Resident Corner www.pjkd.com.pk

Statistics and Biostatistics-2 (Basic Concepts)

Afshan Khanum, M Phil (Statistics)
Fatima Memorial College of Medicine & Dentistry, Lahore, Pakistan

Email: afshankhanam16@hotmail.com

PJKD 2021;5(3):RC1-3

Some basic concepts:

Data: Collection of raw facts and figures is called data. For example: weight of a patient, temperature of a patient or number of patients discharged from hospital is data. Each of these three numbers is datum and three taken together is data.

Types of Data:

Primary data: A real time data originated by the researcher for the first time for addressing the problem at hand is known as primary data.

Secondary data: A second hand data that has already been collected by someone and has gone under gone statistical operation. Example of secondary data can be censuses, data collected by government departments or by other organizations.

Variable: any characteristic which can be measured and counted is called variable like age gender and weight.

Quantitative variable: a variable which takes numbers is called quantitative variable, like age, weight and no. of children in a family. There are two types of quantitative variable: continuous and discrete variable.

Qualitative variable a variable which does not take numbers but takes qualities and fits into categories is qualitative variable. Qualitative variables are divided into two type nominal and ordinal variable.

Population: A group under study is called population

Parameter: A numerical quantity calculated from population data is called parameter like population mean

Sample: A representative part of population is called sample.

Statistic: A numerical measure calculated from the subset of the population is statistic like sample mean.

Measurement and measurement scales:

The numbers assigned to the objects or events according to the set of rules is called measurement

Nominal scale:

A scale which refers to the categorical variables without any particular order or ranking. For example, gender has two categories male and female and both of them have no particular order or ranking.

In nominal scale, categories of variables are assigned numerical numbers which do not hold true significance. In this case one cannot calculate descriptive statistics like mean median or mode of those numbers as this would be of no meaning.

Suppose we have a nominal variable named gender we can give male a value of 1, female a value of 2, and transgender a value of 3. Use of these values to calculate descriptive statistics like mean, median, mode, etc. would be of no real meaning. Hence consider each number a label to that category and each label is mutually exclusive which have no true significance.

Examples of nominal scale:

What is your gender?

- Male
- Female

What is your marital status?

- Single
- Married
- Widow
- Divorced

Which is your residential area?

- Urban
- Rural

Nominal data can be presented in form of frequency percentages and graphs. Mode is appropriate descriptive statistics for nominal data. But calculating mean, median and mode for these kind of variable is meaningless.

Ordinal Scale:

A scale which refers to the variables which have categories or options with any particular order or ranking. In this scale order of the values is important but the difference between them is not measureable.

In this scale we cannot quantify the difference between the categories of the variable but we know that the one category is higher than the other. This is the main difference between nominal and ordinal scale. In nominal scale we cannot quantify the difference between categories of the variable as well as the order of the categories can not be decided. Both scales ignore the value of the variable.

Nominal scale takes label of the variables into consideration but ignores the order or ranking of the variables. Ordinal scale takes into consideration the label of the different categories of the variables as well as the order or ranking of those categorical variables.

Example of ordinal scale:

Level of education

- illiterate
- Primary
- Secondary
- Higher secondary
- graduate

Ordinal data can be presented in form of frequency percentages and graphs. Median is appropriate descriptive statistics to calculate in ordinal data.

Interval scale:

This is a scale which incorporates both the nominal and ordinal scale. In this scale order of the value is known and the difference between them is meaningful. It is a scale in which the order of the variable is known and the difference between the values of the variable has some meaning. Time, temperature and credit score are the examples of interval scale. In these example difference between the value is meaningful and easily calculated. Or simply a variable which has no meaningful zero can be measured in interval scale measurement level. Time is a perfect example for interval scale since you can't say when the time started. One important thing to remember is that time and duration are two different variables.

Example of an interval scale:

Temperature, credit score

It is important to note that in interval scale there is no true zero. True zero means something which does not exist. Like if we say a paper weighs zero means paper's weigh does not exist. but if we say temperature is zero then it does not mean that temperature does not exist but it can go to negative values too.

Ratio Scale:

Ratio scales are the most superior among other scales in terms of statistical analysis. It has everything an order, absolute zero and a set of values between the units. Weight, age, height and duration are the examples of ratio scale. All the descriptive and inferential statistics can easily be applied on this scale data as it has true zero in it.

Appropriate descriptive statistics for nominal scale is mode, for ordinal scale is median and for interval scale is mean median and mode whereas ratio scales can use all of the above.

Interval and ratio scales are closely related and can cause confusion sometimes. There are two points which can differentiate between both the scales.

1. In interval scale categories of the variable have order with meaningful difference between them but the ratio between them is not meaningful. For example, if temperature increases 5 to 10 degrees, it's not twice as hot. Whereas in ratio scale categories of the variable have order with meaningful difference between them but the ratio between them is also meaningful. For example, if weight increases 20kg to 40 kg, its twice as double.

2. The other main difference between the two scales is true zero. True zero means if something is zero, it doesn't exist. If something weighs zero, then its weight doesn't exist and ratio scale has true zero in it. In interval scale zero may exist but this is not true zero. As temperature is an example of interval scale and temperature zero does not mean that temperature does not exist because it can even go into negative values.

In conclusion all scales: nominal and ordinal, interval and ratio have labels. Order of these label is important in ordinal, interval and ratio scale but not in nominal scale. In Interval there's no true zero but in ratio scale true zero exists.