The basic concept in statistics

1-Types of data

Data will be collected on individuals from the population and termed as variables, since it changes. Thus, variables **are** some characteristics of the individuals within the population. Variables can take on various types of values, some of them are numbers and some are categories.
Data is a set of measurements of one or more characteristics or variables of some elements of a population, or of several objects generated by a process.

Data Types are an important concept of statistics, which needs to be understood, to correctly apply statistical measurements to your data and therefore to correctly conclude certain assumptions about it.

Variables can be classified into two groups: Categorical (or qualitative) and Quantitative.

1-1 Categorical or qualitative variables

Categorical or qualitative allow listing individual's characteristics into categories. Categorical data represent characteristics such as a person's gender, marital status, and hometown. Categorical data can take on numerical values (such as "1" indicating male and "2" indicating female), but those numbers do not have mathematical meaning. You could not add them together. The data collected for a categorical variable are qualitative data. They may be further described as either ordinal or nominal:

1-1-1 Nominal variables

Elements of a sample or a population can be classified using a nominal variable: the value of the variable places an element in a certain class or category. Examples of such variables are:

- Gender (male/female).
- Nationality.
- Religion.
- Whether or not one owns a car (yes/no).

Sometimes it can be useful to assign labels, code numbers, or code letters, to the different classes or categories. For example, an Iraqi person may be assigned the code "1", a French person the code "2", and a German person the code "3". It is important to note that these figures do not imply any order and/or quantity.

1-1-2 Ordinal variables

Ordinal variable implies a logical order between the elements of a sample. Typical examples of ordinal variables can be found in all kinds of surveys. There, respondents are typically asked whether they consider the quality of a product or service as “1: very good”, “2: good”, “3: moderate”, “4: bad”, or “5: very bad”. In other surveys, the respondents are asked if they “1: strongly disagree”, “2: rather disagree”, “3: neither agree nor disagree”, “4: rather agree”, or “5: strongly agree” with a particular statement.

1-2 Quantitative variables

A variable that is measured on a quantitative scale can be expressed as a fixed number of measurement units. Examples are length, area, volume, weight, duration, price, income, waiting time, and so on.

1-2-1 Discrete variables

A Discrete variable is a quantitative variable that will assume a finite, or a countable, set of values. A discrete variable cannot take on every possible value in an interval on the real line. Each value can be plotted as a separate point on the real line, with space between any two consecutive points. Examples of discrete variables are the number of passengers on a flight, the number of children in a family, the number of students in a college.

1-2-2 Continuous variables

A Continuous variable is a quantitative variable that has an uncountable number of values. In other words, a continuous variable can assume any value between any two points on the real line, and thus the possible values of a continuous variable can form an interval on the number line, with no spaces between the points. Examples of continuous variables are length, weight. The classification chart of variables is shown in Figure 1.

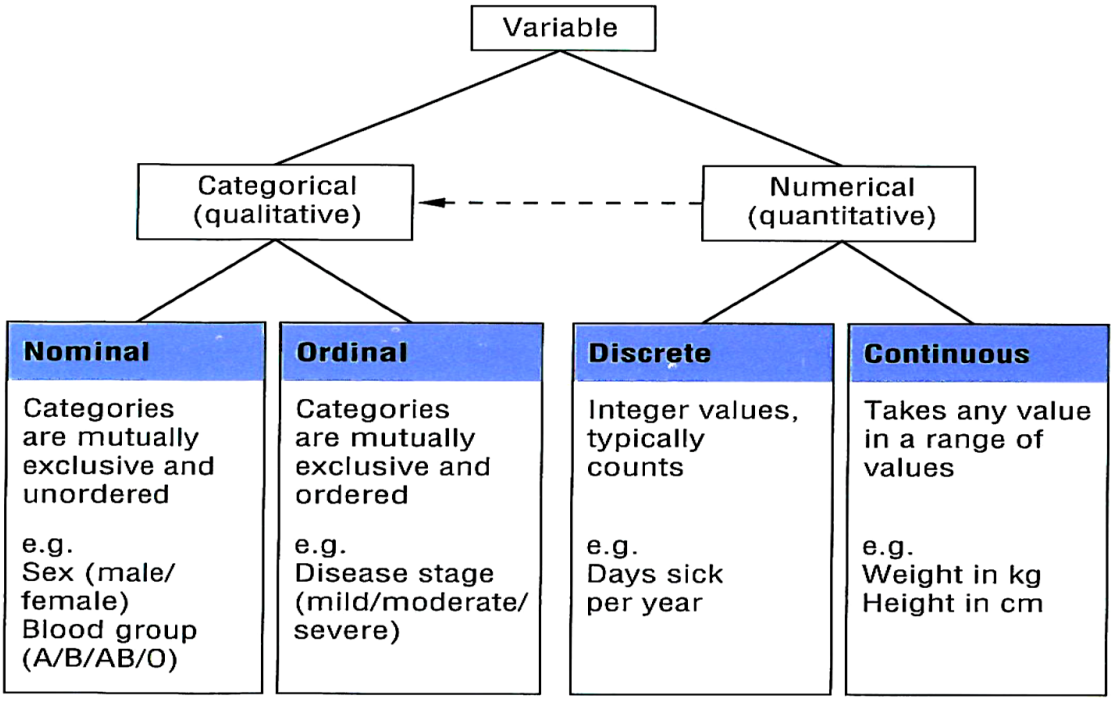

Figure 1: Types of data (variables)

2- Descriptive statistics

Once we have identified our population, and collected the sample data, our goal is to describe the characteristics of the sample in an accurate fashion in such a way that the information will be easily communicated to others. Describing, or just summarizing, the data can be done in two ways: Graphically or Numerically. Graphical description of the data depends on the data type. As we know, there are two types of data: Qualitative and Quantitative data.
Data, in some cases, will be given or presented to the researcher in a table.

3- Frequency distributions

When dealing with large sets of data, a good overall picture and sufficient information can be often conveyed by grouping the data into a number of classes. Tables, like the Table 1, are called frequency distribution. The construction of frequency distributions consists of three steps, particularly for quantitative data:
1. Choosing the classes (intervals, or categories for qualitative data)
2. Tally the data into these classes
3. Count the number of items in each class.

The first step is the most important step, while the others are purely mechanical and depend on step 1. Designing too few classes would obscure the information

in the distribution while, on the other hand, designating too many classes would confuse the reader.

Table 1: a frequency distribution table that lists a set of scores and their distribution

score	tally	frequency (f)
1	𝍸	4
2	𝍸 | 𝍸	9
3	𝍸 |	6
4	𝍸 ||	7
5	|||	3
6	||	2

There are two ways in which frequency distributions can be modified to suit needs. One way is to convert a distribution into a percentage distribution by dividing the frequency in each class by the total number of observations, and express it as a percentage, see column 4 in Table 2. This column has what we call the relative frequency. The other way of modifying a frequency distribution is by presenting it as a cumulative relative frequency distribution by adding the relative frequencies as we go down the classes, and this will generate column 5 in Table 2.

Example 1: The following are the grades of 50 students in a math class. Construct a frequency distribution for the following data.

75	89	66	52	90	68	83	94	77	60
38	47	87	65	97	49	65	70	73	81
85	77	83	56	63	79	69	82	84	70
62	75	29	88	74	37	81	76	74	63
69	73	91	87	76	58	63	60	71	82

Since no grade is less than 20, and no one greater than 100, the following classes are considered, and the summary is given in Table 2:

Table 2: Frequency distribution table for the grades of 50 students in a math class.

Classes	Tally	Frequency	Relative Frequency (%)	Cumulative Rel. Freq (%)
20–29	/	1	2	2
30–39	//	2	4	6
40–49	//	2	4	10
50–59	///	3	6	16
60–69	///// ///// //	12	24	40
70–79	///// ///// ////	14	28	68
80–89	///// ///// //	12	24	92
90–99	////	4	8	100
Total		50	100%	