Statistics الاحصاء

Statistics is the science that is concerned with collecting, displaying, representing and analyzing data. It has its own laws and theories. Statistics is divided into two sections: Descriptive statistics and Inferential statistics.

Biostatistics: (الاحصاء الحياتي) Biostatistics is the application of statistical techniques to scientific research in health-related fields, including medicine, biology, and public health, and the development of new tools to study these areas. In other words, when the data being analyzed are drawn from the biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts.

Sources of Data (مصادر المعلومات)

The performance of statistical activities arises from the need to answer a question. For example, physicians may want answers to questions about the relative merits of competing treatment procedures. When we determine that the appropriate approach to finding an answer to a question requires the use of statistics, we begin to search for appropriate data to serve as the raw material for our investigations. These data are usually available from one or more of the following sources:

1. **Archived Records (السجلات المؤرشفة)**. Hospital medical records contain vast amounts of information about patients.
2. **Surveys (استطلاعات الرأي).** If the data needed to answer a question are not available from routinely kept records, the logical source may be a survey. Suppose, for example, that the administrator of a clinic wishes to obtain information regarding the mode of transportation used by patients to visit the clinic. If admission forms do not contain a question on mode of transportation, we may conduct a survey among patients to obtain this information.

3. **Experiments (التجارب).** Frequently the data needed to answer a question are available only as the result of an experiment. A nurse may wish to know which of several strategies is best for maximizing patient compliance. The nurse might conduct an experiment in which the different strategies of motivating compliance are tried with different patients. Subsequent evaluation of the responses to the different strategies might enable the nurse to decide which is most effective.
4. **External Sources (المصادر الخارجية).** The data needed to answer a question may already exist in the form of published reports, commercially available data banks, or the research literature. In other words, we may find that someone else has already asked the same question, and the answer obtained may be applicable to our present situation

Variable (المتغير). If we find that a property takes on different values for different people, places, or things, we call that property a variable. Examples of variables are diastolic blood pressure, heart rate, heights of adult males, weights of preschool children, and ages of patients seen in a dental office.

Quantitative Variables (المتغيرات الكمية). A quantitative variable is one that can be measured in the usual sense. We can, for example, obtain measurements of the heights of adult males, the weights of preschool children, and the ages of patients.

Qualitative Variables (المتغيرات النوعية). Some characteristics are not capable of being measured.

Random Variable (المتغير العشوائي). It is a variable whose values were obtained as a result of chance factors, so that they cannot be accurately predicted in advance. An example of a random variable is the height of an adult. When a child is born, we cannot accurately predict his height when he matures.

Discrete Variable (المتغير المنفصل). A discrete variable is characterized by gaps or interruptions in the values that it can assume. These gaps or interruptions indicate the absence of values between particular values that the variable can assume. Some examples illustrate the point. The number of daily admissions to a general hospital is a discrete random variable since the number of admissions each day must be represented by a whole number, such as 0, 1, 2, or 3. The number of admissions on a given day cannot be a number such as 1.5, 2.997, or 3.333. The number of decayed, missing, or filled teeth per child in an elementary school is another example of a discrete variable.

Continuous Variable (المتغير المستمر). A continuous random variable does not possess the gaps or interruptions characteristic of a discrete random variable. A continuous random variable can assume any value within a specified relevant interval of values assumed by the variable. Examples of continuous variables include the various measurements that can be made on individuals such as height, weight, and skull circumference. No matter how close together the observed heights of two people, for example, we can, theoretically, find another person whose height falls somewhere in between.

Population(المجتمع). We define the population of entities as the largest set of entities we are interested in at a given point in time. If we measure some variable on each entity in the population, we generate a set of values for that variable. Thus, we might define the population of values as the largest set of values for a random variable we are interested in at a given point in time. For example, if we are interested in the weights of all children enrolled in an elementary school system in a given county, our population consists of all of these weights.

Sample (العينة). A sample may be defined simply as a part of a population. Suppose our population consists of the weights of all the elementary school children enrolled

in a certain county school system. If we collect for analysis the weights of only a fraction of these children, we have only a part of our population of weights, that is, we have a sample.

Statistical Inference (الاستدلال الاحصائي). Is the procedure by which we reach a conclusion about a population on the basis of the information contained in a sample that has been drawn from that population.

A Research Study (الدراسة البحثية). Is a scientific study of a phenomenon of interest. Research studies involve designing sampling protocols, collecting and analyzing data, and providing valid conclusions based on the results of the analyses.

Experiments (التجارب). Are a special type of research study in which observations are made after specific manipulations of conditions have been carried out; they provide the foundation for scientific research.

Simple Sample (العينة البسيطة). If a sample of size n is drawn from a population of size N in such a way that every possible sample of size n has the same chance of being selected, the sample is called a simple sample.

Systematic Sampling (العينة المنهجية). A sampling method that is widely used in healthcare research is the systematic sample. Medical records, which contain raw data used in healthcare research, are generally stored in a file system or on a computer and hence are easy to select in a systematic way. Using systematic sampling methodology, a researcher calculates the total number of records needed for the study or experiment at hand. A random numbers table is then employed to select a starting point in the file system. The record located at this starting point is called record x. A second number, determined by the number of records desired, is selected to define the sampling interval (call this interval k). Consequently, the data set would consist of records $x, x+k, x+2k, x+3k, \cdots$ and so on, until the necessary number of records are obtained.

Stratified Sampling (العينة الطبقية) . It may be desirable to divide a population of interest into groups or strata, where the sample units within a given stratum are more similar to each other than to the sample units that make up the other strata. After dividing the population into strata, it is usual to take an independent random sample from each stratum. Although the benefits of stratified sampling may not be readily apparent, samples taken within a stratum are often much less variable than samples taken across all strata. This is true because the sample units within each stratum tend to have similar characteristics.

Classified Data (البيانات المبوبة)

Organizing and Presenting Data (تنظيم البيانات وعرضها)

Let m be the number of classes. Then $5 \leq m \leq 15$. If there are fewer than five time periods, the data has been summarized too much and the information it contains has been lost. If there are more than 15 time periods, the data has not been summarized adequately.

We will learn how to create a frequency distribution table from the following example:

Example (1). A researcher studied the effectiveness of a certain drug, and 50 patients of different ages agreed to receive the treatment. Their ages were fixed in the following:

35	52	67	25	78	21	44	61	33	57	75	29	48	47	72	36	27
43	36	19	68	46	50	69	52	55	15	64	21	72	56	73	71	63
63	65	27	38	51	42	31	36	12	61	64	45	44	28	52	54	

A frequency distribution table is required.

Solution:

1. Find the range R, which is the difference between the largest value and the smallest value: $R = 78 - 12 = 66$.

2. Determine the number of classes m based on the number of values n=50 from the following formula:

$$m = 1 + 3.322 \log(n)$$

$$m = 1 + 3.322 \log(50) = 6.6 \cong 7$$

3. Determine the length of the class $L = R \div m = 66 \div 7 \cong 10$.

Since the smallest value is 12 and the largest value is 78, we may begin our classes with 10 and end with 79.

4. Determine the Midpoint of a class

$$\text{Midpoint of a class} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

This gives the following table:

Class	10-19	20-29	30-39	40-49	50-59	60-69	70-79	Sum
Midpoint	14.5	24.5	34.5	44.5	54.5	64.5	74.5	
Frequency	3	7	7	8	9	10	6	50

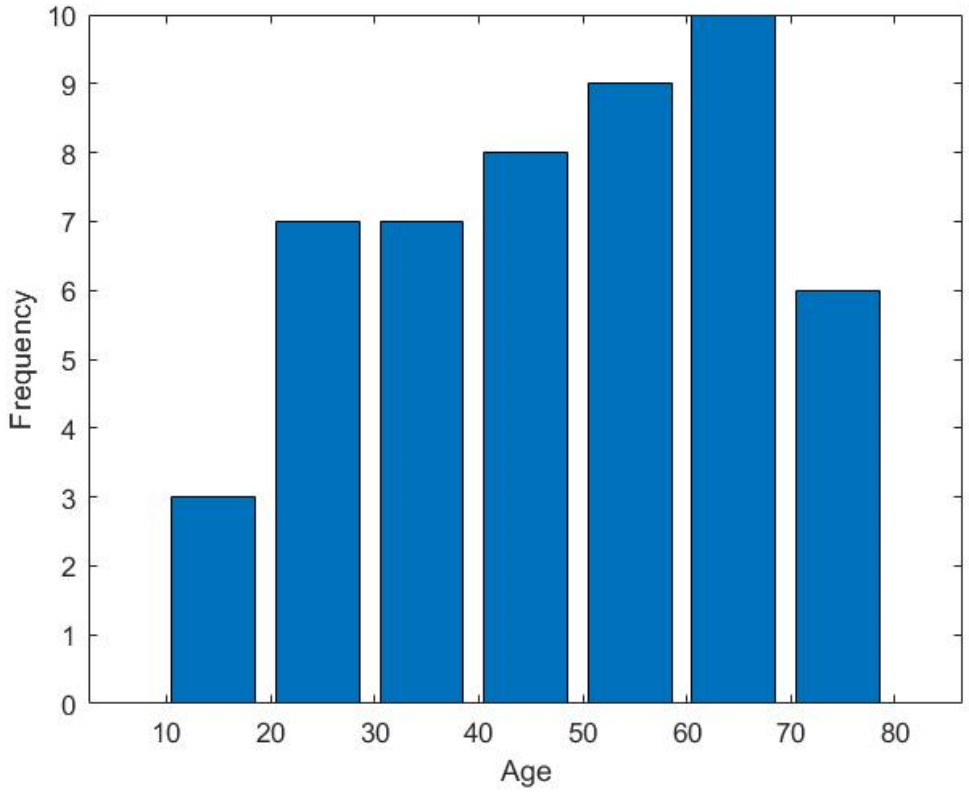
The Histogram

We can display the frequency distribution graphically in the form of a bar graph where the values of the variable under consideration are represented by the horizontal axis, while the vertical axis has a scale of frequency.

These graphs can be created in Excel or Matlab.

To display the values in example 1, we can use the code in Matlab.

```
X=[14.5:10:74.5];
F=[3,7,7,8,9,10,6];
bar(X,F);
xlabel('Age)');
ylabel('Frequency');
```



Data can also be represented using Excel.