

## MÓDULO 4

### SEGUIMIENTO A LAS POLÍTICAS PÚBLICAS SOBRE DESARROLLO SOSTENIBLE



PERIODISMO DE DATOS  
**PARA EL DESARROLLO SOSTENIBLE**

# EN EL MUNDO DE LOS DATOS: ANÁLISIS

El análisis de datos puede usarse para decir verdades alrededor de los datos o también para adecuar la realidad. Toda persona que usa datos, sin importar su propósito, debe ser consciente y cuidadosa al analizar la información, pero sobre todo, tener claridad sobre los conceptos y técnicas mínimas que son necesarias para dicha labor.

Para analizar datos sobre alguna de las temáticas de los ODS es necesario conocer algunos conceptos usados de forma constante en las diversas bases de datos que se pueden encontrar en los temas que los ODS cubren. Aquí algunos de estos conceptos:

- **Conjunto de datos:** un conjunto de datos o también conocido como *dataset* es el archivo base de datos que almacena información asociada a una temática específica. Un conjunto de datos puede ser entregado en un único archivo o en varios, dependiendo de la cantidad de registros y la capacidad informática que se tenga para procesar la información. Por ejemplo, el censo de población de un país es un conjunto de datos ya que habla de una temática específica y puede ser entregado de un archivo o varios.
- **Población (N).** Conjunto de todas las observaciones o elementos de interés en un determinado estudio que tienen propiedades comunes. Puede ser un número finito de datos (número de personas beneficiarias de un programa

en un tiempo determinado) o una colección grande de datos (proyecciones poblacionales de 1950 a 2050).

- **Muestreo.** Una porción representativa de una población o universo.
- **Variable.** Una categoría de datos de un período definido. Estas pueden ser:
  - **Cualitativa.** Cuando la característica no es numérica (preferencia religiosa, estado civil, entre otros).
  - **Cuantitativa.** Cuando se asumen valores numéricos, acompañados de unidades de medida, como cantidad de productos importados a un país, las calificaciones de un examen, etcétera.
  - **Continua.** Aquella que toma cualquier valor dentro de un intervalo; por lo general, proceden de mediciones, como la estatura o la presión de aire.
  - **Discreta.** Aquella que solo puede tomar determinados valores en un intervalo. Son números enteros, resultado de un conteo (número de hijos, número de inasistencias a clase, etcétera).

En el análisis de datos es importante reconocer estas características antes de iniciar cualquier actividad dado que nos permitirá conocer de mejor manera una base de datos antes de iniciar cualquier proceso de análisis. Este

reconocimiento de información con el paso del tiempo se vuelve más sencillo, ya que si se trabaja de forma regular con determinados conjuntos de datos asociados a un ODS se comienza a estar más acostumbrado a la terminología asociada a la materia y también se conoce de mejor forma la información que se presenta en cada base de datos.

Analizar datos representa mucho más que una etapa entre otras del trabajo con ellos; el análisis de datos es la posibilidad de conocer a profundidad la información que una base de datos contiene mediante su interpretación metódica.

El análisis de datos se apoya de forma constante de la estadística descriptiva para interpretar el contenido de una base de datos. De hecho, se puede decir que sin los fundamentos de la estadística descriptiva, no se puede hacer análisis de datos.

### CONCEPTOS COMUNES EN EL ANÁLISIS DE DATOS

A continuación se abordan algunos de los conceptos más comunes:

**Promedio (o media aritmética):** sirve para obtener el valor que representa el término medio de un conjunto de datos. Para calcularlo se suman todos los datos del mismo tipo y luego se dividen entre la cantidad de datos sumados. Por ejemplo: estos son los 6 municipios de El Salvador con menos habitantes, ¿cuál es el promedio de habitantes de estos municipios?

MUNICIPIO	NO. HABITANTES
Municipio 1	1000
Municipio 2	1200
Municipio 3	1100
Municipio 4	1050
Municipio 5	1100
Municipio 6	1005

Para responder esta pregunta es necesario sumar el número de habitantes de los seis Municipios (6470) y dividirlo entre el número de municipios (6). En promedio, los habitantes de los seis municipios de El Salvador es de 1080.

**Mediana:** es el número intermedio de un grupo de números; es decir, la mitad de los números son superiores a la mediana y la mitad tiene valores menores a ella. Para diferenciar la media aritmética de la mediana, imagina que en una empresa con 140 trabajadores, ellos ganan lo siguiente:

EMPLEADOS	SALARIO DEVENGADO
100	\$900
20	\$1100
15	\$1400
4	\$2500
1	\$20000

El promedio de los salarios sería \$5180 mientras que la mediana sería \$1400. Ambos datos son ciertos, pero no reflejan la realidad de la empresa, es decir, un trabajador de esta empresa no gana en promedio \$5180. Pero sí se puede afirmar que la mediana de salarios se ubica en \$1400, un valor que muestra de mejor forma la realidad. Es importante diferenciar la utilidad de ambos conceptos; particularmente aplicado al ejercicio anterior permite ejemplificar también uno de los errores más comunes al interpretar datos.

Por ejemplo con los datos presentados podemos observar que un número importante de trabajadores (100) ganan \$900 mientras que un número mucho menor gana \$2000. Al existir tanta diferencia entre el límite inferior y el límite superior, el promedio se ve afectado porque los valores superiores elevan el resultado sin que esto represente la realidad de los trabajadores del ejemplo.

En cambio, la mediana permite, sin importar las distancias entre los extremos de la serie, tener una idea más clara del salario que los trabajadores reciben. En este caso es normal que dentro de los análisis de datos se confunda constantemente el uso del promedio vs la mediana. La recomendación permanente para estos escenarios es siempre hacer el ejercicio para ambos cálculos y notar las diferencias entre ambos valores.

**Moda:** es el valor que aparece más frecuentemente en un rango de datos. Por ejemplo, en este conjunto de datos relacionados a calificaciones, el valor (calificación) que más se repite es 8 con 3 apariciones. Esto significa que la moda es 8.

ALUMNOS	SALARIO DEVENGADO
Alumno 1	7
Alumno 2	8
Alumno 3	9
Alumno 4	8
Alumno 5	8

## EJEMPLO DE ANÁLISIS CALCULANDO TASAS

### ¿Cómo calcular tasas?

Las tasas se utilizan cuando se calculan valores relacionados con la población de un lugar en específico. Las tasas son necesarias porque ubican en condiciones de igualdad a dos o más lugares sin importar sus diferencias entre el número total de habitantes que cada lugar tenga, ya que al usar números absolutos los hallazgos no serían representativos ni se podrían comparar las localidades entre sí. Por ejemplo, cómo entender el embarazo adolescente en diversas localidades de un mismo país.

Para dar respuesta a esta pregunta es necesario emplear una sencilla fórmula para obtener una tasa por cada 100 mil habitantes de cada municipio o localidad. La fórmula para calcular una tasa por cada 100 mil habitantes es:

$$\text{Tasa} = (100000 * \text{número de hechos}) / \text{población}$$

No es necesario que las tasas siempre estén calculadas por 100 mil personas, además no es correcto hacerlo cuando la población del lugar

es menor de 100 mil personas. Por ejemplo, si estamos calculando la tasa de robos de 10 municipios en los que ninguno tiene una

población de más de 60 mil habitantes, lo correcto sería calcular la tasa por cada 50 mil habitantes. La fórmula cambiaría así:

$$\text{Tasa} = (50000 * \text{número de hechos}) / \text{población}$$

Aplicando el uso de tasas a nuestro ejemplo inicial podemos entonces hacer lo siguiente:

MUNICIPIO	NO. DE EMBARAZOS	HABITANTES	FÓRMULA	TASA
Municipio 1	500	150000	$(100000 * 500) / 150000$	333.33
Municipio 2	200	140000	$(100000 * 200) / 140000$	142.86
Municipio 3	100	130000	$(100000 * 100) / 130000$	76.92
Municipio 4	250	125000	$(100000 * 250) / 125000$	200.00
Municipio 5	50	130000	$(100000 * 50) / 130000$	38.46

Con la presentación de la tabla y después de hacer el cálculo para obtener la tasa por cada municipio podemos entonces entender mejor el fenómeno del embarazo adolescente del ejercicio. Si bien el municipio 1 es el que más embarazos adolescentes tiene por cada 100 mil habitantes, destaca el caso del municipio 4 que con una población menor a otros municipios presenta una tasa superior de embarazos comparado con otros municipios del ejemplo.

Hacer tasas de habitantes para describir un hecho resulta de gran ayuda ya que permite poner en perspectiva cada caso y dimensiona los resultados frente a otros valores. El uso de tasas resulta un gran recurso que vale la pena tener presente para todo ejercicio de análisis de datos.

Si está en México, Colombia o Costa Rica, pase por el ecosistema de datos de [www.datapublica.org](http://www.datapublica.org) y explore un poco el panorama de la disponibilidad de datos para cada Objetivo de Desarrollo Sostenible (ODS) en estos países.