**ESTADISTICA DESCRIPTIVA**

**Introducción**

**Estadística descriptiva**: es la rama que se ocupa de aportar herramientas (mediciones numéricas, gráficos y tablas) a fin de presentar en forma clara y resumida la información más importante contenida en un conjunto de datos.

**Variables, Bases de datos y escalas de medición**

**Definiciones**:

**Variable**: cualquier característica o atributo que se desea medir.

Ejemplo: Preferencia electoral, pH, género, Número de personas en el hogar, Color de ojos, Nacionalidad, Estrato socioeconómico, Estado civil, Peso, Nivel de glucosa, Opinión en una encuesta, etc.

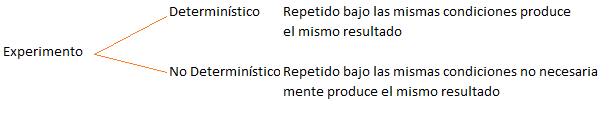
**Experimento**: cualquier actividad o proceso que nos permite generar datos.

Ejemplo 1: Encuestar una persona por su preferencia electoral

Medir el pH de una sustancia

Observar el número de clientes que hacen cola en un banco cada hora.

:



Ejemplos: Modelos matemáticos vs Modelación estadísticos

|  |  |
| --- | --- |
|  |  |

**Base de datos**: nombre dado a la organización de los datos en forma tal que en las columnas se encuentran las variables y las filas contienen la información correspondiente a una misma unidad muestral.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Variable 1 | Variable 2 | …. | Variable k |
| Individuo 1 |  |  | …. |  |
| Individuo 2 |  |  | …. |  |
| :  : |  |  | …. |  |
| Individuo n |  |  | …. |  |

La construcción de una adecuada base de datos es un elemento clave para realizar una investigación. Antes de comenzar a digitar una base de datos es pertinente analizar el nivel de medición de las variables ya que en general de ellos depende en gran parte el tipo de técnicas estadísticas que se pueden o no aplicar.

**Clasificación de Variables en Estadística**:

|  |  |  |
| --- | --- | --- |
| **Estadística Descriptiva** | **Probabilidad** | **Modelos Lineales** |
| Cualitativas:Categóricas-Nominales | Nominal – No admiten orden (Estado Civil – Color preferido) | Independientes  Dependientes |
|  | Ordinal – Si existe orden (Notas de un examen E – B – A – I) |  |
| Cuantitativas – Cardinales numericas | Discretas – Toma un numero finito de valores (Edad – Numero de Hermanos) |  |
|  | Continuas – Toma un numero infinito de valores (Tiempo en recorrer una distancia – Peso de un Balon – Las respuestas pueden tener infinito numero de decimales) |  |

* **Variables discretas: son aquellas que no aceptan un valor entre dos números consecutivos. Si tienes los datos 1, 2, 3, 10, 11 y 15, entre el 1 y 2 no puede aparecer el 1.48, porque del 1 salta directamente al 2. Generalmente, las variables discretas son resultado de un conteo y no permiten los números decimales. Por ejemplo: número de pacientes, número de alumnos, número de motos por modelo.**
* **Variables continuas: son aquellas que pueden tomar cualquier valor entre dos intervalos o números. Por ejemplo, si necesitas escribir la estatura de un grupo de basquetbolistas, seguramente, no podrás utilizar los números 1 y 2, pero si las variables 1.78, 1.65, 1.45, porque la altura suele expresarse de esa manera.**

**Variables según las escalas de Medida**

Medir es un proceso que asigna etiquetas a los diferentes valores que asume una variable. Según la escala de medición que se utilice se ha definido la siguiente clasificación:

Variable Nominal: la variable sólo asume ciertos valores o nombres que crean categorías no importa el orden. Se emplean nombres, símbolos o números para diferenciar las categorías o subgrupos creados. Ejemplo: Lugar de procedencia.

Variable Ordinal: los valores de la variable pueden ser ordenados bajo algún criterio en orden ascendente o descendente. Las etiquetas, símbolos o números, indican jerarquía aunque no es posible conocer la diferencia entre ellas. Ejemplo: Nivel de Inglés medido en la escala A1, A2, B1, Rendimiento académico medido en una escala cualitativa Insuficiente, Aceptable, Bueno, Excelente, etc.

Variable por intervalo (interválica): si las categorías que asume además de poderse ordenar cubren intervalos iguales (la diferencia entre dos categorías sucesivas es la misma cantidad de magnitud de la variable). La adición y diferencia tiene sentido no así la multiplicación y la división. Requiere de un punto cero y una unidad de medida. Ejemplo: La temperatura medida en una escala Fahrenheit, ya que la diferencia entre las temperaturas de 30 y 31 es un grado al igual que la diferencia entre 100 y 101, no importa dónde está ubicado ese uno en la escala, ese un grado representa la misma cantidad de calor. Aquí **el valor de cero no implica ausencia de la característica** o variable medida y puede ser colocado arbitrariamente en algún lugar de la escala, en este caso, Fahrenheit estableció como temperatura 0° al punto de congelación de una mezcla al 50% de sal y hielo.

Variable en escala de razón: además de lo indicado para la interválica, una variable en escala de razón tiene un verdadero punto cero. Ejemplo: Tiempo, peso, velocidad, etc. Para este tipo de variables, todas las operaciones aritméticas tienen sentido, por ejemplo: cuando se trabaja con variables de razón (y no con interválicas) usted puede construir la razón de dos medidas, así un peso de 4 gramos es el doble de uno de 2 en cambio una temperatura de 100ºC no es el doble de una de 50 ºC o un pH de 6 no es el doble de uno de 3 y aquí además el cero no implica no acidez, etc.

Veamos algunos ejemplos:

|  |  |  |
| --- | --- | --- |
| Escalas | Criterio | Ejemplos |
| Nominal | Categorias | Lugar de procedencia, Género, color de ojos, Religión que profesa, Opinón (Si/No), Estado civil, etc. |
| Ordinal | categorias, se pueden ordenadar | Estrato socioeconómico, Nivel en la carrera, Puesto en el ICFES, Nivel de Inglés, Rendimiento académico (Insuficiente, Aceptable, Bueno, Excelente), Nivel de gravedad de una enfermedad, Edad en quinquenios, Status de fumador, cargo, etc. |
| Intervalo | Escala de valores, se pueden ordenadar, intervalos iguales | Temperatura corporal medida en grados F o C.  Sobrepeso respecto a un patrón  Prueba de coeficiente intelectual  Ubicación en una carretera respecto a un punto  Puntaje en el ICFES o test como el de coeficiente intelectual |
| Razón | Escala de valores, se pueden ordenadar, intervalos iguales, un verdadero punto cero | Edad, Peso, Estatura, Tiempo para completar una tarea, ingresos, Nota en un examen en la UIS [0,5], Temperatura medida en grados Kelvin, etc. |

¿Por qué es importante hacer esta consideración?

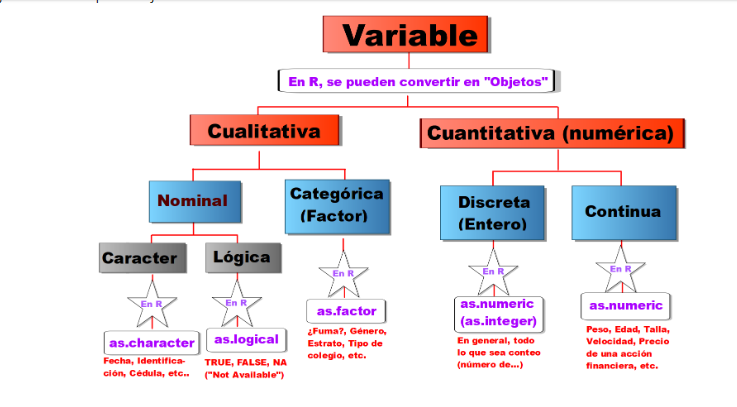
Como ya advertimos dependiendo de la escala de medición de la variable será adecuado o no el uso de ciertas herramientas estadísticas:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Herramienta a utilizar | Nominal | Ordinal | Intervalo | Razón |
| Distribución de frecuencias | SI | SI | SI | SI |
| Mediana y percentiles | No | SI | SI | SI |
| Suma y restas | No | No | SI | SI |
| Media, desviación estándar, percentíles. | No | No | SI | SI |
| Razones, coeficiente de variación | No | No | No | SI |

Notas para el analista:

1. Cada software tiene una definición propia de variables que no siempre se ajusta a lo definido en Estadística, por tanto encontraremos diferentes tipos de variables con las cuales debemos aprender a interactuar:

|  |  |  |  |
| --- | --- | --- | --- |
| **Software** | Excel | SPSS | R |
| **Tipos de Variables** | No usa el término, en cambio alude a argumento en la descripción de las funciones  Números  Valores lógicos o de texto, matrices, referencias | Nominal (cadenas de caracteres)  Ordinal (Numéricas)  Coma, Punto, Notación científica, Fecha, Dólar, Moneda personalizada | Se habla en término de objetos y estos pueden ser:  De valor entero  Numéricos  Carácter  Factor  Lógico  Complejo |

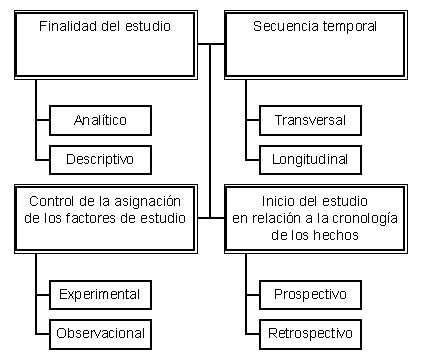


1. Recuerde que usted puede tener una variable Nominal pero utiliza etiquetas numéricas para medirla, ejemplo: La variable Género puede ser codificada como 1 o 0. A veces hay que codificar variables utilizando números para facilitar su lectura o para obtener ciertos gráficos pero es recomendable no perder de vista la naturaleza de la variable.
2. Es muy recomendable documentar las bases de datos, esto implica seleccionar nombre adecuados para las variables y tener registro de información relacionada con unidades de medida y métodos de medición, notación para datos faltantes, etc.

**Métodos de recolección de datos**:

Las estrategias usuales de recolección de datos para análisis estadísticos son: encuestas, entrevistas, test o pruebas estandarizadas, evaluación o medición por un experto (salud), observación directa, , realización de experimentos, Grupos focales, Método Delphi. Se habla de fuentes secundarias para referirse a fuentes donde la información ya existe, entre ellas se ubican: revisión de registros históricos o fuentes bibliográficas, bases de datos, etc. Hoy día también se hace referencia a Información estructurada y no estructurada para diferenciar aquella información que se genera sin que haya un estudio u objetivo de fondo, el ejemplo más claro es la información que se puede extractar a partir de las redes sociales.

Otro aspecto importante para determinar el tipo de herramientas estadísticas a utilizar en un análisis es el diseño de investigación que se va a utilizar, existen una amplia clasificación atendiendo a diferentes criterios, la siguiente es una muy general[[1]](#footnote-1):



1. http://www.uv.es/invsalud/invsalud/disenyo-tipo-estudio.htm [↑](#footnote-ref-1)