

Statistics

Definition/Introduction

Statistics, as a science, is the scientific process of acquisition and management of a given set of data. In the medical field and other life sciences, the term “biostatistics” is often used instead to emphasize its application to medicine and health. Statistics is used to provide information on the given health situation and guide healthcare professionals in the decision-making process, whether as part of the research study or as part of clinical work.

The application of statistics undergoes a series of steps creating a cycle of scientific activities. Usually, it begins with the acquisition of health data. This collection of data involves gathering health-related information through the use of data collection tools (e.g., survey questionnaires) to accurately acquire details pertinent to a given study. Collecting data directly from the respondents is termed as primary sources of data. If the researcher wants to use a given set of data that were collected beyond the scope of the study (e.g., vital statistics and health statistics), then these are termed as secondary sources of data. Before moving on to the next step of the process, the accuracy and reliability of the data collection must be confirmed since any alterations or misinformation during this process would inevitably affect the analysis and interpretation of data on hand.

Data management, on the other hand, employs the organization and analysis of health data. Data can be organized in numerous ways, so every researcher should only use methods depending on the specific goal of the study. For example, if the statistical data must be interpreted as individual units, it can be organized in the form of a raw data or data series (e.g., arranged in arrays or alphabetical order). This is usually done in studies having a small population (e.g., case studies, case series). Otherwise, if the data needs to be described using frequency distribution, it can be organized either as discrete or continuous data series using frequency tables. This collection process is frequently used in studies with a larger study population. It is important to note that the best method of organizing statistical data primarily depends on the type of variable (e.g., qualitative or quantitative) and its level of measurement (e.g., nominal, ordinal, interval, ratio). Arriving at all possible data organizations may not be necessary if, and only if, this will give the best information to the researchers about the objectives of the study.

The use of appropriate methods to organize data will lead to its accurate analysis. In descriptive data analysis, the use of narratives, tables, graphs, and charts can be sufficient to describe the study variables. In the inferential analysis, the researcher needs to either estimate specific clinical or health parameters or perform hypothesis testing. Several versions of data analysis software are available for use according to the type of research work.^[1]

Eventually, accurate and reliable interpretation follows from properly conducted data analysis. This step focuses on generating correct information based on the findings while relating it to the context of the topic under study. The current generation of discoveries, conclusions, and hypotheses will make future researchers capable of studying its

underlying issues and restarting the statistical process, creating a continuous cycle of collecting, organizing, analyzing, and interpreting data.