

Scales of Measurement and Presentation of Statistical Data

Abstract

Measurement scale is an important part of data collection, analysis, and presentation. In the data collection and data analysis, statistical tools differ from one data type to another. There are four types of variables, namely nominal, ordinal, discrete, and continuous, and their nature and application are different. Graphs are a common method to visually present and illustrate relationships in the data. There are several statistical diagrams available to present data sets. However, their use depends on our objectives and data types. We should use the appropriate diagram for the data set, which is very useful for easily and quickly communicating summaries and findings to the audience. In the present study, statistical data type and its presentation, which are used in the field of biomedical research, have been discussed.

Keywords: Graphical presentation, qualitative variable, quantitative variable, scales of measurement

**Prabhaker Mishra,
CM Pandey,
Uttam Singh,
Anshul Gupta¹**

*Departments of Biostatistics
and Health Informatics and
¹Haematology, Sanjay Gandhi
Post Graduate Institute of
Medical Sciences, Lucknow,
Uttar Pradesh, India*

Introduction

Statistics is a branch of mathematics dealing with the collection, analysis, presentation, interpretation, and conclusion of data, while biostatistics is a branch of statistics, where statistical techniques are used on biomedical data to reach a final conclusion.^[1] Measurement scale (data type) is an important part of data collection, analysis, and presentation. In the data collection, the type of questionnaire and the data recording tool differ according to the data types. Similarly, in the data analysis, statistical tests/methods differ from one data type to another.

Data presentation is an important step to communicate our information and findings to the audience and readers in an effective way. If done properly, they not only reduce word count but also convey an important message in a meaningful way so that the readers can grasp it easily.^[2] There are various tabulation and graphical methods used to present the data, which are not possible without proper knowledge of data types.

The objective of this paper is to discuss the statistical data type (Section A) and its presentation (Section B), which is an important part of biomedical research.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

Section A

Scales of measurement

As data are the heart of the statistics, and at the time of data analysis and presentation, many people are confused about what type of statistical tools to be used on a set of data and the relevant forms of presentation or data display. Its decision is taken by looking the types of data and the objectives of the research.

Data

Data are a collection of facts such as values or measurements. It can be numbers, words, measurements, observations, or even just descriptions of things. Basically, data are two types: constant and variable. Constant is a situation/value that does not change, while a characteristic, number, or quantity that increases or decreases over time or takes different values in different situations is called variable. Due to unchangeable property, constant is not used and only variable is used for summary measures and analysis.^[1,3,4]

Types of variables

There are four types of variables: nominal, ordinal, discrete, and continuous. The first two are called qualitative data and the last two are quantitative data. The first two (nominal and ordinal) are assessed in terms of words or attributes called

How to cite this article: Mishra P, Pandey CM, Singh U, Gupta A. Scales of measurement and presentation of statistical data. *Ann Card Anaesth* 2018;21:419-22.

Address for correspondence:

*Dr. Prabhaker Mishra,
Department of Biostatistics
and Health Informatics,
Sanjay Gandhi Post
Graduate Institute of
Medical Sciences, Raebareli
Road, Lucknow - 226 014,
Uttar Pradesh, India.
E-mail: mishrapk79@gmail.com*

Access this article online

Website: www.annals.in

DOI: 10.4103/aca.ACA_131_18

Quick Response Code:



qualitative data, whereas discrete and continuous variables are part of the quantitative data.^[5]

Qualitative variable

Qualitative variable (also called categorical variable) shows the quality or properties of the data. It is represented by a name, a symbol, or a number code. These scales are mutually exclusive (no overlap) and none of them have any numerical significance. It is two types: nominal and ordinal.

Nominal variable: Nominal data are simply names or properties having two or more categories, and there is no intrinsic ordering to the categories, i.e., data have no natural ranking or ordering. For example, gender (male and female) and marital status (married/unmarried) have two categories, but these categories have no natural order or ranking.

Ordinal variable: An ordinal variable is similar to a nominal variable. The difference between the two is that there is a clear ordering in the data, i.e., ordinal data, unlike nominal data, have some order. For example, ordinal scales are seen in questions that call for ratings of quality (very good, good, fair, poor, very poor), agreement (strongly agree, agree, disagree, strongly disagree), economic status (low, medium, and high), etc. All the ranking data including Likert scales, Bristol stool scale, and all the other scales which are ranked between 0 and 10 are also called ordinal data.

Quantitative variable

Quantitative variable is the data that show some quantity through numerical value. Quantitative data are the numeric variables (e.g., how many, how much, or how often). Age, blood pressure, body temperature, hemoglobin level, and serum creatinine level are some examples of quantitative data. It is also called metric data. It is two types: discrete and continuous.

Discrete variable: Discrete variable is the quantitative data, but its values cannot be expressed or presented in the form of a decimal; for example, number of males, number of females, number of patients, and family size cannot be expressed in decimal in meaningful way.

Continuous data: Data are measured in values and can be quantified and presented in decimals. Age, height, weight, body mass index, serum creatinine, heart rate, systolic blood pressure, and diastolic blood pressure are some examples.

The variables such as heart rate, platelet count, respiration rate, systolic blood pressure, and diastolic blood pressure are in fact discrete (measuring in complete number) but are considered continuous because of large number of possible values. Only those variables which can take a small number of values, say, <10, are generally considered discrete.^[6,7] Summary is that if discrete variables values are at least 10 or more, then discrete variables can be

considered as continuous variable and we analyze them as per the methods applicable on continuous data.

Section B

Data presentation

Data presentation plays a crucial role in research. The researchers can convince their research to the reader by the effective data presentation. Basically, there are two types of data presentation: numerical and graphical.^[2,8]

Numerical presentation

There are various types of numerical presentation of the data including arranging them into ascending order, descending order, and classification of the data in the tabular form.

Graphical presentation

Graphs are a common method to visually illustrate relationships in the data. A chart, also called a graph, is a graphical representation of the data, in which the data are represented by symbols, such as bars in a bar chart, lines in a line chart, or slices in a pie chart. Graphs enable us in studying the cause-and-effect relationship between two variables. Graphs help measure the extent of change in one variable when another variable changes by a certain amount.^[8,9]

There are various types of graphical presentation given below.^[10,11]

Bar graph

A bar graph is the presentation of data using rectangular bars, with heights or lengths proportional to the values that they represent. The reader can easily compare the quantity by observing the length of the bar. In bar graph, the bars may be plotted either horizontally or vertically. In the x-axis, use categorical variable, while in y-axis, use numerical values. Bar graph is three types: simple, adjacent, and cumulative. The last two are also called multiple bar graph. In simple bar graph, maximum two variables (one categorical and one quantitative) are used, while in multiple bar diagram, maximum three variables (two categorical and one quantitative) are used. Multiple bar graphs are useful when a researcher wants to compare figures of two or more different data [Figure 1a and b].

Line graph

It is alternative graph of the bar graph. A line graph is a kind of graph which represents data in a way that a series of points are to be connected by segments of straight lines. Difference between bar and line graph is that bar represented by rectangle while line graph showing by line, although both used for the same purpose [Figure 1c].

Pie charts

A pie chart is defined as a graph which contains a circle and is divided into sectors. The arc lengths of the sectors are proportional to the numerical value they represent. It is used only for the categorical data [Figure 1d].

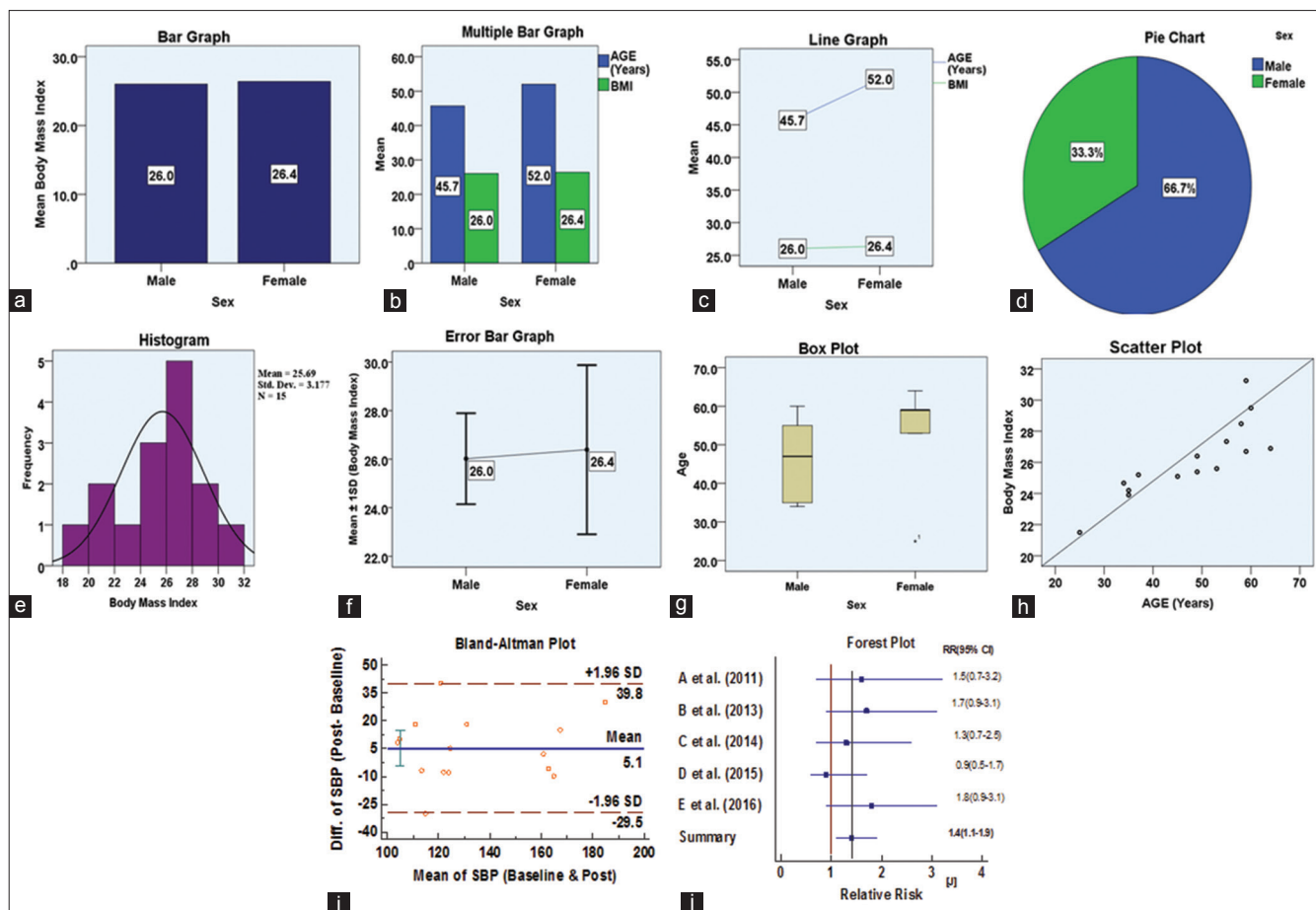


Figure 1: (a) Simple bar graph showing mean age between sex, (b and c) multiple bar graph and line graph showing mean age and BMI between sex, (d) sex distribution of the data, (e) histogram showing nature of the BMI distribution, (f) error bar graph showing mean BMI as per sex, (g) box plot showing age distribution as per sex, (h) scatter diagram showing linear relationship between age and BMI, (i) Bland–Altman plot showing relationship between SBP difference and their average values, (j) Forest plot showing relative risk and associated confidence intervals observed from five different studies for the exposure group as compared to nonexposure group. BMI: Body mass index, SBP: Systolic blood pressure

Histogram and frequency polygon

A histogram represents the frequency distribution of a continuous variable whose areas are proportional to the corresponding frequencies. A histogram is quite similar to the bar graph and both are made up of rectangular bars. The difference is that there is no gap between any two bars in the histogram. The histogram is used to check the normal distribution of continuous data and have only one continuous variable, and no categorical variables are used to plot it, while in bar graph, we have required at least two variables including one quantitative and one categorical variables. Frequency polygons serve the same purpose as histograms but are particularly helpful for comparing sets of data. Frequency polygons are also a good choice for displaying cumulative frequency distributions. When the midpoints of tops of the rectangular bars in histogram are joined together, the frequency polygon is made [Figure 1e].

Error bar

Error bars are graphical representations of the variability of data and used on graph to indicate the error or uncertainty (standard deviation/standard error/confidence

interval) in a reported measurement (mean). They give a general idea of how precise a measurement is or conversely, how far from the reported value [Figure 1f].

Box plot

Box plots characterize a sample using the minimum, 25th, 50th, and 75th percentiles, maximum values. The interquartile range (IQR = Q3 – Q1, where Q1 is first quartile or 25th percentile while Q3 is third quartile or 75th percentile) which covers the central 50% of the data. Quartiles are insensitive to outliers and preserve information about the center and spread (variation). If a data point is below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$, it is viewed as being too far from the central values (median), which are called outliers [Figure 1g].

Scatter plot

A scatter plot (also called scatter diagram) is a graph in which the values of two quantitative variables are plotted along two axes, the pattern of the resulting points revealing any correlation present between variables for a set of data. Scatter plots show how much one variable is affected by another. The relationship between two variables is called

their correlation. If the data points make a straight line going from the origin out to high x- and y-values, then the variables are said to have a positive correlation. If the line goes from a high value on the y-axis down to a high value on the x-axis, the variables have a negative correlation. In case no trend was shown, it is called no correlation [Figure 1h].

Bland–Altman plot

A Bland–Altman plot (difference plot) is a method of data plotting used in analyzing the agreement between two different assays. In the Bland–Altman plot, the differences (between the two methods) are plotted against the averages of the two methods. Alternatively, we can choose to plot the differences (between the two methods) against one of the two methods, if this is a reference method of both methods [Figure 1i].

Forest plot

A forest plot, also known as a blobbogram, is a graphical display of estimated results from a number of scientific studies addressing the same question, along with the overall results. It is a graphical representation of a meta-analysis. It is usually accompanied by a table listing references (author and date) of the studies with their estimated result included in the meta-analysis^[12] [Figure 1j].

Other graphical methods

Besides above, there are some other graphical methods, used in the research studies, although they are less popular including stem and leaf plot, area chart, polar plot, youden plot, and high-low graph.

Relationship between Scales of Measurement, Statistical Methods, and Graphical Presentation of Statistical Data

Statistical methods are varying according to the scales of measurements. For example, when the data are a continuous variable, then we can use the parametric methods (including *t*-test, ANOVA test, linear regression, and Pearson correlation). When the data are a discrete variable/qualitative variable, we cannot use parametric testing and only nonparametric methods (including Mann–Whitney U-test, Kruskal–Wallis H-test, Wilcoxon test, Friedman test, Chi-square test, logistic regression, and Spearman correlation) are used. Similarly, graphical methods are varying according to the scales of measurements. For example, histogram, error bar graph, scatter plot, boxplot, and Bland–Altman graph can be drawn for continuous variables, but not for qualitative variables. In contrast, the pie chart is a graph that is only for qualitative data. There are many diagrams those are used for either categorical variable(s) or mix of the categorical and quantitative variables including bar graph and line graph. In brief, it is not possible to use appropriate statistical method and graphical presentation without proper knowledge of the concepts and properties of data types.

Conclusions

Data type is an important concept of statistics, which should be understood to implement statistical tools correctly. Proper knowledge of data types is necessary to analyze data sets with appropriate statistical methods. This not only enhances our ability to decide its summary measures but helps us to analyze data sets with proper statistical methods. There are several statistical diagrams available to display summaries and finding of data sets. There are several statistical diagrams available to display summaries and finding of data sets, although their use depends on our objectives and data types. We should use appropriate diagrams for our data sets, which is very much useful to communicate the summary and findings to the viewers with easily and quickly.

Acknowledgment

We would like to express their deep and sincere gratitude to Dr. Prabhat Tiwari, Professor, Department of Anaesthesiology, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, for his encouragement to write this article. His critical reviews and suggestions were very useful for improvement in the article

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

References

1. Sundaram KR, Dwivedi SN, Sreenivas V. Medical Statistics: Principles and Methods. 2nd ed. India: Wolters Kluwer; 2014.
2. Praharaj SK. Illustrate well to get noticed: Graphs and figures in research papers. Kerala J Psychiatry 2017;30:115-9.
3. Armitage P, Berry G. Statistical Methods in Medical Research. 2nd ed. London: Blackwell Scientific Publications; 1987.
4. Machin D, Campbell MJ, Walters SJ. Medical Statistics: A Textbook for the Health Sciences. 4th ed. Chichester: John Wiley & Sons Ltd, 2007.
5. Indrayan A, Sarmukaddam SB. Medical Bio-Statistics. New York: Marcel Dekker, Inc.; 2000.
6. Altman DG. Practical Statistics for Medical Research Chapman & Hall/CRC Texts in Statistical Science. London: CRC Press; 1999.
7. Indrayan A, Satyanarayana L. Essentials of biostatistics. Indian Pediatr 1999;36:1127-34.
8. Slutsky DJ. The effective use of graphs. J Wrist Surg 2014;3:67-8.
9. Franzblau LE, Chung KC. Graphs, tables, and figures in scientific publications: The good, the bad, and how not to be the latter. J Hand Surg Am 2012;37:591-6.
10. Durbin CG Jr. Effective use of tables and figures in abstracts, presentations, and papers. Respir Care 2004;49:1233-7.
11. McDonald JC. Charts, graphs and tables – Reporting the data. Radiat Prot Dosimetry 2001;95:291-3.
12. Wikipedia Contributors. Forest Plot. Wikipedia, The Free Encyclopedia, 16, 46UTC; 07 June, 2018. Available from: https://www.en.wikipedia.org/w/index.php?title=Forest_plot&oldid=844855630. [Last accessed on 2018 Jun 16].