ABC of Statistics

Statistics in Medical Research

Dinesh Kumar

Medicine is a science with chance playing a very significant role. Statistics as science help to quantify the contribution of chance and as an art helps individual clinician make valid diagnostic, prognostic or therapeutic decisions. It also helps programme managers and policy planners to plan, monitor and evaluate public health initiatives.

As would be evident in the present series, very few mathematical skills are needed to understand and handle elementary statistics. Author intends to initiate the readers to the subject in a simplistic manner and without the use of extensive formulae, whose need has been cut drastically by the availability of user-friendly statistical software. This doesn't mean that the fundamentals and conceptual framework should be ignored. However, this series would be of limited use to those researchers who wish to pursue advanced level statistics. The overall aim is to make the readers realize its need in clinical/ population medicine based on hard evidence with the clever use of the knowledge of epidemiological principles.

Learning statistical principles is not seen as the most glamorous exercise as far as medical fraternity is concerned. There seems no valid reasons why we all are afraid of numbers even though all of us keep using them in various forums to substantiate our viewpoint. All of us at some point in our career are exposed to core statistical principles but we keep relying on statisticians to decipher the data that we have painstakingly collected. There is nothing wrong seeking their help but our over dependence on them don't allow us to make any effort to reduce data to some meaningful form.

To put the learning into proper perspective, I am reminded of an old adage. "Give a man a fish and you feed him for a day, Teach a man to fish and you feed him for life".

Let's begin by defining Statistics. Definitions of Statistics abound but the components that capture the essence of statistics are listed below :

1. Collection of data. 2. Organization of data. 3. Presentation of data. 4. Summarization of data. 5. Significance testing. 6. Drawing inferences and Interpreting data. 7. Finally, Communicating data.

Statistics mean aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to reasonably standards of accuracy, collected in a systematic manner, for a predetermined purpose and placed in relation to each other.

Based on the components outlined above, statistics is broadly classified into :

Descriptive statistics include collection, organiza-tion, presentation and summarization of data. Whereas, *inferential statistics* is concerned with procedures used to draw inferences and conclusions about a large body of data (called population) from a smaller set of data (called sample).

Remember descriptive statistics generally preceeds deferential statistics. The techniques used for these purposes do not necessarily require an extensive knowledge of mathematics.

Data : Refers to observations made on individuals. In many instances, the individuals are people but they need not be e.g. clinical specimens, hospital records, discharge summaries etc. Data can be Primary or Secondary.

Primary Data : Collected and recorded by the investigator/s themselves by observation, interviews or measuring instruments usually systematically and for defined purposes.

Secondary Data : Collected by somebody else or for other purposes e.g. information derived from hospital records.

From the Post-Graduate Department of Preventive and Social Medicine, Govt. Medical College, Jammu (J&K) - 180001
Correspondence to: Dr. Dinesh Kumar, Statistical Editor, JK Science, Journal of Medical Education & Research.

Variable : One of the most commonly encountered words in statistical parlance. Requires thorough understanding since the choice of Statistical tests to be used depend on the kind of variable studied. It is defined as an attribute, quality, characteristic or property of a persons or things being studied that can be quantitatively measured or enumerated. It varies from person to person or from time to time in the same person. Height, weight, age , gender, blood pressure, pulse rate, smoking status are commonly studied variables.

Variable can be Qualitative or Quantitative

Qualitative or Attributable or Categorical : Classify subjects according to whether he/she belongs to or doesn't belong to or he/she possesses or doesn't possess a quality or attribute e.g., male/female, died/recovered, blood group A/AB/O, white/black, rural/urban etc. The qualitative variable can be dichotomous or poly-chotomous i.e. may be classified in two or more than two categories. The categories may be ordered e.g. severity of disease specified as mild, moderate or severe.

Quantitative variable is further classified into discrete and continuous

Discrete : A variable that can have only finite number of values in any given interval. These values are invariably whole numbers e.g., no. of pregnant women, no. of households in a community, no. of episodes of an illness, WBC/ RBC counts, no. of beds in a hospital etc.

Continuous : A continuous variable can assume infinite number of values in any given interval i.e., the data is generated in whole numbers or fractional numbers. A continuous variable can be divided and subdivided into smaller and smaller values along a continuum, the limiting factor being the accuracy of the measuring instrument e.g., height, weight, blood pressure, plasma level of drugs, hormones etc.

Simply put Qualitative vs. Quantitative Distinction is between what type vs. how much

Interestingly while some variables/characteristics can be dealt only in one way, others are amenable to transformation.

Variable	Qualitatively described as	Quantitative measured
Weight (kgs)	Overweight/ Underweight	70*R 70.1**, 70.12**, 70.123**
Height (meters)	Tall/ Short	1.6* or 1.61**, 1.62**, 1.623**

*discrete **continuous

Transformation from Quantitative to Qualitative entails a loss of information and can seriously impact precision of the estimate being made. When an association or relationship between variables is being studied, they are referred to as Independent variables and Dependent variables.

Independent (stimulus) variable : A variable that is manipulated or applied by the investigator *e.g.* maternal age, age at marriage, spacing between successive outcome pregnancies, pre-pregnancy, weight, weight gain during pregnancy, low birth weight, age and hypertension.

Dependent (response) variable : Resulting response or behavior that is observed when exposed to independent variable *e. g.* dependent variable/s low birth weight, perinatal outcome, anthropometric failure, predisposition to chronic diseases in later life, hypertension, and coronary artery disease.

Remember an independent variable in one situatin may become dependent in other situation and vice versa

Composite variable : A variable based on two or more variables is known as Composite variable e.g. Ponderal and Quetlet's index for measuring adiposity are calculated by taking into account height and weight, Apgar score for assessing well-being of newborn etc. *Contd.....*

Suggested reading

1. Editorial. Health statistics: time to get serious. Bulletin of World Health Organization Oct 2005; 83(10) 722.
2. Editorial. Stumbling around in the dark. *Lancet* 2005; 365: 1983.
3. Last JM, ed. A dictionary of epidemiology, 3rd ed. New York, Oxford University Press, 1995.
4. The role of research and information systems in decision-making for the development of human resources for health. Report of a WHO Study Group. Geneva, World Health Organization, 1990 (WHO Technical Report Series, No. 802).
5. Bland M. An introduction to medical statistics. Oxford, Oxford University Press, 1987.
6. Bradford Hill A. A short Textbook of medical statistics, 11th ed. Hodder and Stoughton, London, 1984.
7. Knapp RG. Basic Statistics for Nurses, 2nd ed. New York, John Wiley and Sons, 1985.
8. Lwanga SK, Tye Cho-Yook. Ayeni O. Teaching health statistics, Lesson and seminar outlines, 2nd ed. Geneva, World Health Organization, 1999.