SOME BASIC CONCEPTS

Like all fields of learning, statistics has its own vocabulary. Some of the words and phrases encountered in the study of statistics will be new to those not previously exposed to the subject. Other terms, though appearing to be familiar, may have specialized meanings that are different from the meanings that we are accustomed to associating with these terms. The following are some terms that we will use extensively in this book.

Como todos los campos del aprendizaje, la estadística tiene su propio vocabulario. Algunas de las palabras y frases que se encuentran en el estudio de la estadística serán nuevas para quienes no hayan estado previamente expuestos al tema. Otros términos, aunque parezcan familiares, pueden tener significados especializados que son diferentes de los significados que estamos acostumbrados a asociar con estos términos. Los siguientes son algunos términos que usaremos extensamente en este libro.

**Data** The raw material of statistics is data. For our purposes we may define data as numbers. The two kinds of numbers that we use in statistics are numbers that result from the taking—in the usual sense of the term—of a measurement, and those that result from the process of counting. For example, when a nurse weighs a patient or takes a patient’s temperature, a measurement, consisting of a number such as 150 pounds or 100 degrees Fahrenheit, is obtained. Quite a different type of number is obtained when a hospital administrator counts the number of patients—perhaps 20—discharged from the hospital on a given day. Each of the three numbers is a datum, and the three taken together are data.

La materia prima de la estadística son los datos. Para nuestros propósitos podemos definir los datos como números. Los dos tipos de números que utilizamos en estadística son los que resultan de tomar (en el sentido habitual del término) una medida y los que resultan del proceso de contar. Por ejemplo, cuando una enfermera pesa a un paciente o le toma la temperatura, se obtiene una medida que consta de un número como 150 libras o 100 grados Fahrenheit. Un tipo de número bastante diferente se obtiene cuando el administrador de un hospital cuenta el número de pacientes (quizás 20) dados de alta del hospital en un día determinado. Cada uno de los tres números es un dato y los tres juntos son datos.

**Statistics** The meaning of statistics is implicit in the previous section. More concretely, however, we may say that statistics is a field of study concerned with (1) the collection, organization, summarization, and analysis of data; and (2) the drawing of inferences about a body of data when only a part of the data is observed.

El significado de estadística está implícito en la sección anterior. Sin embargo, de manera más concreta, podemos decir que la estadística es un campo de estudio que se ocupa de (1) la recopilación, organización, resumen y análisis de datos; y (2) la extracción de inferencias sobre un conjunto de datos cuando solo se observa una parte de los datos.

The person who performs these statistical activities must be prepared to interpret and to communicate the results to someone else as the situation demands. Simply put, we may say that data are numbers, numbers contain information, and the purpose of statistics is to investigate and evaluate the nature and meaning of this information.

La persona que realiza estas actividades estadísticas debe estar preparada para interpretar y comunicar los resultados a otra persona según lo requiera la situación. En pocas palabras, podemos decir que los datos son números, que los números contienen información y que el propósito de las estadísticas es investigar y evaluar la naturaleza y el significado de esta información.

**Sources of Data** The performance of statistical activities is motivated by the need to answer a question. For example, clinicians may want answers to questions regarding the relative merits of competing treatment procedures. Administrators may want answers to questions regarding such areas of concern as employee morale or facility utilization. When we determine that the appropriate approach to seeking an answer to a question will require the use of statistics, we begin to search for suitable data to serve as the raw material for our investigation. Such data are usually available from one or more of the following sources:

La realización de actividades estadísticas está motivada por la necesidad de responder una pregunta. Por ejemplo, los médicos pueden querer respuestas a preguntas sobre los méritos relativos de procedimientos de tratamiento competitivos. Es posible que los administradores quieran respuestas a preguntas relacionadas con áreas de preocupación como la moral de los empleados o la utilización de las instalaciones. Cuando determinamos que el enfoque apropiado para buscar una respuesta a una pregunta requerirá el uso de estadísticas, comenzamos a buscar datos adecuados que sirvan como materia prima para nuestra investigación. Estos datos suelen estar disponibles en una o más de las siguientes fuentes:

**1. Routinely kept records.** It is difficult to imagine any type of organization that does not keep records of day-to-day transactions of its activities. Hospital medical records, for example, contain immense amounts of information on patients, while hospital accounting records contain a wealth of data on the facility’s business activities. When the need for data arises, we should look for them first among routinely kept records.

Es difícil imaginar cualquier tipo de organización que no lleve registros de las transacciones diarias de sus actividades. Los registros médicos de los hospitales, por ejemplo, contienen inmensas cantidades de información sobre los pacientes, mientras que los registros contables de los hospitales contienen una gran cantidad de datos sobre las actividades comerciales de la instalación. Cuando surge la necesidad de datos, debemos buscarlos primero entre los registros que se llevan de forma rutinaria.

**2. Surveys.** 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.

Si los datos necesarios para responder una pregunta no están disponibles en los registros que se llevan de forma rutinaria, la fuente lógica puede ser una encuesta. Supongamos, por ejemplo, que el administrador de una clínica desea obtener información sobre el medio de transporte utilizado por los pacientes para visitar la clínica. Si los formularios de admisión no contienen una pregunta sobre el modo de transporte, podremos realizar una encuesta entre los pacientes para obtener esta información.

**3. Experiments.** 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.

Con frecuencia, los datos necesarios para responder una pregunta sólo están disponibles como resultado de un experimento. Es posible que una enfermera desee saber cuál de varias estrategias es mejor para maximizar el cumplimiento del paciente. La enfermera podría realizar un experimento en el que se prueben diferentes estrategias para motivar el cumplimiento con diferentes pacientes. La evaluación posterior de las respuestas a las diferentes estrategias podría permitir a la enfermera decidir cuál es la más eficaz.

**4. External sources.** 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.

Los datos necesarios para responder una pregunta pueden existir ya en forma de informes publicados, bancos de datos disponibles comercialmente o literatura de investigación. En otras palabras, podemos encontrarnos con que alguien ya ha hecho la misma pregunta y la respuesta obtenida puede ser aplicable a nuestra situación actual.

**Biostatistics** The tools of statistics are employed in many fields—business, education, psychology, agriculture, and economics, to mention only a few. When the data analyzed are derived from the biological sciences and medicine, we use the term biostatistics to distinguish this particular application of statistical tools and concepts. This area of application is the concern of this book.

Las herramientas de la estadística se emplean en muchos campos: negocios, educación, psicología, agricultura y economía, por mencionar sólo algunos. Cuando los datos analizados se derivan de las ciencias biológicas y la medicina, utilizamos el término bioestadística para distinguir esta aplicación particular de herramientas y conceptos estadísticos. Esta área de aplicación es el objetivo de este libro.

**Variable** If, as we observe a characteristic, we find that it takes on different values in different persons, places, or things, we label the characteristic a variable. We do this for the simple reason that the characteristic is not the same when observed in different possessors of it. Some examples of variables include diastolic blood pressure, heart rate, the heights of adult males, the weights of preschool children, and the ages of patients seen in a dental clinic

Si, al observar una característica, encontramos que adquiere diferentes valores en diferentes personas, lugares o cosas, etiquetamos la característica como variable. Hacemos esto por la sencilla razón de que la característica no es la misma cuando se observa en diferentes poseedores de la misma. Algunos ejemplos de variables incluyen la presión arterial diastólica, la frecuencia cardíaca, la altura de los hombres adultos, el peso de los niños en edad preescolar y las edades de los pacientes atendidos en una clínica dental.

**Quantitative Variables** A quantitative variable is one that can be measured in the usual sense. We can, for example, obtain measurements on the heights of adult males, the weights of preschool children, and the ages of patients seen in a dental clinic. These are examples of quantitative variables. Measurements made on quantitative variables convey information regarding amount.

Una variable cuantitativa es aquella que se puede medir en el sentido habitual. Podemos, por ejemplo, obtener mediciones de la altura de varones adultos, el peso de niños en edad preescolar y las edades de los pacientes atendidos en una clínica dental. Estos son ejemplos de variables cuantitativas. Las mediciones realizadas sobre variables cuantitativas transmiten información sobre la cantidad.

**Qualitative Variables** Some characteristics are not capable of being measured in the sense that height, weight, and age are measured. Many characteristics can be categorized only, as, for example, when an ill person is given a medical diagnosis, a person is designated as belonging to an ethnic group, or a person, place, or object is said to possess or not to possess some characteristic of interest. In such cases measuring consists of categorizing. We refer to variables of this kind as qualitative variables. Measurements made on qualitative variables convey information regarding attribute.

Algunas características no se pueden medir en el sentido en que se miden la altura, el peso y la edad. Muchas características sólo pueden categorizarse, como, por ejemplo, cuando a una persona enferma se le da un diagnóstico médico, se designa a una persona como perteneciente a un grupo étnico, o se dice que una persona, lugar u objeto posee o no posee alguna característica de interés. En tales casos medir consiste en categorizar. A las variables de este tipo nos referimos como variables cualitativas. Las mediciones realizadas sobre variables cualitativas transmiten información sobre los atributos.

Although, in the case of qualitative variables, measurement in the usual sense of the word is not achieved, we can count the number of persons, places, or things belonging to various categories. A hospital administrator, for example, can count the number of patients admitted during a day under each of the various admitting diagnoses. These counts, or frequencies as they are called, are the numbers that we manipulate when our analysis involves qualitative variables.

Aunque en el caso de variables cualitativas no se logra la medición en el sentido habitual de la palabra, podemos contar el número de personas, lugares o cosas que pertenecen a diversas categorías. Un administrador de hospital, por ejemplo, puede contar el número de pacientes admitidos durante un día según cada uno de los distintos diagnósticos de admisión. Estos recuentos, o frecuencias como se les llama, son los números que manipulamos cuando nuestro análisis involucra variables cualitativas.

**Random Variable** Whenever we determine the height, weight, or age of an individual, the result is frequently referred to as a value of the respective variable. When the values obtained arise as a result of chance factors, so that they cannot be exactly predicted in advance, the variable is called a random variable. An example of a random variable is adult height. When a child is born, we cannot predict exactly his or her height at maturity. Attained adult height is the result of numerous genetic and environmental factors. Values resulting from measurement procedures are often referred to as observations or measurements.

Siempre que determinamos la altura, el peso o la edad de un individuo, el resultado suele denominarse valor de la variable respectiva. Cuando los valores obtenidos surgen como resultado de factores aleatorios, de modo que no pueden predecirse exactamente de antemano, la variable se denomina variable aleatoria. Un ejemplo de variable aleatoria es la altura adulta. Cuando nace un niño, no podemos predecir exactamente su altura en la madurez. La altura adulta alcanzada es el resultado de numerosos factores genéticos y ambientales. Los valores resultantes de los procedimientos de medición a menudo se denominan observaciones o mediciones.

**Discrete Random Variable** Variables may be characterized further as to whether they are discrete or continuous. Since mathematically rigorous definitions of discrete and continuous variables are beyond the level of this book, we offer, instead, non-rigorous definitions and give an example of each.

Las variables se pueden caracterizar aún más en cuanto a si son discretas o continuas. Dado que las definiciones matemáticamente rigurosas de variables discretas y continuas están más allá del nivel de este libro, ofrecemos, en cambio, definiciones no rigurosas y damos un ejemplo de cada una.

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.

Una variable discreta se caracteriza por lagunas o interrupciones en los valores que puede asumir. Estos espacios o interrupciones indican la ausencia de valores entre valores particulares que la variable puede asumir. Algunos ejemplos ilustran este punto. El número de ingresos diarios a un hospital general es una variable aleatoria discreta ya que el número de ingresos cada día debe representarse por un número entero, como 0, 1, 2 o 3. El número de ingresos en un día determinado no puede ser un número como 1,5, 2,997 o 3,333. El número de dientes cariados, perdidos o empastados por niño en una escuela primaria es otro ejemplo de variable discreta.

**Continuous Random Variable** 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.

Una variable aleatoria continua no posee los espacios o interrupciones características de una variable aleatoria discreta. Una variable aleatoria continua puede asumir cualquier valor dentro de un intervalo relevante específico de valores asumidos por la variable. Ejemplos de variables continuas incluyen las diversas mediciones que se pueden realizar en individuos, como la altura, el peso y la circunferencia del cráneo. No importa cuán cercanas sean las alturas observadas de dos personas, por ejemplo, podemos, en teoría, encontrar otra persona cuya altura se encuentre en algún punto intermedio.

Because of the limitations of available measuring instruments, however, observations on variables that are inherently continuous are recorded as if they were discrete. Height, for example, is usually recorded to the nearest one-quarter, one-half, or whole inch, whereas, with a perfect measuring device, such a measurement could be made as precise as desired.

Sin embargo, debido a las limitaciones de los instrumentos de medición disponibles, las observaciones de variables que son inherentemente continuas se registran como si fueran discretas. La altura, por ejemplo, normalmente se registra con precisión de un cuarto, media o pulgada entera, mientras que, con un dispositivo de medición perfecto, dicha medición podría hacerse con la precisión deseada.

**Population** The average person thinks of a population as a collection of entities, usually people. A population or collection of entities may, however, consist of animals, machines, places, or cells. For our purposes, we define a population of entities as the largest collection of entities for which we have an interest at a particular time. If we take a measurement of some variable on each of the entities in a population, we generate a population of values of that variable. We may, therefore, define a population of values as the largest collection of values of a random variable for which we have an interest at a particular time. If, for example, we are interested in the weights of all the children enrolled in a certain county elementary school system, our population consists of all these weights. If our interest lies only in the weights of first-grade students in the system, we have a different population—weights of first-grade students enrolled in the school system. Hence, populations are determined or defined by our sphere of interest. Populations may be finite or infinite. If a population of values consists of a fixed number of these values, the population is said to be finite. If, on the other hand, a population consists of an endless succession of values, the population is an infinite one.

La persona promedio piensa en una población como un conjunto de entidades, generalmente personas. Sin embargo, una población o conjunto de entidades puede consistir en animales, máquinas, lugares o células. Para nuestros propósitos, definimos una población de entidades como la colección más grande de entidades por las que tenemos interés en un momento particular. Si tomamos una medida de alguna variable en cada una de las entidades de una población, generamos una población de valores de esa variable. Por lo tanto, podemos definir una población de valores como la colección más grande de valores de una variable aleatoria por la que tenemos interés en un momento particular. Si, por ejemplo, estamos interesados ​​en los pesos de todos los niños matriculados en un determinado sistema de escuelas primarias del condado, nuestra población se compone de todos estos pesos. Si nuestro interés reside únicamente en el peso de los estudiantes de primer grado en el sistema, tenemos una población diferente: el peso de los estudiantes de primer grado matriculados en el sistema escolar. Por tanto, las poblaciones están determinadas o definidas por nuestra esfera de interés. Las poblaciones pueden ser finitas o infinitas. Si una población de valores consta de un número fijo de estos valores, se dice que la población es finita. Si, por el contrario, una población consta de una sucesión infinita de valores, la población es infinita.

**Sample** 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.

Una muestra puede definirse simplemente como una parte de una población. Supongamos que nuestra población está formada por el peso de todos los niños de escuela primaria matriculados en un determinado sistema escolar del condado. Si recopilamos para el análisis los pesos de sólo una fracción de estos niños, tenemos sólo una parte de nuestra población de pesos, es decir, tenemos una muestra.