be educated about the need for better high-quality science, and we have a long way to go here. Perhaps we need multi-dimensional reform in this department.

➢ *Data Correlation*

Data correlation, in the field of statistics, refers to the claimed or established relationship between two or more variables, and the strengths of association between such variables. Correlation analysis sets out to indicate how changes in one variable are associated with ripple or cascading changes in another. Correlation does not automatically imply causation, meaning that just because two variables are correlated, it does not mean that a change in one variable is responsible in causing changes to the other variable. There are different types of correlation, including positive correlation (the values of both variables either rise or fall together), negative, (the value of one variable rises, and the other falls, or vice versa) and zero correlation (no correlation at all), each indicating a specific relationship between the variables. The correlation coefficient, or correlation of coefficient whose value ranges from -1 to +1, indicates the strength and direction of the relationship. A value of +1 or -1 indicates a perfect positive or negative correlation, respectively, while a value of 0 indicates no correlation. Common techniques to measure correlation include Pearson correlation, Spearman correlation, and scatter diagrams. Sometimes, rank correlation is used, and this technique is particularly useful when rank is more important than absolute value. Sometimes, regression analysis is also used to identify a relationship between an independent variable and a dependant variable. Types of regression analysis may include simple linear regression – relationship between two variables only, multiple regression – multiple independent variables and one independent variables, polynomial regression, - uses polynomial equations, etc.

We must also differentiate between causal laws versus statistical laws. Causal laws in general describe relationships where one event directly determines or causes another, and always so, while statistical laws describe probabilistic relationships, often showing correlations between events without implying a direct cause-and-effect link, or a sure fire recipe for change or impact. In other words, statistical laws are normally tied to likelihoods and probabilities, while causal laws are tied to certainties and inevitabilities. Extrapolation of data is also often carried out, and in all such cases, researchers must assess the reliability of data, methods or techniques involved before using it in his or own research. Extrapolation of data is a statistical method used to estimate or predict values that fall outside the range of observed data points. It involves extending an observed pattern or general trend in existing data to forecast future or unknown outcomes. This is a way of making educated guesses about what might happen in future based on available data. However, there is an assumption of linearity involved, and as such this method is prone to risks and errors. To overcome this, polynomial extrapolation is sometimes used. Extrapolation is different from intrapolation because the latter plots values within the predetermined or predefined range, and is resultantly much less useful.

➢ *Qualitative Versus Quantitative Research*

Quantitative research must also be differentiated from qualitative techniques. The former is more common in the physical sciences, and the latter is more common in the social sciences, though this is by no means a hard and a fast rule. Quantitative research mostly deals with mathematical and statistical values. Quantitative research can be broadly categorized into descriptive research, correlational research, quasi-experimental, and experimental research. In case of experimental research, there is a random assignment of subjects to control and experimental groups, while in the case of quasi-experimental design, the assignment is not random. It must be stated here at the very outset, that in an experiment, the control group is some kind of a masters baseline, receiving no treatment at all, while the experimental group receives the treatment whose effect is being tested. It is only by comparing the outcomes of these two groups that researchers will be able to determine the true effect of the treatment.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

Qualitative research is a type of research that focuses on understanding the nature of phenomena by gathering and analyzing non-numerical data. In such cases, quantities or numerical values are seldom directly dealt with. Types of qualitative research techniques include interviews, questionnaire, focus group discussions, emic perspectives of various kinds, etic perspectives of various kinds, etmic perspectives, ethnography. Let us now attempt to answer the question: How reliable is qualitative research? The answer to this question depends of what the objectives of the research is. In most fields of social sciences research, absolute precision may prove to be elusive – such fine tuned and granular precision may also not be necessary for the most part. Quantification techniques such as Likert's scale, Thurstone's scale, and rank order scales may also be used, though they may not entirely be objective or cut and dry. Mixed methods research is also often applied and this is a judicious and harmonious blend of the two basic streams of research.

- *Logic*

Logic refers the study of reasoning, (and making valid and reasoned conclusions) encompassing and covering the principles, concepts and methods that are used to distinguish between valid and invalid arguments, and shades in-between. It involves analyzing the structure of reasoning, identifying patterns, making arguments and inferences, and evaluating the correctness of inferences. The validity and soundness of conclusions reached must be clearly established at all times, and this is an extremely important point to consider under any circumstances. We have different types of logic such as deductive logic which focuses on inferences that guarantee the truth of the conclusion as long as the premises are valid and true. Inductive logic deals with inferences in which the conclusion is made probable by the premises, but not completely guaranteed. Abductive logic provides the best explanation to a situation, but one that is not guaranteed. Logic may also be classified into formal logic and informal logic based on the formality of rules involved. Logic is used in many fields of science, and in many walks of everyday life, but most people in the street are not formally trained on logic. This is why they sometimes falter and arrive at absurd conclusions. Logic and reasoning skills must be properly taught in schools as a part of critical thinking skills, and as a part of scientific method. Logical sequitors must be adhered to under all circumstances, and logical non-sequitors avoided at any rate – in other words, conclusions must logically emanate from the premises. Logical fallacies and invalid arguments must also be explained to students from a young age, and convoluted or needlessly complex logic avoided. Therefore, we believe that the field of pedagogy itself needs a radical overhaul, and all these aspects and components accommodated, embedded and suitably incorporated into school syllabi.[39] [40]

[39] Rocci, Andrea (8 March 2017). *Modality in Argumentation: A Semantic Investigation of the Role of Modalities in the Structure of Arguments with an Application to Italian Modal Expressions*. Springer. p. 26.

- *Epistemology*

Epistemology refers to that branch of philosophy which examines the nature, origin, and the natural limits of knowledge, also known as justified true belief, and the knowledge building or the knowledge creation process. Epistemology is also called "the theory of knowledge", as explores different types of knowledge exhaustively and intricately, including forms of knowledge such as propositional knowledge about facts, practical knowledge in the form of usable and practical skills, and knowledge acquired through familiarity or experience. Epistemologists study the concepts of truth, belief, and justification, along with sources of justification such as reason, introspection, memory, perception, and testimony to understand the true nature of knowledge, and its limits. The concepts of epistemology must be taught from a young age, and all students must be made aware of them. This is necessary to avoid dogma and prejudice and build up a rational mindset. [41] [42]

We must also realize at the same time, that some things are unprovable, and some things are unknowable. For example, on what date and time did Indo-European migrants arrive on Indian soil? This is limited by the quantum and range of data available. Godel's incompleteness theorems, conceived, developed and released by the mathematician Kurt Godel in 1931, are two fundamental results in mathematical logic that show inherent limitations in formal axiomatic systems. According to the first of his theorems, there will always be true statements that cannot be proven within the system even in a sufficiently strong and consistent formal system. According to the second theorem such a system cannot even prove its own consistency. Natural limits to knowledge also include sensory constraints, language limitations, brain processing speed, and the possibility of unknowable truths. There may be some cognitive biases present in the mind of the researcher, and the independence of researcher, and the independence of tests formalized and executed. There may be vested interests and conflicts of interests present, and we have discussed all these concepts previously.

Researchers may also lack objectivity in mindset, a concept that we have been mulling upon off and on. Marxist historians apparently do not possess this in adequate measures as can be understood from their double standards as far as the reconstruction of Indian history between 1900 BC and 600 BC – they virtually love to proclaim that nothing existed during this period, but narratives magically

[40] Schreiner, Wolfgang (2021). *Thinking Programs: Logical Modeling and Reasoning About Languages, Data, Computations, and Executions*. Springer Nature. p. 22

[41] Sharpe, Matthew (2018). "The Demise of Grand Narratives? Postmodernism, Power-knowledge, and Applied Epistemology". In Coady, David; Chase, James (eds.). *The Routledge Handbook of Applied Epistemology*. Routledge. pp. 318–331

[42] Wheeler, Gregory R.; Pereira, Luís Moniz (2004). "Epistemology and Artificial Intelligence". *Journal of Applied Logic*. **2** (4): 469–493

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

reappear when they want to project it in negative light. A right-leaning Indian individual (a highly educated individual to boot) told the Author, that he would not accept anything from anyone if he disputed the literal interpretation of Indian scriptures and other sacrosanct theories such as the divine origin of Indian scriptures theory, dating them on impossible epistemological grounds to millions of years ago. This alone shows that we still have a very long way to go. This problem is not limited to India along; we have dubious concepts in the West such as the theory of intelligent design proposed by members of the discovery institute among others who have developed concepts such as the concept of irreducible complexity and specified complexity. Researchers must possess skepticism, but not skeptopathy, and this is a fine and a delicate balance to tread. Interdisciplinary research and corroboration, transdisciplinary research and corroboration, and multidisciplinary research and corroboration must be pursued, but indeed, this may not be done. Other innovative and creative thinking techniques such as lateral thinking, out of the box thinking may also not be used, thereby limiting the diversity and breadth of the human experience.

Errors may occur in the processing of in evidence or data, and error handling in data processing may also occur. Issues must also be flagged and qualified as necessary, though this may not always be done. Third party review of claims must be done carefully, and attestation of evidence may be performed as required. Ideologies, which we defined and explained multiple times in the past along with techniques to identify and isolate ideologies, must be shown the door, and scientific ideologies, a concept first proposed by Georges Canguilhem must be quickly booted out too. Dogma and identity-driven interpretations must also be booted out and shown the door. The quality and direction of research in many cases still depends on the researchers cultural background, mind-orientation, and cultural-orientation, and this is rather unfortunate indeed. Researchers resort to cherry picking, selective obfuscation and amnesia as well, and confirmation bias is also often present. Overt agendas, hidden agendas, or ulterior motives sometimes still rule the roost. We often have evidence denial, and there are many climate change deniers even in the USA to this very day. Researchers still often make sweeping assertions, tall and unsubstantiated claims, grandiose claims, make statements out of context, sweeping generalizations, gross exaggerations, and over simplifications, and in this respect, are not much better than the laity. For example, some people state that artificial intelligence will wipe out global jobs, though this statement is not based on robust data or proper predictive modeling. This is akin to the 1980's paranoid fear that computerization would. kill jobs. This never materialized or came to pass.

Sampling, refers to the process of selecting a small set of data from a larger group of data (also known as a population or sampling frame) to study and draw conclusions about the entire group of data which it may not be practically possible to study owing to time and cost constraints and considerations. In the field of statistics, sampling is a oft-used technique to analyze data from a large and composite population by studying a smaller, but highly representative sample. This allows researchers to make reliable and generalizable inferences about the whole population without having to collect data from every individual in an elaborate or a time consuming manner. Sampling techniques are broadly categorized into probability and non-probability sampling. Probability sampling includes simple random sampling where Every member of the population has an equal and known chance of being selected, systematic sampling where samples are selected on a systematic and a consistent basis, stratified sampling which is carried out by dividing the population into subgroups and then taking a random sample from each subgroup, cluster sampling, which involves dividing the population into clusters and randomly selecting some clusters. Non-probability sampling includes convenience sampling which involves selecting individuals who are readily available and accessible and easy to contact and connect with, quota sampling which ensures that the sample is reflective of the proportions of specific characteristics in the population such as age, weight and height. Snowball sampling is not so often used, but it involves using referrals from initial participants to identify and quiz more participants. Sampling errors occur from an inappropriate sampling strategy or size, and must be avoided at any costs. Other techniques such as analysis of variance are also used in statistics, but are not so important from our point of view.

Also refer out paper, "Advocating output criteria based scientific and research methodologies: Why the reliability of scientific and research methods must be measured based on output criteria and attributes" published by us in IJISRT in August 2023, where all the output attributes of high quality research such as reliability, objectivity, precision, accuracy, rigour, and consistency were discussed in great detail. We will not repeat or reproduce the entire list here, but it would suffice here to state, and at this juncture, that high quality research can only stem from objectivity in mindset. Objectivity in mindset can stem only if clarity and objectivity begin at the beginning of the research lifecycle, and that is the collection and analysis of the data phase itself. Objectivity in mindset will also arise if researchers acknowledge the fact that their research will be used by downstream researchers to draw additional conclusions, and researchers as such need to be very careful. Use of reliable dating techniques must be employed such as radio carbon dating and thermo luminescence dating. Hermeneutical techniques must also be used, along with glotto chronological techniques and other bona fide techniques. People must be taught the importance of search for the truth and intellectual honest. This will put an end to data forgery as attributed to NS Rajaram, when a seal falsified to represent a horse. [43] [44]

[43] Advocating output criteria based scientific and research methodologies: Why the reliability of scientific and research methods must be measured based on output criteria and attributes Sujay Rao Mandavilli IJISRT, August 2023

[44] Recognizing "Non self-cancelling contradictory evidence" as and when it occurs or arises: Delineating its special place in twenty-first scientific method

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

Traditional sources of data must be used with great care and utmost caution. Uses of traditional sources of data must be limited and must be backed up with corroborative evidence. Limitations of traditional sources of data must also be understood and laid bare for the general public to see and hear. The pertains to traditional dating methods adopted for the great Indian epics and older Indian literature that cannot hold up to the mildest of scrutiny, and do not stand up to the litmus test of transdisciplinarity. They push their dates to impossibly earlier periods that have not found wider acceptance among a larger body of Indian and international scientists. Likewise, the notion of flying sages and flying ancient Indian aircraft or even idols drinking milk also must be taken with a large lump of salt. Most certainly, they are pseudoscience. This is also true of the notion of the tower of Babel, and divine origin theory of languages from ancient India. Some claim that a heavier than air flight was made in India towards the end of the nineteenth century, but the first reliable claim of an aircraft in India was in 1911 –February 18th, 1911 to be exact and precise – from Allahabad to Naini, carrying mail – it was a Humber biplance and was flown by Henri Pequet. We also know that the Indus valley civilization was pre-Vedic because it lay in an area west of the Aryavartha, and because it went out of business by 1400-1300 BC. It originated from Maehrgarh in Baluchistan unlike the Vedic culture which originated from the Vedic homeland. Objects in the Indus valley civilization are also different from post-Harappan artifacts. All this counterweighs so called traditional and mythological data convincingly.

Likewise, all ancient and current systems of medicines must be reexamined and revalidated from time to time if they are to remain valid and current. The power of tradition alone will not work. Of late, mainstream publications like the Frontline magazine have resorted to this mumbo jumbo. The Iron age in Tamil Nadu has been pushed back to impossibly early periods based on jingoistic parochialism. All their other contemporaries subscribe to much higher standards of journalism. The journal in question is highly left-leaning. Some Indian nationalists want to claim Rakhigarhi as the epicenter of the Indus valley civilization based on supposed dating of Individual items. This may not be in order as the roots of the Indus valley civilization lie in Mehrgarh in Baluchistan. There supposedly was a sunken city found in the Gulf of Cambay some twenty years ago, and this was dated to 7000 BC. During a visit to Lepakshi in Andhra Pradesh in 1995, the author was informed by the caretaker of the local monument that it could be dated to the early India epics. As per archeological evidence, the monument was only five hundred years old and datable to the Vijayanagara empire. His narratives were lapped up eagerly by a gullible (Mostly educated) public. Some may see this as a revivial of tradition, but this is ludicrous, and outright egregious to say the least

A certain Mr Raman from the Indian state of Tamil Nadu supposedly converted water to petrol in the 1990's. There have been miracle busters too in India on the lines of Mr James Randi of the west. We have the Rationalist association of India for example, and other noted individuals such as Abraham Kovvur and H Narasimhaiah. They have not unfortunately garnered wide support because humans are always humans, and are more often than not drenched in tradition and steeped in cultural narrative. This is as worldwide phenomenon not just limited to India. The Beatles made a trip to India in 1968 supposedly for salvation, and so did Steve Jobs and Alanis Morisette. Scientism cannot work, and that is why we need emic perspectives and ethnography of enculturation. The Titanic was hailed as a marvel of modern engineering but was brought down by hubris. It did not have fully sealed bulkheads or double sided hulls. The Boeing 737 was a catastrophe and disaster in the making. As the new engine had the potential to scrape the runway, it was moved up front, leading to a loss of balance. It was equipped with an MCAS system, but this was not communicated to the pilots. There were two crashes in Indonesia and Ethiopia where hundreds were killed, leading to a temporary grounding of the aircraft. The financial position of Boeing was also a contributing factor. A culture of quality is also extremely important, not just technology. Likewise, Indian automobile manufacturers took life easy during the License Raj turning out shoddy products. This was because the Indian car market was a seller's market at the time, not a buyer's market, and one that encouraged a customer-friendly attitude. Many factors contribute to technological progress, and we must not adopt a myopic and an overly simplistic view. While this may appear to be a digression, it is as a matter of fact, not. People-centric and culture-centric approaches must be adopted to drag people upwards into the scientific age. This aspect has sadly not been completely or fully understood.

Archeoastronomy is another fiasco which must be taken with a liberal pinch of salt as it throws up highly erroneous results. It found acceptance among Indian nationalists not too long ago. Paranormal claims such as extrasensory perception, clairvoyance, telepathy and intuition such as those attributed to Edgar Cayce and others must be investigated carefully and impartially. The same goes for claimed or supposed miracles, reincarnation and out of the body experiences that are claimed from time to time. Methods for testing and evaluating paranormal claims are not robust, reliable or foolproof. Extraordinary claims require extraordinary evidence. This is popularly referred to as the Sagan standard. We need methodological inductivism, epistemic coherentism, institutional coherentism, foundationalism, integrationism, confirmation holism, grounded research, and inductive approaches in science. We also need pragmatism and practicalism in every facet of scientific endeavour and activity. We need b**ona** fide science, not pseudo science, proto science, subpar science, and non science. Beliefs, opinions, and attitudes with cognitive, behavioural and affective components factored in, though important in their own way, do not matter a great deal for core scientific activity. People will continue to use what we called fuzzy logic or improper logic, but this must be surmounted and overcome in the long

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

term through the medium of better education. Fallacies in science continue to about and we had listed all of them previously. One of the most common fallacies is the appeal to authority fallacy which Indians believe in a great deal. Everything is attributed to "some great man", and even highly educated people use this. The appeal to ignorance fallacy is also widely used as the general public is perceived to be ignoramuses.

- *Assumptions, Postulates and Axioms*

An assumption is something very common in science. It may be defined as something that is accepted as being true or as absolutely certain to happen, without any extra or additional proof. It is something that is taken for granted to word it differently. An assumption is also accepted as a basis for reasoning or for action, even if it is not definitively or incontrovertibly established. We must also discuss the concept of Occam's razor at the same breadth. Occam's razor is a principle which states that the simplest explanation for a phenomenon is the best or the most plausible one. Therefore, the hypothesis with the fewest number of assumptions is the most likely, and should be taken forward. In science, a proposition is a statement that seeks to establish a clear and lucid relationship between concepts or variables. It also serves as a basis and a building block for further complex theories. An axiom, which is another important concept in scientific method**, is** a statement or proposition which is regarded as being established, accepted, beyond the shadow of any reasonable doubt within the concept of the research problem or the research question. Axioms may be either strong or weak, and we have the concept of strong axiom or weak axiom too. A postulate is defined as an advanced and a highly complex hypothesis or a set of hypothesis that is agreed upon by a vast majority of scholars or scholarly people, and one that is useful in providing proof or concrete substance to untested propositions and hypotheses, thereby taking it forward for more advanced research.

Paradigm**s** and frameworks are far more complex and may be comprised of multiple entities. We had discussed these multiple times previously. Also read our paper, "Continuous zero-based reassessment of assumptions, hypotheses and methods": A vital tool and technique in the interests of better science" which was published over a year ago, and our paper on irreducible simplicity as well. We must also distinguish between strong hypotheses versus weak hypotheses. A strong hypothesis is one which is corroborated with maximum lines of evidence, and possesses maximum explanatory power. A preliminary hypothesis is a working or an initial hypothesis, and one that is further tested and ratified as more evidence is obtained. A more refined form of a hypothesis is known as a theory, and once incontrovertibly established becomes a law. However, the distinction between all these three concepts is not properly maintained even by mainstream scholars and researchers, with rampant miscategorizations galore, and this is a rather unfortunate trend. This only goes on to show the amount of clutter that has built up in the field of scientific research over the years. The reliability of hypotheses and theories would also depend on the quality of data used, and these would likewise need to be graded. For example, theories on the origin of spoken language would be notoriously unreliable because of the limited amount and unreliable nature of data available. We may talk about static pace of linguistic change, and accelerating pace of linguistic change, but these are also dubious and iffy**.**

We still have a long way to go, and we must build a more rational world. Student must be taught scientific method. This is because scientific method is far more important than rote learning. We must bring about a change. Future generations must be different from present ones, and there must be a generational change. Even educated people don't know to differentiate science from pseudo science. Even educated people don't know to differentiate high quality science from subpar science. Even educated people do not know to distinguish high quality evidence from dubious or low-quality evidence. This is the pathetic state of affairs today. We must move towards an evidence rating scale. We may propose a 1 to 10 scale to begin with. Composite rating must be employed (with all the individual components rated individually before aggregation), judgment must be exercised, as this cannot be a blind and a senseless exercise. Along with this, the entire scientific activity or undertaking must also be rated. We look forward to more work and new research here, and more meaningful contributions by other erudite and knowledgeable researchers and scholars. We hope that this paper will be a useful starting point for this process.

VIII. CONCLUSION

This paper was as usual tied to many of our earlier papers in the business. The core objective of this paper was to show why techniques to isolate and differentiate strong evidence from weak evidence needed to be formulated, orchestrated and developed and communicated to the wider public and masses, and why this was in the best interest of science and society i.e. why it had, and still has the potential to catapult science and scientific temper to an altogether higher league and trajectory in many different parts of the world. We began this paper by reviewing what data was, and examined the different types of data in common and widespread use. The different types of evidence were also critically and thoroughly examined, and the differences between data, information and evidence suitably brought out. The importance of data collection, data evaluation and examination data modeling, data correlation, and data synthesis were also brought out, and the importance of strong methods in this regard duly stressed and emphasized. A host and plethora of related mathematical and statistical techniques were also probed and investigated and we explained how they added meat and substance to the paper. The importance of research design was also emphasized, and different types of research design were also suitably probed and investigated. All the concepts in this paper were suitably interlinked in a continuous chain, and the essential requirements of rock solid and high-quality research laid bare. While also providing suitable examples to bolster our case, we also stressed and emphasized the need to rank and rate different aspects of scientific activity on the basis of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1831

their inherent strengths. As such, we expect this paper to be a crucial cog in the wheel of our globalization of science movement, and bolster the tempo and magnitude of scientific activity around the globe.