US06DZOO26

Introduction about biostastic

Biostatistics is the application of statistical principles to questions and problems in medicine, public health or biology. One can imagine that it might be of interest to characterize a given population (e.g., adults in Boston or all children in the United States) with respect to the proportion of subjects who are overweight or the proportion who have asthma, and it would also be important to estimate the magnitude of these problems over time or perhaps in different locations. In other circumstances in would be important to make comparisons among groups of subjects in order to determine whether certain behaviors (e.g., smoking, exercise, etc.) are associated with a greater risk of certain health outcomes. It would, of course, be impossible to answer all such questions by collecting information (data) from all subjects in the populations of interest. A more realistic approach is to study samples or subsets of a population. The discipline of biostatistics provides tools and techniques for collecting data and then summarizing, analyzing, and interpreting it. If the samples one takes are representative of the population of interest, they will provide good estimates regarding the population overall. Consequently, in biostatistics one analyzes samples in order to make inferences about the population. This module introduces fundamental concepts and definitions for biostatistics.

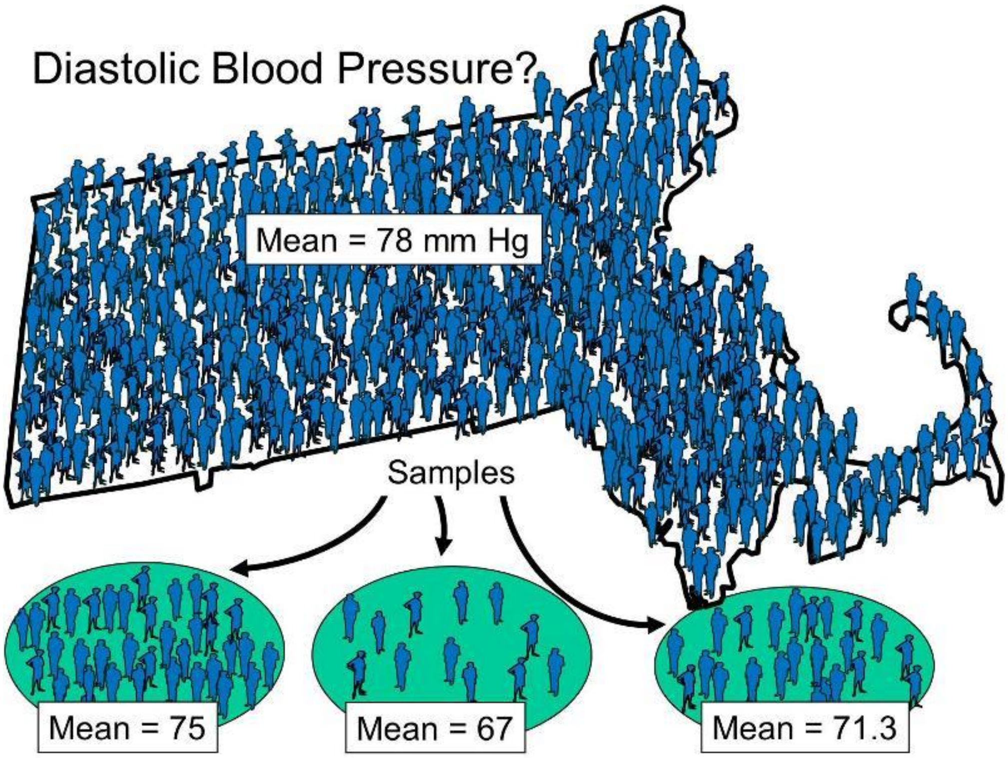
Learning Objectives

After completing this module, the student will be able to:

1. Define and distinguish between populations and samples.
2. Define and distinguish between population parameters and sample statistics.
3. Compute a sample mean, sample variance, and sample standard deviation.
4. Compute a population mean, population variance, and population standard deviation.
5. Explain what is meant by statistical inference.

Population Parameters versus Sample Statistics

As noted in the Introduction, a fundamental task of biostatistics is to **analyze samples in order to make inferences about the population from which the samples were drawn**. To illustrate this, consider the population of Massachusetts in 2010, which consisted of 6,547,629 persons. One characteristic (or variable) of potential interest might be the diastolic blood pressure of the population. There are a number of ways of reporting and analyzing this, which will be considered in the module on Summarizing Data. However, for the time being, we will focus on the mean diastolic blood pressure of all people living in Massachusetts. It is obviously not feasible to measure and record blood pressures for of all the residents, but one could take samples of the population in order estimate the population's mean diastolic blood pressure.



Despite the simplicity of this example, it raises a series of concepts and terms that need to be defined. The terms *population*, *subjects*, *sample*, *variable*, and *data elements* are defined in the tabbed activity below.

It is possible to select many samples from a given population, and we will see in other learning modules that there are several methods that can be used for selecting subjects from a population into a sample. The simple example above shows three small samples that were drawn to estimate the mean diastolic blood pressure of Massachusetts residents, although it doesn't specify how the samples were drawn. Note also that each of the samples provided a different *estimate* of the mean value for the population, and none of the estimates was the same as the *actual* mean for the overall population (78 mm Hg in this hypothetical example). In reality, one generally doesn't know the true mean values of the characteristics of the population, which is of course why we are trying to estimate them from samples. Consequently, it is important to define and distinguish between:

- population size versus sample size
- parameter versus sample statistic.

Sample Statistics

In order to illustrate the computation of sample statistics, we selected a small subset (n=10) of participants in the Framingham Heart Study. The data values for these ten individuals are shown in the table below. The rightmost column contains the body mass index (BMI) computed using the height and weight measurements. We will come back to this example in the module on Summarizing Data, but it provides a useful illustration of some of the terms that have been introduced and will also serve to illustrate the computation of some sample statistics.

Data Values for a Small Sample

Participant ID	Systolic Blood Pressure	Diastolic Blood Pressure	Total Serum Cholesterol	Weight	Height	Body Mass Index
1	141	76	199	138	63.00	24.4
2	119	64	150	183	69.75	26.4
3	122	62	227	153	65.75	24.9
4	127	81	227	178	70.00	25.5
5	125	70	163	161	70.50	22.8
6	123	72	210	206	70.00	29.6
7	105	81	205	235	72.00	31.9
8	113	63	275	151	60.75	28.8
9	106	67	208	213	69.00	31.5
10	131	77	159	142	61.00	26.8

The first summary statistic that is important to report is the sample size. In this example the sample size is n=10. Because this sample is small (n=10), it is easy to summarize the sample by inspecting the observed values, for example, by listing the diastolic blood pressures in ascending order:

62 63 64 67 70 72 76 77 81 81

Simple inspection of this small sample gives us a sense of the center of the observed diastolic pressures and also gives us a sense of how much variability there is. However, for a large sample, inspection of the individual data values does not provide a meaningful summary, and summary statistics are necessary. The two key components of a useful summary for a continuous variable are:

- a description of the center or 'average' of the data (i.e., what is a typical value?) and
- an indication of the variability in the data.

Sample Mean

There are several statistics that describe the center of the data, but for now we will focus on the sample mean, which is computed by summing all of the values for a particular variable in the sample and dividing by the sample size. For the sample of diastolic blood pressures in the table above, the sample mean is computed as follows:

$$\textbf{Sample Mean} = \frac{62+63+64+67+70+72+76+77+81+81}{10} = \mathbf{71.3}$$

To simplify the formulas for sample statistics (and for population parameters), we usually denote the variable of interest as "X". X is simply a placeholder for the variable being analyzed. Here X=diastolic blood pressure.

The general formula for the sample mean is:

$$\bar{X} = \frac{\Sigma X}{n}$$

The X with the bar over it represents the sample mean, and it is read as "X bar". The Σ indicates summation (i.e., sum of the X's or sum of the diastolic blood pressures in this example).

When reporting summary statistics for a continuous variable, the convention is to report one more decimal place than the number of decimal places measured. Systolic and diastolic blood pressures, total serum cholesterol and weight were measured to the nearest integer, therefore the summary statistics are reported to the nearest tenth place. Height was measured to the nearest quarter inch (hundredths place), therefore the summary statistics are reported to the nearest thousandths place. Body mass index was computed to the nearest tenths place, summary statistics are reported to the nearest hundredths place.

Biostatistics

Vital statistics: probably the major source of information about the health of population is its vital statistics. By vital statistics we mean the data collected from ongoing recording, or registration of all "vital events", deaths, marriages, adoptions, and fetal deaths, divorces, legal separations and annulments. We will discuss only those aspects of registration with which physician are much concerned certification of deaths, births and fetal deaths.

Epidemiology: Epidemiology is the study of the distribution and Determinant of diseases and injuries in human population.

Statistics: The term statistics is concerned with scientific methods for collecting, organizing, presenting and analysis data, as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

Bio Statistics: Is the branch of statistics, which deal with collecting, presenting, summarizing, analyzing and interpreting sets of data in health field.

Scale of measurements Nominal scale Ordinal or raking scale
Interval scale Ratio scale

A variable: It is a value of characteristics that changes from one subject to another, from thing to thing

variables can be divided to:
1.Qualitative variable2.Quantitative variable. It is divided to :A. Continuous variable B. Discrete variable

Collection of data: 1.Regular or routine system: 2.AD HOC System:
Consist of registration procedure for collecting data as the become a variable.2.AD HOC System: A survey to collect information that is not available on a regular basis. It is done for local administrative or research purpose.

Data Presentation

(a) Presentation of qualitative data:
1.Tabulation form2.Graphical form

1- Tabulation form: Frequency Sex 30 Male 20 Female 50 Other
The basic rule for displaying qualitative data is to classify them into categories and then count the number of observations in each category of the variable and present the number and percentage in a table1-Tabulation form:FrequencySex30Male20Female50Other

Cross tabulation: Total Disease No Disease Test Result 90 80 10
A Frequency table involving at least two variable that have been cross classified. Example Total Disease No Disease Test Result90 80 10Positive21020190Negative300100200

1.Graphical form: 1.Bar chart:
Is used with categories data, with nominal data

2- Pie chart: Sectors of a circle with areas proportional to class frequencies, used to present data in nominal classes Example :Fig (1) the number of students according to the sex

(B) Presentation of quantitative data:
1.Tabulation form2.Graphic form

Number of observations
(1) Tabulation form: Example (1):Serum albumin value in g/ liter of blood of 50 women seen in survey are as follow :42 , 41 , 42 , 44 , 44 , 36 , 38 , 41 , 42 , 44 , 42 , 39 , 49 , 40 , 45 , 32 , 34 , 43 , 37, 39 , 41 , 39 , 48 , 42 , 43 , 33 , 43 , 35 , 32 , 34 , 39 , 35 , 43 , 44 , 47 , 40 , 39 , 42 , 41 , 46 , 37 , 49 , 41 , 39 , 43 , 42 , 47 , 48 , 51 , 52 .Number of observations Serum Albumin (g/ liter)330-734-1438-1742-46-250-5350Total

2- Graphic form:1- Histogram2- Frequency polygon (line chart)3- Frequency curve4- Cumulative frequency polygon (ogive)

Measure of central tenderly and location
Arithmetic's mean: The sum of all values of a set of observations divided by the number of observations .The median :When a set of observation is the value that divides a distribution into two equal halves. The mode :The most frequently occurring in series of observations

Example :For the following data 10 , 8 , 7 , 2 , 8 , 7 find the mean median and mode .

Solution: The mean: x =∑x n = 10 + 8 + 7 + 2 + 8 + 7 =7 6 The median:
= =76The median :Array all the observations in a sending order2,7,7,8,8,10X = 8+7 / 2 = 7.5

The mode: There are two modes :X=7 & 8

if the mean, the mode and the median are equal the distribution of observation is symmetric

If the mean larger than the median, the distribution is skewed to the right.
modemean

If the mean is smaller than the median , the distribution skewed to the left.
mode mean

Biological control or **biocontrol** is a method of controlling pests such as insects, mites, weeds and plant diseases using other organisms.[1] It relies on predation, parasitism, herbivory, or other natural mechanisms, but typically also involves an active human management role. It can be an important component of integrated pest management (IPM) programs.

There are three basic strategies for biological pest control: classical (importation), where a natural enemy of a pest is introduced in the hope of achieving control; inductive (augmentation), in which a large population of natural enemies are administered for quick pest control; and inoculative (conservation), in which measures are taken to maintain natural enemies through regular reestablishment.[2]

Natural enemies of insect pests, also known as biological control agents, include predators, parasitoids, pathogens, and competitors. Biological control agents of plant diseases are most often referred to as antagonists. Biological control agents of weeds include seed predators, herbivores, and plant pathogens.

Biological control can have side-effects on biodiversity through attacks on non-target species by any of the above mechanisms, especially when a species is introduced without a thorough understanding of the possible consequences.

(Courtesy by google)