DOI:https://doi.org/10.52206/jsmc.2023.13.3.815
OPEN ACCESS

What is Data, Its Type and Summary Statisitcs

Umar Hussain

This article may be cited as: Hussain U. What is Data, Its Type and Summary Statisitcs.
J Saidu Med Coll Swat 2023;13(3):132-133. DOI:https://doi.org/10.52206/jsmc.2023.13.3.815

INTRODUCTION

In medical research, data is collected from various sources, including patient records, clinical trials, surveys, and experiments. Researchers need to accurately classify and analyze this data to draw meaningful conclusions about diseases, treatments, and patient outcomes. Knowing the types of data (nominal, ordinal, interval, and ratio) helps researchers choose appropriate statistical methods for analysis.

Data

Data refers to a collection of information, facts, or observations that can be used for various purposes, including analysis, decision-making, and research.[1]

Type of data

Data can come in many forms and can be classified into different types. Types of data can nominal, ordinal, interval or ratio. *Nominal data* are categorical data that represent categories or labels with no inherent order or ranking. Examples include gender, colors, or types of fruit. Nominal data can only be used for qualitative analysis. *Ordinal data* also represent categories, but they have a meaningful order or ranking. However, the intervals between categories are not necessarily uniform. Examples include education levels (e.g., high school, bachelor's degree, master's degree) or patient's satisfaction ratings (e.g., very dissatisfied, dissatisfied, neutral, satisfied, very satisfied). *Interval data* have a meaningful order, and the intervals between values are consistent and equal, but they lack a true zero point. Common examples include temperature measured in degrees Celsius or Fahrenheit, where zero does not represent the absence of temperature. *Ratio* data have a meaningful order, consistent intervals, and a true zero point, indicating the absence of the measured quantity. Examples include height, weight, age, income, and the number of items in a collection. Arithmetic operations like addition and multiplication can be applied to ratio data.[2]

Summary Statistics

Summary statistics are numerical measures that provide a concise overview of a dataset's main characteristics. They help researchers and analysts understand the central tendency, variability, and distribution of data.

When we talk about central tendency, we are referring to the heart of the data, and three common measures come into play. The mean, or average, provides a numerical representation of typical values within the dataset. The median, on the other hand, identifies the middle value when data is ordered, helping to mitigate the impact of outliers. Lastly, the mode represents the most frequently occurring value.

Variability measures, such as *variance and standard deviation(SD)*, help us gauge the extent to which data points deviate from the central tendency, offering insights into the data's spread. Standard deviation quantifies how much individual data points differ from the mean. A high standard deviation indicates that data points are spread out over a wide range, while a low standard deviation suggests that data points are clustered closely around the mean. SD helps researchers assess the reliability and consistency of data, identify outliers, and make inferences about populations based on sample data. Standard deviation is the square root of the variance. Mathematically, it is represented as: Standard Deviation (σ) = ? Variance (var^2)

Distribution metrics like *skewness* and *kurtosis* help us understand the shape and symmetry of the data's distribution, offering crucial context for further analysis. These summary statistics are vital tools, enabling researchers and analysts to make informed decisions, detect underlying patterns, and gain a holistic view of the data's behavior. According to Bryne (2010), data is typically regarded as conforming to a normal distribution when skewness falls within the range

Saidu College of Dentistry, Saidu Sharif Swat

Correspondence: Umar Hussain
Lecturer, Saidu College of Dentistry, Swat, Saidu Sharif Swat.
Email: drumarhussain@gmail.com

Received: September 18th, 2023 Accepted: September 18st, 2023

of -2 to +2, and kurtosis falls within the range of -7 to +7.[3] If data is not normally distributed, assessed from summary statistics (skewness and kurtosis), median with interquartile range should be reported instead of the mean and standard deviation.[4]

In coming issue we will discuss how to report descriptive statistics in the form of tables and graph.

REFERENCES

1. Kirkwood BR, Sterne JAC. Essential medical statistics. 2nd ed. Oxford, United Kingdom: Blackwell; 2003. p. 14-41.
2. Pandis N. Statistics for orthodontists. Am J Orthod Dentofac Orthop. 2015;147(3):405-8.
3. Byrne BM, Van de Vijver FJ. Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. Int J Testing. 2010;10(2):107-32.
4. Van Belle G, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: a methodology for the health sciences: John Wiley & Sons; 2004.