

**PRIMERA PARTE: CARACTERÍSTICAS, RECOLECCIÓN, REVISIÓN Y
CÓMPUTO DE DATOS.**

1. CARACTERÍSTICAS DE LOS DATOS.

CARACTERÍSTICAS DE LOS DATOS.

Los datos con los cuales se realizan los trabajos estadísticos varían entre individuos y se obtienen a partir de elementos que en conjunto conforman un universo.

Con frecuencia **dato** e **información** se utilizan como sinónimos. Sin embargo, por información entendemos los datos procesados en forma significativa para el receptor, con valor real y perceptible para tomar decisiones presentes y futuras, los cuales se nos presentan en forma de indicadores. La información así planteada se obtiene como resultado o producto del proceso que se muestra en la Figura 1-1.



Figura 1-1. Generación de información a partir de datos.

Los datos no son útiles o significativos como tales, sino hasta que son procesados y convertidos en información. De alguna manera, la información es el conocimiento derivado del análisis de los datos. Esta es la diferencia básica entre datos e información. Hay que hacer notar que la información obtenida en un proceso puede servir como dato para otro proceso.

Universo.

En estadística se define como **universo** o **población** al conjunto de valores por los cuales existe algún interés. El total del universo o población se representa con la letra mayúscula N .

Las poblaciones pueden definirse especificando una regla (o reglas). Éstas pueden ser: características de individuos, límites geográficos, grupos ya existentes, límites de tiempo, etc., por ejemplo: residentes de Guadalajara, asistentes a un paseo escolar, derechohabientes del IMSS, enfermos de cólera.

Elementos del universo.

Los **elementos del universo** pueden ser personas, lugares o cosas, sean éstos individuos únicos o agrupados. Por ejemplo: los pacientes encamados son elementos que conforman parte del universo definido como hospital, pero también el personal, el mobiliario y los diversos servicios que en él se prestan pueden ser elementos del mismo conjunto. Para individualizarlo, cada elemento de la población se identifica con un número progresivo, que inicia en 1 y termina en N . Esta identificación de cada sujeto se representa como un subíndice y se le conoce como valor " i " o valor i -ésimo.

Variables.

Normalmente, el interés del investigador se dirige a las características de los elementos que conforman el universo. A dichas características se les designa como **variables**.

Ejemplos de variables pueden ser: la talla, el peso, el sexo, la temperatura corporal, la condición social y la escolaridad. Si el valor de la variable no puede predecirse con anticipación se le denomina **variable aleatoria** y para representarla se utilizan letras mayúsculas (X, Y, Z). Así pues, la variable aleatoria “edad” se puede representar con la letra X , y las variables aleatorias “sexo” y “escolaridad” se pueden definir como Y y Z . Los valores individuales de una variable aleatoria se representan con letras minúsculas (x, y, z) y un subíndice i -ésimo que identifica el elemento del conjunto que posee la característica.

Ejemplo explicativo 1-1 -----

Un investigador está interesado en los valores de hemoglobina en sangre de los trabajadores de un taller dedicado al mantenimiento de automóviles. Para ello, decide estudiar a todos los trabajadores que laboran en el taller. Después de recabar los datos de interés decide organizarlos de la manera que se presenta en el cuadro 1-1.

De acuerdo con los datos reportados, el universo se compone de cinco trabajadores; por tanto, $N = 5$. En el Cuadro 1-1 el investigador está representando a la variable “sexo” con la letra X , a la variable “edad en años cumplidos” con la letra Y , y a la variable “hemoglobina en sangre” con la letra Z , mientras que ha identificado a Miguel Domínguez con el valor i -ésimo 1, siendo su “sexo” $x_1 =$ masculino, su “edad en años cumplidos” $y_1 = 45$, y su “hemoglobina en sangre (mg/dL)” $z_1 = 12$. Para el caso de Domitila Hernández, los valores x_2, y_2, z_2 son, respectivamente, femenino, 23, 13. Y así para los restantes hasta completar todos los elementos que conforman el universo definido por el investigador.

Cuadro 1-1. Algunas características de trabajadores que laboran en un taller de automóviles

Nombre del trabajador	Número progresivo que identifica al trabajador	Sexo	Edad en años cumplidos	Hemoglobina en sangre (mg/dl)
	i	X	Y	Z
Miguel Domínguez	1	Masculino	45	12
Domitila Hernández	2	Femenino	23	13
Manuel Benítez	3	Masculino	32	11
Jesús Ortiz	4	Masculino	18	15
Sergio Martínez	5	Masculino	21	14

Definición operativa.

Todas las variables que se utilicen en cualquier trabajo de estadística han de ser definidas con claridad, de tal manera que se eviten confusiones, se facilite la búsqueda y análisis de los datos, y se garantice la comparabilidad de los resultados con los obtenidos con otros estudios previos o posteriores. Esto es particularmente importante cuando las variables a estudiar pueden ser definidas de maneras diferentes.

Ejemplo explicativo 1-2 -----

El sarampión es una enfermedad viral caracterizada por síntomas prodrómicos (fiebre, conjuntivitis, coriza, tos y manchas de Koplik en la mucosa bucal) que del tercer al séptimo días presenta en cara una erupción exantemática que se generaliza al resto del cuerpo y que desaparece de 4 a 7 días después. Durante el periodo de incubación, enfermedad y convalecencia se presentan modificaciones inmunes características del proceso morboso. El término “enfermo de sarampión” puede definirse operativamente de diversas maneras, entre las que se pueden encontrar las siguientes:

- Pacientes con presencia de manchas de Koplik en mucosa bucal.
- Pacientes con exantema maculopapular de tres o más días de duración, fiebre y cualquiera de las tres siguientes: tos, coriza o conjuntivitis.
- Pacientes con IgM específica para virus del sarampión.

La selección de una de éstas como definición operativa de sarampión tiene implicaciones importantes en el desarrollo de la investigación. El equipo de trabajo tendrá que utilizar una de ellas, u otra diferente que se adapte a sus necesidades, y precisarla de tal manera que quienes conozcan de su trabajo sepan a qué se refiere con el término “enfermo de sarampión”.

Una vez que la variable ha sido definida operativamente se debe especificar la escala de valores que se utilizará para clasificar los elementos en estudio. Cuando se trata de variables cuantitativas la decisión suele ser sencilla, pero no cuando la variable es cualitativa o cuando una variable cuantitativa se presenta agrupada. En cualquier caso, se espera que la escala de la variable permita clasificar a todos los elementos, sin excepción, y que cada elemento sea contado sólo una vez en relación con esa variable: en otras palabras, la escala ha de ser exhaustiva y excluyente.

Tipos de variables.

Las variables se pueden clasificar en cualitativas y cuantitativas. Cuando sus características se expresan como categorías se dice que se trata de variables cualitativas, mientras que cuando se expresan como valores se les identifica como variables cuantitativas.

1. Las **variables cualitativas** proporcionan **datos nominales** (en los que se tiene, o no se tiene, la característica de interés) y **datos ordinales** (en los que la característica es graduable).
 - a) Una **variable nominal** es aquella cuya característica se define por un nombre, y al ser definida por uno no implica ser más o menos que la característica definida por otro nombre diferente. Por ejemplo: “sexo” es una variable nominal, ya que ser “masculino” no significa ser más o menos “femenino”; la ocupación también es una variable nominal, ya que ser “ingeniero” o “abogado” no significa ser más o menos que “médico”. Aquellas variables nominales que se conforman de dos categorías (nacional, extranjero; con diarrea, sin diarrea, etc.) se designan como dicotómicas.

- b) Las **variables ordinales** son aquéllas cuyas características pueden recibir algún orden subjetivo. Su característica principal es que al ser clasificadas de una manera se puede asumir que se es más o menos que las otras, aunque se desconozca qué tanto más o qué tanto menos. En relación con el dolor, por ejemplo, el paciente puede decir que le duele “poco” o “mucho” y quien lo interroga puede asumir con seguridad que “mucho” significa más dolor que “poco”, aunque no se podría saber qué tanto es “mucho” ni qué “distancia” existe entre “poco” y “mucho”. Aún se podría asignarle una graduación subjetiva más detallada y no por eso dejaría de ser ordinal. Tal es el caso cuando se le pide al paciente que ubique, entre el “0” y el “100”, qué tanto dolor siente. En este caso, al igual que al clasificarlo como “mucho” o “poco”, se le asigna un orden en el cual “25” significa más que “15” y menos que “35”, pero se desconoce qué tan grande es la distancia entre “15” y “25” y no se puede asumir que sea la misma que existe entre “25” y “35”, ni que el “15” en una persona corresponda al “15” en otra. Las variables ordinales, al igual que las nominales, también pueden dicotomizarse sin que cambie su escala de medición. Por ejemplo, al definir el comportamiento como “bueno” o “malo” se divide la variable en dos categorías, en las cuales “bueno” representa algo más deseable (o indeseable) que “malo”.
2. Las **variables cuantitativas** también permiten diferenciar entre los individuos, pero además señalan cuán grandes son las diferencias observadas. Las observaciones cuantitativas brindan **datos discretos** (en los que sólo se admiten valores individuales en números enteros) y **datos continuos** (en los cuales es posible un número infinito de fracciones entre dos puntos de la escala).
- a) Se definen como **variables discretas** aquéllas cuyos valores en la escala están separados entre sí por una cantidad determinada. Ejemplos son el número de consultas otorgadas por médico en un día o el conteo de linfocitos en sangre. A diferencia de las variables ordinales, la “distancia” absoluta existente entre 5 y 7 consultas es la misma que entre 105 y 107 consultas, y también se puede saber qué tantas más son 100 consultas en relación con 10 consultas. Distintivo de estas variables es que la unidad no puede fraccionarse porque pierde su naturaleza. Así, si se parte por la mitad un paciente, éste deja de serlo para convertirse en dos mitades de cadáver.
- b) Las **variables continuas** son aquéllas en las cuales la escala de medición se puede dividir en una cantidad infinita de valores entre dos puntos cualquiera. Entre éstas se encuentran las medidas de longitud, peso, tiempo y volumen. Por ejemplo: entre 0 y 100 metros existe un número infinito de valores que pueden caracterizar al elemento en estudio, pero también es infinito el número de valores que se encuentran entre 0 y 10 metros o entre 0 y 10 centímetros. Al igual que con las variables discretas, la “distancia” absoluta entre dos puntos se mantiene a lo largo de la escala. Algunas variables parecen no respetar la última característica; tal es el caso de las titulaciones que se reportan como 1:1, 1:2, 1:4, 1:8, etc. Sin embargo, mediante una transformación matemática (logarítmica para el caso) se puede observar que sí se apegan a las características señaladas para las variables continuas.

Las variables cuantitativas también pueden clasificarse según tengan en su escala un valor de cero absoluto en variables de **intervalo** (no tienen cero absoluto) y variables de **razón** (si tienen cero absoluto). Por ejemplo, la temperatura medida en grados centígrados tiene un valor de 0° C, pero este es arbitrario y no es un valor absoluto porque existen otros valores por debajo de esa temperatura. La edad en años cumplidos, por otra parte, si tiene un valor de 0 absoluto porque nadie llega a tener menos de cero años de edad. La temperatura en grados Kelvin también tiene una escala de razón porque no existe una temperatura por debajo de 0° Kelvin. La distinción de estas dos escalas resulta importante para la interpretación de una razón. Por ejemplo: supongamos un niño que pesaba 50 kg y ahora tiene 60 kg. Podemos decir correctamente que el incremento del peso fue del 20%. Por otra parte, si un líquido que tenía una temperatura de 50° C llega a los 60° C no podremos afirmar lo mismo porque la escala de temperatura en grados Celsius no tiene un cero absoluto.

Ejemplo explicativo 1-3 -----

Una investigadora estaba interesada en identificar algunas condiciones de la madre que pudieran relacionarse con el bajo peso de los niños al nacer. Para ello, decidió estudiar las variables que se muestran en el Cuadro 1-2. En el mismo cuadro, las variables son clasificadas según su escala de medición.

Cuadro 1-2. Algunas características de mujeres durante el embarazo.

Variable	Escala de medición
• Edad de la madre (en años cumplidos)	Discreta, de razón
• Talla de la madre (en centímetros)	Continua, de razón
• Estado civil (soltera, casada, unión libre, otro)	Nominal
• Escolaridad (menos de primaria, primaria, secundaria, preparatoria o más)	Ordinal
• Temperatura (en grados Celsius)	Continua, de intervalo
• Ocupación durante el embarazo (según la Clasificación Mexicana de Ocupaciones)	Nominal
• Exposición al humo de tabaco durante el embarazo (no, fumadora pasiva, fumadora activa)	Ordinal
• Hemoglobina en sangre (mg/dl) en la primera consulta prenatal	Continua, de razón
• Número de consultas prenatales antes del tercer trimestre del embarazo	Discreta, de razón

Regularmente, la clasificación de las variables es una tarea sencilla, pero en ocasiones genera algunas dificultades menores. Por ejemplo, el tiempo es una variable continua porque entre un instante y otro cualquiera existe una cantidad infinita de divisiones, pero la edad (al menos como frecuentemente es registrada) es una variable discreta: un niño que cumple 10 años sigue reportando la misma edad durante todo el año hasta su

siguiente cumpleaños, fecha a partir de la cual empezará a decir que tiene 11 años. Por otra parte, el número de moléculas de glucosa (o de cualquier otra sustancia) es una variable discreta, porque al dividirse la molécula deja de ser glucosa, pero la concentración de glucosa en sangre es una variable continua, porque la dilución (mg/dL) está medida en una escala que tiene un número infinito de divisiones.

EJERCICIOS

- En las variables que se presentan a continuación, indique de qué tipo de variable se trata mediante las siguientes abreviaturas: N para una variable cualitativa nominal; O para una variable cualitativa ordinal; D para una variable cuantitativa discreta; y, C para una variable cuantitativa continua.

Variable	
1. Concentración de triglicéridos en sangre (mg/dl)	()
2. Diabetes (si, no)	()
3. Colesterol total (mg/dl)	()
4. Creatinina ($\mu\text{mol/L}$)	()
5. Número de cigarrillos fumados el día anterior	()
6. Índice de masa corporal (peso/talla ²)	()
7. Número de consultorios en la clínica	()
8. Presión arterial diastólica (mm Hg)	()
9. Intensidad del dolor (escala de 0 a 10)	()
10. Opinión del servicio (bueno, regular, malo)	()
11. Edad en años cumplidos	()
12. Uso de drogas anti-hipertensivas (si, no)	()
13. Presencia de cefalea (si, no)	()
14. Sexo (masculino, femenino)	()
15. Porcentaje de grasa corporal	()
16. Depósito de agua (aljibe, balde, alberca, tina, otro)	()

- Observe el cuadro que se le presenta a continuación. Los datos que contiene se refieren a las características de seis pacientes.

Identificación (número progresivo)	Índice de masa corporal (IMC)	Colesterol HDL	Glucosa en sangre	Fumador
1	24	52	110	si
2	27	50	95	no
3	25	57	103	no
4	27	48	115	si
5	26	49	100	no
6	29	56	120	no

Primero identifique la variable IMC con la letra *W*, la variable Colesterol-HDL con la letra *X*, la variable glucosa en sangre con la letra *Y*, y la variable Fumador con la letra *Z*. Luego escriba cuales son los valores individuales que representan las siguientes referencias:

Referencia	Valor	Referencia	Valor
w_1	()	y_3	()
w_2	()	y_4	()
w_3	()	y_5	()
x_2	()	z_4	()
x_3	()	z_5	()
x_4	()	z_6	()

EJERCICIOS

Ejercicio 1

Observe el cuadro que se le presenta a continuación. Los datos que contiene se refieren a las características de seis pacientes.

Identificación (número progresivo)	Índice de masa corporal (IMC)	Colesterol HDL	Glucosa en sangre	Fumador
1	24	52	110	si
2	27	50	95	no
3	25	57	103	no
4	27	48	115	si
5	26	49	100	no
6	29	56	120	no

Ejercicio 2

Primero identifique la variable IMC con la letra *W*, la variable Colesterol-HDL con la letra *X*, la variable glucosa en sangre con la letra *Y*, y la variable Fumador con la letra *Z*. Luego escriba cuales son los valores individuales que representan las siguientes referencias:

Referencia	Valor	Referencia	Valor
w_1	()	y_3	()
w_2	()	y_4	()
w_3	()	y_5	()
x_2	()	z_4	()
x_3	()	z_5	()
x_4	()	z_6	()

REFERENCIAS

Camel F: Estadística Médica y Planificación de la Salud. Tomo I, 1ª edición. Universidad de los Andes. Mérida, Venezuela: Consejo de Publicaciones, 1991.

Nunnally JC, Bernstein I: Psychometric Theory. 3ª edición. McGraw-Hill. Estados Unidos de América: 1997.